

Comparing Within- and Between-Family Polygenic Score Prediction

Saskia Selzam,^{1,*} Stuart J. Ritchie,¹ Jean-Baptiste Pingault,^{1,2} Chandra A. Reynolds,³ Paul F. O'Reilly,^{1,4} and Robert Plomin¹

Polygenic scores are a popular tool for prediction of complex traits. However, prediction estimates in samples of unrelated participants can include effects of population stratification, assortative mating, and environmentally mediated parental genetic effects, a form of genotype-environment correlation (rGE). Comparing genome-wide polygenic score (GPS) predictions in unrelated individuals with predictions between siblings in a within-family design is a powerful approach to identify these different sources of prediction. Here, we compared within- to between-family GPS predictions of eight outcomes (anthropometric, cognitive, personality, and health) for eight corresponding GPSs. The outcomes were assessed in up to 2,366 dizygotic (DZ) twin pairs from the Twins Early Development Study from age 12 to age 21. To account for family clustering, we used mixed-effects modeling, simultaneously estimating within- and between-family effects for target- and cross-trait GPS prediction of the outcomes. There were three main findings: (1) DZ twin GPS differences predicted DZ differences in height, BMI, intelligence, educational achievement, and ADHD symptoms; (2) target and cross-trait analyses indicated that GPS prediction estimates for cognitive traits (intelligence and educational achievement) were on average 60% greater between families than within families, but this was not the case for non-cognitive traits; and (3) much of this within- and between-family difference for cognitive traits disappeared after controlling for family socio-economic status (SES), suggesting that SES is a major source of between-family prediction through rGE mechanisms. These results provide insights into the patterns by which rGE contributes to GPS prediction, while ruling out confounding due to population stratification and assortative mating.

Introduction

The recent influx of well-powered genome-wide association (GWA) studies has led to substantial advances in our ability to detect genetic associations between single base pair variants (single-nucleotide polymorphisms [SNPs]) across the genome and a myriad of complex traits. Although individual SNP effect sizes are extremely small,¹ the surge in GWA power has improved the ability to predict complex traits through the genome-wide polygenic score (GPS) approach.^{2,3} GPSs are indices of individuals' genetic propensity for a trait and are derived as the sum of the total number of trait-associated alleles across the genome, weighted by their respective association effect size estimated through GWA analysis.⁴ GPS can be calculated in any sample with genotype data that is independent from the discovery GWA study, and have permeated research in the social, behavioral, and biomedical sciences.⁵ In this paper, we use within-family analysis to investigate an important potential source of prediction in polygenic score analysis: passive genotype-environment correlation.

Currently one of the largest GWA meta-analyses with a sample size of 1.1 million was performed on educational attainment (years of schooling).⁶ A GPS derived from this study is the most predictive GPS for any behavioral trait to date, explaining 10.6% of the variance in years of education⁶ and 14.8% in tested educational achievement.⁷ The predictive power of the educational attainment GPS

(EA GPS) is considerable in contrast to other GPS for behavioral traits. Notably, cross-trait analyses have revealed that EA GPS is widely associated with traits other than educational achievement, including intelligence,^{2,6,7} socioeconomic status (SES),^{8–11} behavior problems,¹² mental health,¹³ physical health,¹³ and personality,^{14,15} in some cases accounting for as much as or more than the variance in cross-trait associations explained by the target GPSs themselves.^{15,16}

However, GWA analyses, and the GPSs derived from them in independent samples, are naive to the pathways that lead from SNPs to trait outcomes.¹⁷ With a focus on prediction, the mechanisms by which polygenic scores relate to phenotypes are left largely unexplored. Given the popularity and widespread use of the GPS approach, the interpretation of GPS prediction estimates requires more careful consideration. Potentially, *passive genotype-environment correlation* (prGE)¹⁸ effects could be one source of prediction. Parents generate family environments consistent with their own genotypes, which in turn facilitate the development of the offspring trait, thus inducing a correlation between offspring genotype and family environment.^{19–21} Although these effects are also genetic in origin, they stem from the parents and are thus environmentally mediated. Therefore, GPS prediction among unrelated individuals may include contributions from both direct genetic effects and also indirect effects due to prGE.

Within-family analysis of siblings is a powerful approach to disentangle these potential sources of prediction. The

¹Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, UK; ²Division of Psychology and Language Sciences, University College London, London WC1H 0AP, UK; ³Department of Psychology, University of California Riverside, Riverside, CA 92521, USA; ⁴Icahn School of Medicine, Mount Sinai, New York, NY 10029, USA

*Correspondence: saskia.selzam@kcl.ac.uk
<https://doi.org/10.1016/j.ajhg.2019.06.006>

© 2019 The Authors. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

additive genetic correlation between siblings is on average 0.50,²² and the transmission of alleles from parents to offspring is randomized during meiosis, such that siblings have equal probability of inheriting any given allele.²³ The variability around the average genetic relationship between siblings due to random segregation is generally independent of the environment, so any genetic difference between siblings is free of shared environmental influence.²⁴ A relationship between their genetic differences and trait differences provides evidence for a causal effect of the measured genetic difference, since (1) siblings are well matched on all shared familial genetic influences that shape the environment, and (2) potential bias due to population stratification and assortative mating is completely eliminated within families.^{6,25,26} Such within-family analyses account for prGE effects that are related to common family environments that are correlated with the transmitted alleles shared between siblings, but also environmental effects related to non-transmitted parental alleles that contribute to offspring similarity within a family. The use of DZ co-twins strengthens this design further as all shared environmental influences are time-invariant between twins (e.g., pregnancy risk factors, parental age, family income).

Indeed, previous within-family analyses have revealed substantial reductions in individual SNP effect sizes. For example, there was an effect size attenuation of ~40% compared to between-family associations in the most recent GWA study on educational attainment.⁶ Most of this reduction has been attributed to prGE; no similar deflation of effect sizes was found for height,⁶ indicating that prGE is not likely at play. A novel method relying on closely and distantly related individuals, and that is applied to very large populations, detected a similar reduction of SNP-heritability estimates of educational achievement (~40%).²⁴ Moreover, studies that tested the effect between non-transmitted alleles from parental to offspring genotypes on offspring outcomes reported a significant association for educational attainment^{20,21}—an effect of so-called *genetic nurture*—but not for height and BMI.^{20,21} In contrast, one study that tested within-family predictions of educational attainment using the EA GPS found no noteworthy difference in comparison to between-family estimates.²⁷ However, this GPS was based on the first GWA study for educational attainment²⁸ and may have been underpowered to pick up prGE-driven effects. Indeed, a more recent study found that using the latest GPS for educational attainment, there was an attenuation of ~55% in the prediction of years of schooling within families in comparison to between-family estimates.²⁹

Overall, relatively little research has been conducted on within-family GPS prediction, mostly focusing on educational and anthropometric traits. This study adds substantially to this literature by systematically comparing within-family GPS prediction to between-family GPS prediction across eight life outcomes (height, BMI, self-rated

health, intelligence, educational achievement, neuroticism, attention-deficit/hyperactivity symptoms, and schizophrenia symptoms). Educational achievement is both phenotypically and genetically correlated with many life outcomes.^{30–36} It is also highly genetically correlated with family SES,^{8,37,38} and EA GPS predicts 7.3% of the variance in SES.⁹ Therefore, it is possible that the effects identified in the GWA studies for educational attainment related to family environment (e.g., SES) also contribute to the development of other behavioral traits through prGE mechanisms. Although it has been suggested that the widespread cross-trait associations between the EA GPS and various outcomes may be partly driven by prGE effects,^{15,39} to our knowledge no study to date has tested this hypothesis.

It is the aim of this study to investigate potential influences of prGE in a range of life outcomes through the comparison of within- and between-family polygenic score prediction estimates. First, we predict that within-family estimates will be disproportionately lower than between-family estimates for EA GPS predictions of educational achievement in contrast to other GPS predictions of their target trait. Second, we predict that cross-trait associations between the EA GPS and other outcomes will be smaller within families than between families, in comparison to the cross-trait associations of other GPSs.

Material and Methods

Our hypotheses, measures, and analysis plan were preregistered with the Open Science Framework (for more details, see [Web Resources](#)), except where indicated below. The non-preregistered analyses should be considered exploratory.

Sample

Participants were drawn from the Twins Early Development Study (TEDS). Between 1994 and 1996, TEDS recruited 16,810 twin pairs born in England and Wales, who have been assessed in multiple waves across development until the present. The demographic characteristics of TEDS participants and their families closely match those of families in the UK.^{9,40} Written informed consent was obtained from parents prior to data collection and from TEDS participants themselves past the age of 18. Project approval was granted by King's College London's ethics committee for the Institute of Psychiatry, Psychology and Neuroscience PNM/09/10–104. Only DZ co-twins with complete data were included in this study.

Phenotypic Data

Height

Self-reported height was assessed at the average age of 22.1 (SD = 0.86) in 1,463 twin pairs.

Body Mass Index (BMI)

BMI was calculated using self-reported weight in kg and height in meters (kg/m^2) at age 22.1 (SD = 0.86) in 1,353 twin pairs.

Self-Rated Health

Twins rated their health on the reduced RAND Short-Form Health Survey.⁴¹ Individuals scored their health on a five-point Likert

scale for five questions such as “In general, would you say your health is?” (“Poor” to “Excellent”), or “I am as healthy as anybody I know” (“Strongly Disagree” to “Strongly Agree”). Data were available on 1,494 twin pairs at age 22.1 (SD = 0.86).

Intelligence

At age 11.4 (SD = 0.65), twins were assessed on their non-verbal abilities (Raven’s Standard Progressive Matrices;⁴² WISC-III-UK Picture Completion⁴³) and on their verbal abilities (WISC-III-PI Vocabulary Multiple-Choice;⁴⁴ WISC-III-PI Information Multiple-Choice⁴⁴). A composite variable was calculated as the arithmetic mean of the z-standardized scales for 1,569 twin pairs.

Educational Achievement

Results for standardized tests taken at the end of compulsory education in the United Kingdom (General Certificate of Secondary Education; GCSE) were obtained for twins at age 16.3 (SD = 0.29) via self-report. Grades were coded from 4 (G; the minimum pass grade) to 11 (A*; the highest possible grade). Self-reported GCSE grades in TEDS highly correlate with grades obtained for a subsample of individuals from the National Pupil Database ($r = 0.98$ for English, $r = 0.99$ for mathematics, $r > 0.95$ for all sciences).³¹ A composite was calculated as the arithmetic mean of the compulsory core subjects—Maths, English, and Science—for 2,366 twin pairs.

Neuroticism

At age 16.5 (SD = 0.27), twins were assessed on their Big Five personality traits on a five-point Likert scale.⁴⁵ For this study, we used the six neuroticism items (e.g., anxiousness, vulnerability) to form a composite score by taking the arithmetic mean for 789 twin pairs.

Attention-Deficit Hyperactivity Disorder (ADHD) Symptoms

At age 11.5 (SD = 0.69) and 16.3 (SD = 0.69), parents reported on twins’ ADHD symptoms via the Strength and Difficulties Questionnaire⁴⁶ hyperactivity subscale (three-point Likert scale) and the Conners’ rating scales (CPRS-R; four-point Likert scale)⁴⁷ on hyperactivity and inattention. Although self-report ratings were available, it has been shown that informant-based ratings are more reflective of objective measures of ADHD symptoms.⁴⁸ A composite score was created as the arithmetic mean of the sex and age z-standardized scales. Where ratings were available at one assessment only, this value was used to maximize sample size, leading to a sample of 2,469 twin pairs.

Schizophrenia Symptoms

At age 22.7 (SD = 0.85), paranoia and hallucinations were assessed through self-reported ratings on the Specific Psychotic Experiences Questionnaire (SPEQ; six-point Likert scale),^{49,50} and parent-reported negative symptoms using the Scale for the Assessment of Negative Symptoms (SANS; four-point Likert scale).⁵¹ Data were available for 1,140 twin pairs.

Family Socio-economic Status (SES)

This measure was calculated as the mean of the z-standardized maternal age at birth of the first child, maternal and paternal highest education level (coded from 1 = “no qualifications” to 8 = “postgraduate qualifications”), and maternal and paternal occupation (coded from 1 = “Other Occupations – dockers, porters, labourers,...” to 9 = “Managers and Administrators”). These measures were assessed at first contact at age 1.8 (SD = 1.13). Data were available for 2,962 twin pairs.

Measures were selected based on largest sample sizes available, and ages at phenotype assessment matching most closely the ages of GWA study samples to maximize predictive power. None of the measures were significantly associated with birth order, but most showed sex and age differences (see Table S1) and were

therefore adjusted for these effects using the regression method, and z-standardized residuals (mean = 0, SD = 1) were used for all subsequent analyses.

Genotypic Data

Two different genotyping platforms were used because genotyping was undertaken in two separate waves, 5 years apart. AffymetrixGeneChip 6.0 SNP arrays were used to genotype 3,665 individuals. Additionally, 8,122 individuals (including 3,607 DZ co-twin samples) were genotyped on Illumina HumanOmniExpressExome-8v1.2 arrays. After quality control, 635,269 SNPs remained for AffymetrixGeneChip 6.0 genotypes, and 559,772 SNPs for HumanOmniExpressExome genotypes.

Genotypes from the two platforms were separately phased and imputed into the Haplotype Reference Consortium (release 1.1) through the Sanger Imputation Service⁵² before merging. Genotypes from a total of 10,346 samples (including 3,320 DZ twin pairs and 7,026 unrelated individuals) passed quality control, including 3,057 individuals genotyped on Affymetrix and 7,289 individuals genotyped on Illumina. The identity-by-descent (IBD) between individuals was < 0.05 for 99.5% in the sample excluding the DZ co-twins (range = 0.00 – 0.12) and ranged between 0.36 and 0.62 for the DZ twin pairs (mean = 0.49). The final data contained 7,363,646 genotyped or well-imputed SNPs (for full genotype processing and quality control details, see Selzam et al.⁵³). To ease high computational demands of the software that generates polygenic scores, we further excluded SNPs with $\text{info} < 1$, leaving 515,000 SNPs for analysis.

We performed principal component analysis on a subset of 39,353 common (MAF $> 5\%$), perfectly imputed ($\text{info} = 1$) autosomal SNPs, after stringent pruning to remove markers in linkage disequilibrium ($r^2 > 0.1$) and excluding high linkage disequilibrium genomic regions to ensure that only genome-wide effects were detected.

Polygenic Scores

We calculated polygenic scores, which are the SNP effect size weighted sums of the number of trait-associated alleles, based on summary statistics for the largest GWA studies available for key developmental outcomes, including height,⁵⁴ body mass index (BMI),⁵⁴ self-rated health,⁵⁵ intelligence,⁵⁶ educational attainment,⁶ neuroticism,⁵⁷ ADHD,⁵⁸ and schizophrenia.⁵⁹ These GWA studies were selected because their respective GPS yield the highest predictive accuracy within their trait category (details about the studies, reported SNP heritabilities, and GPS predictions can be found in Table S2). To calculate the polygenic scores, we used the software LDpred⁶⁰ which re-weights the SNP effect sizes based on a prior on the effect size and the LD in the sample. Here, we applied a prior on the fraction of causal markers of 1 for all analyses, based on the assumption that all genetic markers contribute to trait development (see Supplemental Material and Methods for details on polygenic score calculation). All polygenic scores were statistically adjusted for the first ten principal components, chip and plate using the regression method, and were z-standardized (mean = 0, SD = 1).

Statistical Analysis

Mixed-Effects Modeling

We applied a random intercept mixed-effects model on DZ data, including two fixed effects to separate the total effect between

the polygenic score predictor and the outcome into within- and between-family effects.⁶¹

$$Y_{ij} = \alpha_0 + \beta_W(GPS_{ij} - \overline{GPS}_j) + \beta_B \overline{GPS}_j + \gamma_j + \varepsilon_{ij}, \quad (\text{Equation 1})$$

where Y denotes the outcome and GPS the polygenic score, $i = \{1,2\}$ corresponds to the individual twins that are clustered within family j , and \overline{GPS} refers to the mean GPS value in family j . The i^{th} value represents birth order, where twin 1 is the elder twin. The notation α_0 represents the intercept and γ_j the random effect with $\gamma_j \sim N(0, \sigma_\gamma^2)$, which corresponds to a change in the intercept for both twins in family j , and ε_{ij} with $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, which denotes the independent random error for each individual i in family j . The between-family effect β_B represents the expected change in the outcome Y given a one unit change in the family GPS average, and the within-family effect β_W represents the expected change given a one unit change in the difference between the individual GPS and the family average GPS. By including both β_W and β_B in the same model, the individual estimates are adjusted for, and independent of, the effect of the other estimate. The random effect term σ_γ^2 , which estimates the difference between each group intercept γ_j and the overall intercept α_0 , accounts for the residual structure in the data corresponding to all unaccounted familial factors (both genetic and environmental) that contribute to the trait similarity of the twins.^{61,62}

The use of a mixed-effects model is only justified if co-twins within a family correlate in the outcome, which can be estimated through the intraclass correlation coefficient (ICC). The ICC is the ratio of the between-family (i.e., random intercept) variance over the total variance and is an estimate of how much of the total variation in the outcome is accounted for by family:

$$\text{Cor}(Y_{1j}, Y_{2j}) = \frac{\sigma_\gamma^2}{(\sigma_\gamma^2 + \sigma_\varepsilon^2)}, \quad (\text{Equation 2})$$

where σ_γ^2 is the covariance between the family variable, in this case family ID, and the outcome, and σ_ε^2 indicates the residual variance capturing within-twin pair differences. The total effect of the relationship between GPS and outcome is the ICC weighted sum of the within- and between-family effects:⁶²

$$\text{Total effect} = \beta_W(1 - \text{ICC}) + \beta_B \text{ICC}. \quad (\text{Equation 3})$$

It follows from Equation 3 that the total effect ranges between β_W and β_B . If the relationship between GPS and outcome is mostly due to individual-level variation, the ICC approximates 0 and the total effect will be close to β_W . In contrast, if the association is mostly due to family effects, the ICC approximates 1 and the total effect will be close to β_B .⁶² To calculate the total effect, we used ICC estimates adjusted for the fixed effects described in Equation 1.

Performing a regression corresponding to Equation 1, we estimated the β_W and β_B parameters using each of the eight polygenic scores in turn as predictors of each of the eight measured outcomes. To estimate potential SES effects, we repeated these analyses including the SES composite as a covariate in the model (these latter analyses were not preregistered). For the fixed effects, we calculated 95% bootstrap percentile intervals. These were based on 10,000 bootstrap samples with random resampling of DZ twin pairs with replacement.

To empirically test the statistical difference between β_W and β_B , we divided the difference between the fixed effect coefficients by the standard deviation of the sampling distribution of the estimate differences.^{63,64} We also applied this approach to statistically test

the significance of the difference between the β_B coefficients before and after the inclusion of family SES in Equation 1. To evaluate the effect size change between the coefficients, we calculated the beta differences with 95% bootstrap percentile intervals, as well as the percentage change (e.g., $((\beta_B - \beta_W)/\beta_B)$).

Quantile Analysis of Within-DZ Pair Differences

To illustrate the extent to which within-DZ pair GPS differences result in differences in developmental outcomes, we performed quantile analysis. First, we generated twin-GPS difference scores by subtracting the twin 2 score from the twin 1 score, and then split this variable into ten equal quantiles based on absolute GPS differences, ranging from the lowest to the highest GPS differences. Birth order did not explain any statistically significant amount of variance (Table S1), so no randomization of twin order was required. We tested mean differences in outcome variables between individuals in the lowest and highest decile. We performed quantile analysis on variables with scales that are easily interpretable: that is, BMI, height, intelligence, and educational achievement. For this purpose, the z-standardized and cleaned variables were transformed back to their original scale, and intelligence values were scaled to have a mean of 100 and a standard deviation of 15.

Multiple Testing Correction

Multiple testing correction of the α significance threshold was performed using the Benjamini Hochberg false discovery rate (FDR) adjustment.⁶⁵ In contrast to more conservative corrections, this method has higher statistical power to detect true positives while controlling for false positives. Based on an α threshold of 0.05, the corrected α in this study was 0.01, defined as the maximum raw p value that is smaller than or equal to the FDR critical value ($p_{\text{raw}} \leq (\text{rank of } p_{\text{raw}} / \text{total number of } p_{\text{raw}} \text{ values}) \times \alpha$).

Sensitivity Analyses

We performed additional, non-pre-registered sensitivity analyses to evaluate the robustness of our findings. Mixed-effects models were run separately for same-sex and for opposite-sex twin pairs (for twin pair N , see Table S3), as well as for twin pairs where both twins were genotyped on the OEE chip, and twin pairs where one twin was genotyped on OEE and the other twin genotyped on Affymetrix (for twin pair N , see Table S4). Analyses were also performed using GPSs that were constructed applying a causal fraction of 0.1.

To control for any unaccounted relatedness between families, we estimated the fixed effects including a SNP-kinship matrix as random effect. Here, Equation 3 becomes $y = \alpha + \beta_W(GPS_{ij} - \overline{GPS}_j) + \beta_B \overline{GPS}_j + g + \varepsilon$, where g is the random effect with $g \sim N(0, A\sigma_g^2)$ and A being a genetic relationship matrix between individuals. A pairwise genetic relationship between individual m and n is estimated as $A_{mn} = 1/N \sum_{i=1}^N (x_{im} - 2p_i)(x_{in} - 2p_i)/2p_i(1 - p_i)$, where N is the number of SNPs, x_{im} is the number of copies of the reference allele for the i^{th} SNP of the m^{th} individual and p_i is the reference allele frequency. These analyses were performed using the GCTA software (v.1.90.0).⁶⁶

Due to the large study population of the UK Biobank, there may be relatedness between this sample and the UK target sample TEDS. The UK Biobank sample was included in the GWA meta-analysis of height, BMI, and educational attainment, and relatedness between discovery and target sample could lead to GPS prediction estimate inflation in the target sample.^{67,68} We therefore calculated an additional set of height and BMI polygenic scores based on GWA meta-analyses published before UK Biobank data became available.^{69,70} We also calculated an additional GPS

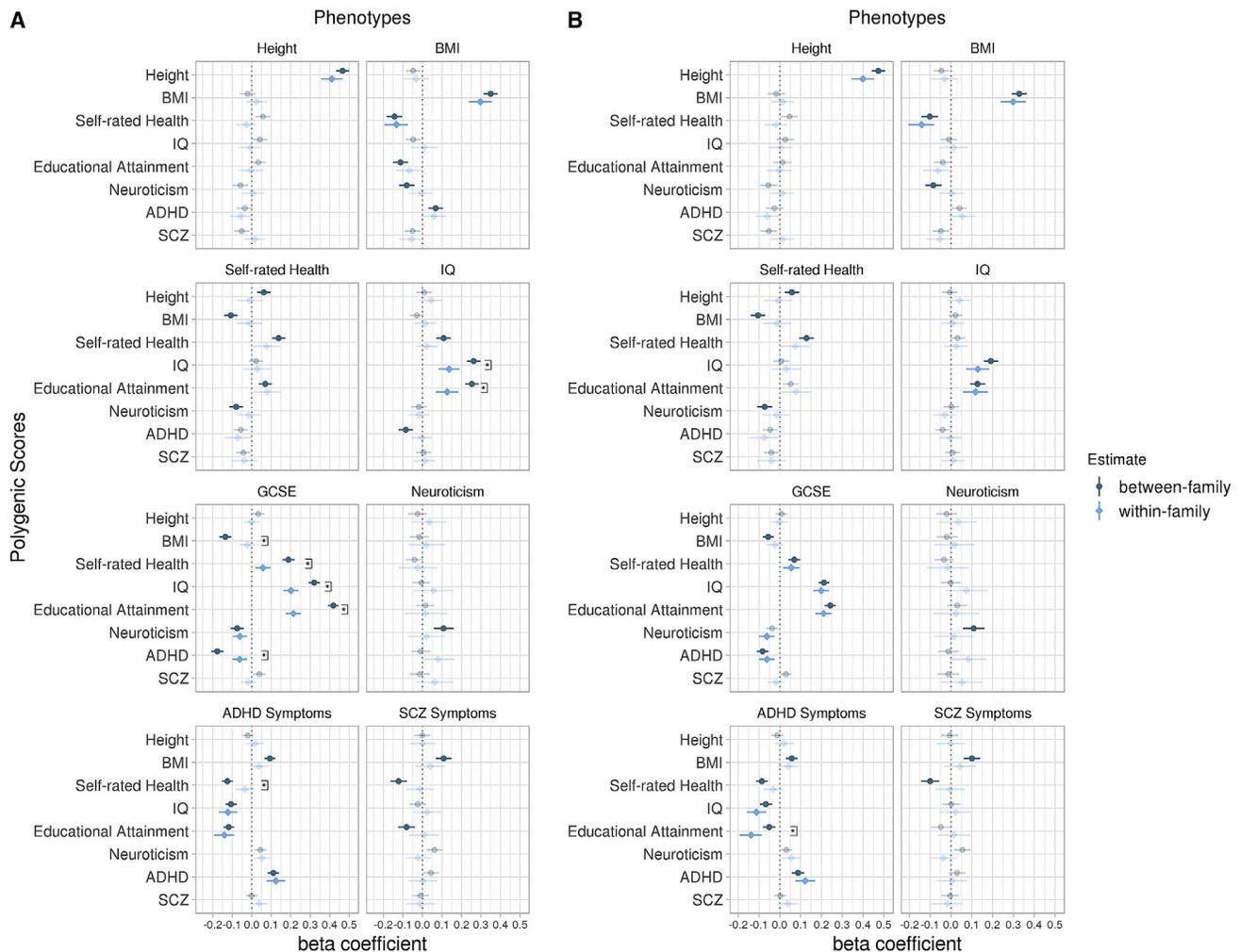


Figure 1. Within- and Between-Family Prediction Estimates of Eight Outcomes using Eight Genome-wide Polygenic Scores

Findings before (A) and after (B) statistical correction for family socio-economic status (SES). The genome-wide polygenic scores (GPS) are presented on the y axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimates. Significant differences are only shown where at least one of the estimates is statistically significant at the false discovery rate corrected threshold of $p < 0.01$ (for all prediction estimates and p values, see [Tables S6 and S7](#)). The dotted line represents a beta coefficient of zero. BMI, body mass index; IQ, intelligence; GCSE, general certificate of secondary education (educational achievement); ADHD, attention-deficit/hyperactivity disorder; SCZ, schizophrenia.

for educational attainment based on a GWA analysis that had all British cohorts removed.⁶ While this rules out discovery and target sample relatedness, it also controls for effect size inflation due to population stratification.

Results

Phenotypic resemblance between DZ twins within a family varied across traits, with Pearson's correlation coefficients ranging from 0.10 to 0.59 ([Figure S1](#), and [Table S5](#) for ICCs). Twins were least alike in their neuroticism levels and self-rated health, and most alike in IQ, height, and educational achievement. Within-twin pair polygenic score correlations were close to expectations (range $r = 0.49$ – 0.57), as the expected shared additive genetic variance between siblings is 50% of the total additive

genetic variance based on quantitative genetic theory.²² Given the 95% confidence intervals of the within-twin pair correlations ([Figure S1](#)), there was a significant difference from the expected correlation coefficient of 0.50 for the self-rated health GPS ($r = 0.53$), the IQ GPS ($r = 0.54$), and the educational attainment GPS ($r = 0.57$), indicating assortative mating for these traits.

Within-Family Polygenic Score Predictions

[Figure 1A](#) depicts the within- and between-family polygenic score prediction estimates of the eight outcomes from the mixed-effects model analyses. Within-family target-trait predictions were statistically significant for height, BMI, intelligence, educational achievement, and ADHD symptoms, indicating that polygenic variation within twin pairs was related to these outcome differences.

Specifically, phenotypic differences in height were significantly positively correlated with height GPS twin differences ($\beta = 0.41$, $p = 5.72e^{-53}$) and differences in BMI were significantly correlated with BMI GPS differences ($\beta = 0.30$, $p = 1.76e^{-21}$) such that twins with a higher height GPS and BMI GPS were taller and heavier than their co-twin, respectively. IQ GPS differences predicted intelligence differences ($\beta = 0.14$, $p = 1.32e^{-6}$) and EA GPS differences were significantly associated with GCSE grade differences ($\beta = 0.21$, $p = 2.22e^{-26}$), indicating that those twins with a higher GPS also scored higher on intelligence measures and in their GCSE tests than their co-twin. For behavior problems, twins with higher ADHD GPS had higher phenotypic ADHD symptoms than their co-twins ($\beta = 0.12$, $p = 1.50e^{-7}$).

We also investigated cross-trait relationships (Figure 1A). For example, self-rated health GPS differences were negatively correlated with differences in BMI, such that twins with a higher self-rated health GPS had a lower BMI ($\beta = -0.13$, $p = 3.56e^{-5}$). EA GPS differences significantly related to phenotypic intelligence differences ($\beta = 0.13$, $p = 2.15e^{-5}$), and IQ GPS predicted GCSE grade differences ($\beta = 0.20$, $p = 7.24e^{-25}$), suggesting that those with higher GPSs also had higher IQ and GCSE grades than their co-twin. GCSE grade differences were also negatively predicted by ADHD GPS twin differences ($\beta = -0.07$, $p = 2.20e^{-4}$), indicating that twins with a higher ADHD GPS obtain lower GCSE results. Notably, IQ GPS differences ($\beta = -0.12$, $p = 6.38e^{-7}$) and EA GPS differences ($\beta = -0.14$, $p = 3.09e^{-8}$) were just as predictive of ADHD symptoms as the ADHD GPS itself, and the direction of effect sizes indicates that the twin with a higher GPS had lower ADHD symptoms than their co-twin (all prediction estimates and total effects are presented in Table S6).

Comparing Within-Family and Between-Family Polygenic Score Prediction

By simultaneously and independently estimating within- and between-family GPS predictions, it was possible to compare these estimates. Between-family estimates (Figure 1A) are mostly consistent with GPS correlations reported for unrelated individuals (Table S2). Figure 1A also shows that between-family associations are generally greater than within-family associations. Significant associations were found for 46.9% of the between-family associations and only 20.3% for within-family associations. On average, magnitudes of within-family associations were almost half (44.1% reduction) that compared to significant between-family estimates (for all prediction estimates, beta difference values and their 95% confidence intervals, and significance of differences, see Table S6).

Notably, significant differences in associations within and between families for polygenic scores predicting their target traits were almost exclusively found for IQ and educational achievement (Figure 1A). The within-family prediction was significantly lower than between-family prediction for both IQ ($p = 6.27e^{-4}$, $\Delta = 48.0\%$) and

GCSE grades ($p = 8.45e^{-14}$, $\Delta = 48.9\%$). Despite not reaching statistical significance, we also observed attenuation of the within-family prediction relative to the between-family prediction for height ($\Delta = 11.8\%$), BMI ($\Delta = 15.1\%$), self-rated health ($\Delta = 45.2\%$), and neuroticism ($\Delta = 80.4\%$).

Also, for cross-trait associations, differences in within- and between-family polygenic score predictions were most pronounced for IQ and educational achievement. For IQ, there were significant differences for the EA GPS ($p = 7.57e^{-4}$, $\Delta = 50.1\%$). For educational achievement, there were significant differences for the BMI GPS ($p = 8.10e^{-5}$, $\Delta = 83.3\%$), the self-rated health GPS ($p = 4.60e^{-6}$, $\Delta = 69.5\%$), the IQ GPS ($p = 1.79e^{-5}$, $\Delta = 37.2\%$), and the ADHD GPS ($p = 4.95e^{-5}$, $\Delta = 65.4\%$). In addition, there was a significant difference in within- and between-family prediction for the self-rated health GPS ($p = 4.00e^{-3}$, $\Delta = 71.7\%$) predicting ADHD symptoms. Although not significant, effect size attenuations were also sizeable for other cross-trait predictions, such as for the neuroticism GPS predicting BMI ($\Delta = 91.2\%$) or the self-rated health GPS predicting schizophrenia symptoms ($\Delta = 90.1\%$) (Table S6). However, for these comparisons, between-family coefficients may not be as reliable as the between-family coefficients that showed a significant difference to their within-family estimate, as estimates were considerably smaller to begin with.

The finding that polygenic score prediction estimates of our measured traits are substantially smaller within families suggests that the corresponding between-family associations are mediated by some combination of family-specific (i.e., shared family) effects, population stratification, and potentially assortative mating. Family SES, which is the same for members of a family, is a predictor not only of educational achievement and IQ but also physical and mental health outcomes. Therefore, we repeated our analyses including family SES as a covariate in the model to interrogate its role in between-family GPS prediction. As noted above, this analysis was not pre-registered. As shown in Figure 1B, between-family predictions were greatly reduced and magnitudes approached those of within-family prediction estimates, which did not change (because any shared family effects are already controlled for in within-family estimates; for all prediction estimates, beta difference values and their 95% confidence intervals, and significance of differences, see Table S7).

Formal testing of the between-family estimate differences before and after correcting for SES indicated significant differences only for cognitive traits (Table S8, Figure S2). For example, there was an average attenuation of 60.9% across the within- and between-family comparisons for the GPSs that showed a statistically significant difference in their prediction of GCSE grades, which was reduced to 25.8% after accounting for SES. Although this is a substantial attenuation, these findings show that family SES does not account for all of the observed differences.

We performed additional contrasts, controlling for the SES components parental education and parental occupation separately in an attempt to identify more specific potential sources of prGE. For GCSE grades and IQ, between-family beta coefficients showed greater attenuation when controlling for parental education in comparison with parental occupation (Tables S9 and S10, Figure S3). However, only for the educational attainment GPS predicting GCSE grades was the difference between the attenuation due to parental education ($\Delta = 18.9\%$) and parental occupation ($\Delta = 37.7\%$) statistically significant ($p = 6.40e^{-3}$) (Table S11), indicating that parental education may present a stronger prGE effect.

As a further set of analyses, we applied a multiple regression approach to predict family SES using the within- and between-family estimates of the eight GPSs. Family SES acts as a control trait as there should be no direct genetic effects from the offspring to family SES, as indicated by the within-family effect. Results confirmed that all within-family beta coefficients were zero, while between-family estimates were related to family SES (Table S12).

Sensitivity analyses (not pre-registered) were performed by repeating all analyses separately for same-sex and opposite-sex twins (Tables S3, S13, S14, S15, and S16, and Figures S4 and S5), and for twin pairs grouped by genotyping chip (Tables S4, S17, S18, S19, and S20, and Figures S6 and S7). In addition, we estimated the fixed effects using a SNP-kinship matrix as a random effect to control for any unaccounted between-family relatedness (Table S21 and Figure S8). For the different sets of sensitivity analyses described, no substantial deviations from the results using the combined sample were found.

We also repeated analyses using GPSs that were calculated based on a fraction of causal markers of 0.1 (Tables S22 and S23, and Figure S9), and using GPSs that had the UK Biobank sample (height; BMI) or all British samples (educational attainment) removed at the GWA analysis stage (Table S24 and Figure S10). Although prediction estimates were smaller in some cases likely due to reduced power, the pattern of within- versus between-family effect size changes remained unchanged as indicated by the mostly overlapping 95% confidence intervals of the beta difference values.

Quantile Analysis

To illustrate within-family differences further, quantile analysis demonstrated how within-family polygenic score differences related to differences in height, BMI, IQ, and GCSE grades (Figure 2). There was an 8.7 cm height mean difference ($p = 1.28e^{-11}$) between the lowest absolute difference decile versus the highest difference decile. For BMI, the difference was 2.9 BMI points ($p = 8.33e^{-6}$) between the lowest and the highest absolute GPS difference deciles. Mean GCSE grade differences (0.40) were also statistically significant ($p = 7.13e^{-5}$) when comparing the lowest and the highest absolute GPS difference deciles. In contrast, IQ point differences

(1.9 points) were not statistically different ($p = 0.26$) between the lowest and the highest absolute GPS difference quantiles (for trait and GPS means at each difference decile, see Table S25).

Discussion

Polygenic score prediction of complex traits is now a common approach in genomics research, but the potential pathways by which polygenic score variation predicts phenotypic variation remain largely unexplored. In this study, we contrasted within- and between-family polygenic prediction estimates to quantify the extent to which environmentally mediated genetic effects (i.e., passive genotype-environment correlation) are picked up in polygenic score analyses. By systematically performing target- and cross-trait analyses across eight life outcomes using eight corresponding GPS, we found evidence that prGE might be a mechanism explaining a considerable proportion of the GPS prediction in cognitive traits (intelligence and educational achievement), but not as much for non-cognitive traits. We also found that for between-family GPS predictions of cognitive traits—but, again, not as much for other traits—family SES is likely to be the major source of prGE.

For the prediction of IQ and educational achievement, within-family estimates were on average 60% smaller than between-family estimates. The within- versus between-family attenuation for the EA GPS prediction was 49%, which is close to the 40% estimate in GWA study effect sizes for years of education⁶ and the 55% estimate using the same EA GPS in a different target sample.²⁹ These findings highlight the influence of prGE in the development of IQ and educational achievement and demonstrate the extent to which between-family GPS prediction may be partly driven by prGE effects. Results from our study are also in line with adoption studies showing evidence of between-family prGE in that correlations between home environment and children's IQ is twice as great in non-adoptive families than in adoptive families.⁷¹ Our findings are compatible with recent research on *genetic nurture*, using non-transmitted alleles from parental genotypes to assess prGE^{20,21} in terms of GPS target trait prediction of educational achievement and anthropometric traits. Our findings also extend to cross-trait associations using a wide range of GPSs. Contrary to our prediction that within- and between-family EA GPS associations would be significantly different across many associated life outcomes, results from cross-trait analysis suggest that within- and between-family predictions were only significantly different across a range of GPS for the prediction of cognitive outcomes.

A possible explanation for these results is that IQ and educational achievement show more shared environmental influences (24% and 27%, respectively) relative to other traits used in this study such as height (10%), BMI (10%), ADHD (2%), or schizophrenia (0%), as estimated

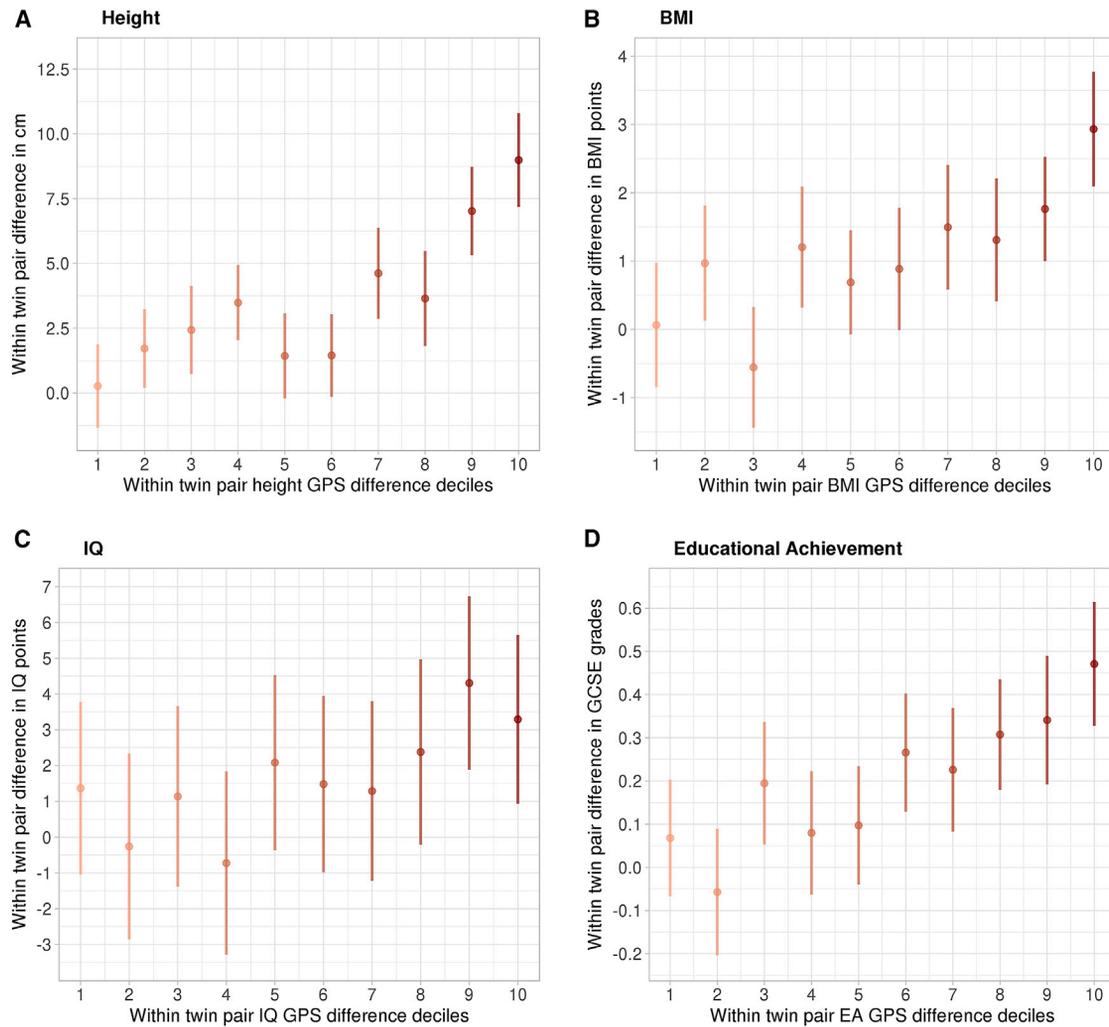


Figure 2. The Relationship between Absolute Dizygotic (DZ) Twin Pair Polygenic Score Decile Differences and Trait Outcome Differences

Lower deciles represent small absolute genome-wide polygenic score (GPS) differences and higher deciles represent large GPS differences between DZ co-twins. Error bars indicate 95% confidence intervals. Each GPS decile included the following numbers of twin pairs: height = 146; BMI = 135; IQ = 157; GCSE = 236. Regression through origin analysis (fixed intercept of zero) using the continuous GPS difference values to predict outcome differences were significant for height ($B = 4.42$, $p = 3.73e^{-53}$, $R^2 = 0.148$), BMI ($B = 1.34$, $p = 1.73e^{-21}$, $R^2 = 0.064$), IQ ($B = 2.1$, $p = 4.53e^{-7}$, $R^2 = 0.015$), and GCSE grades ($B = 0.26$, $p = 3.04e^{-26}$, $R^2 = 0.046$).

through a large twin study meta-analysis.⁷² The type of rGE that we assessed in this study—defined as the exposure to a family environment that is correlated with both parental and offspring genotypes and which contributes to sibling similarity in their outcomes—is absorbed by the shared environment variance component (C) in classical twin analyses.⁷³ Therefore, it may be more likely that genetic effects related to cognitive traits as estimated through GWA studies partly contain prGE effects—in contrast to other traits tested in our study—because the shared environmental component is larger to begin with for cognitive traits. In TEDS, the C component for the same IQ and educational achievement measures used in this study were estimated around 35%⁷⁴ and 29%,⁷⁵ respectively.

As known from the existing literature, family SES is strongly genetically correlated with offspring cognitive

traits,^{8,37,38} rendering it a likely source of prGE. Indeed, our results showed that between-family effects were considerably more similar in magnitude to within-family effects when holding SES constant, suggesting that SES is a source of the majority of the within-between discrepancy, rather than residual population stratification or assortative mating. When controlling for parental education and parental occupation separately, we found that between-family effect sizes were closer to within-family coefficients for parental education than for occupation. However, this difference was significant only for the educational attainment GPS predicting GCSE grades, suggesting that parental education is likely a stronger source of prGE than parental occupation influencing offspring educational achievement.

Despite the sizeable attenuations after controlling for family SES, we still observed some effect size differences

when comparing within- and between-family coefficients. For example, there was still a 32.5% difference for the IQ GPS predicting IQ and a 13.4% difference for the educational attainment GPS predicting GCSE scores. The within-twin pair correlations for these GPS indicated assortative mating, which could explain some of this remaining discrepancy. Indeed, previous research on genetic nurture indicated that a small proportion of the direct genetic effect of the educational attainment polygenic score predicting educational attainment captures assortative mating-related effects.²⁰ The same research also showed genetic nurture effects between siblings using the educational attainment GPS.²⁰ Such effects may further contribute to the within-family effect attenuation, potentially accounting for some of the residual difference after controlling for family SES.

The results showed that more distantly related GPS captured considerable prGE effects in cross-trait GPS predictions of cognitive traits. For instance, within-family effect sizes for the ADHD GPS predicting educational achievement were significantly smaller (65% reduction), in contrast to the ADHD GPS predicting ADHD symptoms, where no difference was detected. This suggests that the GWAS for ADHD captures genetic variation that is correlated with aspects of the family environment that contribute to the co-development of ADHD symptoms and educational achievement, although it is unclear why these effects do not appear to contribute to the development of ADHD symptoms themselves.

It is important to go beyond GPS predictions of traits in unrelated individuals to consider prGE mechanisms by comparing within- and between-family predictions in order to explain the sources of predictions in polygenic score analysis. However, finding between-family prGE does not diminish the usefulness of GPS predictions for cognitive traits in unrelated individuals, because these prGE effects help maximize the prediction of trait variance. Although within-family genetic effects do not include prGE effects due to between-family factors such as SES, within-family genetic effects are not free of *all* kinds of rGE, as demonstrated by twin studies showing that correlations between putative measures of the environment and children's specific outcomes are genetically influenced.⁷¹ Within-family GPS prediction estimates can be interpreted as direct genetic effects in the sense that they stem from the individual level and not the family level. Children select, modify, and create experiences (active rGE) or evoke responses in their environment (evocative rGE) that are correlated with their genetic propensities. Therefore, within-family genetic differences can relate to trait differences through active or evocative rGE pathways but are free of any passive rGE effects.

Implications

The results from this study have three important implications for the interpretation of the existing polygenic score literature as well as for future genetic research. First, the

finding that between-family predictions pick up effects due to prGE mostly and substantially in cognitive traits is informative for causal inference studies that use designs such as Mendelian randomization.^{76,77} Here, a genetic instrument that is related to a predictor (in the form of a single genetic marker or GPS) is used to assess the causal relationship between the predictor and an outcome. At a population level, genotypes are not inherited randomly: individuals with particular genotypes are not born into environmental conditions at chance. If family environment is associated with the genetic instrument as well as the predictor and the outcome, this opens a backdoor path whereby predictor and outcome are related through the prGE mechanisms.¹⁹ This could lead to an assumption violation, therefore biasing causal inference in between-family analysis. Only in a within-family design is it ensured that Mendelian randomization meets its assumptions because transmission of alleles is randomized at meiosis within families, and because prGE effects due to shared environment are held constant.^{19,26,78,79} Although genetic data for siblings are often not available, our results provide a useful guideline for the GPS-outcome combinations that are unlikely to suffer from this assumption violation when applying designs such as Mendelian randomization to unrelated samples. For example, our results indicate that caution should be warranted due to prGE mechanisms if applying Mendelian randomization to cognitive traits, even if family SES is included as a confounder in the analyses as confounding effects might not be captured perfectly. In contrast, other traits such as BMI and ADHD (with the possible exclusion of the self-rated health GPS) should be less problematic, because within- and between-family effect sizes match closely, ruling out potential confounding due to prGE.

Second, our results provide evidence that location-related population stratification is not a large bias in GPS prediction of complex traits when controlling for genetic principal components in samples from white European backgrounds. As it has been shown that the GPS prediction of height is affected by population stratification,⁸⁰ we also find an attenuation of around 12% of the within-family coefficient, which is by necessity free of population stratification since stratification is constant within a family. When we performed our analyses using a GPS for height based on a discovery sample that did not include UK Biobank, the attenuation decreased to 5%. This may indicate that the inclusion of a large discovery sample genetically similar to the target sample could have resulted in a GPS that is more strongly confounded by population stratification—although it is noteworthy that the 95% confidence intervals of the beta difference values overlap for the two height GPSs. For those traits where within- and between-family estimate differences were large and significant, differences were greatly reduced after accounting for SES, indicating that SES was the main source of the discrepancy, as opposed to location-related population stratification. Our additional analyses using a GPS based on GWA analysis

that had all British samples removed did not show less attenuation, which would be expected if population stratification strongly influenced GPS prediction.

Third, our study illustrates the usefulness of obtaining genotypic data on family members, since it makes it possible to identify mechanisms of polygenic prediction. Our results demonstrate that by analyzing DZ co-twins' genetic data jointly, prGE mechanisms due to shared environment (and in this case associated with SES) can be revealed.

Limitations

There are some limitations to this study. The GWA studies used to generate the eight GPS for this study had different statistical power to discover genetic effect sizes due to sample size variations and different underlying genetic architectures of the GWA study traits. As a result, each of the eight GPSs were differently powered to detect target- and cross-trait associations, making it difficult to draw direct comparisons across the within- and between-family prediction effect sizes. Lack of power may also lead to an inability to detect small prGE effects that would become visible with (1) more powerful GPS and (2) the availability of larger DZ twin pair samples. However, we detected prGE effects in cross-trait analysis using the ADHD GPS, which is based on the smallest GWAS study sample (~55,000 individuals), indicating that we had sufficient power to detect at least some of the prGE effects.

It is also possible that some important within- and between-family effect differences did not reach statistical significance due to insufficient statistical power. While the effect size differences in cognitive traits are large, it may be that effects due to prGE, population stratification and/or assortative mating are more subtle in other traits. Therefore, our study sample, which ranged between 789 and 2,469 DZ twin pairs, may have not had enough power to establish the statistical significance of small effect size differences. Notably, GPS predictions were generally small where no significant difference was found between large within- versus between-family effect size attenuations. With the availability of more powerful GPS in the future, it may be possible to detect such differences statistically.

Another limitation was that we did not have parental genotypes available to directly test the influence of non-transmitted parental alleles on offspring outcomes (genetic nurture).²⁰ Although the within-family design used in this study accounts for the effects of both transmitted and non-transmitted parental alleles on offspring outcomes, it is not possible to disentangle these two sources of prGE. Future studies would benefit from incorporating parental and sibling genotypes to disentangle the prGE effects through the joint analysis of parental and sibling genotypes, which will shed light on how both non-transmitted parental and non-co-inherited sibling alleles contribute to trait development.

Conclusion

This study provided strong evidence for prGE mechanisms in polygenic score prediction mainly for cognitive traits

across a range of different polygenic scores. The implications of these findings for future studies depend on their aims. If maximizing trait prediction is the goal, the use of unrelated samples is valid even in the presence of prGE effects because these influences are informative nonetheless. However, if the goal is causal inference and explanation, a within-family genetic design is recommended to avoid prGE-related confounding. The increasing availability of genotypic data in relatives will become a crucial element in genetics research, allowing researchers to disentangle the mechanisms of polygenic prediction of complex human traits.

Supplemental Data

Supplemental Data can be found online at <https://doi.org/10.1016/j.ajhg.2019.06.006>.

Acknowledgments

We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a program grant to R.P. from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938). The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/grant agreement no. 602768 and ERC grant agreement no. 295366. R.P. is supported by a Medical Research Council Professorship award (G19/2). S.S. is supported by the MRC/IoPPN Excellence Award and by the US National Institutes of Health (AG046938). P.F.O. received funding from the UK Medical Research Council (MR/N015746/1). High-performance computing facilities were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980).

Declaration of Interests

The authors declare no competing interests.

Received: April 11, 2019

Accepted: June 6, 2019

Published: July 11, 2019

Web Resources

OSF pre-registration link, https://osf.io/eq8ga/?view_only=768f42366d134eb50f5999763c3fce

References

1. Gratten, J., Wray, N.R., Keller, M.C., and Visscher, P.M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat. Neurosci.* *17*, 782–790.
2. Plomin, R., and von Stumm, S. (2018). The new genetics of intelligence. *Nat. Rev. Genet.* *19*, 148–159.
3. Martin, A.R., Daly, M.J., Robinson, E.B., Hyman, S.E., and Neale, B.M. (2018). Predicting polygenic risk of psychiatric disorders. *Biol. Psychiatry*, S0006-3223(18)32119-X.

4. Wray, N.R., Lee, S.H., Mehta, D., Vinkhuyzen, A.A.E., Dudbridge, F., and Middeldorp, C.M. (2014). Research review: Polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry* *55*, 1068–1087.
5. Plomin, R. (2018). *Blueprint* (Penguin UK).
6. Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., et al.; 23andMe Research Team; COGENT (Cognitive Genomics Consortium); and Social Science Genetic Association Consortium (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* *50*, 1112–1121.
7. Allegrini, A.G., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J.B., and Plomin, R. (2019). Genomic prediction of cognitive traits in childhood and adolescence. *Mol. Psychiatry* *24*, 819–827.
8. Hill, W.D., Hagenaars, S.P., Marioni, R.E., Harris, S.E., Liewald, D.C.M., Davies, G., Okbay, A., McIntosh, A.M., Gale, C.R., and Deary, I.J. (2016). Molecular Genetic Contributions to Social Deprivation and Household Income in UK Biobank. *Curr. Biol.* *26*, 3083–3089.
9. Selzam, S., Krapohl, E., von Stumm, S., O'Reilly, P.F., Rimfeld, K., Kovas, Y., Dale, P.S., Lee, J.J., and Plomin, R. (2017). Predicting educational achievement from DNA. *Mol. Psychiatry* *22*, 267–272.
10. Belsky, D.W., Domingue, B.W., Wedow, R., Arseneault, L., Boardman, J.D., Caspi, A., Conley, D., Fletcher, J.M., Freese, J., Herd, P., et al. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proc. Natl. Acad. Sci. USA* *115*, E7275–E7284.
11. Belsky, D.W., Moffitt, T.E., Corcoran, D.L., Domingue, B., Harrington, H., Hogan, S., Houts, R., Ramrakha, S., Sugden, K., Williams, B.S., et al. (2016). The Genetics of Success: How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development. *Psychol. Sci.* *27*, 957–972.
12. de Zeeuw, E.L., van Beijsterveldt, C.E.M., Glasner, T.J., Bartels, M., Ehli, E.A., Davies, G.E., Hudziak, J.J., Rietveld, C.A., Groen-Blokhuys, M.M., Hottenga, J.-J., et al.; Social Science Genetic Association Consortium (2014). Polygenic scores associated with educational attainment in adults predict educational achievement and ADHD symptoms in children. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* *165B*, 510–520.
13. Hagenaars, S.P., Harris, S.E., Davies, G., Hill, W.D., Liewald, D.C.M., Ritchie, S.J., Marioni, R.E., Fawns-Ritchie, C., Cullen, B., Malik, R., et al.; METASTROKE Consortium, International Consortium for Blood Pressure GWAS; SpiroMeta Consortium; and CHARGE Consortium Pulmonary Group, CHARGE Consortium Aging and Longevity Group (2016). Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112c151) and 24 GWAS consortia. *Mol. Psychiatry* *21*, 1624–1632.
14. Möttus, R., Realo, A., Vainik, U., Allik, J., and Esko, T. (2017). Educational Attainment and Personality Are Genetically Intertwined. *Psychol. Sci.* *28*, 1631–1639.
15. Smith-Woolley, E., Selzam, S., and Plomin, R. (2019). Polygenic score for educational attainment captures DNA variants shared between personality traits and educational achievement. *J. Pers. Soc. Psychol.*
16. Krapohl, E., Euesden, J., Zabaneh, D., Pingault, J.-B., Rimfeld, K., von Stumm, S., Dale, P.S., Breen, G., O'Reilly, P.F., and Plomin, R. (2016). Phenome-wide analysis of genome-wide polygenic scores. *Mol. Psychiatry* *21*, 1188–1193.
17. Belsky, D.W., and Harden, K.P. (2019). Phenotypic Annotation: Using Polygenic Scores to Translate Discoveries From Genome-Wide Association Studies From the Top Down. *Curr. Dir. Psychol. Sci.* *28*, 82–90.
18. Plomin, R., DeFries, J.C., and Loehlin, J.C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol. Bull.* *84*, 309–322.
19. Pingault, J.-B., O'Reilly, P.F., Schoeler, T., Ploubidis, G.B., Rijdsdijk, F., and Dudbridge, F. (2018). Using genetic data to strengthen causal inference in observational research. *Nat. Rev. Genet.* *19*, 566–580.
20. Kong, A., Thorleifsson, G., Frigge, M.L., Vilhjalmsdottir, B.J., Young, A.I., Thorgeirsson, T.E., Benonisdottir, S., Oddsson, A., Halldorsson, B.V., Masson, G., et al. (2018). The nature of nurture: Effects of parental genotypes. *Science* *359*, 424–428.
21. Bates, T.C., Maher, B.S., Medland, S.E., McAloney, K., Wright, M.J., Hansell, N.K., Kendler, K.S., Martin, N.G., and Gillespie, N.A. (2018). The Nature of Nurture: Using a Virtual-Parent Design to Test Parenting Effects on Children's Educational Attainment in Genotyped Families. *Twin Res. Hum. Genet.* *21*, 73–83.
22. Fisher, R.A. (1918). The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans. R. Soc. Edinb.* *52*, 399–433.
23. Fletcher, J.M. (2011). The promise and pitfalls of combining genetic and economic research. *Health Econ.* *20*, 889–892.
24. Young, A.I., Frigge, M.L., Gudbjartsson, D.F., Thorleifsson, G., Bjornsdottir, G., Sulem, P., Masson, G., Thorsteinsdottir, U., Stefansson, K., and Kong, A. (2018). Relatedness disequilibrium regression estimates heritability without environmental bias. *Nat. Genet.* *50*, 1304–1310.
25. Benyamin, B., Visscher, P.M., and McRae, A.F. (2009). Family-based genome-wide association studies. *Pharmacogenomics* *10*, 181–190.
26. Brumpton, B., Sanderson, E., Pires Hartwig, F., Harrison, S., Abergé Vie, G., Cho, Y., Hughes, A., Boomsma, D., Havdahl, A., Hopper, J., et al. (2019). Within-family studies for Mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases. *bioRxiv*. <https://doi.org/10.1101/602516>.
27. Domingue, B.W., Belsky, D., Conley, D., Harris, K.M., and Boardman, J.D. (2015). Polygenic Influence on Educational Attainment: New evidence from The National Longitudinal Study of Adolescent to Adult Health. *AERA Open* *1*, 1–13.
28. Rietveld, C.A., Esko, T., Davies, G., Pers, T.H., Turley, P., Benyamin, B., Chabris, C.F., Emilsson, V., Johnson, A.D., Lee, J.J., et al. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl. Acad. Sci. USA* *111*, 13790–13794.
29. Trejo, S., and Domingue, B.W. (2019). Genetic Nature or Genetic Nurture? Quantifying Bias in Analyses Using Polygenic Scores: Appendix. *bioRxiv*. <https://doi.org/10.1101/524850>.
30. Spinath, B., Spinath, F.M., Harlaar, N., and Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence* *34*, 363–374.
31. Krapohl, E., Rimfeld, K., Shakeshaft, N.G., Trzaskowski, M., McMillan, A., Pingault, J.-B., Asbury, K., Harlaar, N., Kovas, Y., Dale, P.S., and Plomin, R. (2014). The high heritability of educational achievement reflects many genetically influenced

- traits, not just intelligence. *Proc. Natl. Acad. Sci. USA* *111*, 15273–15278.
32. Briley, D.A., Domiteaux, M., and Tucker-Drob, E.M. (2014). Achievement-Relevant Personality: Relations with the Big Five and Validation of an Efficient Instrument. *Learn. Individ. Differ.* *32*, 26–39.
33. Marques, S.C., Pais-Ribeiro, J.L., and Lopez, S.J. (2011). The Role of Positive Psychology Constructs in Predicting Mental Health and Academic Achievement in Children and Adolescents: A Two-Year Longitudinal Study. *J. Happiness Stud.* *12*, 1049–1062.
34. Pingault, J.-B., Tremblay, R.E., Vitaro, F., Carbonneau, R., Genolini, C., Falissard, B., and Côté, S.M. (2011). Childhood trajectories of inattention and hyperactivity and prediction of educational attainment in early adulthood: a 16-year longitudinal population-based study. *Am. J. Psychiatry* *168*, 1164–1170.
35. De Ridder, K.A.A., Pape, K., Johnsen, R., Holmen, T.L., Westin, S., and Bjørngaard, J.H. (2013). Adolescent health and high school dropout: a prospective cohort study of 9000 Norwegian adolescents (the Young-HUNT). *PLoS ONE* *8*, e74954.
36. Zuffianò, A., Alessandri, G., Gerbino, M., Luengo Kanacri, B.P., Di Giunta, L., Milioni, M., and Caprara, G.V. (2013). Academic achievement: The unique contribution of self-efficacy beliefs in self-regulated learning beyond intelligence, personality traits, and self-esteem. *Learn. Individ. Differ.* *23*, 158–162.
37. Krapohl, E., and Plomin, R. (2016). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Mol. Psychiatry* *21*, 437–443.
38. Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., Dale, P.S., and Plomin, R. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence* *42*, 83–88.
39. Koellinger, P.D., and Harden, K.P. (2018). Using nature to understand nurture. *Science* *359*, 386–387.
40. Haworth, C.M.A., Davis, O.S.P., and Plomin, R. (2013). Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Res. Hum. Genet.* *16*, 117–125.
41. Ware, J.E., Jr., and Sherbourne, C.D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med. Care* *30*, 473–483.
42. Raven, J., and Court, J.H. (1996). *Manual for Raven's Progressive Matrices and Vocabulary Scales* (Oxford: Oxford University Press).
43. Wechsler, D. (1992). *Wechsler Intelligence Scale for Children - Third Edition UK (WISC-III-UK) Manual* (London: The Psychological Corporation).
44. Kaplan, E., Fein, D., Kramer, J., Delis, D., and Morris, R. (1999). *WISC-III as a Process Instrument (WISC-III-PI)* (New York: The Psychological Corporation).
45. Mullins-Sweatt, S.N., Jamerson, J.E., Samuel, D.B., Olson, D.R., and Widiger, T.A. (2006). Psychometric properties of an abbreviated instrument of the five-factor model. *Assessment* *13*, 119–137.
46. Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *J. Child Psychol. Psychiatry* *38*, 581–586.
47. Conners, C.K. (2003). *Conners' Rating Scales-Revised: Technical Manual* (Multi-Health Systems).
48. Rietz, Du, E., Cheung, C.H.M., McLoughlin, G., Brandeis, D., Banaschewski, T., Asherson, P., and Kuntsi, J. (2016). Self-report of ADHD shows limited agreement with objective markers of persistence and remittance. *J. Psychiatr. Res.* *82*, 91–99.
49. Fenigstein, A., and Vanable, P.A. (1992). Paranoia and self-consciousness. *J. Pers. Soc. Psychol.* *62*, 129–138.
50. Bell, V., Halligan, P.W., and Ellis, H.D. (2006). The Cardiff Anomalous Perceptions Scale (CAPS): a new validated measure of anomalous perceptual experience. *Schizophr. Bull.* *32*, 366–377.
51. Andreasen, N.C. (1989). Scale for the Assessment of Negative Symptoms (SANS): conceptual and theoretical foundations. *British J. Psychiatry* *155*, 49–52.
52. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al.; Haplotype Reference Consortium (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* *48*, 1279–1283.
53. Selzam, S., McAdams, T.A., Coleman, J.R.I., Carnell, S., O'Reilly, P.F., Plomin, R., and Llewellyn, C.H. (2018). Evidence for gene-environment correlation in child feeding: Links between common genetic variation for BMI in children and parental feeding practices. *PLoS Genet.* *14*, e1007757.
54. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M.; and GIANT Consortium (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* *27*, 3641–3649.
55. McInnes, G., Tanigawa, Y., DeBoever, C., Lavertu, A., Olivieri, J.E., Aguirre, M., and Rivas, M.A. (2018). Global Biobank Engine: enabling genotype-phenotype browsing for biobank summary statistics. *Bioinformatics* *9*, 1612.
56. Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C.A., Nagel, M., Awasthi, S., Barr, P.B., Coleman, J.R.I., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* *50*, 912–919.
57. Luciano, M., Hagenaars, S.P., Davies, G., Hill, W.D., Clarke, T.-K., Shireli, M., Harris, S.E., Marioni, R.E., Liewald, D.C., Fawns-Ritchie, C., et al. (2018). Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat. Genet.* *50*, 6–11.
58. Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., et al.; ADHD Working Group of the Psychiatric Genomics Consortium (PGC); Early Life-course & Genetic Epidemiology (EAGLE) Consortium; and 23andMe Research Team (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* *51*, 63–75.
59. Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., et al.; GERAD1 Consortium; and CRESTAR Consortium (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* *50*, 381–389.
60. Vilhjálmsson, B.J., Yang, J., Finucane, H.K., Gusev, A., Lindström, S., Ripke, S., Genovese, G., Loh, P.-R., Bhatia, G., Do, R., et al.; Schizophrenia Working Group of the Psychiatric Genomics Consortium, Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE) study (2015). Modeling

- Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* 97, 576–592.
61. Carlin, J.B., Gurrin, L.C., Sterne, J.A., Morley, R., and Dwyer, T. (2005). Regression models for twin studies: a critical review. *Int. J. Epidemiol.* 34, 1089–1099.
 62. Genser, B., Teles, C.A., Barreto, M.L., and Fischer, J.E. (2015). Within- and between-group regression for improving the robustness of causal claims in cross-sectional analysis. *Environ. Health* 14, 60.
 63. Paternoster, R., Brame, R., Mazerolle, P., and Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology* 36, 859–866.
 64. Clogg, C.C., Petkova, E., and Haritou, A. (1995). Statistical Methods for Comparing Regression Coefficients Between Models. *Am. J. Sociol.* 100, 1261–1293.
 65. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57, 289–300.
 66. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82.
 67. Choi, S.W., Mak, T.S.H., and O'Reilly, P. (2018). A guide to performing Polygenic Risk Score analyses. *bioRxiv*. <https://doi.org/10.1101/416545>.
 68. Wray, N.R., Yang, J., Hayes, B.J., Price, A.L., Goddard, M.E., and Visscher, P.M. (2013). Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* 14, 507–515.
 69. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al.; Electronic Medical Records and Genomics (eMEMERGE) Consortium; MIGen Consortium; PAGEGE Consortium; and LifeLines Cohort Study (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46, 1173–1186.
 70. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al.; LifeLines Cohort Study; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investi-
gators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; and International Endogene Consortium (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518, 197–206.
 71. Plomin, R. (1994). *Genetics and experience: The interplay between nature and nurture* (Sage Publications, Inc).
 72. Polderman, T.J.C., Benyamin, B., de Leeuw, C.A., Sullivan, P.F., van Bochoven, A., Visscher, P.M., and Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* 47, 702–709.
 73. Rijdsdijk, F.V., and Sham, P.C. (2002). Analytic approaches to twin data using structural equation models. *Brief. Bioinform.* 3, 119–133.
 74. Kovas, Y., Voronin, I., Kaydalov, A., Malykh, S.B., Dale, P.S., and Plomin, R. (2013). Literacy and numeracy are more heritable than intelligence in primary school. *Psychol. Sci.* 24, 2048–2056.
 75. Shakeshaft, N.G., Trzaskowski, M., McMillan, A., Rimfeld, K., Krapohl, E., Haworth, C.M.A., Dale, P.S., and Plomin, R. (2013). Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *PLoS ONE* 8, e80341.
 76. Davey Smith, G., and Ebrahim, S. (2005). What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ* 330, 1076–1079.
 77. Davey Smith, G., and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* 23 (R1), R89–R98.
 78. Smith, G.D., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* 32, 1–22.
 79. Davey Smith, G. (2007). Capitalizing on Mendelian randomization to assess the effects of treatments. *J. R. Soc. Med.* 100, 432–435.
 80. Berg, J.J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A.M., Mostafavi, H., Field, Y., Boyle, E.A., Zhang, X., Racimo, F., Pritchard, J.K., and Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *eLife* 8, 68.

The American Journal of Human Genetics, Volume 105

Supplemental Data

Comparing Within- and Between-Family

Polygenic Score Prediction

Saskia Selzam, Stuart J. Ritchie, Jean-Baptiste Pingault, Chandra A. Reynolds, Paul F. O'Reilly, and Robert Plomin

Supplementary Figures

Figure S1. Within-twin pair Pearson's correlation coefficients.

Figure S2. Effect sizes and significance of differences comparing between-family coefficients before and after controlling for family socio-economic status

Figure S3. Within- and between-family prediction estimates accounting for separate measures of family socio-economic status

Figure S4. Within- and between-family prediction estimates for same-sex twin pairs

Figure S5. Within- and between-family prediction estimates for opposite-sex twin pairs

Figure S6. Within- and between-family prediction estimates based on twin pairs genotyped on OEE chip

Figure S7. Within- and between-family prediction estimates based on twin pairs genotyped on OEE chip and the co-twin genotyped on the Affymetrix chip

Figure S8. Within- and between-family coefficients estimated using a SNP-kinship matrix as random effect

Figure S9. Within- and between-family prediction estimates using polygenic scores calculated with causal fraction of markers of 0.1

Figure S10. Within- and between-family prediction estimates based on polygenic scores with UK Biobank or all UK discovery samples removed

Supplementary Tables

Table S1. Descriptive statistics, age and sex effects for phenotypes

Table S2. GWAS used for polygenic score calculation

Table S3. Intraclass coefficients for same-sex and opposite-sex twin pairs

Table S4. Intraclass coefficients for twin pairs split by genotyping chip

Table S5. Intraclass coefficients

Table S12. Within- and between-family prediction estimates of family socio-economic status

Table S24. Within- and between-family prediction estimates based on polygenic scores with UK Biobank or all UK discovery samples removed

Table S25 Phenotypic and polygenic score mean differences by polygenic score difference quantiles

Supplementary Methods

Methods S1. Polygenic score calculation

References

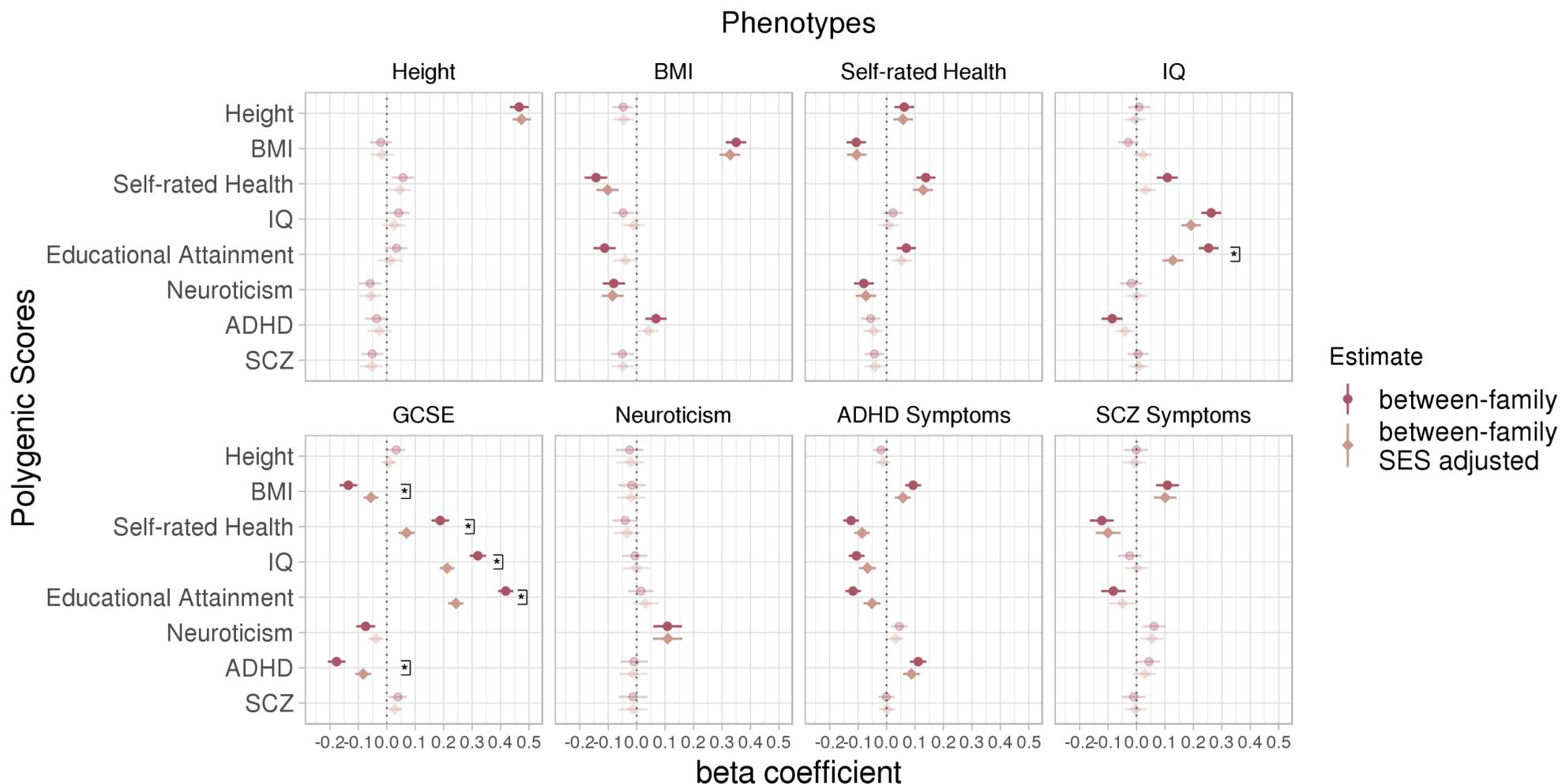
Supplementary Figures

Figure S1. Within-twin pair Pearson's correlation coefficients.

SCZ symptoms 1	0.02 [-0.04, 0.07]	0.07 [0.01, 0.12]	-0.11 [-0.16, -0.06]	0.00 [-0.05, 0.06]	-0.08 [-0.13, -0.03]	0.04 [-0.01, 0.09]	0.01 [-0.04, 0.06]	0.03 [-0.02, 0.08]	-0.04 [-0.1, 0.02]	0.10 [0.04, 0.15]	-0.14 [-0.2, -0.09]	-0.06 [-0.12, -0.01]	-0.12 [-0.18, -0.07]	0.02 [-0.07, 0.1]	0.20 [0.15, 0.26]	0.25 [0.2, 0.3]
ADHD symptoms 1	-0.03 [-0.07, 0.01]	0.06 [0.02, 0.1]	-0.10 [-0.14, -0.06]	-0.06 [-0.1, -0.02]	-0.05 [-0.08, -0.01]	0.03 [-0.01, 0.07]	0.07 [0.03, 0.11]	0.00 [-0.04, 0.04]	-0.05 [-0.1, 0]	0.05 [0, 0.11]	-0.13 [-0.18, -0.08]	-0.11 [-0.16, -0.06]	-0.13 [-0.17, -0.09]	0.03 [-0.04, 0.1]	0.30 [0.26, 0.34]	***
Neuroticism 1	0.00 [-0.07, 0.06]	-0.02 [-0.09, 0.04]	0.00 [-0.07, 0.06]	0.00 [-0.07, 0.06]	0.05 [-0.01, 0.12]	0.05 [-0.02, 0.11]	-0.04 [-0.1, 0.03]	-0.02 [-0.09, 0.05]	-0.03 [-0.11, 0.05]	0.01 [-0.07, 0.1]	-0.09 [-0.16, -0.01]	0.05 [-0.03, 0.13]	0.04 [-0.03, 0.11]	0.10 [0.03, 0.17]	*	
GCSE 1	0.02 [-0.02, 0.06]	-0.09 [-0.13, -0.05]	0.12 [0.08, 0.15]	0.19 [0.15, 0.23]	0.25 [0.22, 0.29]	-0.04 [-0.08, 0]	-0.11 [-0.14, -0.07]	0.03 [-0.01, 0.07]	0.05 [0, 0.1]	-0.13 [-0.18, -0.08]	0.07 [0.02, 0.12]	0.33 [0.29, 0.38]	0.58 [0.55, 0.61]		***	***
IQ 1	0.02 [-0.03, 0.06]	-0.02 [-0.07, 0.03]	0.06 [0.01, 0.11]	0.16 [0.11, 0.21]	0.14 [0.09, 0.18]	-0.03 [-0.07, 0.02]	-0.06 [-0.11, -0.02]	-0.04 [-0.09, 0.01]	-0.03 [-0.08, 0.03]	-0.08 [-0.14, -0.02]	0.01 [-0.05, 0.07]	0.43 [0.39, 0.47]	***		***	
SRH 1	0.05 [0, 0.1]	-0.06 [-0.11, -0.02]	0.10 [0.05, 0.14]	-0.03 [-0.07, 0.02]	0.02 [-0.03, 0.07]	-0.01 [-0.06, 0.03]	-0.03 [-0.07, 0.02]	-0.03 [-0.07, 0.02]	0.02 [-0.03, 0.07]	-0.04 [-0.09, 0.01]	0.14 [0.09, 0.19]			*		***
BMI 1	-0.01 [-0.05, 0.04]	0.21 [0.16, 0.25]	-0.07 [-0.12, -0.02]	0.00 [-0.05, 0.04]	-0.08 [-0.13, -0.04]	-0.07 [-0.12, -0.02]	0.06 [0.01, 0.1]	-0.05 [-0.1, 0]	0.01 [-0.04, 0.06]	0.32 [0.27, 0.36]	*	*	***		***	**
Height 1	0.26 [0.22, 0.3]	-0.03 [-0.07, 0.02]	0.04 [0, 0.09]	0.02 [-0.03, 0.07]	0.00 [-0.05, 0.04]	-0.04 [-0.08, 0.01]	-0.01 [-0.06, 0.04]	-0.04 [-0.09, 0.01]	0.44 [0.4, 0.48]				*	*		
SCZ GPS 1	-0.04 [-0.07, 0]	-0.04 [-0.08, -0.01]	-0.03 [-0.06, 0]	-0.08 [-0.11, -0.05]	0.02 [-0.02, 0.05]	0.06 [0.03, 0.1]	0.03 [0, 0.06]	0.53 [0.5, 0.55]								
ADHD GPS 1	0.00 [-0.04, 0.03]	0.10 [0.06, 0.13]	-0.12 [-0.15, -0.09]	-0.07 [-0.11, -0.04]	-0.13 [-0.16, -0.1]	0.04 [0.01, 0.07]	0.51 [0.48, 0.53]				*	*	***			
Neuroticism GPS 1	-0.05 [-0.08, -0.01]	0.01 [-0.02, 0.05]	-0.16 [-0.2, -0.13]	-0.04 [-0.07, 0]	-0.10 [-0.13, -0.06]	0.49 [0.46, 0.51]	**	**			*		*	*	*	*
EA GPS 1	0.09 [0.05, 0.12]	-0.15 [-0.18, -0.12]	0.25 [0.21, 0.28]	0.26 [0.23, 0.29]	0.57 [0.54, 0.59]	***	***		*	*	*	***	***		**	*
IQ GPS 1	0.05 [0.01, 0.08]	-0.07 [-0.1, -0.04]	0.15 [0.12, 0.18]	0.54 [0.51, 0.56]	***	*	***	**	*	*		***	***		*	
SRH GPS 1	0.06 [0.02, 0.09]	-0.20 [-0.24, -0.17]	0.53 [0.51, 0.56]	***	***	***	***	**		*		***	***		**	
BMI GPS 1	-0.05 [-0.08, -0.01]	0.52 [0.49, 0.54]	***		***		***	*		***	*		***		*	*
Height GPS 1	0.52 [0.49, 0.54]	*	*	*	*				***	*						
	Height GPS 2	BMI GPS 2	SRH GPS 2	IQ GPS 2	EA GPS 2	Neuroticism GPS 2	ADHD GPS 2	SCZ GPS 2	Height 2	BMI 2	SRH 2	IQ 2	GCSE 2	Neuroticism 2	ADHD symptoms 2	SCZ symptoms 2

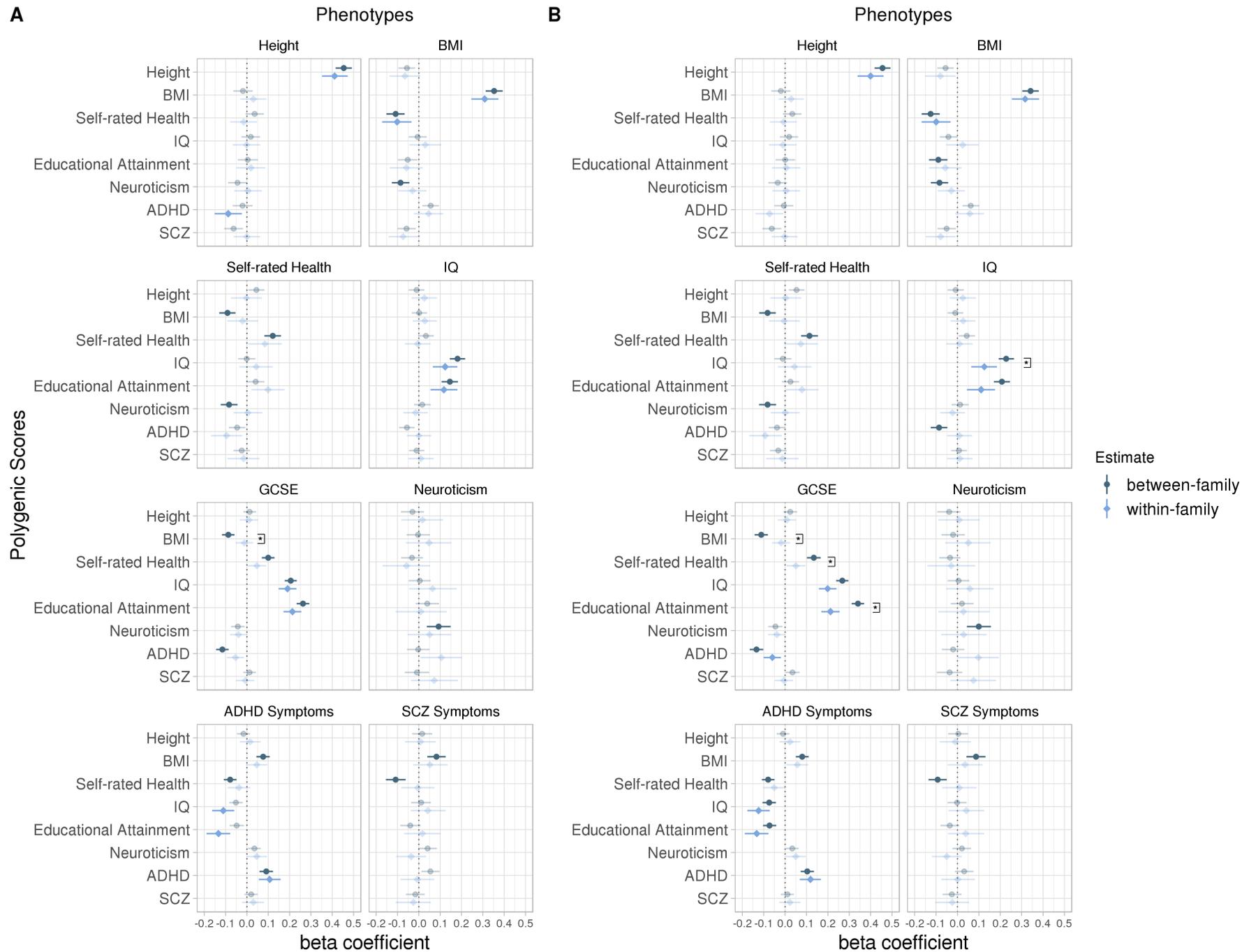
Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia; EA = Educational Attainment; SRH = Self-rated Health; 1 = Twin 1; 2 = Twin 2.

Figure S2. Effect sizes and significance of differences comparing between-family coefficients before and after controlling for family socio-economic status



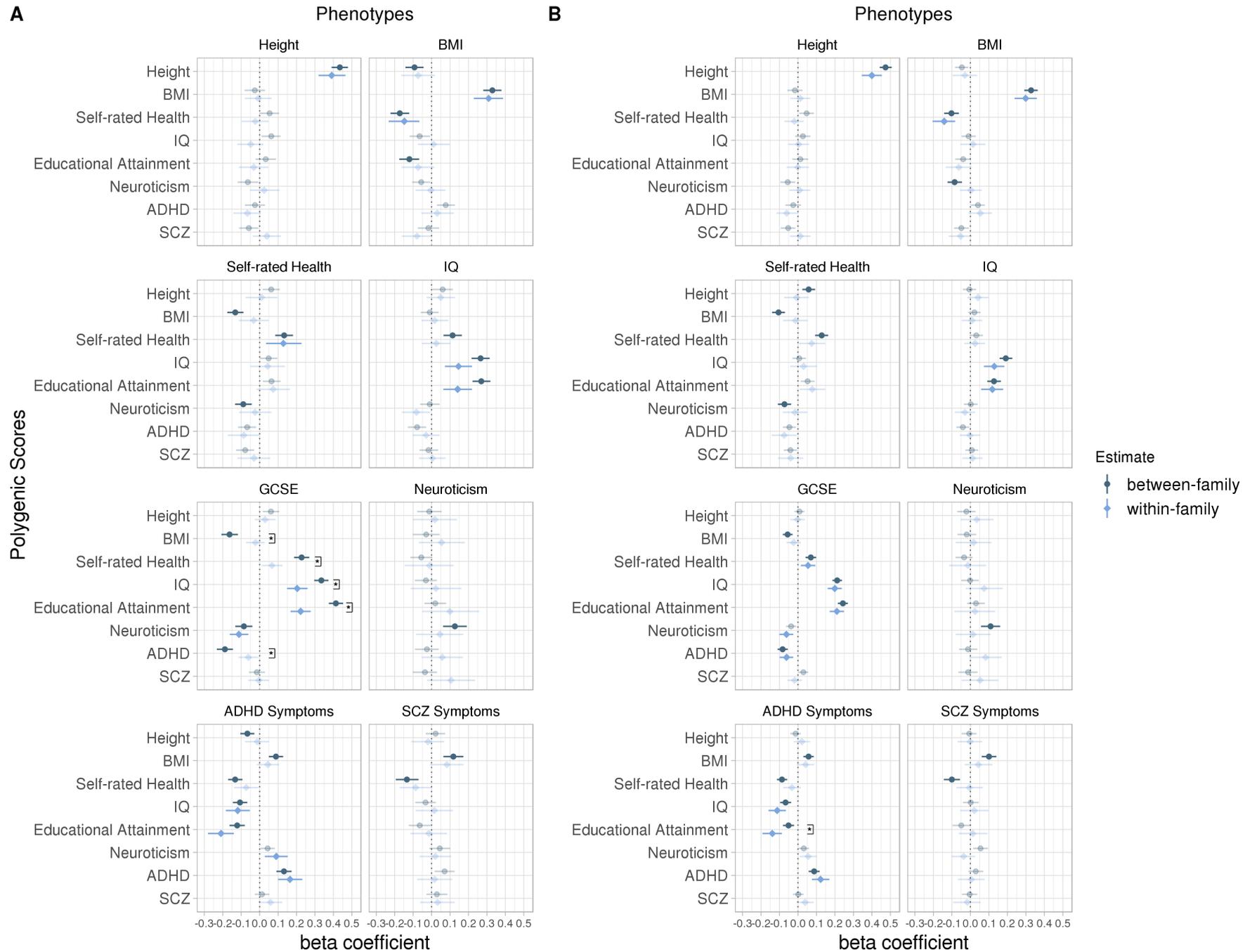
Note. Between-family prediction estimates of eight developmental outcomes using eight polygenic scores before and after statistical correction for family socio-economic status. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between the two between-family prediction estimates, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia.

Figure S3. Within- and between-family prediction estimates accounting for separate measures of family socio-economic status



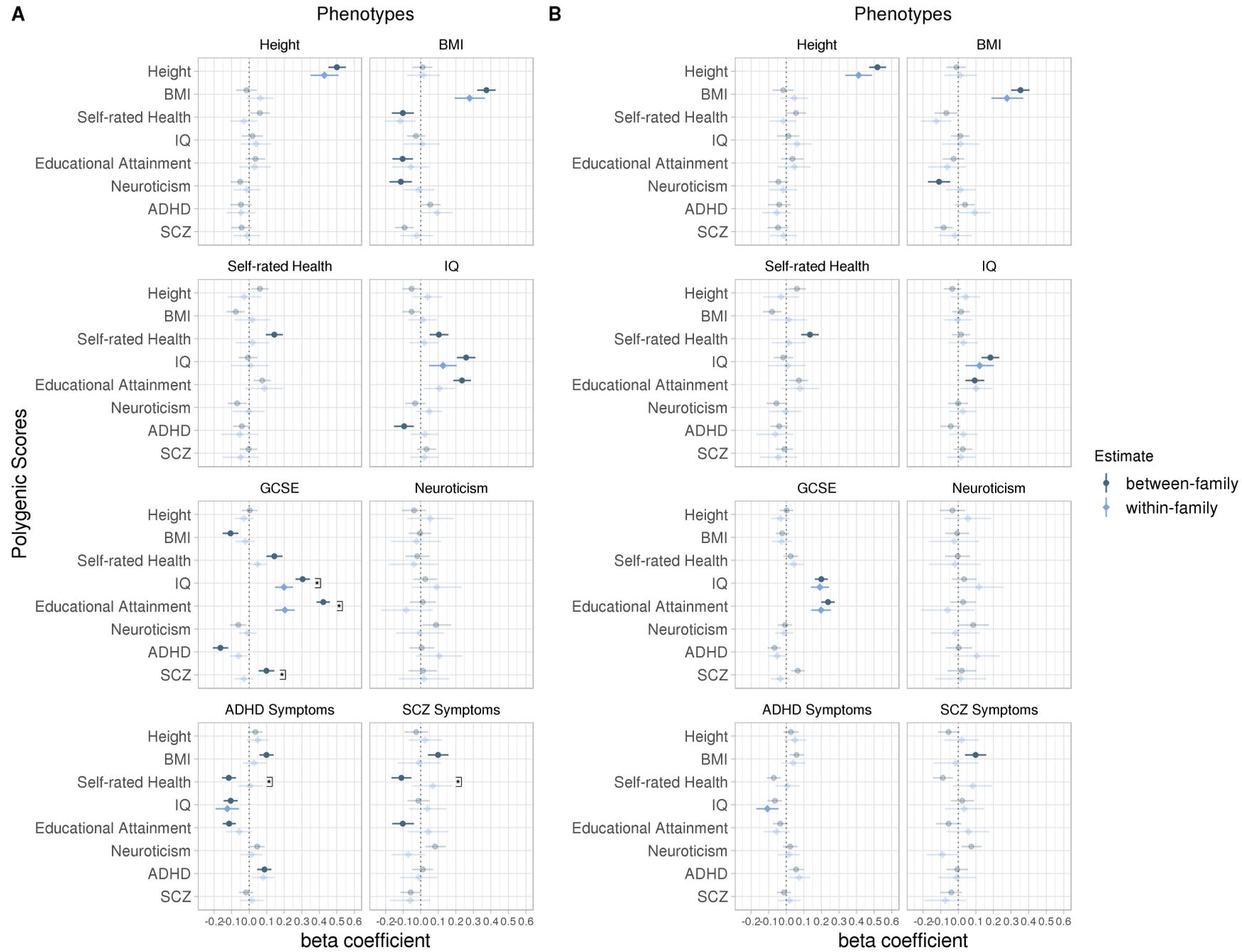
Note. Within- and between-family prediction estimates of eight developmental outcomes using eight polygenic scores after statistically correcting for (A) parental education and (B) parental occupation. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia

Figure S4. Within- and between-family prediction estimates for same-sex twin pairs



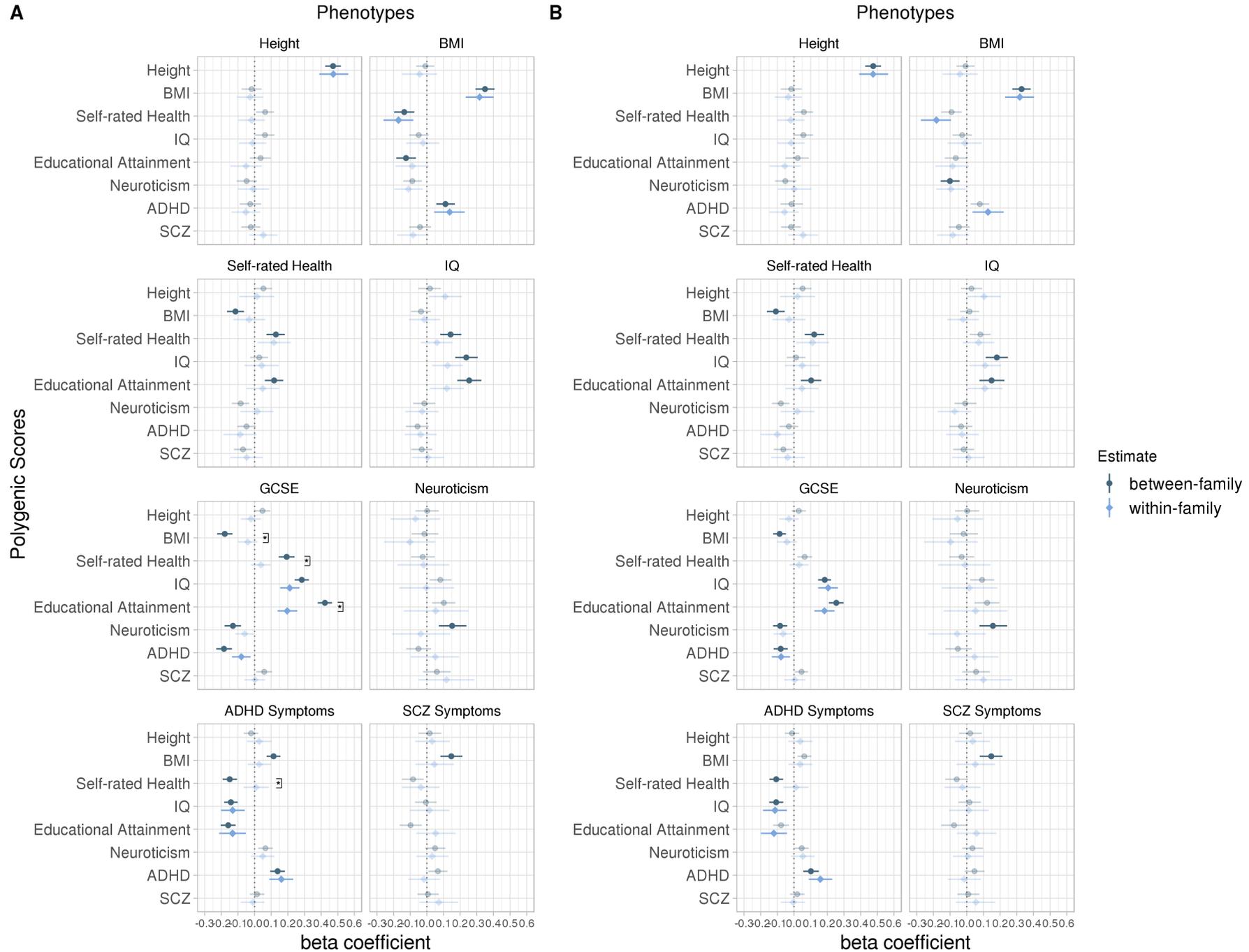
Note. Within- and between-family prediction estimates of eight developmental outcomes using eight polygenic scores before (A) and after (B) statistical correction for family socio-economic status, based on same-sex twin pairs only. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia.

Figure S5. Within- and between-family prediction estimates for opposite-sex twin pairs



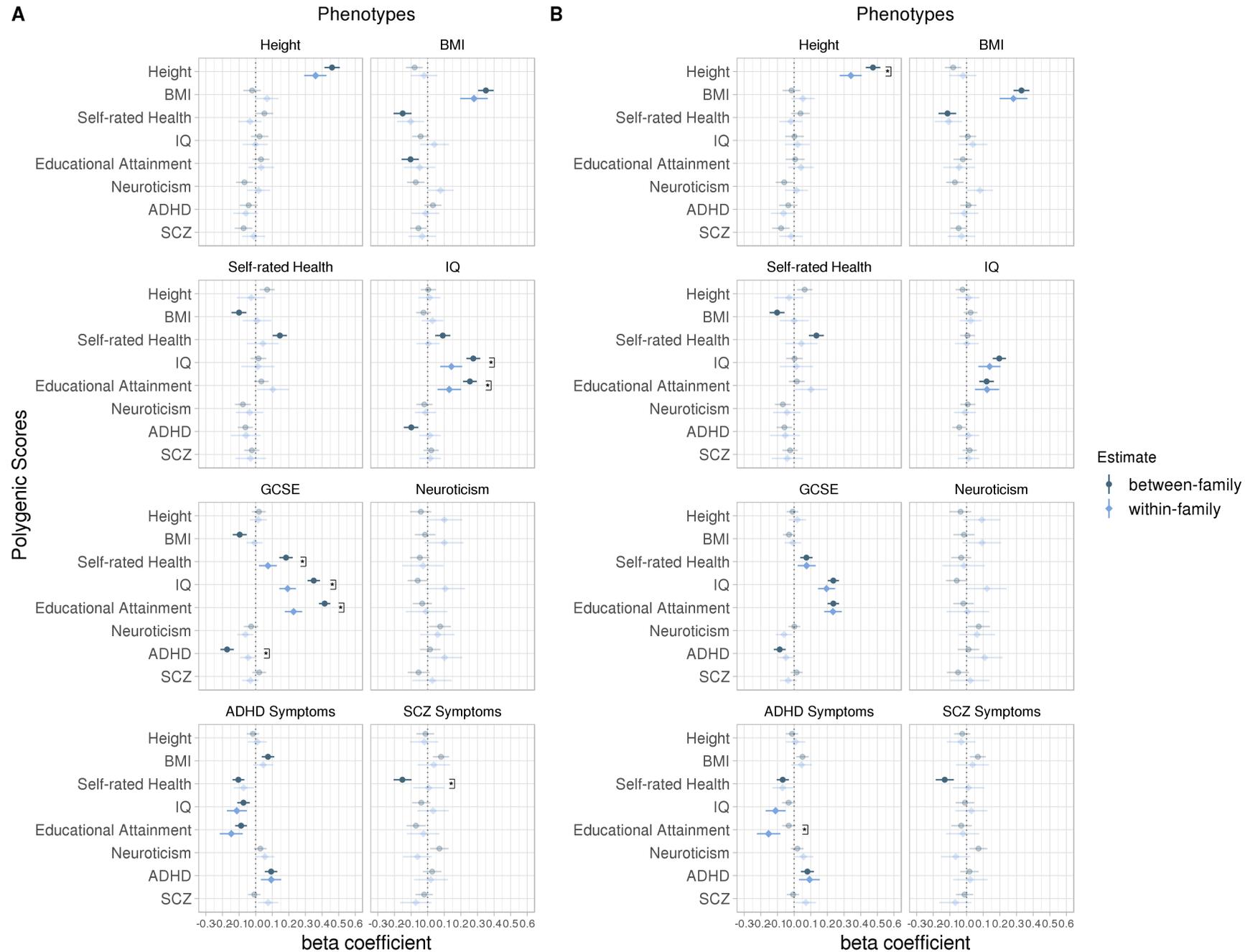
Note. Within- and between-family prediction estimates of eight developmental outcomes using eight polygenic scores before (A) and after (B) statistical correction for family socio-economic status, based on opposite-sex twin pairs only. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia.

Figure S6. Within- and between-family prediction estimates based on twin pairs genotyped on OEE chip



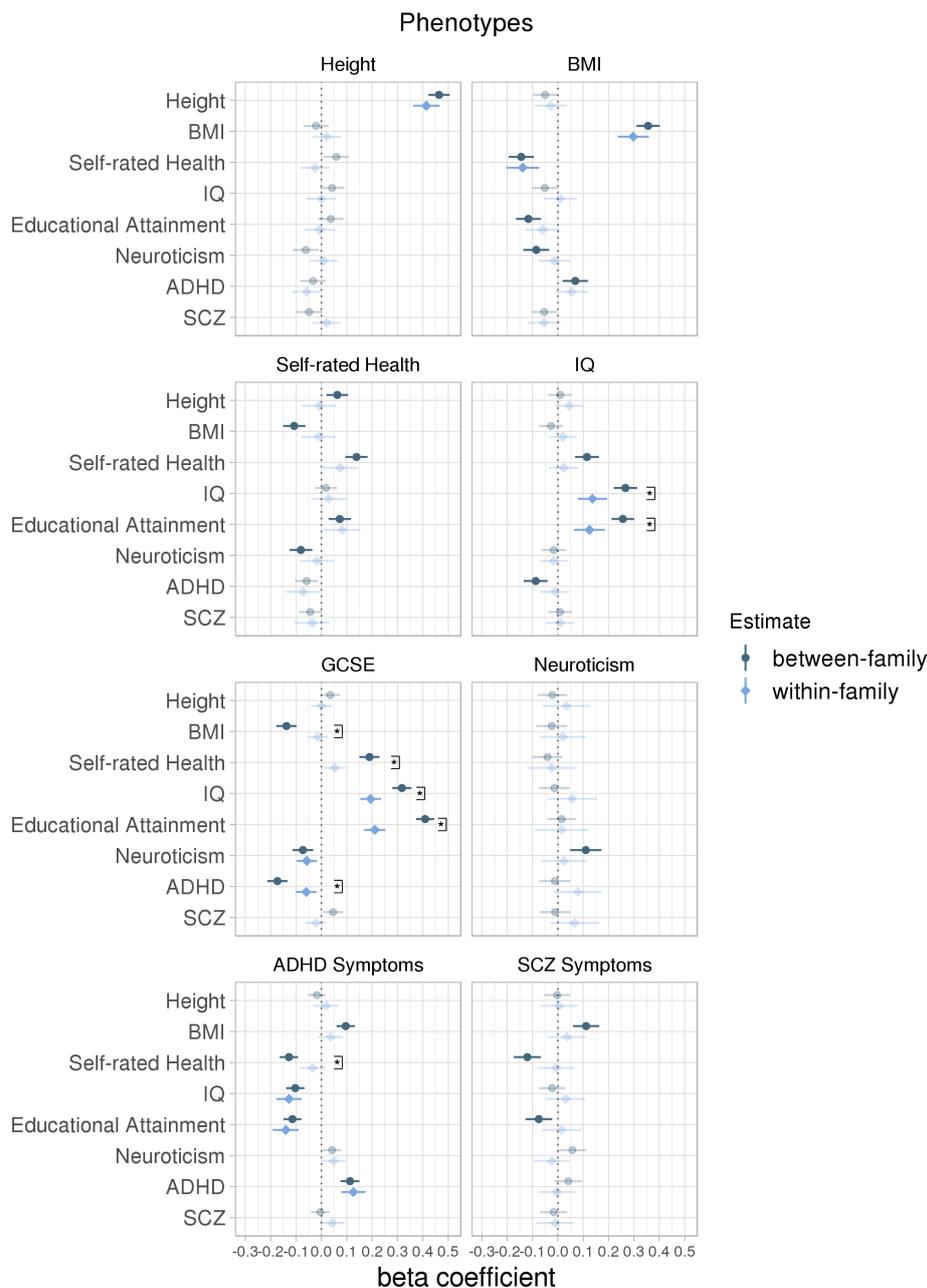
Note. Within- and between-family prediction estimates of eight developmental outcomes using eight polygenic scores before (A) and after (B) statistical correction for family socio-economic status, based on twin pairs genotyped on OEE chip. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia.

Figure S7. Within- and between-family prediction estimates based on twin pairs genotyped on OEE chip and the co-twin genotyped on the Affymetrix chip



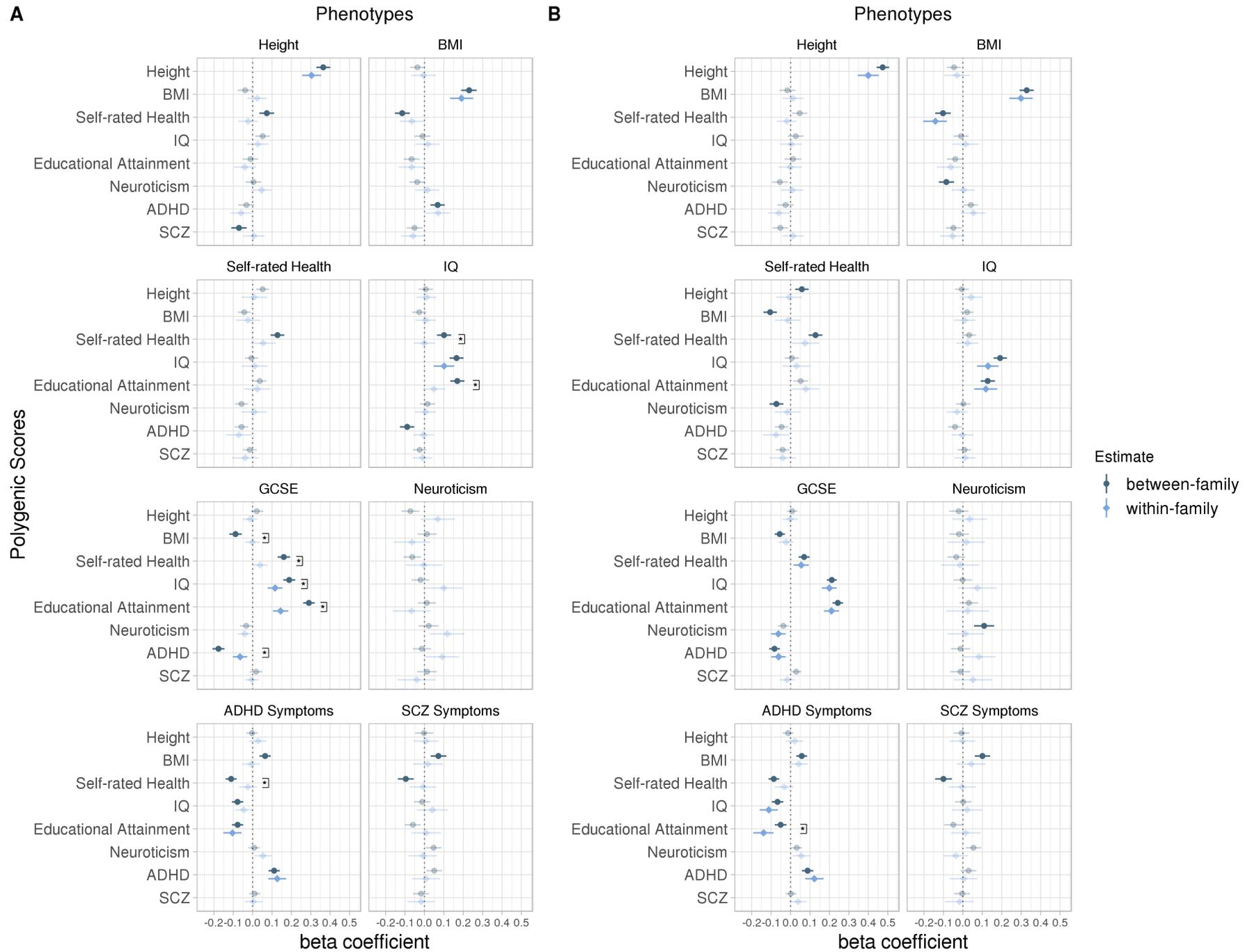
Note. Within- and between-family prediction estimates of eight developmental outcomes using eight polygenic scores before (A) and after (B) statistical correction for family socio-economic status, based on twin pairs with one twin genotyped on OEE chip and the other twin genotyped on Affymetrix chip. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia

Figure S8. Within- and between-family coefficients estimated using a SNP-kinship matrix as random effect



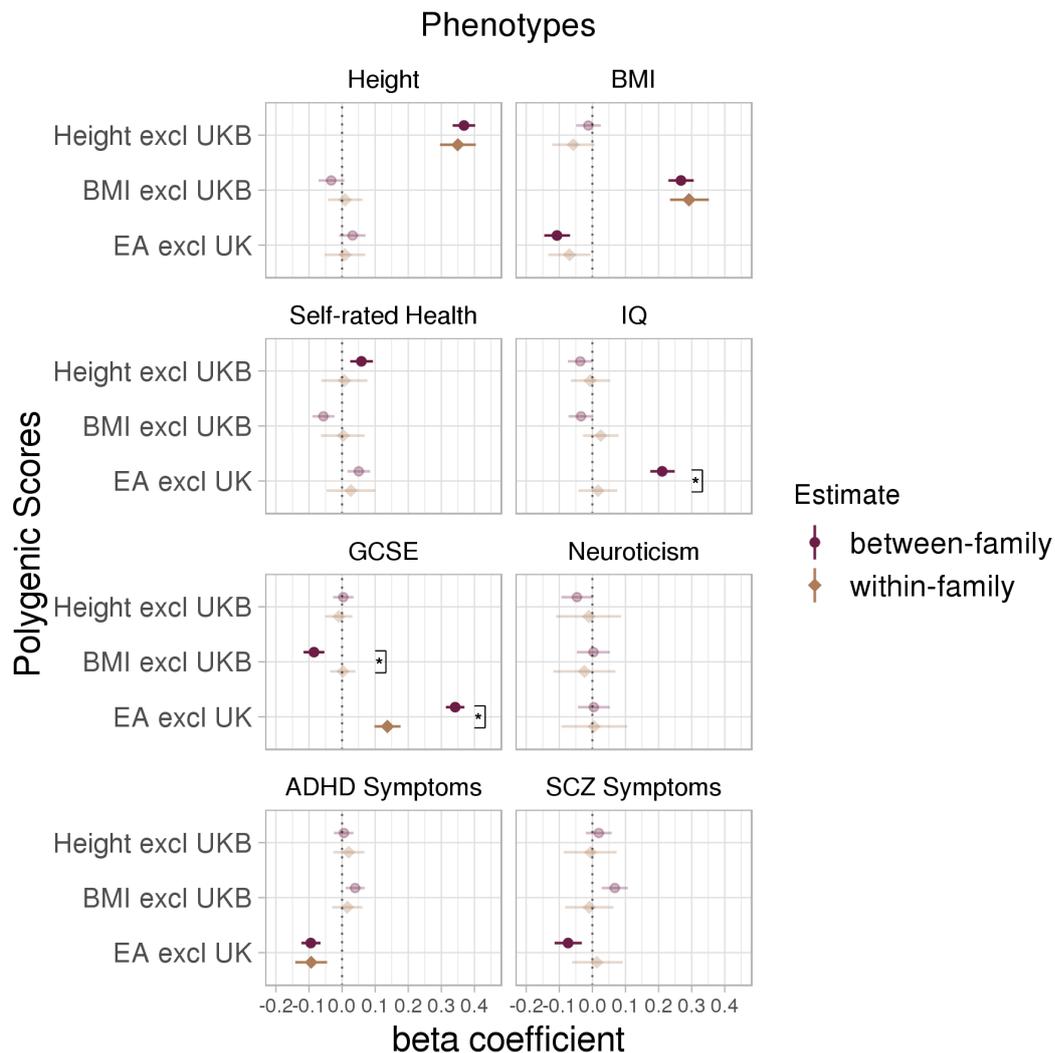
Note. Within- and between-family prediction estimates of eight developmental outcomes using eight polygenic scores. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% confidence intervals calculated using the standard errors of the fixed effects. Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia.

Figure S9. Within- and between-family prediction estimates using polygenic scores calculated with causal fraction of markers of 0.1



Note. Within- and between-family prediction estimates of eight developmental outcomes using eight polygenic scores before (A) and after (B) statistical correction for family socio-economic status, using polygenic scores calculated with causal fraction of markers of 0.1. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia.

Figure S10. Within- and between-family prediction estimates based on polygenic scores with UK Biobank or all UK discovery samples removed



Note. Within- and between-family prediction estimates of eight developmental outcomes in using polygenic scores with UK Biobank or all UK samples removed at discovery stage. Genome-wide Polygenic Scores are presented on the y-axis, predicting each of the eight phenotypic traits. Error bars are 95% bootstrap percentile intervals based on 10,000 bootstrap samples (random resampling of DZ twin pairs with replacement). Opaque estimates indicate statistical significance at the false discovery rate corrected threshold of $p < 0.01$. Brackets indicate a significant difference between within- and between-family prediction estimate, and only significant differences are shown where at least one of the estimates is significant at the false discovery rate corrected threshold of $p < 0.01$. The dotted line represents a beta coefficient of zero. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia; EA = Educational Attainment; UKB = UK Biobank.

Supplementary Tables

Table S1. Descriptive statistics, age and sex effects for phenotypes

	N Pairs	Mean	SD	Skew	Min	Max	F sex	P sex	R² sex	F age	P age	R² age	P order	R² order
Height	1,463	171.997	10.457	0.142	132	211	1,460.639	< 0.001	0.500	0.855	0.355	0.001	0.174	<0.001
BMI	1,353	23.495	4.676	1.520	12.061	47.477	0.002	0.964	<0.001	16.470	< 0.001	0.012	0.918	<0.001
Self-rated Health	1,494	3.480	0.672	-0.330	1.000	5.000	8.188	0.004	0.005	1.457	0.228	0.001	0.987	<0.001
IQ	1,569	0.117	0.954	-0.242	-3.441	3.040	7.277	0.007	0.005	104.278	< 0.001	0.062	0.125	<0.001
GCSE	2,366	8.952	1.194	-0.330	4.670	11.000	1.814	0.178	<0.001	4.49	0.034	0.001	0.487	<0.001
Neuroticism	789	2.583	0.655	0.280	1.000	5.000	31.894	< 0.001	0.039	2.202	0.138	0.003	0.740	<0.001
ADHD Symptoms	2,469	0.063	1.002	1.357	-1.371	5.066	159.896	< 0.001	0.061	16.877	< 0.001	0.007	0.073	<0.001
SCZ Symptoms	1,140	-0.026	0.705	1.560	-0.816	4.093	2.041	0.153	0.002	6.122	0.013	0.005	0.858	<0.001
SES	2,962	0.209	0.994	0.046	-2.351	2.495	--	--	--	--	--	--	--	--

Note. Means and standard deviations for individual measures are calculated based on raw data. Height, BMI, self-reported health, GCSE grades and neuroticism means and standard deviations are reported on their original scale. IQ, ADHD symptoms, schizophrenia symptoms and socioeconomic status are reported on the z-scale as standardization was required to form the composite. Sex, age and birth order tests were performed on one randomly selected twin per pair. R²= proportion of variance explained. Order = birth order; BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia; SES = family socio-economic status.

Table S2. GWAS used for polygenic score calculation

Trait	Year	SNP-h ²	GPS R ²	Cases	Controls	GWAS sample size	Overlapping SNPs ^a	Notes
BMI ²	2018	22.4% (3.7%) ²	10.2% ²	-	-	681,275	424,816	-
Height ²	2018	48.3% (3.7%) ²	24.4% ²	-	-	693,529	427,030	-
Self-rated health ³	2018	13% (0.6%) ⁴	--	-	-	337,199	499,422	-
Intelligence ⁵	2018	19% (1%) ⁵	6.7% ⁶	-	-	266,453	497,059	GWAS re-run excluding TEDS sample (3,414)
Educational Attainment ⁷	2018	12.2% (0.3%) ⁷	11.4% ⁷	-	-	766,345	496,633	-
Neuroticism ⁸	2017	10.8% (0.5%) ⁸	2.8% ⁸	-	-	329,821	500,849	-
ADHD ⁹	2019	21.6% (1.4%) ⁹	5.5% ⁹	20,183	35,191	55,374	469,352	-
Schizophrenia ¹⁰	2018	20% (0.6%) ¹⁰	5.7% ¹⁰	40,675	64,643	105,318	487,423	-

Note. H² = heritability; R² = phenotypic variance explained; ^a = number of overlapping SNPs between GWA studies and TEDS individual-level genotype data.

Table S3. Intraclass coefficients for same-sex and opposite-sex twin pairs

Phenotype	Same-sex twin pairs				Opposite-sex twin pairs			
	N pairs	ICC	ICC 95% CI L	ICC 95% CI U	N pairs	ICC	ICC 95% CI L	ICC 95% CI U
Height	789	0.435	0.365	0.518	674	0.443	0.367	0.534
BMI	733	0.339	0.271	0.424	620	0.286	0.215	0.381
Self-rated Health	805	0.182	0.124	0.266	689	0.083	0.036	0.193
IQ	824	0.451	0.381	0.532	745	0.386	0.316	0.47
GCSE	1,220	0.579	0.517	0.647	1,146	0.585	0.521	0.656
Neuroticism	429	0.084	0.025	0.282	360	0.123	0.054	0.281
ADHD Symptoms	1,285	0.328	0.275	0.391	1,184	0.317	0.262	0.382
SCZ Symptoms	613	0.308	0.235	0.403	527	0.194	0.124	0.302

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; ICC = Intraclass coefficient; CI = Confidence Interval.

Table S4. Intraclass coefficients for twin pairs split by genotyping chip

Phenotype	Twin pairs with both twins genotyped on OEE				Twin pairs with one twin genotyped on OEE and the other twin genotyped on Affy			
	N pairs	ICC	ICC 95% CI L	ICC 95% CI U	N pairs	ICC	ICC 95% CI L	ICC 95% CI U
Height	650	0.401	0.326	0.492	813	0.468	0.398	0.549
BMI	603	0.331	0.257	0.426	750	0.297	0.232	0.381
Self-rated Health	665	0.172	0.111	0.267	829	0.107	0.056	0.201
IQ	510	0.391	0.308	0.496	1059	0.437	0.376	0.508
GCSE	1050	0.57	0.504	0.644	1316	0.59	0.531	0.656
Neuroticism	304	0	0	0	485	0.167	0.098	0.284
ADHD Symptoms	1047	0.312	0.255	0.382	1422	0.329	0.278	0.388
SCZ Symptoms	490	0.241	0.165	0.35	650	0.263	0.195	0.356

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; ICC = Intraclass coefficient; CI = Confidence Interval; OEE = Illumina HumanOmniExpressExome-8v1.2 chip; Affy = AffymetrixGeneChip 6.0 SNP chip.

Table S5. Intraclass coefficients

Phenotype	N pairs	ICC	ICC 95% CI L	ICC 95% CI U
Height	1,463	0.439	0.386	0.498
BMI	1,353	0.317	0.265	0.379
Self-rated Health	1,494	0.136	0.093	0.199
IQ	1,569	0.422	0.371	0.479
GCSE	2,366	0.582	0.537	0.63
Neuroticism	789	0.103	0.055	0.193
ADHD Symptoms	2,469	0.323	0.285	0.366
SCZ Symptoms	1,140	0.254	0.201	0.32

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; ICC = Intraclass coefficient; CI = Confidence Interval.

Table S12. Within- and between-family prediction estimates of family socio-economic status

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	P.Diff
SES	ADHD	-0.168	-0.214	-0.13	1.32e-28	0	0	0	1	2.69e-08
SES	BMI	-0.19	-0.216	-0.134	3.80e-38	0	0	0	1	1.80e-10
SES	EA	0.433	0.409	0.479	2.35e-204	0	0	0	1	5.47e-49
SES	Height	0.069	0.02	0.102	6.51e-06	0	0	0	1	0.025
SES	IQ	0.231	0.213	0.291	4.28e-55	0	0	0	1	7.65e-14
SES	Neurot	-0.06	-0.129	-0.043	5.35e-05	0	0	0	1	0.049
SES	SCZ	0.04	-0.015	0.066	0.008	0	0	0	1	0.201
SES	SRH	0.302	0.238	0.318	3.86e-89	0	0	0	1	2.24e-23

Note. SES = family socio-economic status; BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; P.diff = statistical significance of difference between within- and between-family estimates. Fixed effects are estimated using multiple regression analysis as parental SES is invariant within DZ twin pairs and thus mixed-effects modelling is not possible.

Table S24. Within- and between-family prediction estimates based on polygenic scores with UK Biobank or all UK discovery samples removed

pheno	GPS	beta.B	L.CI.B	U.CI.B	P.B	beta.W	L.CI.W	U.CI.W	P.W	TotEff	PercRed	BetaDiff	Diff.L.CI	Diff.U.CI	P.diff
ADHD	BMI excl UKB	0.039	0.011	0.068	3.64e-02	0.016	-0.03	0.061	5.04e-01	0.023	60.52	0.024	-0.029	0.076	4.26e-01
ADHD	EA excl all UK	-0.095	-0.123	-0.065	2.93e-07	-0.093	-0.141	-0.045	2.44e-04	-0.094	1.15	-0.001	-0.057	0.056	9.72e-01
ADHD	Height excl UKB	0.005	-0.024	0.034	7.73e-01	0.02	-0.026	0.067	4.14e-01	0.015	-274.51	-0.015	-0.07	0.041	6.33e-01
BMI	BMI excl UKB	0.268	0.23	0.306	2.71e-26	0.292	0.234	0.352	2.18e-21	0.285	-9.07	-0.024	-0.096	0.047	5.34e-01
BMI	EA excl all UK	-0.107	-0.145	-0.067	2.15e-05	-0.069	-0.133	-0.006	4.56e-02	-0.081	35.51	-0.038	-0.111	0.037	3.73e-01
BMI	Height excl UKB	-0.012	-0.049	0.025	6.18e-01	-0.058	-0.121	0.007	8.12e-02	-0.043	-370.21	0.046	-0.028	0.118	2.70e-01
GCSE	BMI excl UKB	-0.085	-0.117	-0.053	7.53e-05	0.003	-0.036	0.04	8.87e-01	-0.048	103.15	-0.088	-0.136	-0.039	2.08e-03
GCSE	EA excl all UK	0.342	0.314	0.37	3.23e-64	0.137	0.098	0.177	2.15e-11	0.249	59.88	0.205	0.155	0.254	4.46e-13
GCSE	Height excl UKB	0.004	-0.027	0.034	8.58e-01	-0.01	-0.051	0.031	6.16e-01	-0.002	369.05	0.014	-0.037	0.064	6.33e-01
SRH	BMI excl UKB	-0.056	-0.09	-0.023	1.27e-02	0.004	-0.064	0.068	9.06e-01	-0.004	106.98	-0.06	-0.132	0.016	1.34e-01
SRH	EA excl all UK	0.051	0.016	0.085	2.51e-02	0.027	-0.047	0.1	4.66e-01	0.03	46.46	0.023	-0.057	0.105	5.88e-01
SRH	Height excl UKB	0.058	0.025	0.093	7.49e-03	0.007	-0.063	0.076	8.48e-01	0.014	88.27	0.052	-0.025	0.129	2.18e-01
Height	BMI excl UKB	-0.033	-0.071	0.007	2.04e-01	0.009	-0.042	0.062	7.39e-01	-0.009	127.78	-0.042	-0.108	0.024	2.65e-01
Height	EA excl all UK	0.032	-0.009	0.071	2.16e-01	0.009	-0.053	0.07	7.77e-01	0.019	73.04	0.023	-0.05	0.095	5.59e-01
Height	Height excl UKB	0.368	0.334	0.402	6.29e-54	0.35	0.296	0.404	4.03e-35	0.358	5.02	0.019	-0.047	0.083	6.05e-01
IQ	BMI excl UKB	-0.034	-0.072	0.002	1.65e-01	0.025	-0.028	0.079	3.44e-01	0	173.67	-0.06	-0.125	0.007	1.01e-01
IQ	EA excl all UK	0.211	0.175	0.249	1.27e-18	0.017	-0.042	0.075	5.63e-01	0.095	91.83	0.194	0.126	0.265	3.41e-07
IQ	Height excl UKB	-0.037	-0.074	0.001	1.33e-01	-0.007	-0.065	0.054	8.21e-01	-0.019	82.19	-0.03	-0.1	0.038	4.25e-01
Neurot	BMI excl UKB	0.003	-0.047	0.053	9.15e-01	-0.024	-0.118	0.07	6.11e-01	-0.021	825.01	0.027	-0.08	0.134	6.27e-01
Neurot	EA excl all UK	0.004	-0.044	0.053	8.88e-01	0.005	-0.093	0.105	9.17e-01	0.005	-25.51	-0.001	-0.112	0.109	9.85e-01
Neurot	Height excl UKB	-0.047	-0.093	0.001	1.12e-01	-0.011	-0.11	0.086	8.27e-01	-0.015	76.63	-0.036	-0.143	0.073	5.36e-01
SCZ	BMI excl UKB	0.068	0.028	0.107	1.40e-02	-0.01	-0.083	0.064	7.88e-01	0.01	114.1	0.077	-0.01	0.163	8.55e-02

SCZ	EA excl all UK	-0.073	-0.114	-0.032	6.90e-03	0.014	-0.061	0.092	7.31e-01	-0.008	118.91	-0.087	-0.178	0.002	7.22e-02
SCZ	Height excl UKB	0.019	-0.02	0.058	4.57e-01	-0.005	-0.086	0.073	8.92e-01	0.001	126.82	0.025	-0.062	0.112	5.96e-01

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); ADHD = Attention-Deficit/Hyperactivity Disorder; SCZ = Schizophrenia symptoms; EA = Educational Attainment; Neurot = Neuroticism; SRH = Self-rated Health; UKB = UK Biobank; B = Between-family estimate; W = Within-family estimate; L.CI = Lower 95% bootstrap percentile interval; U.CI = Upper 95% bootstrap percentile interval; P = statistical significance of beta estimate; TotEff = Total effect derived as the intra-class correlation weighted sum of the within- and between family effect. PercRed = Reduction of prediction estimates when comparing within- to between-family estimates in percentage. BetaDiff = Difference of between-family and within-family beta coefficient; P.diff = statistical significance of difference between within- and between-family estimates.

Table S25 Phenotypic and polygenic score mean differences by polygenic score difference quantiles

phenotype	GPS quant	mean phen	CI.L	CI.U	mean GPS	CI.L	CI.U
Height	1	0.270	-1.339	1.880	-0.004	-0.016	0.008
Height	2	1.718	0.203	3.233	-0.003	-0.033	0.027
Height	3	2.435	0.738	4.132	0.020	-0.030	0.071
Height	4	3.489	2.039	4.939	-0.030	-0.102	0.041
Height	5	1.433	-0.209	3.075	0.066	-0.025	0.158
Height	6	1.452	-0.144	3.048	0.019	-0.101	0.139
Height	7	4.620	2.862	6.379	-0.007	-0.154	0.141
Height	8	3.649	1.813	5.485	-0.038	-0.218	0.142
Height	9	7.021	5.316	8.727	0.065	-0.165	0.295
Height	10	8.989	7.181	10.798	-0.090	-0.422	0.242
BMI	1	0.064	-0.843	0.972	0.001	-0.012	0.014
BMI	2	0.968	0.126	1.810	0.004	-0.029	0.037
BMI	3	-0.555	-1.439	0.329	-0.011	-0.066	0.044
BMI	4	1.204	0.319	2.089	-0.027	-0.105	0.051
BMI	5	0.689	-0.073	1.451	0.055	-0.047	0.158
BMI	6	0.885	-0.009	1.778	-0.031	-0.159	0.097
BMI	7	1.495	0.584	2.406	0.075	-0.084	0.235
BMI	8	1.309	0.410	2.208	-0.084	-0.281	0.113
BMI	9	1.762	0.999	2.526	0.043	-0.200	0.286
BMI	10	2.933	2.092	3.773	-0.026	-0.378	0.327
IQ	1	1.369	-1.040	3.778	0.005	-0.006	0.016
IQ	2	-0.260	-2.860	2.341	-0.011	-0.042	0.019
IQ	3	1.139	-1.385	3.662	0.010	-0.041	0.061
IQ	4	-0.725	-3.287	1.838	-0.009	-0.081	0.063
IQ	5	2.081	-0.367	4.529	0.023	-0.069	0.115
IQ	6	1.483	-0.982	3.949	0.019	-0.099	0.137
IQ	7	1.288	-1.222	3.799	-0.051	-0.193	0.092
IQ	8	2.379	-0.215	4.973	0.092	-0.083	0.268
IQ	9	4.304	1.881	6.728	-0.101	-0.324	0.121
IQ	10	3.291	0.934	5.647	0.023	-0.293	0.340
GCSE	1	0.068	-0.067	0.203	0.005	-0.004	0.013
GCSE	2	-0.057	-0.204	0.090	0.001	-0.023	0.024
GCSE	3	0.195	0.053	0.336	-0.012	-0.051	0.026
GCSE	4	0.080	-0.063	0.223	0.021	-0.033	0.074
GCSE	5	0.097	-0.039	0.234	-0.022	-0.094	0.049
GCSE	6	0.266	0.129	0.403	0.054	-0.039	0.146

GCSE	7	0.226	0.083	0.369	0.084	-0.030	0.199
GCSE	8	0.307	0.180	0.435	0.060	-0.079	0.200
GCSE	9	0.341	0.192	0.490	0.013	-0.159	0.185
GCSE	10	0.471	0.327	0.614	-0.202	-0.445	0.041

Note. BMI = Body Mass Index; IQ = Intelligence; GCSE = General Certificate of Secondary Education (educational achievement); GPS = genome wide polygenic score; quant = quantile; phen = phenotype; CI.L = 95% lower confidence interval; CI.U = 95% upper confidence interval; GPS quant 1 = lowest absolute GPS twin pair difference quantile; GPS quant 10 = highest absolute GPS twin pair difference quantile.

Supplementary Methods

Methods S1. Polygenic score calculation

To calculate polygenic scores, we used a Bayesian approach to polygenic score calculation, implemented in the software *LDpred*¹. In comparison with conventional clumping and p value thresholding approaches, LDpred has demonstrated an improvement in predictive accuracy¹. Through this method, a posterior effect size is calculated for each single SNP that is present in both the GWA study summary statistics and the target genotype sample (see Table S2 for number of overlapping SNPs). To calculate the posterior effect size, the original summary statistic effect size estimates are adjusted based on two factors: (a) the relative influence of a SNP given its level of LD with surrounding SNPs in the target sample (here TEDS), and (b) a prior on the effect size of each SNP. To account for LD, we set the radius to a 2 megabase window. The effect size prior depends on the SNP-heritability of the discovery (i.e., GWA study) trait and an assumption on the fraction of causal markers believed to influence the discovery trait. Using the prior, the beta effect sizes are reweighted such that the effects are spread out among the SNPs across the whole genome in proportion to the LD present among these SNPs. To accommodate the high computational demands of these calculations, we reduced our genotype data set to SNPs that had perfect imputation scores (info = 1), leaving 515,100 SNPs for analysis. In the next step, all trait-associated alleles were counted (0, 1, or 2 for each SNP), weighted by the posterior SNP effect size obtained through LDpred, and summed across the genome to calculate a GPS for each individual in TEDS.

References

1. Vilhjalmsson, B.J., Yang, J., Finucane, H.K., Gusev, A., Lindstrom, S., Ripke, S., Genovese, G., Loh, P.-R., Bhatia, G., Do, R., et al. (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* 97, 576–592.
2. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in similar to 700 000 individuals of European ancestry. *Hum. Mol. Genet.* 27, 3641–3649.
3. McInnes, G., Tanigawa, Y., DeBoever, C., Lavertu, A., Olivieri, J.E., Aguirre, M., and Rivas, M.A. (2018). Global Biobank Engine: enabling genotype-phenotype browsing for biobank summary statistics. *Bioinformatics* 9, 1612.
4. Harris, S.E., Hagenaars, S.P., Davies, G., Hill, W.D., Liewald, D.C.M., Ritchie, S.J., Marioni, R.E., Sudlow, C.L.M., Wardlaw, J.M., McIntosh, A.M., et al. (2017). Molecular genetic contributions to self-rated health. *Int J Epidemiol* 46, 994–1009.
5. Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C.A., Nagel, M., Awasthi, S., Barr, P.B., Coleman, J.R.I., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* 50, 912–919.
6. Allegrini, A., Selzam, S., Rimfeld, K., Stumm, von, S., Pingault, J.-B., and Plomin, R. (2019). Genomic prediction of cognitive traits in childhood and adolescence. *Molecular Psychiatry*.
7. Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghziyan, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Linner, R.K., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* 50, 1112–1121.
8. Luciano, M., Hagenaars, S.P., Davies, G., Hill, W.D., Clarke, T.-K., Shirali, M., Harris, S.E., Marioni, R.E., Liewald, D.C., Fawns-Ritchie, C., et al. (2018). Association analysis in over

329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat. Genet.* 50, 6–11.

9. Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* 51, 63–75.

10. Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* 50, 381–389.