Title

Estimating the sensitivity of associations between risk factors and outcomes to shared genetic effects

Byline

Jean-Baptiste Pingault[1,2], PhD; Frühling Rijsdijk[2], PhD; Tabea Schoeler[1], PhD; Shing Wan Choi[2], PhD; Saskia Selzam[2], MsC; Eva Krapohl[2], PhD; Paul F. O'Reilly[2], PhD; Frank Dudbridge[3], PhD.

[1] Department of Clinical, Educational and Health Psychology, University College London, London, UK.

[2] Social, Genetic, and Developmental Psychiatry, King's College London, London, UK.

[3]Department of Health Sciences, University of Leicester, Leicester.

Address correspondence to Dr. Pingault (j.pingault@ucl.ac.uk), Division of Psychology and Language Sciences, 26 Bedford Way, London, WC1H 0DS.

Abstract

Objective. Countless associations between risk factors and outcomes are reported in epidemiological research, but often without estimating the contribution from genetics. However most outcomes and risk factors are substantially heritable, and genetic influences can confound these associations. Here we propose a two-stage approach for evaluating the role of shared genetic effects in explaining these observed associations.

Method. Genotyped unrelated participants from the Twins Early Development Study are included (N from 3,663 to 4,693 depending on the outcome) in our analyses. As an example for our proposed approach, we focus on maternal educational attainment, a risk factor known to associate with a variety of offspring social and health outcomes, including child educational achievement, Body Mass Index, and Attention Deficit Hyperactivity Disorders (ADHD). In the first stage of our approach we estimate how much of the phenotypic associations between maternal education and child outcomes can be attributed to shared genetic effects via regressions controlling for increasingly powerful polygenic scores. In the second stage, we estimate shared genetic effects using heritability estimates and genetic correlations equal to those derived in both SNP-based and twin-based studies. Finally, evidence from the two stages are evaluated in conjunction to provide an overall assessment of the likelihood that the association is explained by genetics.

Results. Associations between maternal education and the three developmental outcomes are highly significant. The magnitude of these associations decrease when using polygenic scores to account for shared genetic effects, explaining between 14.3% and 24.3% of the original associations. For the three outcomes, the magnitude of these associations further decrease under a SNP-based heritability scenario and are almost entirely or entirely explained by genetics under a twin-based heritability scenario.

Conclusions. Observed association between maternal education and child educational attainment, BMI and ADHD symptoms may be largely explained by genetics. To the extent that available estimates of SNP-based and twin-based heritabilities are accurate, the present findings represent a call for caution

when interpreting non-genetically informed epidemiology studies of the role of maternal education or other 'environmental' risk factors. The two-stage approach that we propose adds a new tool to probe the robustness of findings regarding the role of a range of risk factors. Our approach, akin to a genetically informed sensitivity analysis, only requires a genotyped cohort with adequate phenotypic measurements, and has the potential to be widely applied across the life and social sciences.

Introduction

Associations between risk factors and outcomes are commonly reported in epidemiological research, but often without estimating the contribution from genetics. However most outcomes and risk factors are substantially heritable, and genetic influences can confound these associations. Here, we propose a new sensitivity analysis, which we call *Gsens*, to assess to what extent shared genetic effects can account for observed associations. The genetically informed sensitivity analysis we propose has the potential to be widely implemented across the life and social sciences.

Genetic confounding and sensitivity analysis

Identifying risk factors that can be targeted in effective interventions is a fundamental objective shared across the life and social sciences. To this end, identifying *causal* risk factors is essential as interventions that target non-causal risk factors will likely fail. To establish causation, it is necessary to account for confounding, which happens when a third variable causally influences both the risk factor and the outcome, thereby generating a non-causal association between them. Genetic confounding is a special case when genetic factors play the role of the third variable. The concept of genetic confounding was introduced during the controversy regarding the effect of smoking on lung cancer. In a letter entitled 'alleged dangers of cigarette-smoking', Ronald Fisher qualified smoking as ‘possibly and entirely imaginary cause’ for lung cancer[1]. He argued that genetic factors could directly influence both smoking and lung cancer, generating a non-causal association between them. Although Fisher was mistaken in this particular instance, the notion of genetic confounding remains relevant, in his words ‘a common cause, in this case the individual genotype’. During this controversy, Jerome Cornfield argued against this ‘constitutional hypothesis’[2,3]. He contended that implausibly large genetic effects (or other unobserved confounders) would be required to explain away all of the observed association. This led to the birth of the approach now called *sensitivity analysis*, which consists in estimating how strong an unknown confounder needs to be in order to explain away an observed association, providing insights

into the robustness of that association (i.e. how sensitive it is to confounding and whether it is likely causal or not)[2]. Since then, sensitivity analyses became common epidemiological tools to probe the robustness of findings under alternative scenarios. However, sensitivity analysis using genetic data has not progressed. We recently[4] proposed to use polygenic scores – individual-level scores that summarize genetic risk (or protection) for a given phenotype – to estimate the proportion of observed associations explained by shared genetics effects. However, because polygenic scores currently capture only a small part of heritability, controlling for polygenic scores cannot entirely capture shared genetic effects. We therefore proposed a sensitivity analysis using polygenic scores to gauge how likely it is that shared genetic effects account, in part or entirely, for a given risk factor-outcome association. Here, we implement this proposition in two stages. First, we test to what extent associations of interest are accounted for by observed polygenic scores. Second, in the sensitivity analysis per se, we create scenarios examining how a gradual increase in the predictive power of polygenic scores based on heritability estimates would affect association estimates. This can be thought of as adjusting for polygenic scores that would effectively capture as much of the variance in the risk factor and outcome as suggested by available heritability estimates.

Maternal education and child developmental outcomes

To illustrate our approach, we focus on maternal education as the risk factor of interest. Maternal education is associated with child developmental outcomes in several key domains : social development (e.g. better educational attainment), physical health (e.g. lower Body Mass Index, BMI), and mental health (e.g. lower levels of Attention-Deficit Hyperactivity Disorder symptoms).[5–8] However, observed associations between maternal education and developmental outcomes are not free from confounding, in particular genetic confounding as both maternal educational attainment and developmental outcomes are heritable.[9–14]

Here, we illustrate the use of a new method, *Gsens,* to estimate the role of shared genetic effects in explaining the associations between maternal education and three developmental outcomes: educational attainment in the child, BMI, and ADHD. Each of these analyses will provide new insights on the likelihood that maternal education is a determinant or a mere correlate of these developmental outcomes. Importantly, the sensitivity analysis we propose has a wide scope of applications as it only requires genome-wide data on large samples and a focus on outcomes for which polygenic scores are available. Its applicability will further expand with the steady increase in the number and the power of available polygenic scores.

Method

Participants

Participants were drawn from the Twins Early Development Study (TEDS), a longitudinal study of twin pairs born in England and Wales, between 1994 and 1996. Detail regarding TEDS, the recruitment process, and representativeness can be found elsewhere.[15] A total of 7,026 unrelated individuals have been genotyped in TEDS. For each individual analysis, sample size comprised between 3,663 to 4,693 individuals with data for maternal education and each outcome. Written consent was obtained from all the families who agreed to take part in the study. This study was approved by the Institute of Psychiatry, Kings College London, Ethics Committee.

Measures

Maternal education was reported by mothers at first contact, when children were on average 18 months old, with 8 levels: 1= no qualifications; 2 = CSE grade 2-5 or O-level/GCSE grade D-G; 3 = CSE grade 1 or O-level/GCSE grade A-C; 4 = A-level or S-level; 5 = HNC; 6 = HND; 7= undergraduate degree; 8 = postgraduate qualification.

Child educational achievement was operationalized as performance on the standardized UK-wide examination, the General Certificate of Secondary Education (GCSE), at 16 years. We computed a mean of the three compulsory core subjects, mathematics, English, and science (further details in [11]). A total of 3,785 genotyped individuals had data on both maternal education and child GCSE.

Body Mass Index (BMI) was derived from parent reported (ages 11 and 14 years) and self-reported weight and height (age 16 years). Extreme BMI values (<1% and >99% quantiles) were winsorized and resulting values were averaged across ages 11 to 16 years. A total of 3,663 genotyped individuals had data on maternal education and the resulting BMI score.

The DSM-IV ADHD symptom subscale, taken from the Conners' Parent Rating Scales–Revised,[16] was completed by mothers to assess inattentive and hyperactive/impulsive symptoms (9 for hyperactivity/impulsivity and 9 for inattention). Each item was rated on a 4-point Likert scale ranging from 0 (not at all true) to 3 (very much true). A total ADHD score was created by averaging scores across the following mean ages of participants at assessments: 8, 11, 14, and 16 years. The score measures population symptoms dimensionally and not the clinical disorder. A total of 4,693 genotyped individuals had data on maternal education and the ADHD score.

Analyses

Genotyping, quality control procedures and principal component analysis are detailed in the supplementary material. A total sample of 7,026 participants with European ancestry remained after quality control. Single Nucleotide Polymorphisms (SNPs) were excluded if the minor allele frequency was <5%, if more than 1% of genotype data were missing, or if the Hardy Weinberg p-value was lower than 10^{-5}. Non-autosomal markers and indels were removed.

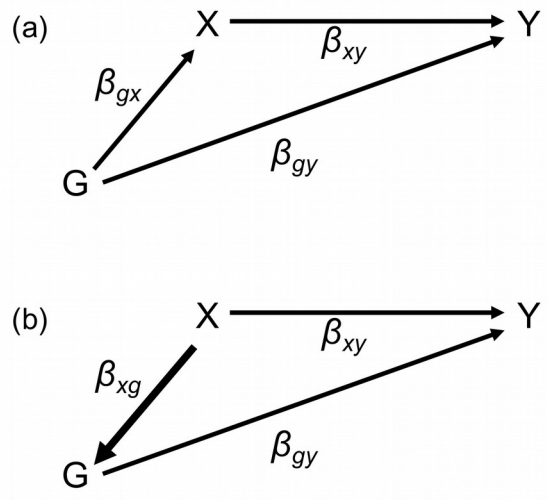
We computed genome-wide polygenic scores based on summary statistics from the following genome-wide association studies (GWAS): (i) years of education[17]; (ii) ADHD[14]; and (iii) BMI[12]. Polygenic

scores for all TEDS participants and all traits were computed using PRSice software[18], with prior clumping to remove SNPs in linkage disequilibrium ($r^2 > 0.10$). PRSice allowed us to select the best-fitting polygenic score for each trait, e.g. maximizing the amount of variance explained by the polygenic score for BMI in TEDS participants' BMI. To this end, we computed a series of polygenic scores including an increasing number of SNPs corresponding to increasing p-value thresholds (e.g. all SNPs associated to BMI at $p < .0001$ and $p < .10$) as illustrated in eFigures 1, 2, & 3 in the supplementary material. Using linear regression analyses, we estimated the proportion of variance explained by each generated polygenic score in the corresponding trait in TEDS. The following covariates were included in regression analyses: sex, age (for GCSE), and 10 principal components of ancestry.

Genetic confounding

Akin to third variable confounding, genetic confounding is represented in Figure 1a: genetic factors (G) – here measured by polygenic score(s) – influence both the risk factor (X) and the outcome (Y). MacKinnon et al. demonstrated that mediation and confounding are statistically identical in linear structural equation modelling[19]. Therefore, genetic confounding can be estimated using structural equation modelling by treating the confounder – here the polygenic score G – as a mediator of the effect of X and Y (Figure 1b). The confounding effect is the indirect effect of X on Y through G: $\hat{\beta}_{XG}\hat{\beta}_{GY}$. We also calculated the proportion of the observed effect of X on Y that is accounted for by shared genetic effects, i.e. $\frac{\hat{\beta}_{XG}\hat{\beta}_{GY}}{\hat{\beta}_{XG}\hat{\beta}_{GY}+\hat{\beta}_{XY}}$. Further comments on the interpretation of 'genetic confounding' as shared genetic effects can be found in the discussion.

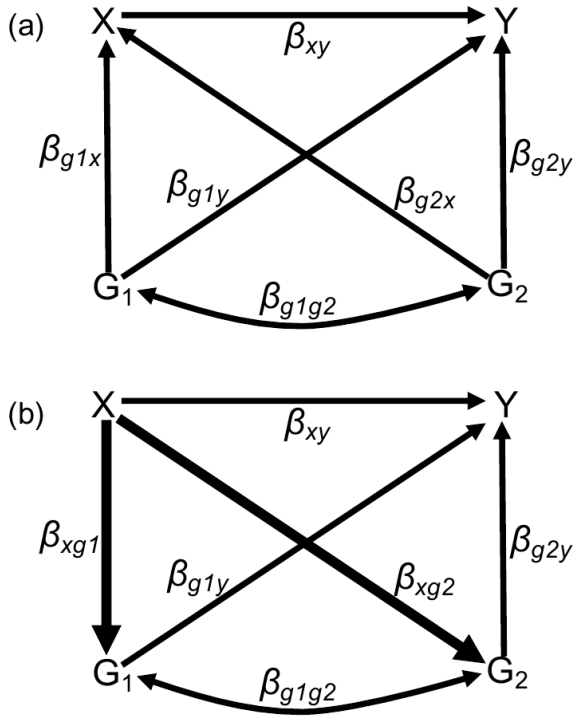
Figure 1. Genetic confounding, one polygenic score case.



Caption. Figure 1 (a) represents the underlying causal model. (b) represents the model to calculate the confounding effect by treating G as a 'mediator'. Of note is that the commonly-used terminology 'genetically-mediated' can be confusing. Although 'genetically-mediated' makes sense statistically, conceptually, a mediator is on the causal pathway from the predictor to the outcome. However, because germline genetic variants are set at conception and do not change throughout the lifespan, posterior risk factors (e.g. individual alcohol intake) cannot influence health outcomes (e.g. depression) through modifying germline genetic variants.[20] Although statistically treated as a mediator here to estimate confounding, conceptually G does not qualify as a true mediator.

When the polygenic scores for the predictor (G1) is different from the polygenic score for the outcome (G2), the confounding effect is estimated in a similar fashion as the sum of all the indirect effects from X to Y through G1 and/or G2 (Figure 2a and 2b).

Figure 2. Genetic confounding, two polygenic score.



Caption. Figure 2a represents the underlying causal model. Figure 2b represents the model to calculate the confounding effect, which is equal to: $\hat{\beta}_{XG_1}\hat{\beta}_{G_1Y}+\hat{\beta}_{XG_2}\hat{\beta}_{G_2Y}$

Genetic confounding effects were calculated for all three developmental outcomes:

- Maternal education to child educational achievement using the best-fitting polygenic score for years of education (as in Figure 1).
- Maternal education to child BMI using best-fitting polygenic scores for years of education (G1) and BMI (G2) (as in Figure 2)
- Maternal education to child ADHD symptoms using best-fitting polygenic scores for years of education (G1) and ADHD symptoms (G2) (as in Figure 2)

In these analyses (Figures 1 & 2), the effect size of X on Y decreases as a function of the strength of shared genetic effects. However, this approach does not account for all shared genetic effects. This is because polygenic scores based on current GWAS capture a relatively small amount of all genetic influences. For example, the current polygenic score for BMI explains around 6% of the variance in BMI in TEDS, far less that SNP-based and twin heritability estimates of BMI heritability. The sensitivity analysis we propose contributes to address this issue.

Sensitivity analysis

The sensitivity analysis aims to answer the following question: is it likely that X is associated with Y after we control for all shared genetic effects? To say it otherwise, to what extent would $\hat{\beta}_{XY}$ decrease if we were to control for 'perfect' polygenic scores capturing all genetic influences on X and Y rather than a small fraction. This is done by estimating $\hat{\beta}_{XY}$ under plausible scenarios that combine information on: 1) existing polygenic scores; 2) heritability estimates; 3) genetic correlations.

Single polygenic score. When predictor and outcome are of similar nature – here maternal education and child educational attainment – one polygenic score is used in the sensitivity analysis. In the present

case, a polygenic score for the child, derived from the GWAS of years of education, predicts a substantial amount of variance both in child GCSE but also in maternal education, confounding the effect of maternal education. The effect of maternal education on child educational attainment can be first adjusted for the observed best-fitting polygenic score. However, this polygenic score does not capture all the heritability of the outcome and therefore incompletely adjusts for genetic confounding. The sensitivity analysis consists in re-examining the effect of maternal education under scenarios where the polygenic score could capture additional variance in GCSE, i.e. up to SNP-based and twin-based estimates of heritability.

Figure 1a shows the underlying model of relationships between the polygenic score (G), the predictor (X) and the outcome (Y). We can obtain an estimate of the adjusted estimate of X on Y based on the observed paths available with the following expression:

$$\hat{\beta}_{XY} = \left(\hat{r}_{XY} - \hat{r}_{GX}\hat{r}_{GY}\right) / \left(1 - \hat{r}_{GX}^2\right) \quad (1)$$

Where $\hat{\beta}_{XY}$ stands for the adjusted estimate and $\hat{r}$ for observed associations. Details are presented in the supplementary material eFigure 4.

When using the best-fitting polygenic score $\hat{r}_{GY}$ is simply the observed standardized association between the polygenic score and Y. Under the sensitivity analysis scenarios $\hat{r}_{GY}$ is replaced by increasing values reflecting the additional variance captured in the outcome, for example √(0.30), the path value corresponding to 30% of the variance explained by genetic influences on Y. The path to the predictor ($\hat{r}_{GX}$) is also assumed to increase to k*√(0.30), where k reflects the ratio between the path to the predictor and the path to the outcome $k = \hat{r}_{GX} / \hat{r}_{GY}$. The value of k is obtained from $\hat{r}_{GX}$ and $\hat{r}_{GY}$ derived from the observed best-fitting polygenic score. Note that if X and Y are measured in the same individuals and the polygenic score for the outcome is used in the sensitivity analysis then: the minimum value for k is 0 (when the polygenic score for the outcome explains no variance in the predictor leading to no induced genetic confounding), and the maximum is k = 1 in the unlikely case that the polygenic score for Y explains as much variance in the predictor as it does for the outcome.

Next, to estimate the quantity of interest, $\hat{\beta}_{XY}$, under these different scenarios, two approaches are possible. First, as shown in eFigure 4, it is straightforward to express the model-based adjusted path estimates as a function of the observed bivariate estimates and the paths under different sensitivity scenarios. These model-based adjusted paths can then be used in a model simulating both X and Y and estimating $\hat{\beta}_{XY}$ and confidence intervals. Second, using structural equation modelling, the model-based paths can be estimated from a correlation matrix. Consequently, a series of correlation matrices are created under the different scenarios, prior to estimating corresponding adjusted βxy values. This matrix approach is retained throughout as it easily generalizes to the two polygenic scores case (detail in Supplementary material eFigure 4).

Complete genetic confounding. In equation (1), the association between X and Y is completely genetically confounded when the adjusted effect $\hat{\beta}_{XY} = 0$. We can then express the observed standardized association as a function of the heritabilities of X and Y under complete genetic confounding as:

$$\hat{r}_{XY} = \hat{r}_{GX}\,\hat{r}_{GY} = \sqrt{\left(h_x^2\right)}\sqrt{\left(h_y^2\right)} \quad (2)$$

Logically, we find that when the adjusted effect of X on Y is null, then $\hat{r}_{XY}$ is equal to the indirect path through G (i.e. genetic confounding). In the special case when X and Y are the same trait in parent and child (e.g. BMI in the mother and in the child), and assuming constant heritability across generations (i.e. equal heritability of X and Y),we thus obtain:

$$\hat{r}_{XY} = 0.5\,h_y^2 \quad (3)$$

This means that the adjusted effect of X on Y is likely to be null whenever the observed association does not exceed half of the trait heritability. As such, reported associations between maternal (or paternal) and child traits can be assessed against Figure 3 and if they lie in the shaded area, it is likely that they can be entirely accounted by genetic confounding. Of note is that associations not in the shaded area can still be confounded by environmental risk factors. See supplementary material for additional details on

equations (2) and (3). Note that in this special case when X and Y are measured in the mother and the child, then the minimum for *k* is still 0; however, the maximum if X and Y measure the same trait, should be *k* = 0.5 (detail in supplementary material).

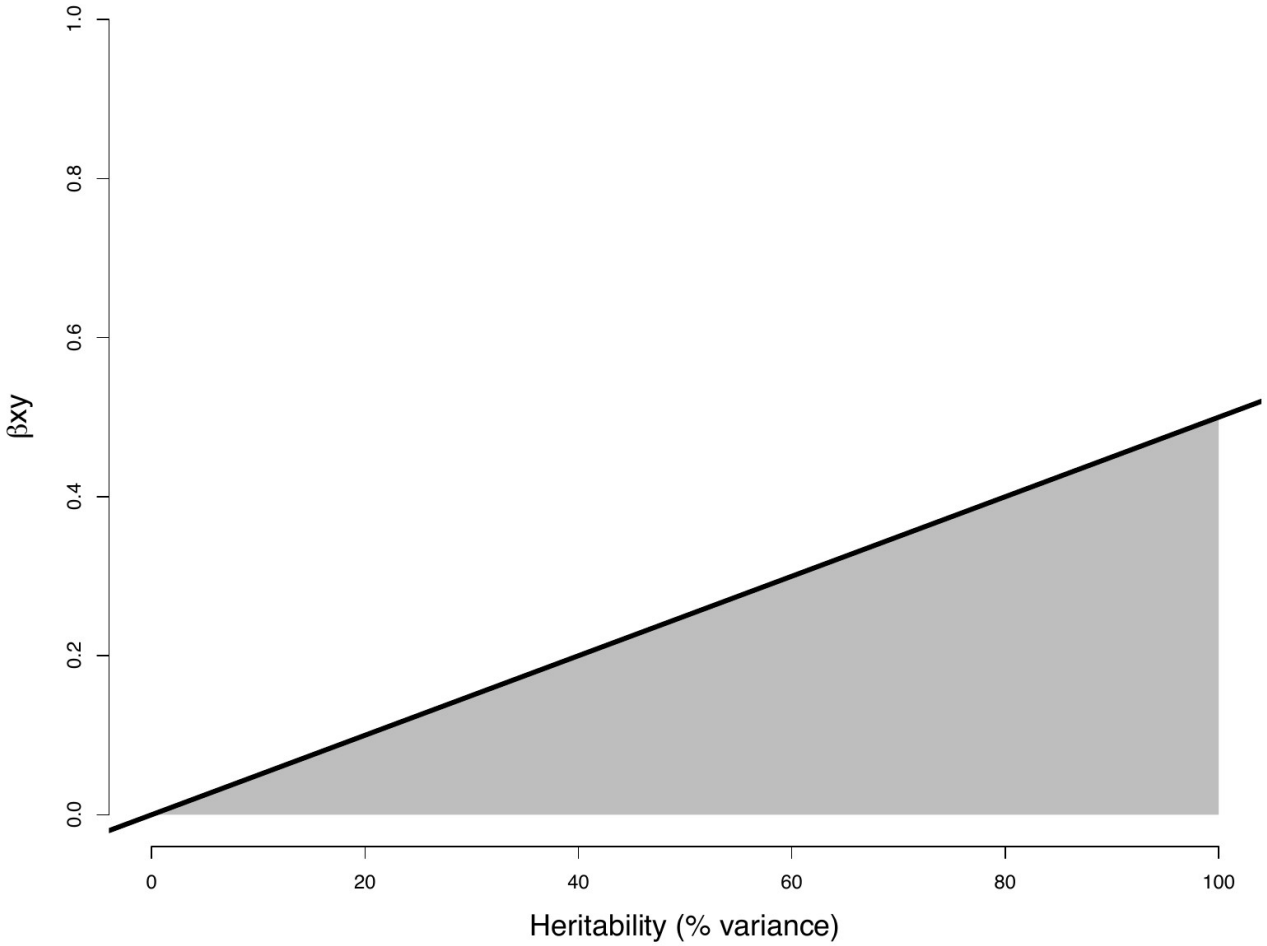


Figure 3. The role of genetics in explaining phenotypic associations between parent and child
Caption. Standardized observed associations between the same traits in the mother (or father) and the child are represented as a function of trait heritability. An observed association of 0.20 with trait heritability of 0.60 is likely entirely explained by genetic confounding. Conversely, an association of 0.40 with heritability of 0.40 is unlikely to be entirely explained by genetics.

The two polygenic scores case

When predictor and outcome are different variables – for example maternal education and child BMI – two polygenic scores are used in the sensitivity analysis, as shown in Figure 3. In theory, if we had a polygenic score capturing all genetic influences for Y, this score would also capture all the shared genetic variance between Y and X, and we could simply use the one polygenic score case above. In

practice, polygenic scores do not capture all genetic influences on their respective phenotypes and are differentially powered, which is why we use two polygenic scores in the sensitivity analysis. In the two polygenic scores case, new parameters are introduced: (i) the genetic correlation between the two polygenic scores $\hat{\beta}_{G_1G_2}$; (ii) the cross paths, i.e. the paths from each polygenic score to the other phenotype ($\hat{\beta}_{G_1Y}$ and $\hat{\beta}_{G_2X}$). Due to these new parameters, the derivation of $\hat{\beta}_{XY}$ becomes considerably more complex than for the single case polygenic score. A simplifying assumptions is to assume that the cross paths are null. This assumption is plausible to the extent that the influences of one polygenic score on the other phenotype is entirely captured by the genetic correlation and the direct genetic influences. For example, if the polygenic score for BMI explained the entire heritability of BMI (i.e. perfect $\hat{\beta}_{G_2Y}$), then the polygenic score for maternal education would not add to the prediction of BMI, so that cross path ($\hat{\beta}_{G_1Y}$) would be entirely accounted for through the genetic correlation ($\hat{\beta}_{G_1G_2}\hat{\beta}_{G_2Y}$) leading to a null adjusted cross path ($\hat{\beta}_{G_1Y}$ = 0, note that the *observed* cross path is not expected to be null). Empirical findings presented below support this assumption to a certain extent. eFigure 5 presents the expression of $\hat{\beta}_{XY}$ under this assumption. However, current polygenic scores do not capture all genetic influences and can be differentially powered (e.g. leading to a situation where $\hat{\beta}_{G_1Y}$ is superior even to $\hat{\beta}_{G_2Y}$). In this situation, we cannot expect cross paths to be entirely null. Consequently, we adopted the structural equation modelling approach based on a correlation matrix, as it does not impose null cross paths. A maximum likelihood estimator is then used to estimate model-based adjusted paths based on the correlation matrix. In the two polygenic score cases, we also need parameters similar to *k* to derive the two cross paths to input in the correlation matrix; the values of these parameters *m* depend on genetic correlation and on the relevant heritability estimate, such that :$m_{G_2X}=\hat{r}_{G_2X}/\left(\hat{r}_{G_1X}\hat{r}_{G_1G_2}\right)$ and $m_{G_1Y}=\hat{r}_{G_1Y}/\left(\hat{r}_{G_2Y}\hat{r}_{G_1G_2}\right)$. Details on *m* and how to specify the correlation matrices can be found in the supplementary material.

Results

Observed and heritability-based scenarios

Observed scenarios were based on polygenic scores. As shown in Table 1, the best-fitting polygenic scores derived from GWAS for years of education, BMI and ADHD, explained a substantial amount of the variance in TEDS for educational achievement (for a threshold of p = .158), BMI (threshold: p = .20) and ADHD symptoms (threshold: p = 0.358). All three were highly significant (larger p value for ADHD = 1.6e-20).

Two main heritability-based scenarios were used: (i) SNP-based heritability; (ii) Twin-based heritability. Table 1 shows parameters for SNP-based and twin-based scenarios. SNP-based heritability estimates were obtained through LD score regression[21,22], based on LD Hub for years of education and BMI and the latest ADHD GWAS for ADHD[14]. Twin estimates were derived from TEDS and from the literature (see Table 1 note). Table 1 also shows genetic correlations between years of education and BMI, and years of education and ADHD.

Table 1: Heritability and genetic correlation under different scenarios

	Heritability (% variance)			Genetic correlation	
	Education	BMI	ADHD	Education-BMI	Education-ADHD
Best-Fitting Polygenic score	11.9	6.3	1.3	-0.185	-0.184
SNP-based scenario	12.4	18.6	21.6	-0.279	-0.535
Twin scenario	63.0[1]	64.0	62.0[2]	-0.045[3]	-0.444[3]

[1]Heritability of the GCSE score estimated in TEDS was used. [2]Twin estimates for ADHD in TEDS are superior to > .80.[9] However, a twin meta-analysis has argued that commonly reported heritability estimates for ADHD are biased, and estimated broad-sense heritability to be 62%, value that is used here[23]. [3]As maternal education does not vary within family, it is not possible to directly estimate the genetic correlation between maternal education and child BMI and ADHD in TEDS. When using GCSE as a proxy, both twin estimates of genetic correlations were lower than SNP based estimates using LD score regression. Power was especially low for education-BMI given the low phenotypic correlation. Therefore, in the sensitivity analyses, we used SNP-based instead of twin-based genetic correlations for our two main scenarios.

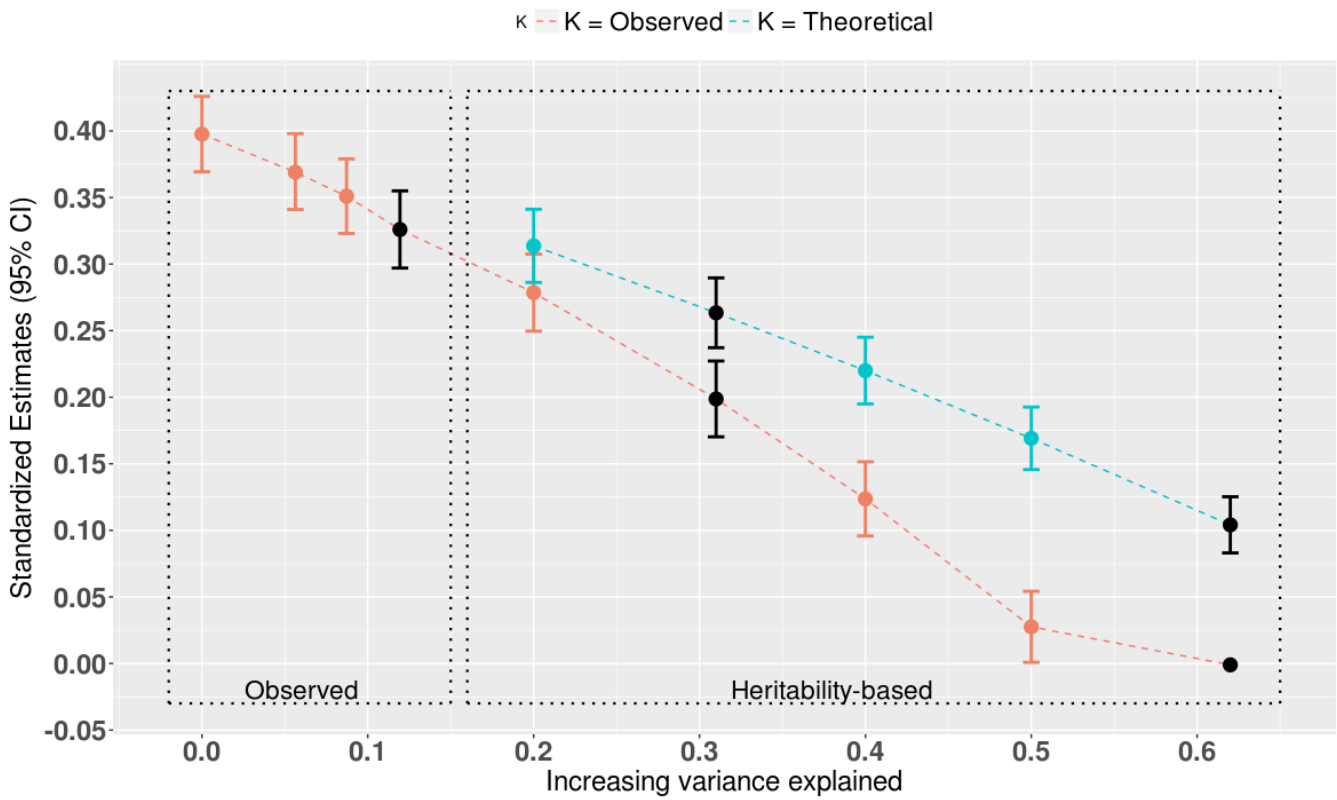
Genetic confounding and sensitivity analyses

Single polygenic score: child educational attainment

The observed standardised estimate of the relationship between maternal education and child educational achievement was β_{xy} = 0.40 (95% Confidence Intervals [CI]: 0.37, 0.43). Using the best fitting polygenic score for years of education, the effect explained by genetics was estimated as 0.07 (0.06, 0.08), corresponding to 18.1% of the total effect. After taking this genetic confounding effect into account, the relationship between maternal education and child educational achievement was reduced to 0.33 (0.30,0.36).

The sensitivity analysis is represented in Figure 4, including the standardized estimates of the effect of maternal education on child achievement as a function of the variance explained by observed polygenic scores on GCSE and under heritability-based scenarios. We first re-estimated the effect of maternal education on achievement by adjusting for observed polygenic scores at different p value thresholds (i.e. explaining different amounts of variance in GCSE). We then estimated what the effect of maternal education on achievement would be under scenarios where polygenic scores could capture additional variance in educational outcomes (see methods). Note that the SNP-heritability scenario for years of education (12.4%) is very close to the variance effectively captured in TEDS by the polygenic score (11.9%) (See Table 1). Therefore, we estimated SNP-heritability scenario based on the SNP-heritability of GCSE, which was previously estimated in TEDS to be 31%[11]. Under this scenario the effect further decreased to 0.20 (0.17;0.23). The effect estimate was null under the twin-heritability scenario when using *k* estimated from the best-fitting polygenic score.

Figure 4: Gsens analysis of the effect of maternal education and child educational attainment



Caption. Estimated standardized effect of maternal education on child educational attainment (Y axis) after accounting for shared genetic effects using observed polygenic scores and heritability-based scenarios explaining an increasing percentage of variance (X axis). Each estimate is computed independently using the procedure described in methods. Therefore, estimates under the SNP-heritability or twin-heritability scenarios are not based on a direct extrapolation of what was observed for polygenic scores. Point estimates and confidence intervals in black represent estimates of interest, from left to right: 1: the best-fitting polygenic score; 2: SNP-heritability of educational achievement as assessed by GCSE in TEDS; 3: twin-heritability scenario. A lower bound of 0 was imposed on the estimate, which is reached for the twin estimate of heritability (63%). The line *k* = Observed corresponds to heritability-based scenario using *k* values derived from observed polygenic scores (i.e. average of the values of *k* calculated for each of the four polygenic scores). *k* = theoretical corresponds to the value of *k* if the same trait was measured in parents and children and the heritability was the same in parents and children, i.e. k = 0.5.

Note that the average *k* derived from the observed polygenic score was *k* = 0.75, superior to the expected value of 0.5 expected when X and Y are the same trait measured in parents and children. In addition to sample-specific findings, this can be because the polygenic score used was derived from the GWAS which measured years of education in adults, i.e. closer to the maternal education phenotype than to child achievement phenotype used for Y. A similar finding was observed by Bates et al.[24] When

using k = 0.5 under the twin-heritability scenario, the value of $\hat{\beta}_{XY}$ is still considerably reduced compared to the phenotypic estimate but remains significant, with 0.10 (0.08;0.12).

Two polygenic scores: BMI and ADHD

The observed estimate of the relationship between maternal education and child BMI was β = -0.087 (-0.119, -0.055). Using the best fitting polygenic scores for years of education and BMI, the genetic confounding effect was estimated at -.021 (-.028, -0.014), corresponding to 24.3% of the total effect. After taking this genetic confounding effect into account, the relationship between maternal education and child BMI was -0.066 (-.098,-.035). Table 2 presents model parameters under different scenarios, with increasingly predictive observed polygenic scores as well as two heritability-based scenarios. The first scenario relied on SNP-based heritability estimates for years of education and BMI and their genetic correlation (see Table 1). In that scenario, the relationship between maternal education and child BMI further decreased to -0.047 (-0.076,-0.017). In the twin heritability scenario, the estimate was null, meaning that, under this scenario, the entire association between maternal education and child BMI is accounted for by genetic confounding (Table 2).

Table 2 Sensitivity analysis for BMI

	β_{xy}[1]	β_{g2y}[1]	β_{g1x}[1]	β_{g1g2}[1]
Unadjusted for PS scores[2]	-0.087 (-0.119, -0.055)	-	-	-
Observed polygenic scores				
PS[3] (p=0.000001)	-0.075 (-0.107,-0.043)	0.219 (0.191,0.247)	0.170 (0.143,0.197)	-0.122 (-0.145,-0.098)
PS (p=0.0001)	-0-0.073 (-0.104,-0.041)	0.237 (0.209,0.265)	0.211 (0.184,0.238)	-0.155 (-0.178,-0.131)
PS(p = 0.01)	-0.070 (-0.101,-0.039)	0.241 (0.213,0.269)	0.265 (0.238,0.292)	-0.183 (-0.207,-0.159)
Best-Fitting PS[4]	-0.066 (-0.098,-0.035)	0.247 (0.219,0.275)	0.283 (0.256,0.31)	-0.185 (-0.208,-0.161)
Heritability-based with *m* from best-fitting				
SNP scenario[5]	-0.047 (-0.076,-0.017)	0.415 (0.384,0.445)	0.167 (0.134,0.2)	-0.279 (-0.313,-0.245)
Twin scenario[5]	0 (0,0)	0.774 (0.754,0.794)	0.376 (0.345,0.407)	-0.279 (-0.313,-0.245)
Heritability-based with *m* = 1				
SNP scenario	-0.066 (-0.095,-0.037)	0.427 (0.398,0.457)	0.176 (0.144,0.208)	-0.279 (-0.313,-0.245)
Twin scenario	0 (0,0)	0.800 (0.781,0.819)	0.397 (0.366,0.428)	-0.279 (-0.313,-0.245)

Note. [1]βxy: effect of maternal education on BMI; βg2y: BMI polygenic score on observed BMI, the squared estimate yields the variance explained; βg1x: Years of education polygenic score on maternal education. For the heritability-based scenarios, the path entered in the simulation is based on the heritability of years of education adjusted as the child genotype is used (see Supplementary Material). Bg1g2: genetic correlation estimated by the correlation between the polygenic scores or by the LD score regression in simulated scenarios. [2]Initial bivariate observed estimate of βxy before accounting for genetic confounding. [3]All observed scenarios are based on polygenic scores computed for different pvalue thresholds given in brackets (the same for BMI and years of education), which explain increasing percentages of variance. [4]Based on observed values for the best fitting score for years of education (p = 0.158) and BMI (p = 0.20). [5]Genetic correlation is equal in both scenarios, see Table 1 note).

The observed estimate of the relationship between maternal education and child ADHD was β = -0.127 (-0.155, -0.098). Using the best fitting polygenic scores for years of education and ADHD, the genetic confounding effect was estimated to -0.018 (-0.026, -0.010), corresponding to 14.3% of the total effect. After taking shared genetic effects into account, the relationship between maternal education and child ADHD was -0.109 (-0.138; -0.079). Table 3 presents model parameters under different observed and simulated scenarios. In the heritability-based scenario, the relationship between maternal education and child ADHD was further reduced and null in the twin-based scenario.

Table 3: Sensitivity analysis for ADHD

	β_{xy}[1]	β_{g2y}[1]	β_{g1x}[1]	β_{g1g2}[1]
Unadjusted for PS scores[2]	-0.127(-0.155, -0.098)			
Observed polygenic scores				
PS[3] (p=0.000001)	-0.119 (-0.148,-0.089)	0.003 (-0.022,0.028)	0.175 (0.148,0.202)	-0.081 (-0.104,-0.057)
PS (p=0.0001)	-0.115 (-0.144,-0.085)	0.028 (0.003,0.053)	0.218 (0.191,0.244)	-0.112 (-0.135,-0.088)
PS(p = 0.01)	-0.112 (-0.141,-0.082)	0.066 (0.041,0.091)	0.271 (0.244,0.298)	-0.160 (-0.184,-0.136)
Best-Fitting PS[4]	-0.109 (-0.138,-0.079)	0.105 (0.08,0.13)	0.289 (0.262,0.316)	-0.184 (-0.208,-0.16)
Heritability-based with *m* from best-fitting[5]				
SNP scenario	0 (0,0)	0 (0,0)	0.162 (0.129,0.195)	-0.535 (-0.567,-0.502)
Twin scenario[6]	-	-	-	-
Heritability-based with *m* = 1				
SNP scenario	-0.084 (-0.109,-0.059)	0.457 (0.432,0.482)	0.176 (0.143,0.209)	-0.535 (-0.567,-0.502)
Twin scenario	0 (0,0)	0.787 (0.77,0.805)	0.397 (0.366,0.428)	-0.535 (-0.567,-0.502)

Note. [1-3]See Table 2. [4]Based on observed values for the best fitting score for ADHD (best threshold p = 0.351). [5]Genetic correlation is equal in both scenarios (see Table 1 note). [6]The twin scenario did not converge for *m* values based on the best-fitting polygenic score, see section 'additional parameters and constraints' for comments.

Two polygenic scores: additional parameters and constraints

Compared with the one polygenic score scenario, the model with two polygenic scores described above includes new parameters. First the size of the genotypic correlation between the two traits plays a role. The genetic correlation effect size increases with increasingly predictive polygenic scores. As shown in Table 3, the genetic correlation between the polygenic scores for years of education and ADHD goes from -0.081 for less predictive polygenic scores to -0.184 for the best fitting polygenic scores and to -0.535 when using LD score regression. The cross paths also represent new parameters (i.e. direct path from the polygenic score of BMI to education and vice versa). Effect sizes for cross paths are rather small but precision also increases with more accurate polygenic scores, for example observed bivariate cross paths from polygenic scores of years of education to BMI phenotype ranged between 0.009 and -.027 (see Supplementary eTable 1). These increasingly negative observed cross paths are consistent with the expected negative correlation between the polygenic score for education and BMI. Increasingly negative observed cross paths are observed in all cases (eTable 1 to eTable 4). On the other hand, the adjusted cross paths are close to 0 and not increasing, consistent with the fact that the observed cross paths are largely generated by the genetic correlation and the relevant heritability estimates. In some

cases (eTable 1 and 4), cross paths even flip to a positive sign, which is implausible (e.g. higher polygenic score for education predicting higher BMI or higher ADHD). To avoid such implausible cases, an upper bound of 0 was imposed on parameters for the cross paths ($\hat{\beta}_{G_1Y}$, $\hat{\beta}_{G_2X}$). Similarly, a bound of 0 was imposed on the main estimate ($\hat{\beta}_{XY}$), corresponding to the case where shared genetic effects accounts for all the association between X and Y (Table 2 & 3).

The heritability-based scenarios depend substantially on the values of *m*. Whether for BMI or ADHD, the value of $\hat{\beta}_{XY}$ was null under twin-based heritability, whatever the value of *m*. However, for SNP-heritability, the point estimate of $\hat{\beta}_{XY}$ changed from -0.047 when using *m* values based on the best-fitting polygenic scores to -0.066 when using the minimum values of *m* = 1 (Table 3). For ADHD, the difference was more substantial as the point estimates for $\hat{\beta}_{XY}$ varied from -0.084 to 0. This is because, in the case of ADHD, *m* values were imprecisely estimated as the polygenic score for ADHD predicts little variance in ADHD. The resulting large *m* value leads to an impossible value of the cross path to input in the heritability-based scenario (i.e. standardized absolute cross path superior to 1). This explains the non-convergence of the model in the twin heritability-based scenario based on *m* values from the best-fitting polygenic score reported in Table 3. Further considerations on the estimation of *m* and its consequences on the estimates are proposed in the supplementary material. Importantly, the method we propose here, as most sensitivity analyses, thus offers a range of possible values of a parameter of interest (here $\hat{\beta}_{XY}$) under different scenarios, rather than a unique point estimate. In addition, this range will become narrower as the power of GWAS increases, which will lead to better estimated heritabilities, genetic correlations and *m* values. In the meanwhile, results should be considered with caution in particular for polygenic scores with very low predictive power (e.g. ADHD).

Discussion

In the present study, we combined polygenic scores with heritability estimates in a sensitivity analysis – *Gsens* – aiming to gauge to what extent shared genetic effects can account for observed epidemiological associations. The genetic sensitivity analysis we propose adds a new tool to probe the robustness of findings regarding the role of genetics in associations between risk factors and outcomes. This approach only requires a genotyped cohort with adequate phenotypic measurements, which is increasingly the rule rather than the exception. It is therefore possible to envisage that such sensitivity to genetic confounding analysis becomes routine in the not too distant future. Below, we first discuss empirical findings regarding the associations between maternal education and child educational attainment, BMI, and ADHD. We then discuss the interpretation and applicability of *Gsens.*

Maternal education and developmental outcomes

Findings show that the association between maternal education and both child educational attainment and BMI were still significant under a SNP-heritability scenario but were null under a twin-heritability scenario. The association between maternal education and ADHD was null even when assuming only SNP-heritability. Overall, the observed association between maternal education and these three developmental outcomes may largely be due to shared genetic effects.

Relevant to our findings is previous research using causal inference designs to investigate the effect of parental educational attainment on child educational attainment (although note that our outcome was educational achievement rather than attainment). In particular, a systematic review on the topic has summarized evidence from twin and adoption designs, as well as non-genetic instrumental variable estimations[25]. The systematic review of findings from such designs suggest that intergenerational associations between parent and child educational attainment are largely driven by selection effects, including genetic confounding; it suggests only small but still significant causal effects. A new method –

the ‘virtual-parent design’ – has recently emerged, which consists in splitting parental genetic variants associated with a parental risk factor into variants transmitted and nontransmitted to the child[24,26]. Parental polygenic scores including only nontransmitted variants are free from genetic confounding and index plausible causal effects of the parental risk factor on the child outcome. Empirical findings implementing this method in education research suggests substantial genetic confounding and small but still significant causal effects of parental attainment on child attainment (as index by nontransmitted polygenic scores). Our findings on educational attainment are overall consistent with this literature. Although shared genetic effects accounted entirely for the association between parental education attainment and child achievement, we detected a small but significant effect when using the upper theoretical limit of the *k* value, consistent with previous findings. In addition, scenarios based on slightly lower heritability estimates also yielded small but significant effects, raising the possibility that null findings resulted from overestimated twin-heritability estimates, a possibility further discussed below. Taken together, this set of findings represent a clear call for caution when interpreting non-genetically informed epidemiology studies on the role of maternal education.

Interpreting the sensitivity to genetic confounding analysis

Two key points regarding the interpretation of *Gsens* findings must be highlighted. First, the reliability of findings depends on the accuracy of heritability estimates. Overestimated heritability would lead to underestimating the plausible causal role of risk factors. Whether heritability estimates are accurate remains an open debate. The notion of ‘missing heritability’ is central in this debate (i.e. the double gap between the variance explained by polygenic scores versus SNP-heritability and by SNP-heritability versus family-based estimates of heritability)[27,28]. Recent research suggests that current heritability estimates include environmentally mediated effects[29]. Once such effects are removed, heritability - defined as the fraction of trait variation due to direct genetic effects –can be substantially lower for traits like education or BMI, thereby affecting the findings from *Gsens*. Increasing power in GWAS and

improved estimation methods will contribute to resolving the issue of missing heritability, and provide more accurate parameters for *Gsens*. In addition, power increase in GWAS will improve the reliability of *Gsens* in the following ways: (i) by increasing the predictive power of polygenic scores and therefore the accuracy of observed scenarios; (ii) by improving the accuracy of parameter estimates for the sensitivity analysis – including values of *k* and *m*, SNP-heritability, and genetic correlation.

Second, *Gsens* is fundamentally different from instrumental variable approaches like Mendelian Randomization (MR). In MR, genetic variants influencing risk factors (i.e. the genetic instrument, corresponding to $\hat{\beta}_{GX}$ in Figure 1a) are used to estimate causal effects of risk factors on outcomes. Conversely, *Gsens* aims to remove all shared genetic effects between risk factors and outcomes ($\hat{\beta}_{XG}\hat{\beta}_{GY}$). Importantly, if MR assumptions are satisfied, the effect of the genetic instrument on Y is entirely mediated by the risk factor X (i.e. exclusion restriction assumption). In other words, there is no direct effect of G on Y and the observed $\hat{r}_{GY}$ reflects the indirect effect of the genetic instrument through the risk factor $\hat{\beta}_{GX}\hat{\beta}_{XY}$. In this case, there is no genetic confounding per se as G predicts X but does not directly predict Y. Our estimate of shared genetic effects thus amalgamates both genetic confounding and indirect genetic effects through the risk factor, which is not per se confounding (it other words, it amalgamates unmediated pleiotropy and mediated pleiotropy).[4] This is not specific to genetic confounding and can happen when controlling for potential environmental confounders, which can include environmental sources of variance in the predictor that behave as instrumental variables. Similarly, in a discordant twins design, in order to strengthen causal inference, we compare identical twins exposed or non-exposed to a given risk factor. In such a design, all shared genetic effects are controlled for, including genetic confounding and genetic effects that directly affect the risk factor but only indirectly affect the outcome. As such, in *Gsens* as in the discordant twins design, we must assume that there will be enough environmental variance left in the risk factor to detect the causal effect on the outcome. A caveat of this approach is that it is theoretically possible that all the variance in the risk

factor is genetic in origin but that this risk factor still has a direct causal effect on the outcome; this causal effect would thus remain undetected in a discordant twins analysis and would be ruled out by *Gsens*. However, genetic influences almost never explain all the variance in risk factors and, when heritability is high, it is likely that adequate genetic instruments can be found to conduct MR analyses.

Applicability of the sensitivity to genetic confounding analysis

A wealth of methods have been developed to account for confounding in observational studies, including genetically informed methods[4]. Among such methods, MR has become an approach of choice in the past decade. Major challenges in MR analyses are: (i) weak instrument(s), i.e. the genetic variant(s) do(es) not strongly predict the risk factor; (ii) unmediated pleiotropy, i.e. the instrument predicts the outcome not only via the risk factor; (iii) the genetic instrument is not independent from confounders. MR can be implemented to test the effect of proteins, such as the C-Reactive Protein (CPR), an inflammatory marker, on a variety of health outcomes[30,31]. In such a case, variants in the CRP gene naturally strongly predict variation in circulating CRP levels, ruling out the problem of weak instruments. As such variants in the CRP gene are directly implicated in the synthesis of CRP, it is reasonable to assume that their effect on later outcomes is mediated by their effect on CRP. Although the third challenge cannot be fully tested, genetic variants in the CRP gene have been shown to be independent of a number of observed potential confounders that are associated with circulating CRP levels[32]. MR has also been implemented to test the causal effect of other molecules, relatively proximal to the genome like lipids, where the path from the genetic sequence to the risk factor is relatively well characterised. The risk of unmediated pleiotropy is greater in this case and MR analyses examining separately the effect of one lipid can be biased. For example, High-Density Level cholesterol (HDL-c), appears protective for cardiovascular diseases when considered separately, but this effect disappears when HDL-c is considered jointly with other lipids, accounting for plausible pleiotropy pathways[4]. More distal from the genome are complex individual traits such as BMI or intelligence, where the pathways

from the many genetic variants used as instruments to their corresponding risk factors are poorly characterised. Considerable methodological developments using multiple instruments to statistically deal with MR challenges have recently been proposed[4]. As shown by the lipid example, the application of such methods may still lead to biased estimates when plausible pleiotropic pathways are not directly modelled (which becomes increasingly difficult as the complexity of the risk factor increases). Thus, the reliability of these improved MR estimators to establish the effect of complex individual risk factors remains difficult to establish. An additional level of complexity arises for 'environmental' risk factors such as education, income, urbanicity or exposure to bullying. These risk factors are by nature 'external' to the individual and largely dependent on the wider social context (e.g. exposure to bullying is not an individual characteristic but implicates acts of others towards the individual; and assessing years of education is only possible in a society with an education system in place). However, such 'environmental' risk factors have been demonstrated to be partly, and often substantially heritable[33]. The concept of gene-environment correlation clarifies how such risk factors can still be heritable: genetic variants first influence a number of 'internal' individual characteristics – biological intermediate phenotypes, cognitive traits, personality –, which in turn contribute to shaping the individual environment and thus influence these 'external' risk factors. Although such dichotomy between internal and external risk factors is reductionist, it is used here simply to highlight that genetic influences on some risk factors can only manifest through indirect pathways through 'internal' risk factors, which are themselves complex (e.g. intelligence). For these risk factors, GWAS are often not available and when they are, implicated genetic variants will only reflect a multitude of indirect pathways. For example, in a recent study, we showed that polygenic scores for BMI, education, depression, ADHD, and risk taking were associated with the likelihood of being exposed to bullying[34]. Using results from a GWAS of bullying victimization to assess its effect would amount to using many unspecific instruments related to other complex individual traits. Genetic influences on such environmental influences are not only complex and indirect but they are also fairly unspecific as, for example, education, depression and risk

taking found to impact exposure to bullying can impact other environmental risk factors such as income or urbanicity. For such risk factors, the aforementioned challenges for MR are thus magnified: (i) genetic effects on such factors are by nature indirect and thus likely to be weak; (ii) top associated SNPs will reflect a number of underlying individual traits, making unmediated pleiotropy a rule rather than an exception; (iii) top associated SNPs will be associated with a number of other environmental risk factors, and therefore be unlikely independent from confounders.

In *Gsens,* whether variants composing the polygenic scores are valid or invalid instruments is unimportant. In addition, Gsens can be implemented even when GWAS for the risk factor are not available, as long as a GWAS for the outcome is. For example, Gsens could be applied to test whether the association between urbanicity and schizophrenia is susceptible to shared genetic effects. In this case, the single polygenic score case would be implemented, using a polygenic score for schizophrenia and testing for its association with urbanicity as a base for the sensitivity analysis. As noted in the method section, genetic factors solely affecting urbanicity and not schizophrenia do not confound the association (and vice versa). As such, only one polygenic score for either the risk factor or the outcome is theoretically needed, to the extent that it captures shared genetic effects appropriately. We therefore propose that Gsens can be conceived as a complementary method, suited for complex environmental risk factors that are of interest for health and social sciences.

Limitations and research avenues

As all sensitivity analyses in observational studies, the sensitivity to genetic confounding analysis cannot provide a definite answer to the question of causality. The convergence of findings across designs and methods – i.e. triangulation – will provide the most robust evidence of causal effects. TEDS does not include maternal (or paternal) genotype, which prevented us from modelling their role directly. Of note is that the sensitivity to genetic confounding analysis can be implemented when both

the risk factor and the outcome are measured for the same individual, without modelling parental effects.

As a first future research avenue, a sensitivity analysis based on jointly modelling shared genetic effects and observed environmental confounders could be envisaged. A second research avenue could consist in using mixed-linear modelling[35] instead of polygenic scores. Genomic Restricted Maximum Likelihood (GREML) estimates SNP-heritability based on a Genomic Relatedness Matrix of distantly related individuals[36,37]. Extensions of GREML enable a better identification of causal SNPs among the many SNPs in LD tagged by GWAS. Mixed linear models are a method of choice to perform a conditional analysis estimating the effect of each SNP, while adjusting for all genotyped markers[35]. We note that the candidate causal SNP can be replaced by any environmental exposure of interest, to examine the effect of this exposure on an outcome, independently of the joint effect of genotyped markers. This can be thought of as adjusting for a polygenic score that would effectively capture as much of the variance in Y as SNP-heritability. Mixed linear models accommodate covariates, enabling the inclusion of key environmental confounders. Contrasting models with and without modelling genotyped markers should help in determining the role of genetic confounding for the association under scrutiny. Sensitivity analyses similar to the ones described above could also be performed. Finally, such models could be extended to examine the intergenerational transmission of risk factors based on multivariate GREML methods[38].

References

1. Fisher, R. A. Alleged dangers of cigarette-smoking. *Br. Med. J.* **2**, 4 & 297–298 (1957).
2. Greenhouse, J. B. Commentary: Cornfield, Epidemiology and Causality. *Int. J. Epidemiol.* **38**, 1199–1201 (2009).
3. Cornfield, J. *et al.* Smoking and lung cancer: recent evidence and a discussion of some questions. 1959. *Int. J. Epidemiol.* **38**, 1175–1191 (2009).
4. Pingault, J.-B. *et al.* Using genetic data to strengthen causal inference in observational research. *Nat. Rev. Genet.* **19**, 566–580 (2018).
5. Magnuson, K. Maternal education and children's academic achievement during middle childhood. *Dev. Psychol.* **43**, 1497–1512 (2007).
6. Rydell, A.-M. Family factors and children's disruptive behaviour: an investigation of links between demographic characteristics, negative life events and symptoms of ODD and ADHD. *Soc. Psychiatry Psychiatr. Epidemiol.* **45**, 233–244 (2010).
7. Galéra, C. *et al.* Early risk factors for hyperactivity-impulsivity and inattention trajectories from age 17 months to 8 years. *Arch. Gen. Psychiatry* **68**, 1267–1275 (2011).
8. Parizkova, J. Impact of education on food behaviour, body composition and physical fitness in children. *Br. J. Nutr.* **99 Suppl 1**, S26-32 (2008).
9. Pingault, J.-B. *et al.* Genetic and environmental influences on the developmental course of attention-deficit/hyperactivity disorder symptoms from childhood to adolescence. *JAMA Psychiatry* **72**, 651–658 (2015).
10. Krapohl, E. *et al.* The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15273–15278 (2014).
11. Krapohl, E. & Plomin, R. Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Mol. Psychiatry* (2015). doi:10.1038/mp.2015.2
12. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
13. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).

14. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for ADHD. *bioRxiv* 145581 (2017). doi:10.1101/145581

15. Haworth, C. M. A., Davis, O. S. P. & Plomin, R. 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 (2013).

16. Conners, C. K. *Conners' Rating Scales-Revised: Technical Manual*. (Multi-Health System Inc., 2003).

17. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).

18. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinforma. Oxf. Engl.* **31**, 1466–1468 (2015).

19. MacKinnon, D. P., Krull, J. L. & Lockwood, C. M. Equivalence of the Mediation, Confounding and Suppression Effect. *Prev. Sci. Off. J. Soc. Prev. Res.* **1**, 173 (2000).

20. Smith, G. D. Mendelian Randomization for Strengthening Causal Inference in Observational Studies: Application to Gene × Environment Interactions. *Perspect. Psychol. Sci. J. Assoc. Psychol. Sci.* **5**, 527–545 (2010).

21. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

22. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

23. Wood, A. C., Buitelaar, J., Rijsdijk, F., Asherson, P. & Kuntsi, J. Rethinking Shared Environment as a Source of Variance Underlying Attention-Deficit/Hyperactivity Disorder Symptoms: Comment on. *Psychol. Bull.* **136**, 331–340 (2010).

24. Bates, T. C. *et al.* 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 (2018).

25. Holmlund, H., Lindahl, M. & Plug, E. The causal effect of parents' schooling on children's schooling: A comparison of estimation methods. *J. Econ. Lit.* **49**, 615–651 (2011).

26. Kong, A. *et al.* The nature of nurture: Effects of parental genotypes. *Science* **359**, 424–428 (2018).

27. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

28. Trzaskowski, M., Dale, P. S. & Plomin, R. No Genetic Influence for Childhood Behavior Problems From DNA Analysis. *J. Am. Acad. Child Adolesc. Psychiatry* **52**, 1048-1056.e3 (2013).

29. Young, A. I. *et al.* Relatedness disequilibrium regression estimates heritability without environmental bias. *Nat. Genet.* **50**, 1304–1310 (2018).

30. Davey Smith, G. *et al.* Association of C-reactive protein with blood pressure and hypertension: life course confounding and mendelian randomization tests of causality. *Arterioscler. Thromb. Vasc. Biol.* **25**, 1051–1056 (2005).

31. Pingault, J.-B., Cecil, C. a. M., Murray, J., Munafò, M. & Viding, E. Causal inference in psychopathology: A systematic review of Mendelian randomisation studies aiming to identify environmental risk factors for psychopathology. *Psychopathol. Rev.* **4**, 4–25 (2017).

32. Burgess, S. & Thompson, S. G. *Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation*. (Chapman and Hall/CRC, 2015).

33. Kendler, K. S. & Baker, J. H. Genetic influences on measures of the environment: a systematic review. *Psychol. Med.* **37**, 615–626 (2007).

34. Schoeler, T. *et al.* Identifying individual vulnerabilities associated with risk of exposure to bullying: a multi-polygenic score approach. *JAMA Psychiatry* (In Press).

35. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed model association methods. *Nat. Genet.* **46**, 100–106 (2014).

36. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

37. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565 (2010).

38. Eaves, L. J., Pourcain, B. S., Davey Smith, G., York, T. P. & Evans, D. M. Resolving the effects of maternal and offspring genotype on dyadic outcomes in genome wide complex trait analysis (‘M-GCTA’). *Behav. Genet.* **44**, 445–455 (2014).