Within-family studies for Mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases

Short: Within-family MR

Ben Brumpton* [1,2,3], Eleanor Sanderson[2,4], Fernando Pires Hartwig[2,5], Sean Harrison[2,4], Gunnhild Åberge Vie[1], Yoonsu Cho[2,4], Laura D Howe[2,4], Amanda Hughes[2,4], Dorret I Boomsma[6], Alexandra Havdahl[2,7,8], John Hopper[9], Michael Neale[10], Michel G Nivard[6], Nancy L Pedersen[11], Chandra A Reynolds[12], Elliot M Tucker-Drob[13], Andrew Grotzinger,[13] Laurence Howe[2,4], Tim Morris[2,4], Shuai Li[14,15], MR within-family Consortium, Wei-Min Chen[17], Johan Håkon Bjørngaard[1], Kristian Hveem[1], Cristen Willer[18,19,20], David M Evans[2,21], Jaakko Kaprio[22,23], George Davey Smith[2,4,^], Bjørn Olav Åsvold[1,24^], Gibran Hemani[2,4,^], Neil M Davies[2,4,^]

[1] K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Norway.
[2] Medical Research Council Integrative Epidemiology Unit, University of Bristol, BS8 2BN, United Kingdom.
[3] Clinic of Thoracic and Occupational Medicine, St. Olavs Hospital, Trondheim University Hospital.
[4] Population Health Sciences, Bristol Medical School, University of Bristol, Barley House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom.
[5]Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, Brazil.
[6] Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.
[7] Nic Waals Institute, Lovisenberg Diaconal Hospital Spångbergveien 25, 0853 Oslo, Norway
[8] Department of Mental Disorders, Norwegian Institute of Public Health, Sandakerveien 24C, 0473 Oslo, Norway
[9] The University of Melbourne, 207 Bouverie St, Carlton, Victoria, 3010, Australia
[10] Virginia Institute for Psychiatric and Behavior Genetics, Virginia Commonwealth University, Richmond, Virginia, USA
[11] Karolinska Institutet, Dept of Medical Epidemiology and Biostatistics, Stockholm, Sweden
[12] Department of Psychology, University of California Riverside, Riverside, CA, USA
[13] Department of Psychology and Population Research Center, University of Texas at Austin, 108 E. Dean Keeton Stop A8000, Austin TX 78712 USA
[14] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne 207 Bouverie Street, Carlton, Victoria 3053, Australia
Shuai Li

[15] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, United Kingdom
[17] Center for public health genomics, Department of public health sciences, University of Virginia, Charlottesville, VA, USA
[18] Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, USA.
[19] Department of Internal Medicine, University of Michigan, Ann Arbor, MI, US.
[20] Department of Human Genetics, University of Michigan, Ann Arbor, USA.
[21] University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia
[22] Department of Public Health, University of Helsinki, Helsinki, Finland
[23] Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
[24] Department of Endocrinology, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.

^ Contributed equally

* Corresponding author (email: ben.brumpton@ntnu.no, tel: +47 46229495)

Key Words: Mendelian randomization, Gene-environment correlation, Population stratification, Assortative mating, Dynastic effects, Educational attainment

Word count: 3,073

Abstract: 154

Abstract

Mendelian randomization (MR) is a widely-used method for causal inference using genetic data. Mendelian randomization studies of unrelated individuals may be susceptible to bias from family structure, for example, through dynastic effects which occur when parental genotypes directly affect offspring phenotypes. Here we describe methods for within-family Mendelian randomization and through simulations show that family-based methods can overcome bias due to dynastic effects. We illustrate these issues empirically using data from 61,008 siblings from the UK Biobank and Nord-Trøndelag Health Study. Both within-family and population-based Mendelian randomization analyses reproduced established effects of lower BMI reducing risk of diabetes and high blood pressure. However, while MR estimates from population-based samples of unrelated individuals suggested that taller height and lower BMI increase educational attainment, these effects largely disappeared in within-family MR analyses. We found differences between population-based and within-family based estimates, indicating the importance of controlling for family effects and population structure in Mendelian randomization studies.

Mendelian randomization is an approach that uses genetic variants as instrumental variables to investigate the causal effects of one trait (the 'exposure') on another (the 'outcome').[1–5] It has gained popularity due to the recent expansion in the scale of genome wide association studies (GWAS) and because it can ameliorate bias due to processes of residual confounding and reverse causation that affect most other observational approaches. In order for Mendelian randomization estimates to be valid, the genetic instrument must meet three assumptions: 1) relevance, it must associate with the exposure, 2) independence, there must be nothing that causes both the instrument and the outcome, and 3) exclusion, the association of the instrument and the outcome must be entirely mediated via the exposure. Methodologists have focused on developing methods to overcome bias in Mendelian randomization studies due to horizontal pleiotropy,[6–11] which would violate the exclusion assumption. However, in this paper our attention is focused on the second assumption: independence. We demonstrate how population and family structures can lead to violations in the second assumption, and that traditional family-based methods are well placed to rectify this problem.

Mendel's laws of genetic inheritance provide a rationale for why much genetic variation will be independent of the environment and genetic variation for other traits.[1,12] However, environmental and social factors such as assortative mating, dynastic effects, and population structure may affect the distribution of genetic variants for specific traits within populations (see **Box 1** in supplement for definitions).[13–16] **Figure 1** illustrates the impact of these processes in the context of Mendelian randomization, the commonality amongst all three processes being that they induce a spurious association between the instrumenting variant and the outcome through confounding. Assortative mating can occur when individuals select a partner based on a selected phenotype.[6,17] For example, couples tend to have more similar age, education, and body mass index than would be expected by chance.[18,19] If assortative mating arises due to individuals with a particular genetic predisposition selecting mates who have a particular genetically influenced phenotype, this can induce spurious genetic associations which can result in biased estimates from Mendelian randomization studies.[6] In addition, social homogamy may lead to people selecting partners who are similar to themselves,[20] which can compound across

generations.[6] Dynastic effects can occur when the expression of parental genotype in the parental phenotype directly affects the offspring phenotype. For example, higher educated parents might support their children's education by providing a stimulating environment, being able to afford tutoring for their child, buying homes in better school districts, or paying for private schools. Finally, residual population structure occurs when there are geographic or regional differences in allele frequency relating to a trait of interest that cannot necessarily be controlled for via principal components.[13] Confounding by population stratification,[1] in which ancestry is correlated with both phenotypes and genotypes, was a major concern during early development of Mendelian randomization.[1] However, this fear was gradually assuaged by a decade of GWAS results that were apparently reliable in the face of population structure.[21] GWAS are now performed on a huge scale; as a consequence the problem of population stratification is again of potential concern because the high statistical power of large studies renders them susceptible to bias from very subtle population structure.[13,22]

Confounding in genetic association estimates, as induced by assortative mating, dynastic effects and population stratification, can and has been resolved by using family-based study designs.[6,23,24] For example, in sibling pair studies, genetic associations at loci can be partitioned into between pair and within pair components.[23] Because genetic differences within sibling pairs reflect random independent meiotic events, within pair effects are unrelated to population stratification and most potential confounders that might influence the phenotype. Similarly, other family-based designs and within-family tests to adjust for or exploit parental genotypes exist, such as estimating maternal and offspring genetic effects using structural equation modelling,[25] quantitative transmission/disequilibrium tests,[26,27] or mother-father-offspring trios to adjust for parental genotypes.[28] Such within-family designs have been used to validate results from genome wide association studies,[29,30] obtain unbiased heritability estimates,[31] and assess causation in the classical twin design.[32,33] Yet, despite the initial extended proposal of Mendelian randomization advising that the only way to ensure true randomization was through a within-family design[1], to-date contemporary implementations using modern genomic methods have rarely been performed. The principal reason for this has

been a lack of genomic data collected from families at a scale sufficient to be suitably powered. However, as we enter the age of national scale biobanks and very large twin studies, this essential extension of Mendelian randomization is becoming feasible.

This paper presents theory and simulations that demonstrate how within-family designs can be coupled with genomic data to perform unbiased Mendelian randomization analyses. We integrate these approaches in a modular fashion alongside other methods that have been developed for pleiotropy-robust inference (i.e. to be resilient to violations of the third assumption of Mendelian randomization).[7–9,34] Using 28,777 siblings from HUNT and 32,231 siblings from the UK Biobank, we illustrate these methods empirically. First, we analyse the influences of BMI on high blood pressure and risk of diabetes as positive controls. Second, we evaluate the influences of height and BMI on educational attainment and demonstrate major discrepancies between population-based and within-family based approaches, indicating the importance of controlling for family effects and population structure in Mendelian randomization studies.

Results

We conducted a simulation to demonstrate how dynastic effects can affect Mendelian randomization estimates from unrelated individuals, and the methods that can be used to reduce these biases. Then we illustrate how these methods can be used with four empirical hypotheses examples using data from the HUNT and UK Biobank studies.

Simulation Study

In simulations where the exposure does not have a causal effect on the outcome (**Figure 2**), the standard Mendelian randomization estimates were biased with high false discovery rates in the presence of dynastic effects (false discovery rate > 0.75 when the confounders were C_x and $C_y = 0.1,\ \beta_{ux} = 0.1$, n> 10,000). The sibling and trio methods were unbiased.

Where we simulated the exposure to have a causal effect on the outcome Mendelian randomization using unrelated individuals had the highest power (**Figure 2**). However, the sibling and trio design also performed well with larger sample and effect size (power > 0.9 when sample sizes >= 10,000, dynastic effect <= 0.2, effect size = 0.05). The within-family models were substantially less powerful than standard Mendelian randomization using unrelated individuals; as usual, controlling bias comes at a cost.

Empirical study

Participants with higher BMI were more likely to have diabetes: each 1kg/m^2 increase in BMI was associated with a 0.60 (95%CI: 0.55 to 0.65, p-value<1.2×10^{-136}) percentage point increase in the diabetes risk. These differences were modestly attenuated after including a family fixed effect (0.46, 95%CI: 0.40 to 0.52, p-value=8.5×10^{-52}). The Mendelian randomization estimate using unrelated individuals suggested that each unit increase in BMI increased the risk of having diabetes by 0.82 (95%CI: 0.71 to 0.93, p-value=3.3×10^{-50}) percentage points. This estimate remained after allowing for the fixed effects of family (1.01 percentage point increase per 1kg/m^2 increase in BMI, 95%CI: 0.58 to 1.44, p-value=3.3×10^{-06}). The summary data Mendelian randomization analysis allowing for family effects estimates were similar (0.75 percentage point increase per 1kg/m^2 increase in BMI, 95%CI: 0.38 to 1.13, p-value=$7.6\text{x}10^{-05}$, $p_{\text{diff unrelated}}$= 0.74). On average the associations of the SNPs and BMI and diabetes were similar before and after allowing for a family fixed effect, falling 7% (95%CI: -5% to 20%, p-value=0.26) and increasing 11% (95%CI: -17% to 40%, p-value=0.42) respectively.

Participants with higher BMI were more likely to have high blood pressure; each 1kg/m^2 increase in BMI was associated with a 2.63 (95%CI: 2.54 to 2.72, p-value<1×10^{-300}) percentage point increase in high blood pressure risk. This association did not attenuate after including a family fixed effect (2.42, 95%CI: 2.30 to 2.54, p-value<1×10^{-300}). The Mendelian randomization estimate using the sample of unrelated individuals suggested that each unit increase in BMI increased the risk of having high blood pressure by 1.59 (95%CI: 1.34 to 1.83, p-value=1.3×10^{-36}) percentage points. The Mendelian randomization estimate was similar after allowing for a family fixed effect (1.13 percentage point increase per 1kg/m^2 increase in BMI, 95%CI: 0.04 to 2.21, p-value=0.04). The summary data Mendelian randomization estimates were similar (0.76 percentage point increase per 1kg/m^2 increase in BMI, 95%CI: -0.19 to 1.70, p-value=0.12, $p_{\text{diff unrelated}}$= 0.10). On average the associations of the SNPs and high blood pressure fell by 51% (95%CI: 23% to 80%, p-value=0.0006) after allowing for family fixed effects.

Taller participants were more educated; each 10cm increase in height associated with an additional 0.45 (95%CI: 0.43 to 0.48, p-value=9.0×10^{-307}) years of education (**Figure 3**). This association was attenuated after including a family fixed effect (0.22, 95%CI: 0.18 to 0.26, p-value=4.6×10^{-25}). The Mendelian randomization estimate using the sample of unrelated individuals implied that each 10cm increase in height caused an increase of 0.17 (95%CI: 0.14 to 0.20, p-value=8.5×10^{-26}) years of education. After allowing for a family fixed effect, the Mendelian randomization estimate was greatly attenuated suggesting little causal effect of height on education (mean difference per 10cm increase in height: 0.002, 95%CI: -0.13 to 0.13, p-value=0.98). When we used two sample Mendelian randomization by estimating the SNP-exposure and SNP-outcome associations in different samples (split sample)[35,36] and then meta-analysed, there was little evidence of a causal effect of height on education (mean difference per 10cm increase in height=0.009, 95%CI: -0.11 to 0.13, p-value=0.87, $p_{\text{diff unrelated}}$= 0.008). On average, the associations of these SNPs and height and education fell by 18% (95%CI: 14% to 22%, p-value=8.5×10^{-24}) and 61% (95%CI: 49% to 73%, p-value=1.5×10^{-21}) after allowing for family fixed effects respectively.

On average, participants with higher BMI were less educated: each 1kg/m^2 increase in BMI was associated with 0.07 fewer years of education (95%CI: 0.06 to 0.07, p-value=8.4×10^{-222}, see **Figure 3**). This association was attenuated after including a family fixed effect (0.02, 95%CI: 0.01 to 0.02, p-value=5.1×10^{-13}). The Mendelian randomization estimate using the sample of unrelated people implied that each unit increase in BMI decreased years of schooling by 0.03 (95%CI: 0.02 to 0.04, p-value=2.6×10^{-06}). This effect was eliminated after allowing for a family fixed effect, suggesting little causal effect of BMI on educational attainment (mean difference per 1kg/m^2 increase in BMI= 0.00, 95%CI: -0.04 to 0.05, p-value=0.89). Again, the effect was also largely attenuated when we used two sample summary data approaches. Using separate samples to estimate the SNP-exposure and the SNP-outcome associations allowing for family fixed effects, there was little evidence of an effect of BMI on educational attainment (mean difference per 1kg/m^2 increase in BMI=-0.01, 95%CI: -0.05 to 0.03, p-value=0.59, $p_{\text{diff unrelated}}$= 1.6×10^{-03}). On average, the association of the 69 BMI SNPs and education fell by 65% (95%CI:

34% to 76%, p-value=1.8×10^{-06}) after allowing for family fixed effects. These results suggest that the estimates in unrelated individuals may be due to dynastic effects or assortative mating. We found little evidence of heterogeneity between the two sample Mendelian randomization estimates from UK biobank and HUNT, except for the effect of BMI on diabetes (p-value=0.027).

We investigated whether our results could be explained by pleiotropy using the weighted median, weighted modal and MR-Egger estimators. These summary data Mendelian randomization estimators use estimates of the SNP-exposure and SNP-outcome associations to estimate the effect of the exposure on the outcome. These estimators are robust to a number of forms of pleiotropy. There was little evidence of differences between the IVW and pleiotropy robust methods, pleiotropy from the MR-Egger intercept, or heterogeneity across the studies (**Supplementary Figure 3**).

Discussion

We have presented within-family methods for Mendelian randomization and demonstrated how confounding due to family structure can bias Mendelian randomization studies using unrelated individuals. The simulations illustrated how bias occurs even if the phenotype of interest has no direct causal effect on the outcome, and these effects can theoretically cause false positive findings. The simulations further demonstrated how family structure can be exploited to control for these effects either using samples of siblings or parent-offspring trios. However, estimates from within-family Mendelian randomization are less precise than estimates using unrelated individuals, which is consistent with those seen for allelic association.[37–39]. Furthermore, in practice, there are fewer relatives than unrelated people in most studies. In samples from HUNT and UK Biobank, we investigated the impact of family structure on four empirical examples; the effects of BMI on the risk of diabetes and high blood pressure and the effects of height and BMI on educational attainment. We found that the effects of BMI on the risk of diabetes and high blood pressure were less precise, but similar when allowing for family effects. Conversely, the effects of height and BMI on educational attainment were almost entirely attenuated after allowing for family fixed effects.

A substantial literature has used Mendelian randomization and samples of unrelated individuals to establish that BMI increases the risk of diabetes and hypertension later in life.[40] Our results suggest that confounding due to family structure is unlikely to be generating these results, and that they are more likely to be due to a causal effect of BMI on individual's risk. Behavioural geneticists have used longitudinal data from samples of twins to understand how different family members affect each other over time.[41,42] Other studies have used animal models to investigate how "social genetic effects" (i.e. indirect or dynastic effects) can affect health outcomes.[43] A rich literature has established that height and BMI are respectively positively and negatively associated with educational attainment and socioeconomic position.[44–46] Consistent with our results, previous studies using twin data have indicated that the relationship between height and educational attainment is likely to be due to non-genetic shared familial factors.[47,48] These findings raise questions about whether height and BMI have causal effects on

socioeconomic outcomes later in life.[49–51] However, larger studies of related individuals with information on socioeconomic outcomes, such as income and occupation later in life are required to provide definitive evidence about the consequences of height and BMI later in life.

In general, within-family Mendelian randomization estimates are less precise than estimates from samples of unrelated individuals. Thus, within-family estimates of a specific association can be considered more robust, but less efficient estimates. Therefore, if there is evidence of differences between the estimates, then generally, the more imprecise but less biased within-family estimates should be preferred. Our estimates of the effect of height and BMI on educational attainment are an example of this situation. If there is little evidence of differences between estimates using unrelated and those allowing for family effects, then the former estimates should be preferred. Our estimates of the effect of BMI on risk of diabetes and high blood pressure are an example of this situation. This is analogous to comparing instrumental variable estimates to multivariable adjusted estimates.[3] While allowing for family fixed effects or using difference estimators will account for dynastic effects or assortative mating, these methods will not address bias due to violations of the third Mendelian randomization assumption (exclusion restriction). This assumption is that the SNPs have no direct effect of the SNPs on the outcome (i.e. no pleiotropy). MR-Egger, weighted median and mode, or Lasso estimators are robust to various forms of violations of this assumption.[7–9,34] It is trivial to use these estimators with the summary data methods we describe above and illustrate in **Supplementary Figure 3**. However, typically these estimators have lower power than the IVW estimator. The within-family summary data SNP-exposure and SNP-outcome associations, which allow for a family fixed effect, can be used with existing summary data estimators. Other proposed approaches allow for sophisticated control and estimation of pleiotropy and can trivially include family fixed effects,[52] but again generally have lower power and require more data than Mendelian randomization approaches using allele scores or IVW. Therefore, given the statistical power of currently available samples of related individuals, investigators may be restricted to estimators that are robust to either family structure or pleiotropy, but not both. A further issue concerns residual population stratification in ancestrally heterogenous GWAS such

as GIANT, which may bias SNP-phenotype associations for height, and affect analyses in other samples using SNPs identified in those GWAS.[22] Within-family Mendelian randomization can control for residual population stratification.

Within-family estimates from samples of siblings, that allow for a family fixed effect are robust to biases due to dynastic effects, assortative mating and fine population structure.[1] Of these approaches, the sibling design is potentially most useful because large amounts of such data are available through biobanks and family-based studies. Phenotypic similarity of siblings may reflect 'passive' sharing of environments or genes, or 'active' imitation or contrast effects arising from interaction between siblings.[53] Contrast effects, which may inflate the estimated contribution of the nonshared environment in twin studies,[54] can be mimicked by parental rating bias.[55,56] However, for biological phenotypes where rating bias is not a concern, Mendelian randomization could be used to study the influence via imitation or contrast of one sibling's genotype on the other's phenotype, sometimes called 'social genetic effects',[43] thereby adding to work on dynamic interplay between siblings. [41,42]

Dynastic effects and assortative mating may cause bias in GWAS.[14] If a GWAS is aiming to estimate the causal effect of variants on a given phenotype, then samples of unrelated individuals may produce biased estimates and potentially spurious findings. Future studies could re-run GWAS on a full range of traits on samples of siblings allowing for family fixed effects. This approach would also address concerns about residual population stratification in GWAS, which may bias SNP-phenotype associations in GWAS especially of heterogeneous ancestry.[22] However, to detect genetic variants that explain 0.1% of the variance of either the offspring or maternal effects (i.e. a 2 df test) will require sample sizes of 50,000 mother-offspring pairs to detect genome-wide significant associations ($\alpha = 5x10^{-08}$). Sample sizes of around 10,000 will be required to partition known loci of similar size to the above into maternal and/or offspring genetic effects ($\alpha = 0.05$).[57] This sample size would provide valuable information about which phenotypes are likely to be most strongly affected by dynastic effects and assortative mating. It is likely that many, particularly biological, traits are relatively

unaffected by these effects and thus GWAS results for these traits are unlikely to be biased due to these factors. Recent GWAS of social traits such as education reported the attenuation after allowing for family effects in their estimates in small samples.[30] Further work in this area should include estimating the consequences of family structure for GWAS and Mendelian randomization estimates.

Conclusions

Family structure can cause bias in Mendelian randomization studies. We found differences between estimates from unrelated individuals and within-family estimates in simulations and empirical analysis. The causal estimates of the effect of height and BMI on educational attainment were almost entirely attenuated after allowing for family fixed effects. Within-family methods, either using individual-level, or summary data Mendelian randomization approaches can be used to obtain unbiased estimates of the causal effects of phenotypes in the presence of dynastic effects, assortative mating and population stratification.

Figure 1. Dynastic effects can cause confounding bias in Mendelian randomization studies. The MR estimate of the causal effect of the exposure on the outcome is biased because of potentially unobserved confounders between the SNPs and the exposure and the outcome. Panel A illustrates how population demography and structure can confound the SNP-outcome association. Panel B illustrates how dynastic effects can confound the SNP-outcome association. The solid red vertical arrow indicates the genetic inheritance of germline DNA. The dotted line indicates the direct (dynastic) effect of the parents on the offspring's outcomes. Mendelian randomization estimates of the effect of the exposure on the outcome in samples of unrelated individuals will be biased, because there is a path between offspring SNP and the outcome via the effect of the parents' phenotypes on their offspring's outcomes (dynastic effects). The presence of dynastic effects would violate one of three key Mendelian randomization (instrumental variable) assumptions – the independence assumption. Estimates that control for mother or father genotype, or sibling genotype will close this path and be unbiased. Panel C illustrates how assortative mating can confound the SNP-outcome association. In this example we present cross-trait assortative mating where there is a pathway between the mother's genotype and offspring's outcome via the father's genotype for the outcome. Dynastic effects and assortative mating can be accounted for by using methods based on within-family contrasts.

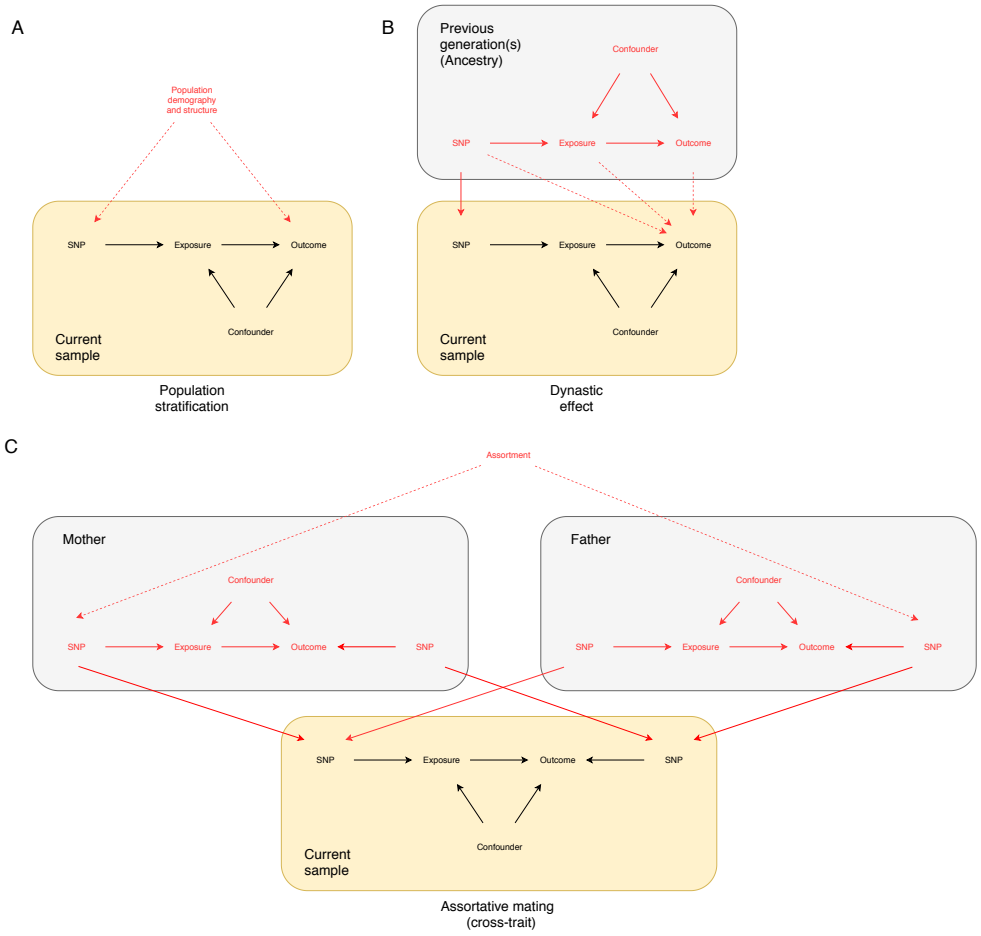


Figure 2: Estimated false discovery rate by power of the studies using different Mendelian randomization designs. A: SNP-exposure r^2 = 0.05; sample size = 10000; simulation involves an influence of parental exposure influencing child's confounder, which explains 10% of variance in child exposures and outcomes. For a simulated causal effect = 0, we expect the false discovery rate to be 0.05. B: Estimated bias by sample size using different Mendelian Randomization designs. The simulations are similar to panel (A) but allow sample size to vary and fixing the causal effect of an exposure x on an outcome y to 1% of variance explained. The bias in within-family Mendelian randomization estimates is small unless the dynastic effects are very small, or the number of observations modest.

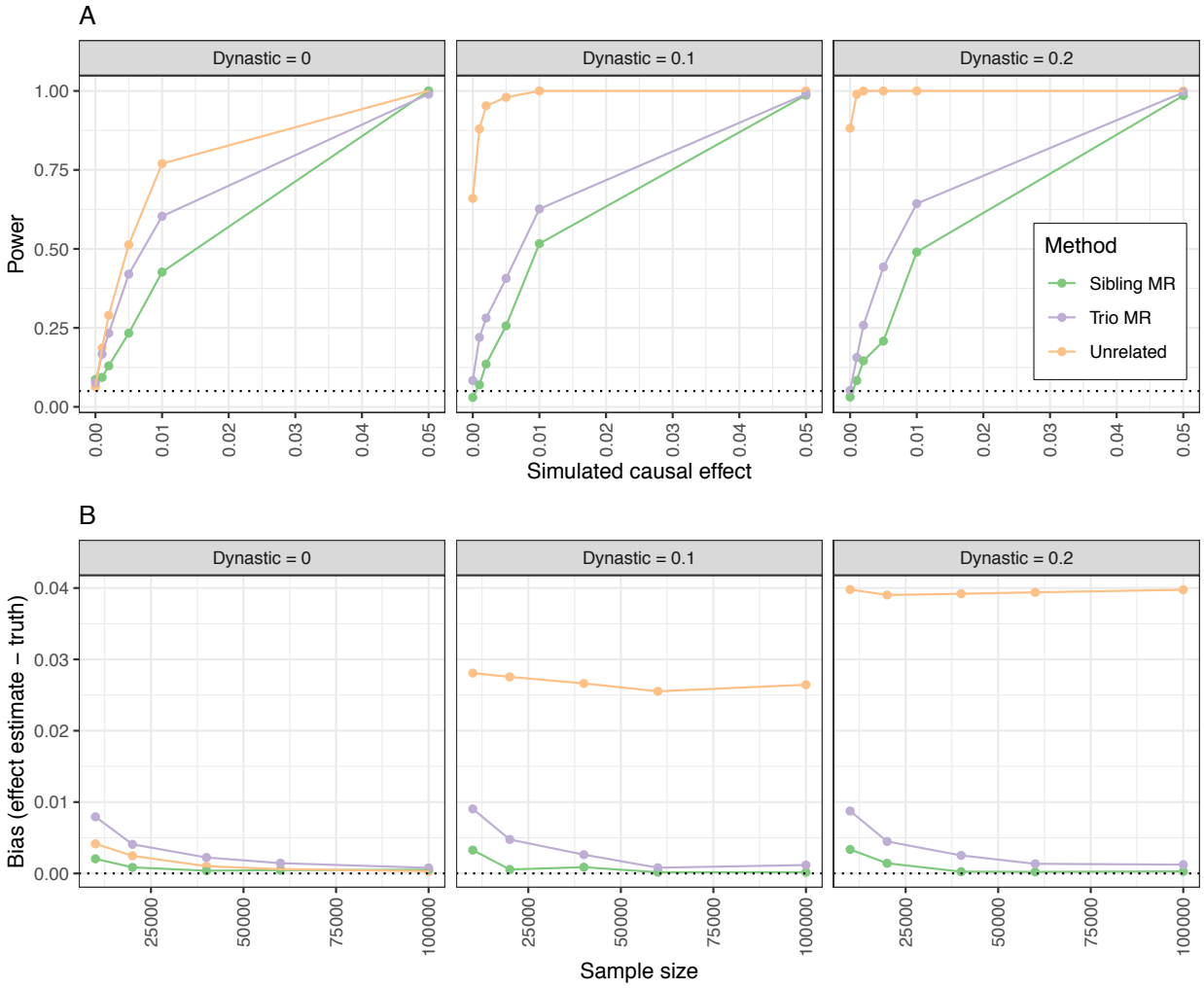
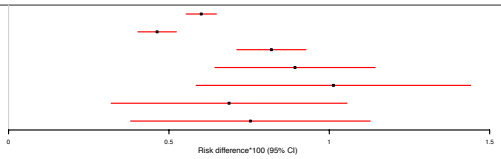
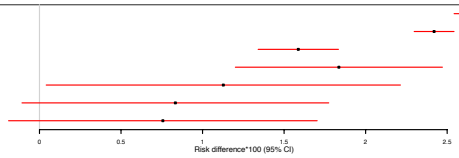


Figure 3: Estimates of the effect of BMI on self-reported diabetes and high blood pressure and height and BMI on educational attainment using ordinary least squares, Mendelian randomization in unrelated individuals and samples of siblings. All methods were consistent with BMI increasing diabetes and high blood pressure risk. Being taller and slimmer were phenotypically associated with higher educational attainment. However, the effects of height and BMI on educational attainment were attenuated but still detected by Mendelian randomization estimates using unrelated individuals from UK Biobank and HUNT. The effects were eliminated after allowing for a family effect using individual-level or summary data Mendelian randomization.

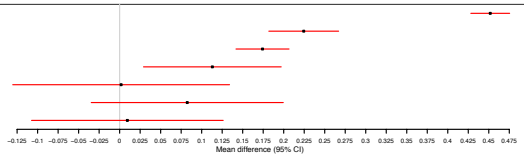
Effect of BMI on diabetes	N	Risk difference*100 (95% CI)	P–value
IPD Ordinary least squares	61008	0.60 (0.55 to 0.65)	1.07e–136
IPD OLS Family FE	61008	0.46 (0.40 to 0.52)	8.48e–52
IPD MR–PRS unrelateds	354836	0.82 (0.71 to 0.93)	3.28e–50
IPD MR–PRS siblings	61008	0.89 (0.64 to 1.14)	2.46e–12
IPD MR–PRS siblings family FE	61008	1.01 (0.58 to 1.44)	3.53e–06
2SMR IVW siblings	61008	0.69 (0.32 to 1.06)	2.47e–04
2SMR IVW siblings – split sample	61008	0.75 (0.38 to 1.13)	7.62e–05



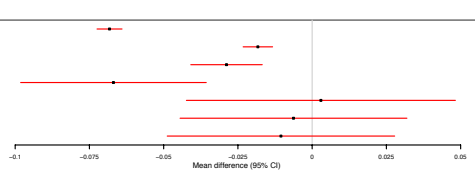
Effect of BMI on HBP	N	Risk difference*100 (95% CI)	P–value
IPD Ordinary least squares	61008	2.63 (2.54 to 2.72)	<1.00e–300
IPD OLS Family FE	61008	2.42 (2.30 to 2.54)	<1.00e–300
IPD MR–PRS unrelateds	354836	1.59 (1.34 to 1.83)	1.26e–36
IPD MR–PRS siblings	61008	1.84 (1.20 to 2.47)	1.44e–08
IPD MR–PRS siblings family FE	61008	1.13 (0.04 to 2.21)	4.19e–02
2SMR IVW siblings	61008	0.83 (–0.11 to 1.77)	8.25e–02
2SMR IVW siblings – split sample	61008	0.76 (–0.19 to 1.70)	1.17e–01



Effect of height on education	N	Mean difference (95% CI)	P–value
IPD Ordinary least squares	61008	0.45 (0.43 to 0.48)	9.05e–307
IPD OLS Family FE	61008	0.22 (0.18 to 0.27)	4.56e–25
IPD MR–PRS unrelateds	354836	0.17 (0.14 to 0.21)	4.50e–26
IPD MR–PRS siblings	61008	0.11 (0.03 to 0.20)	8.30e–03
IPD MR–PRS siblings family FE	61008	0.00 (–0.13 to 0.13)	9.78e–01
2SMR IVW siblings	61008	0.08 (–0.03 to 0.20)	1.67e–01
2SMR IVW siblings – split sample	61008	0.01 (–0.11 to 0.13)	8.74e–01



Effect of BMI on education	N	Mean difference (95% CI)	P–value
IPD Ordinary least squares	61008	–0.07 (–0.07 to –0.06)	8.38e–222
IPD OLS Family FE	61008	–0.02 (–0.02 to –0.01)	5.08e–13
IPD MR–PRS unrelateds	354836	–0.03 (–0.04 to –0.02)	2.59e–06
IPD MR–PRS siblings	61008	–0.07 (–0.10 to –0.04)	2.70e–05
IPD MR–PRS siblings family FE	61008	0.00 (–0.04 to 0.05)	8.99e–01
2SMR IVW siblings	61008	–0.01 (–0.04 to 0.03)	7.47e–01
2SMR IVW siblings – split sample	61008	–0.01 (–0.05 to 0.03)	5.91e–01



Methods

Statistical models

We describe four methods of using family data for Mendelian randomization below. If there are only two siblings, the difference and family fixed effects methods are equivalent, see appendix for proof.

The model to be estimated can be described as:

$$x_{k,i} = \gamma_0 + \gamma_1 g_{k,i} + \gamma_2 C_{k,i} + f_k + v_{k,i} \quad \text{(i)}$$

$$y_{k,i} = \beta_0 + \beta_1 x_{k,i} + \beta_2 C_{k,i} + f_k + u_{k,i} \quad \text{(ii)}$$

Where $y_{k,i}$ and $x_{k,i}$ are the outcome and exposure for individual i from family k. $g_{k,i}$ is a set of genetic variants that are associated with the exposure. f_k is a family-level confounder, modelled in the empirical analysis via a family fixed effect (i.e. an indicator variable for each family). This accounts for all family-level confounders of the genetic variant-outcome association. Both $g_{k,i}$ and f_k are functions of a family-level genetic component. $u_{k,i}$ and $v_{k,i}$ are random error terms. $C_{k,i}$ is a confounder of the association of the exposure and the outcome, γ_2 and β_2 indicate the effect of the confounder on the exposure and the outcome. β_1 is the true causal effect of the exposure on the outcome which we wish to estimate. This model implies that Mendelian randomization using data from unrelated individuals would produce biased estimates of β_1 due to the correlation between $g_{k,i,j}$ and f_k. The effect of the exposure on the outcome can be estimated using individual level data allowing for a family fixed effect, or summary level data using difference methods within families, or by allowing for a family fixed effect. We describe these approaches below.

(1) Siblings difference method

To apply Mendelian randomization to samples of siblings, effect estimates for the SNP-exposure association and SNP-outcome association are based on correlating the phenotypic divergence with the genotypic divergence within sibling pairs. Taking the difference between

siblings removes the effect of the family-level confounder. For any pair of siblings within family k, indicated $k, 1$ and $k, 2$, the genotypic difference at genetic variant j is:

$$\delta_{k,j} = g_{k,1,j} - g_{k,2,j} \quad \text{(iii)}$$

The association between the genotypic differences and phenotypic differences in the exposure, x, and outcome y, for SNP j can be estimated via:

$$\left(x_{k,1} - x_{k,2}\right)^2 = \gamma_j {\delta_{k,j}}^2 + u_{k,j} \quad \text{(iv)}$$

$$\left(y_{k,1} - y_{k,2}\right)^2 = \Gamma_j {\delta_{k,j}}^2 + v_{k,j} \quad \text{(v)}$$

The estimated associations, γ_j and Γ_j, can be used with any summary level Mendelian randomization estimator. Here we apply the inverse variance weighted (IVW) approach. Each pair of siblings can be included as a separate pseudo-independent pair.

(2) Family fixed effect with sibling data

Alternatively, we can estimate the associations using family fixed effects indicated by f_k for each family, which is equivalent to centring the data by subtracting the family mean.

$$x_{k,i} = \gamma_0 + \gamma_{1,j} g_{k,i,j} + f_k + u_{k,i,j} \quad \text{(vi)}$$

and

$$y_{k,i} = \beta_0 + \Gamma_1 g_{k,i,j} + f_k + v_{k,i,j}. \quad \text{(vii)}$$

This estimator accounts for any differences between families, which includes any effect of assortative mating or dynastic effects common to all siblings by including a dummy variable for each family. This provides unbiased estimates of the SNP-exposure and SNP-outcome associations. These estimates can be used with standard summary data Mendelian randomization methods. The difference and family fixed effects methods are identical if there are only two siblings in each family. This fact follows from substituting equations iv and v into equations ii and iii and simplifying (see Appendix for proof). If there are more siblings, then the estimators are non-identical, but likely to be similar, see the appendix for further details.

Cluster robust standard errors can be used to allow for clustering and relatedness within families.

(3) Adjusting for parental genotype with mother-father-offspring trio data

Finally, if data on mother-father-offspring trios are available, the estimates of the SNP-exposure and SNP-outcome associations for each child can be adjusted for their mother's and father's genotypes, indicated by $g_{im,j}$ and $g_{if,j}$ respectively[58]:

$$x_i = \gamma_0 + \gamma_{1,j} g_{i,j} + \gamma_{2,j} g_{im,j} + \gamma_{3,j} g_{if,j} + u_{i,j} \quad \text{(viii)}$$

and

$$y_i = \beta_0 + \Gamma_1 g_{i,j} + \Gamma_2 g_{im,j} + \Gamma_3 g_{if,j} + v_{i,j}. \quad \text{(ix)}$$

Again, these associations can be used to estimate the effect of the exposure on the outcome using summary data Mendelian randomization methods. It is possible to estimate the effect of foetal genotype on the exposure and outcome conditional on the mother and father genotype using summary data.[25,58] The estimated causal effect can be biased if both the SNP-exposure and SNP-outcome associations are estimated in the same sample.[59] This bias can be eliminated by splitting the sample and estimating the associations in separate samples.

(4) Two-stage least squares with sibling data

Many summary data methods assume no measurement error on the SNP-exposure association (NOME).[60] This assumption may lead to underestimation of the standard error of the effect of the exposure on the outcome. Two-stage least squares can estimate the effect of the exposure on the outcome using the individual-level data from siblings. Estimators that use individual-level data can account integrate the estimation error from the SNP-exposure association. We used cluster robust standard errors that allow for clustering and relatedness within family. We used the commands xtivreg and plm.[61]

Simulation of dynastic effects

We simulated a cohort consisting of pairs of unrelated mothers and fathers who had two offspring. All individuals had a genome of 90 SNPs. We set the distribution of identity by descent (IBD) across the 90 SNPs as $N(0.5, 0.037)$ for each sibling pair, as per theory, because there are on average 90 recombination events separating human siblings. Hence, we assume that each SNP has an independent effect on the exposure.

We defined parents' exposure and outcome by defining confounders u, exposure x and outcome y. A directed acyclic graph illustrating these relationships is shown in **Figure 1 (panel B)**. The confounder influences the parents' exposure and outcome. The offspring have the same confounding structure, except the parent's exposure affects their offspring's outcomes, via a dynastic or ancestry effect. The genetic influence of each of the 90 SNPs on the exposure amounts to explaining V_{gx} of the variance in the exposure. We assumed no horizontal pleiotropy. All estimates assume $V_{gx} = 0.1$ and 90 independent causal variants (i.e. somewhat similar to GIANT results for BMI).[62]

To generate the phenotypes under a model of dynastic effects, the offspring outcome was influenced by both the offspring exposure and the parents' exposures. In these simulations all phenotypes had mean of 0 and variance of 1. Differing strengths of dynastic effects, where the parental exposure influenced the offspring outcome were generated ($\beta_{ux} = 0, 0.01, 0.02$) under a set of models with a range of causal effects of the exposure on the outcome ($\beta_{xy} = 0, 0.001, 0.002, 0.005, 0.01, 0.05$). We calculated the false discovery rate (proportion of test with p-value < 0.05) for 100 iterations of each simulation using each of three methods: standard IVW as applied to one of each individual in a set of siblings (i.e. a sample of unrelated individuals), the within-family sibling design, and the within-family trio design. Finally, we calculated bias (estimated effect – simulated effect) for all three study approaches by simulated confounding (C_x and $C_y = 0, 0.1, 0.2$), dynastic bias ($\beta_{Ux} = 0, 0.1, 0.2$) and simulated causal effects ($\beta_{xy} = 0, 0.001, 0.002, 0.005, 0.01, 0.05$). The sample sizes were 10 000, 20 000, 40 000, 60 000 and 100 000 sibling pairs for all simulations.

Empirical Analysis

To demonstrate the approach and assess potential dynastic effects, assortative mating, and residual population stratification, we conducted within-family Mendelian randomization using two illustrative examples in the Nord-Trøndelag Health Study (HUNT) and the UK Biobank.[63–65] We estimated the effects of BMI on hypertension and diabetes and the effects of height and BMI on educational attainment. The effects of BMI on diabetes and hypertension have been well studied and provide a positive control.[40] These effects on clinical outcomes experienced later in life are unlikely to be due to assortative mating or dynastic effects, because parents are less likely to assort on genetic liability for diabetes or high blood pressure. The genetic liability for these conditions was probably unknown when the couples were formed. Previous longitudinal and Mendelian randomization studies using unrelated individuals have suggested that height and BMI may affect educational attainment.[44,49] Such an association, if causal, might be counteracted by changing educational policy. However, the association may be due to parents' education, via dynastic effects or assortative mating, where more educated people select taller and slimmer partners. Assortative mating and dynastic effects can confound the association between genetic variants when data from the offspring generation are used. Therefore, the ratio of individual-level causal effects to family-level effects is likely to be higher for the effects of BMI on clinical end points than for the effects of height and BMI on education.

HUNT descriptive data

There were 56,374 genotyped individuals in HUNT, including 11,448 families with at least two siblings comprising of 28,777 siblings (14,718 women) with complete data on genotype, height and education, diabetes and blood pressure. On average the participants in the full unrelated sample were 48.7 (SD = 15.1) years old, had a BMI of 26.3 kg/m^2 (SD = 3.9), were 177.5cm tall (SD = 6.6) and 164.3cm tall (SD = 6.11) for men and women respectively, 2.5% of them had diabetes, and 42.5% had high blood pressure and had 12.0 years (SD = 2.1) of education. High blood pressure was defined as either currently taking anti-hypertensive medication or having systolic or diastolic blood pressure above 140mmHg or 90mmHg on average across up to three

measurements in HUNT2. Descriptive statistics of the sample can be found in **Supplementary Tables 1** and **2**.

UK Biobank descriptive data

There were up to 370,180 genotyped individuals in the UK Biobank, among whom were 16,847 families with at least two siblings, with 33,642 siblings (19,445 women) with complete data on genotype, height and education, diabetes and blood pressure. On average, the participants without siblings were 57.5 (SD = 7.4) years old, had a BMI of 27.4 kg/m^2 (SD = 4.8), were 175.0cm tall (SD = 6.8) and 162.8cm tall (SD = 6.2) for men and women respectively, had 14.1 (SD = 2.3) years of education. 4.5% of them had diabetes and 54.0% had high blood pressure. High blood pressure was defined as either having a diagnosis of high blood pressure or having systolic or diastolic blood pressure above 140mmHg or 90mmHg respectively on average across up to two clinic measurements.

Selection of genotypes for instruments

We selected 385 independent (r^2<0.01 within 10,000kb) SNPs associated with height ($p<5x10^{-08}$) from Wood et al. and 79 associated with BMI in Locke et al.[62,66] Neither HUNT or UK Biobank were included in these studies. We clumped variants using the TwoSampleMR package.[67] We harmonized the alleles' effect sizes across the two samples and constructed weighted polygenic scores which were sums of the phenotype increasing alleles and weighted each variant by its effect on the phenotype in the published GWAS.

The Nord-Trøndelag Health Study

The Nord-Trøndelag Health Study (HUNT) is a population-based cohort study. The study was carried out at four time points over approximately 30 years (HUNT1 [1984-1986], HUNT2 [1995-1997] and HUNT3 [2006-2008] and HUNT4 [2017-2019]). A detailed description of HUNT is available.[63] We include 71,860 participants from HUNT2 and HUNT3 as they have been recently genotyped using one of three different Illumina HumanCoreExome arrays (HumanCoreExome12 v1.0, HumanCoreExome12 v1.1 and UM HUNT Biobank v1.0). For a flow chart of participants

inclusion and exclusion from the study see **Supplementary Figure 1**. Imputation was performed on samples of recent European ancestry using Minimac3 (v2.0.1, http://genome.sph.umich.edu/wiki/Minimac3) from a merged reference panel constructed from the Haplotype Reference Consortium (HRC) panel (release version 1.1) and a local reference panel based on 2,202 whole-genome sequenced HUNT participants [12-14]. Ancestry of all samples was inferred by projecting all genotyped samples into the space of the principal components of the Human Genome Diversity Project (HGDP) reference panel (938 unrelated individuals; downloaded from http://csg.sph.umich.edu/chaolong/LASER/)[15, 16], using PLINK. We defined recent European ancestry as samples that fell into an ellipsoid spanning exclusively the European population within the HGDP panel. We restricted the analysis to individuals of recent European ancestry who passed quality control. Among these, 17,329 pairs of siblings comprising of 28,777 siblings, were inferred using KING[17], where an estimated kinship coefficient between 0.177 and 0.355, the proportion of the genomes that share two alleles IBD > 0.08, and the proportion of the genome that share zero alleles IBD > 0.04 corresponded to a full sibling pair.

Questionnaires, clinical measurements and hospitalizations

Participants attended a health survey which included comprehensive questionnaires, an interview and clinical examination. The participants' height and weight were measured with the participant wearing light clothes without shoes to the nearest centimetre and half kilogram, respectively. Education was defined using the question "What is your highest level of education". Participants answered one of five categories 1) primary school (=10 years), 2) high school for 1 or 2 years (=12 years), 3) complete high school (=13 years), 4) college or university less than 4 years (=16 years), and 5) college or university 4 years or more (=17 years). Participants with university degrees were assigned to 16 years of education, those who completed high school were assigned 13 years, those who attended high school for 1 or 2 years were assigned to 12 years, and those who only attended primary school were assigned to 10 years. Diabetes was defined using responses to the question "Have you had or do you have diabetes" which has high validity.[68] High blood pressure was defined as those with systolic or

diastolic blood pressure equal to or more than 140 or 90, respectively, or reported use of antihypertensive medication.

Ethics

This study was approved by the Regional Committee for Medical and Health Research Ethics, Central Norway and all participants gave informed written consent (REK Central application number 2016/21296).

The UK Biobank

The UK Biobank invited over 9 million people and sampled 503,317 participants from March 2006 to October 2010. The study sampled individuals from 21 study centres across Great Britain. A detailed description of the study can be read elsewhere.[64,65,69] We included 33,642 related participants who were born in England to ensure that they experienced a similar school system. For a flow chart of participants inclusion and exclusion from the study see **Supplementary Figure 2**. The participants gave blood samples, from which DNA was extracted. Full details of the quality control process are available elsewhere. [70]Briefly, we excluded participants who had mismatched genetic and reported sex, or those with non-XX or XY chromosomes, extreme heterozygosity or missingness. We limited the analysis to 11,554,957 SNPs on the Haplotype Reference Consortium (HRC) panel. We selected SNPs that were reported to be associated with height and BMI in GWAS.[62,66] We used these GWAS because they did not include UK Biobank. We selected independent SNPs associated $r^2<0.01$ within 10,000kb by selecting the SNP most strongly associated with the trait of interest. We created weighted allele scores for height and BMI by calculating a weighted average of the number of height or BMI increasing alleles each participant had.[59]

Questionnaires, clinical measurements and hospitalizations

Weight (ID:21002) and standing height (ID:50) were measured using standardised instruments the baseline assessment centre visits. We defined education using the participants response to the touch screen questionnaires about their educational qualifications (ID= 6138). We defined

educational attainment using the participants' highest reported educational qualification at either measurement occasion. We assigned participants with university degrees to 17 years of education, those with professional qualifications such as teaching or nursing to 15 years, those with A-levels to 14 years, those with National Vocational Qualifications (NVQs), Higher National Diplomas (HNDs) to 13 years, General Certificate of Secondary Education (GCSEs), Certificate of Secondary Education (CSEs) or O-levels to 12 years, and those who reported no qualifications to 11 years, which was the legal minimum length of education for this cohort. Diabetes and high blood pressure were defined using responses to the self-reported touch screen questionnaire (ID=6150 and ID=2443). We used self-reported measures because measured blood pressure is affected by medication use. Missing values at the baseline visit were replaced by measures from subsequent visits if available.

Covariates and standard errors

All analyses included age, sex, the first 20 principal components of genetic variation. Cluster robust standard errors were used to allow for heteroskedasticity and allow for clustering and relatedness across siblings within families. Inclusion of the covariates age, sex, and principal components did not meaningfully affect the within family estimates, as they are independent of genotype conditional on sibling genotype.

Ethics

UK Biobank received ethical approval from the Research Ethics Committee (REC reference for UK Biobank is 11/NW/0382). This research was approved as part of application 8786.

Empirical analyses

We compared seven empirical estimates of the BMI on self-reported diabetes and high blood pressure and the effects of the height and BMI on educational attainment:

1. **IPD Ordinary least squares**: The phenotypic association using ordinary least squares
2. **IPD OLS Family FE**: The phenotypic association using ordinary least squares allowing for a family fixed effect across siblings.

3. **IPD MR-PRS unrelateds**: An estimate of the effect using a sample of unrelated individuals using a polygenic risk score and two stage least squares.
4. **IPD MR-PRS siblings**: As with 2. above, but restricted to siblings.
5. **IPD MR-PRS siblings family fixed effects**: An estimate using individual level data from the full sample of siblings family fixed effects. The standard errors allow for clustering by family and integrate imprecision from the SNP-exposure association. The results from HUNT and UK Biobank were combined via meta-analysis. Estimated using the xtivreg and plm packages.[61]
6. **2SMR IVW siblings**: An estimate using SNP summary data for Mendelian randomization including family fixed effects. The SNP-exposure and SNP-outcome associations were estimated on the same sample. The SNP level estimates of the effect of the exposure on the outcome were estimated separately in HUNT and UK Biobank and the overall Mendelian randomization (Wald) estimates are calculated for each SNP. For each SNP the Wald estimate is the ratio of the SNP-outcome and SNP-exposure association. We combine the estimate using Inverse Variance Weighted (IVW) meta-analysis.
7. **2SMR IVW siblings – split sample**: As with 5. above, but the SNP-exposure and SNP-outcome associations were estimated in separate samples (i.e. split sample approach). The overall Wald estimates were combined via IVW meta-analysis as above.

Sensitivity analyses

Finally, we tested for difference ($p_{\text{diff unrelated}}$) between the Mendelian randomization estimates using the unrelated individuals and the summary data within-family estimates using the split sample approach (i.e. as in 6. above).[71] We investigated whether our results could be explained by pleiotropy using the weighted median, weighted modal and MR-Egger estimators and the SNP-phenotype associations allowing for a family fixed effect.[8,10,72] We used a split sample approach in which the SNP-exposure and SNP-outcome associations were estimated in separate samples. We estimated the percentage change in the SNP-phenotype coefficients with and without allowing for a family fixed effect.

Data and code availability

The empirical dataset will be archived with the studies and will be made available to individuals who obtain the necessary permissions from the studies' data access committees. The code used to clean and analyse the data and the SNP level summary statistics are available here: https://github.com/nmdavies/within_family_mr.

References

1. Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* **32**, 1–22 (2003).
2. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89-98 (2014).
3. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* k601 (2018). doi:10.1136/bmj.k601
4. Pingault, J.-B. *et al.* Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics* **19**, 566–580 (2018).
5. Katan, M. Apoupoprotein E Isoforms, Serum Cholesterol, And Cancer. *The Lancet* **327**, 507–508 (1986).
6. Hartwig, F. P., Davies, N. M. & Davey Smith, G. Bias in Mendelian randomization due to assortative mating. *Genetic Epidemiology* **42**, 608–620 (2018).
7. Windmeijer, F., Farbmacher, H., Davies, N. & Smith, G. D. On the Use of the Lasso for Instrumental Variables Estimation with Some Invalid Instruments. *Journal of the American Statistical Association* 1–32 (2018). doi:10.1080/01621459.2018.1498346
8. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *International Journal of Epidemiology* **46**, 1985–1998 (2017).

9. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* **44**, 512–525 (2015).

10. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic Epidemiology* **40**, 304–314 (2016).

11. Burgess, S. & Thompson, S. G. Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects. *American Journal of Epidemiology* **181**, 251–260 (2015).

12. Mendel, G. Experiments in plant hybridization. *Verhandlungen des naturforschenden Vereins Brünn.) Available online: www. mendelweb. org/Mendel. html (accessed on 1 January 2013)* (1865).

13. Haworth, S. *et al.* Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nature Communications* **10**, (2019).

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

15. Robinson, M. R. *et al.* Genetic evidence of assortative mating in humans. *Nature Human Behaviour* **1**, 0016 (2017).

16. Beauchamp, J. P., Cesarini, D., Johannesson, M., Lindqvist, E. & Apicella, C. On the sources of the height–intelligence correlation: New insights from a bivariate ACE model with assortative mating. *Behavior Genetics* **41**, 242–252 (2011).

17. Fisher, R. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399–433 (1918).

18. Howe, L. J. *et al.* Alcohol consumption and mate choice in UK Biobank: comparing observational and Mendelian randomization estimates. *bioRxiv* (2018). doi:10.1101/418269

19. Nordsletten, A. E. *et al.* Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. *JAMA Psychiatry* **73**, 354 (2016).

20. Silventoinen, K., Kaprio, J., Lahelma, E., Viken, R. J. & Rose, R. J. Assortative mating by body height and BMI: Finnish twins and their spouses. *Am. J. Hum. Biol.* **15**, 620–627 (2003).

21. Mathieson, I. & McVean, G. Differential confounding of rare and common variants in spatially structured populations. *Nat. Genet.* **44**, 243–246 (2012).

22. Berg, J. J. *et al.* Reduced signal for polygenic adaptation of height in UK Biobank. *bioRxiv* (2018). doi:10.1101/354951

23. Fulker, D. W., Cherny, S. S., Sham, P. C. & Hewitt, J. K. Combined Linkage and Association Sib-Pair Analysis for Quantitative Traits. *The American Journal of Human Genetics* **64**, 259–267 (1999).

24. Spielman, R. S. & Ewens, W. J. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am. J. Hum. Genet.* **62**, 450–458 (1998).

25. Warrington, N. M., Freathy, R. M., Neale, M. C. & Evans, D. M. Using structural equation modelling to jointly estimate maternal and fetal effects on birthweight in the UK Biobank. *Int J Epidemiol* (2018). doi:10.1093/ije/dyy015

26. Abecasis, G. R., Cardon, L. R. & Cookson, W. O. A general test of association for quantitative traits in nuclear families. *Am. J. Hum. Genet.* **66**, 279–292 (2000).

27. Weiner, D. J. *et al.* Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nature Genetics* **49**, 978–985 (2017).

28. Spielman, R. S. & Ewens, W. J. The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.* **59**, 983–989 (1996).

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

30. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).

31. Hemani, G. *et al.* Inference of the Genetic Architecture Underlying BMI and Height with the Use of 20,240 Sibling Pairs. *The American Journal of Human Genetics* **93**, 865–875 (2013).

32. Minică, C. C., Dolan, C. V., Boomsma, D. I., de Geus, E. & Neale, M. C. Extending Causality Tests with Genetic Instruments: An Integration of Mendelian Randomization with the Classical Twin Design. *Behav. Genet.* **48**, 337–349 (2018).

33. Heath, A. C. *et al.* Testing hypotheses about direction of causation using cross-sectional family data. *Behav. Genet* **23**, 29–50 (1993).

34. Kang, H., Zhang, A., Cai, T. T. & Small, D. S. Instrumental Variables Estimation With Some Invalid Instruments and its Application to Mendelian Randomization. *Journal of the American Statistical Association* **111**, 132–144 (2016).

35. Angrist, J. D. & Krueger, A. B. Split-Sample Instrumental Variables Estimates of the Return to Schooling. *Journal of Business & Economic Statistics* **13**, 225–235 (1995).

36. Pierce, B. L. & Burgess, S. Efficient Design for Mendelian Randomization Studies: Subsample and 2-Sample Instrumental Variable Estimators. *American Journal of Epidemiology* **178**, 1177–1184 (2013).

37. Neale, M. C. *et al.* Distinguishing Population Stratification from Genuine Allelic Effects with Mx: Association of ADH2 with Alcohol Consumption. *Behavior Genetics* **29**, 233–243 (1999).

38. *Statistical genetics: gene mapping through linkage and association*. (Taylor & Francis Group, 2008).

39. Ott, J., Kamatani, Y. & Lathrop, M. Family-based designs for genome-wide association studies. *Nature Reviews Genetics* **12**, 465–474 (2011).

40. Holmes, M. V. *et al.* Causal Effects of Body Mass Index on Cardiometabolic Traits and Events: A Mendelian Randomization Analysis. *The American Journal of Human Genetics* **94**, 198–208 (2014).

41. Dolan, C. V., de Kort, J. M., van Beijsterveldt, T. C. E. M., Bartels, M. & Boomsma, D. I. GE Covariance Through Phenotype to Environment Transmission: An Assessment in Longitudinal Twin Data and Application to Childhood Anxiety. *Behavior Genetics* **44**, 240–253 (2014).

42. Moscati, A., Verhulst, B., McKee, K., Silberg, J. & Eaves, L. Cross-Lagged Analysis of Interplay Between Differential Traits in Sibling Pairs: Validation and Application to Parenting Behavior and ADHD Symptomatology. *Behavior Genetics* **48**, 22–33 (2018).

43. Baud, A. *et al.* Genetic Variation in the Social Environment Contributes to Health and Disease. *PLOS Genetics* **13**, e1006498 (2017).

44. Magnusson, P. K. E., Rasmussen, F. & Gyllensten, U. B. Height at age 18 years is a strong predictor of attained education later in life: cohort study of over 950,000 Swedish men. *Int J Epidemiol* **35**, 658–663 (2006).

45. Case, A. & Paxson, C. Stature and status: Height, ability, and labor market outcomes. *J Polit Econ* **116**, 499–532 (2008).

46. Lundborg, P., Nystedt, P. & Rooth, D.-O. Body size, skills, and income: evidence from 150,000 teenage siblings. *Demography* **51**, 1573–1596 (2014).

47. Silventoinen, K., Kaprio, J. & Lahelma, E. Genetic and environmental contributions to the association between body height and educational attainment: a study of adult Finnish twins. *Behav. Genet.* **30**, 477–485 (2000).

48. Silventoinen, K., Krueger, R. F., Bouchard, T. J., Kaprio, J. & McGue, M. Heritability of body height and educational attainment in an international context: Comparison of adult twins in Minnesota and Finland. *American Journal of Human Biology* **16**, 544–555 (2004).

49. Tyrrell, J. *et al.* Height, body mass index, and socioeconomic status: Mendelian randomisation study in UK Biobank. *BMJ* i582 (2016). doi:10.1136/bmj.i582

50. Böckerman, P. *et al.* The effect of weight on labor market outcomes: An application of genetic instrumental variables. *Health Economics* **28**, 65–77 (2019).

51. Davey Smith, G. & Davies, N. M. Can genetic evidence help us understand why height and weight relate to social position? *BMJ* i1224 (2016). doi:10.1136/bmj.i1224

52. DiPrete, T. A., Burik, C. A. P. & Koellinger, P. D. Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data. *Proceedings of the National Academy of Sciences* **115**, E4970–E4979 (2018).

53. Carey, G. Sibling imitation and contrast effects. *Behavior Genetics* **16**, 319–341 (1986).

54. Davey Smith, G. Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice. *International Journal of Epidemiology* **40**, 537–562 (2011).

55. Saudino, K. J., Wertz, A. E., Gagne, J. R. & Chawla, S. Night and Day: Are Siblings as Different in Temperament as Parents Say They Are? *Journal of Personality and Social Psychology* **87**, 698–706 (2004).

56. Simonoff, E. *et al.* Genetic influences on childhood hyperactivity: contrast effects imply parental rating bias, not sibling interaction. *Psychol Med* **28**, 825–837 (1998).

57. Moen, G.-H., Hemani, G., Warrington, N. M. & Evans, D. M. Calculating Power to Detect Maternal and Offspring Genetic Effects in Genetic Association Studies. *Behavior Genetics* (2019). doi:10.1007/s10519-018-9944-9

58. Evans, D. M., Moen, G.-H., Hwang, L.-D., Lawlor, D. A. & Warrington, N. M. Elucidating the role of maternal environmental exposures on offspring health and disease using two-sample Mendelian randomization. *International Journal of Epidemiology* (2019). doi:10.1093/ije/dyz019

59. Hartwig, F. P. & Davies, N. M. Why internal weights should be avoided (not only) in MR-Egger regression. *International Journal of Epidemiology* dyw240 (2016). doi:10.1093/ije/dyw240

60. Bowden, J. *et al.* Improving the accuracy of two-sample summary data Mendelian randomization: moving beyond the NOME assumption. *bioRxiv* (2018). doi:10.1101/159442

61. Schaffer, M. E. *XTIVREG2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models*. (2005).

62. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

63. Krokstad, S. *et al.* Cohort Profile: The HUNT Study, Norway. *International Journal of Epidemiology* **42**, 968–977 (2013).

64. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

65. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology* **186**, 1026–1034 (2017).

66. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* **46**, 1173–1186 (2014).

67. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, (2018).

68. Midthjell, K., Holmen, J., Bjørndal, A. & Lund-Larsen, G. Is questionnaire information valid in the study of a chronic disease such as diabetes? The Nord-Trøndelag diabetes study. *J Epidemiol Community Health* **46**, 537–542 (1992).

69. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* **12**, e1001779 (2015).

70. Mitchell, R., Hemani, G., Dudding, T. & Paternoster, L. UK Biobank Genetic Data: MRC-IEU Quality Control, Version 1. (2017). doi:10.5523/bris.3074krb6t2frj29yh2b03x3wxj

71. Altman, D. G. & Bland, J. M. Interaction revisited: the difference between two estimates. *BMJ* **326**, 219 (2003).

72. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int J Epidemiol* **45**, 1961–1974 (2016).

73. Wooldridge, J. *Econometric analysis of cross section and panel data*. (The MIT press, 2002).

Acknowledgements

Jonathan Beauchamp provided valuable comments and suggestions on an earlier draft of this manuscript. The Medical Research Council (MRC) and the University of Bristol support the MRC Integrative Epidemiology Unit [MC_UU_12013/1, MC_UU_12013/9, MC_UU_00011/1]. The Economics and Social Research Council (ESRC) support NMD via a Future Research Leaders grant [ES/N000757/1]. LDH is supported by a Career Development Award from the UK Medical Research Council (MR/M020894/1). This work is part of a project entitled 'social and economic consequences of health: causal inference methods and longitudinal, intergenerational data', which is part of the Health Foundation's Efficiency Research Programme. The Health Foundation is an independent charity committed to bringing about better health and health care for people in the UK. The Nord-Trøndelag Health Study (The HUNT Study) is a collaboration between HUNT Research Center (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Nord-Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. The K.G. Jebsen Center for Genetic Epidemiology is funded by Stiftelsen Kristian Gerhard Jebsen; Faculty of Medicine and Health Sciences, NTNU; The Liaison Committee for education, research and innovation in Central Norway; and the Joint Research Committee between St. Olavs Hospital and the Faculty of Medicine and Health Sciences, NTNU. The genotyping in HUNT was financed by the National Institute of Health (NIH); University of Michigan; The Research Council of Norway; The Liaison Committee for education, research and innovation in Central Norway; and the Joint Research Committee between St. Olavs Hospital and the Faculty of Medicine and Health Sciences, NTNU. JK has been supported by the Academy of Finland (grants 308248, 312073). RMF and RNB are supported by Sir Henry Dale Fellowship (Wellcome Trust and Royal Society grant: WT104150). No funding body has influenced data collection, analysis or its interpretation. This publication is the work of the authors, who serve as the guarantors for the contents of this paper.

Author contributions

GDS, NMD, KH and BMB obtained funding for this study. BMB, GH and NMD analysed and cleaned the data, interpreted results, wrote and revised the manuscript. GH ran the simulation

study. ES provided the proof comparing the difference and fixed effects estimators, interpreted results and revised the manuscript. GDS conceived of the study, interpreted the results, wrote and revised the manuscript. All other authors interpreted the results and revised the manuscript.

The Mendelian randomization within-family consortium

K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Norway
Ben Brumpton, Gunnhild Åberge Vie, Johan Håkon Bjørngaard, Kristian Hveem, and Bjørn Olav Åsvold

Clinic of Thoracic and Occupational Medicine, St. Olavs Hospital, Trondheim University Hospital
Ben Brumpton

Medical Research Council Integrative Epidemiology Unit, University of Bristol, BS8 2BN, United Kingdom
Ben Brumpton, Eleanor Sanderson, Fernando Pires Hartwig, Sean Harrison, Yoonsu Cho, Laura D Howe, Amanda Hughes, Tim Morris, Laurence Howe, Oliver Davis, Claire Haworth, David Evans, George Davey Smith, Gibran Hemani, and Neil M Davies

Population Health Sciences, Bristol Medical School, University of Bristol, Barley House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom
Eleanor Sanderson, Fernando Pires Hartwig, Sean Harrison, Yoonsu Cho, Laura D Howe, Amanda Hughes, Tim Morris, Laurence Howe, Oliver Davis, Claire Haworth, George Davey Smith, Gibran Hemani, and Neil M Davies

Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, Brazil
Fernando Pires Hartwig

Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
Dorret I Boomsma and Michel G Nivard

Netherlands Twin Register, Vrije Universiteit, Amsterdam, the Netherlands.
Dorret Boomsma

Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit, Amsterdam, the Netherlands
Philipp D Koellinger

Department of Economics, VU University Amsterdam, Amsterdam, the Netherlands
Philipp D Koellinger

German Institute for Economic Research, DIW Berlin, Berlin, Germany
Philipp D Koellinger

Nic Waals Institute, Lovisenberg Diaconal Hospital Spångbergveien 25, 0853 Oslo, Norway
Alexandra Havdahl

Department of Mental Disorders, Norwegian Institute of Public Health, Sandakerveien 24C, 0473 Oslo, Norway
Alexandra Havdahl

The University of Melbourne, 207 Bouverie St, Carlton, Victoria, 3010, Australia
John Hopper

Virginia Institute for Psychiatric and Behavior Genetics, Virginia Commonwealth University, Richmond, Virginia
Michael Neale

Karolinska Institutet, Dept of Medical Epidemiology and Biostatistics, Stockholm, Sweden
Nancy L Pedersen and Sara Hägg

Department of Psychology, University of California Riverside, Riverside, CA, USA
Chandra Renyolds

Department of Psychology and Population Research Center, University of Texas at Austin
Elliot M Tucker-Drob, Andrew Grotzinger and K Paige Harden

Center for public health genomics, Department of public health sciences, University of Virginia, Charlottesville, VA, USA
Wei-Min Chen

Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, USA
Cristen Willer

Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA
Cristen Willer

Department of Human Genetics, University of Michigan, Ann Arbor, USA

Cristen Willer

University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia
David M Evans

Department of Public Health, University of Helsinki, Helsinki, Finland
Jaakko Kaprio

Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
Jaakko Kaprio

Department of Endocrinology, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway
Bjørn Olav Åsvol

School of Psychological Science, University of Bristol, Bristol, United Kingdom
Claire Haworth

University of Edinburgh, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road, Edinburgh, UK EH4 2XU
Archie Campbell

Institute for Molecular Medicine Finland (FIMM), P. O. Box 20 (Tukholmankatu 8) 00014 University of Helsinki
Teemu Palviainen

University of Minnesota, 75 East River Road, Minneapolis, MN 55455
William Iacono

The Danish Twin Registry, Department of Public Health, University of Southern Denmark, J.B. Winsloews Vej 9B, 5000 Odense C, Denmark
Kaare Christensen and Marianne Nygaard

Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne 207 Bouverie Street, Carlton, Victoria 3053, Australia
Shuai Li

Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, United Kingdom
Shuai Li

Italian Twin Register, Center for Behavioural Sciences and Mental Health, Istituto Superiore di Sanità Viale Regina Elena 299, I-00161 Rome, Italy
Maria Antonietta Stazi and Sonia Brescianini

Faculty of Psychology and IMIB-Arrixaca. University of Murcia. Campus de Espinardo. 30100 Murcia (Spain)
Juan R. Ordoñana

The University of Melbourne 207 Bouverie St, Carlton, Victoria, 3010, Australia
Maria Antontietta Stazi

GRIP, École de psychologie, Université Laval Pavillon Félix-Antoine-Savard, 2325 Allée des Bibliothèques, Québec, QC G1V 0A6, Canada
Michel Boivin

Karolinska Institutet Dept of Medical Epidemiology and Biostatistics, Stockholm, Sweden
Nancy L. Pedersen

University of Exeter, College of Medicine and Health, RILD Building, Barrack Road, Exeter, EX2 5DW
Rachel M Freathy and Robin Beaumont

QIMR Berghofer Medical Research Institute
Sarah Medland

University of Michigan, Ann Arbor, Michigan
Patricia Peyser

Department of Epidemiology, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA
Lawrence Bielak

Supplementary materials

Box 1: Possible confounders of genetic associations

Assortative mating: when individuals choose their partners non-randomly, so are more alike or selected on a particular trait than would be expected. This can occur because people select on a specific trait, e.g. if tall women prefer to marry tall men, people who drink alcohol choose partners who also drink,[8] or because of social homogamy where people select partners who have a similar environmental background to themselves, e.g. if educated women select men with a similar amount of education as themselves who happen to be taller.[71]

Dynastic effects: when parental genotype affect offspring outcomes through pathways other than via offspring genotype. For example, if more educated parents support their offspring's education, or if parents smoking positively or negatively affected the likelihood of their offspring smoking. An example of dynastic effects are passive gene-environmental correlations.[11,16,17]

Fine scale population structure: when subtle difference in ancestry are associated with offspring phenotypes. For example, on average individuals from the north and west of England are poorer and are less educated, there are also geographic gradients in the distribution of education associated variants.[13]

Equivalence of the first difference and fixed effects estimators when estimating within-family Mendelian randomization models.

The general model as describe above is:

$$x_{k,i} = \gamma_0 + \gamma_{1,j} g_{k,i,j} + f_k + u_{k,i} \quad \text{(x)}$$

and

$$y_{k,i} = \beta_0 + \beta_1 g_{k,i,j} + f_k + v_{k,i}. \quad \text{(xi)}$$

f_k is a family-level confounder, U_I is an individual-level confounder. $f_k = f(G_k, NG_k)$ where G_k is a family-level genetic component and NG_k is a family-level non-genetic component, and $g_{k,i,j} = f^*(G_k)$. This setup means that Mendelian randomization using data from unrelated individuals is invalid due to the correlation between $g_{k,i,j}$ and f_k.

With two siblings the model can be written as:

$$x_{k,1} = \gamma_0 + \gamma_{1,j} g_{k,1,j} + f_k + u_{k,1} \quad \text{(xii)}$$

$$x_{k,2} = \gamma_0 + \gamma_{1,j} g_{k,2,j} + f_k + u_{k,2} \quad \text{(xiii)}$$

$$y_{k,1} = \beta_0 + \beta_1 x_{k,1,j} + f_k + v_{k,1} \quad \text{(xiv)}$$

$$y_{k,2} = \beta_0 + \beta_1 x_{k,2,j} + f_k + v_{k,2}. \quad \text{(xv)}$$

Fixed effects estimation

Fixed effects estimation estimates the model:

$$x_{k,i} = \gamma_0 + \gamma_{1,j} g_{k,i,j} + cI_k + u_{k,i,j} \quad \text{(xvi)}$$

and

$$y_{k,i} = \beta_0 + \beta_1 x_{k,i,j} + cI_k + v_{k,i,j}. \quad \text{(xvii)}$$

Where I_F is a set of family-level indicator variables that takes one value for each family (i.e. a set of extra indicator variables, the same size as the number of families in the model, each one of which takes 1 for one family and 0 for all other families). This model is estimated by taking

the deviation from the family-level mean for each observation. Under this method of estimation our model becomes:

$$x_{k,i} - \overline{x_k} = \gamma_{1,j} g_{k,i,j} - \gamma_{1,j}\overline{g_{k,J}} + f_k - \overline{f_k} + u_{k,i,j} - \overline{u_{k,J}} \qquad \text{(xviii)}$$
$$= \gamma_{1,j}\left(g_{k,i,j} - \overline{g_{k,J}}\right) + \left(u_{k,i,j} - \overline{u_{k,J}}\right)$$

and

$$y_{k,i} - \overline{y_k} = \Gamma_1 g_{k,i,j} - \Gamma_{1,j}\overline{g_{k,J}} + f_k - \overline{f_k} + v_{k,i,j} - \overline{v_{k,J}}. \qquad \text{(xix)}$$
$$= \Gamma_1\left(g_{k,i,j} - \overline{g_{k,J}}\right) + \left(v_{k,i,j} - \overline{v_{k,J}}\right)$$

In a similar way to the First difference estimator, this estimator can now be consistently estimated using Mendelian randomization as $\left(g_{k,i,j} - \overline{g_{k,J}}\right)$ is independent of the family effect. This can be done by estimating:

$$x_{k,i} - \overline{x_k} = \gamma_{1,j} g_{k,i,j} - \gamma_{1,j}\overline{g_{k,J}} + u_{k,i,j} - \overline{u_{k,J}} \qquad \text{(xx)}$$
$$x^*_{k,i} = \gamma_{1,j} g^*_{k,i,j} + u^*_{k,i,j}$$

and

$$y_{k,i} - \overline{y_k} = \Gamma_j g_{k,i,j} - \Gamma_{1,j}\overline{g_{k,J}} + v_{k,i,j} - \overline{v_{k,J}} \qquad \text{(xxi)}$$
$$y^*_{k,i} = \Gamma_j g^*_{k,i,j} + v^*_{k,i,j}$$

The two-sample MR estimator is then obtained from;

$$\hat{\Gamma}_j = \beta_1 \hat{\gamma}_j + \omega_j$$

First difference (FD) estimation

A special case of the fixed effect estimator is the first difference estimator with two siblings. When the first difference is taken the model becomes:

$$x_{k,1} - x_{k,2} = \gamma_{1,j} g_{k,1,j} - \gamma_{1,j} g_{k,2,j} + f_k - f_k + u_{k,1,j} - u_{k,2,j} \qquad \text{(xxii)}$$
$$= \gamma_{1,j}\left(g_{k,1,j} - g_{k,2,j}\right) + \left(u_{k,1,j} - u_{k,2,j}\right)$$

and

$$y_{k,1} - y_{k,2} = \Gamma_{1,j} g_{k,1,j} - \Gamma_{1,j} g_{k,2,j} + f_k - f_k + v_{k,1,j} - v_{k,2,j} \qquad \text{(xxiii)}$$
$$= \Gamma_{1,j}\left(g_{k,1,j} - g_{k,2,j}\right) + \left(v_{k,1,j} - v_{k,2,j}\right).$$

As $(g_{k,1} - g_{k,2})$ is uncorrelated with $\left(u_{k,1,j} - u_{k,2,j}\right)$ this model can now be consistently estimated in a two-sample MR estimation by:

$$x_{k,1} - x_{k,2} = \gamma_{1,j}\left(g_{k,1,j} - g_{k,2,j}\right) + \left(u_{k,1,j} - u_{k,2,j}\right) \qquad \text{(xxiv)}$$
$$x'_k = \gamma G'_{k,j} + u'_{x,k,j}$$

and

$$y_{k,1} - y_{k,2} = \Gamma_{1,j}\left(g_{k,1,j} - g_{k,2,j}\right) + \left(v_{k,1,j} - v_{k,2,j}\right) \qquad \text{(xxv)}$$
$$y'_k = \Gamma g'_{k,j} + v'_{y,k,j}\ .$$

The two-sample MR estimation of $\hat{\beta}_1$ can then be calculated in the same way as above.

Equivalence of the first difference and fixed effects estimators

When there are exactly two individuals in each family these two methods of estimation will give the same result.

For $i = 1$ the variables in the fixed effects estimator is

$$x^*_{k,1} = x_{k,1} - \overline{x_k} = x_{k,1} - \frac{1}{2}\left(x_{k,1} + x_{k,2}\right) = \frac{1}{2}\left(x_{k,1} - x_{k,2}\right) \qquad \text{(xxvi)}$$

$$y^*_{k,1} = y_{k,1} - \overline{y_k} = y_{k,1} - \frac{1}{2}\left(y_{k,1} + y_{k,2}\right) = \frac{1}{2}\left(y_{k,1} - y_{k,2}\right) \qquad \text{(xxvii)}$$

$$g^*_{k,1,j} = g_{k,1,j} - \overline{g_{k,j}} = g_{k,1,j} - \frac{1}{2}\left(g_{k,1,j} + g_{k,2,j}\right) = \frac{1}{2}\left(g_{k,1,j} - g_{k,2,j}\right) \qquad \text{(xxviii)}$$

Therefore, the fixed effects estimator can be written as:

$$\frac{1}{2}\left(x_{k,1} - x_{k,2}\right) = \gamma_{1,j}\frac{1}{2}\left(g_{k,1,j} + g_{k,2,j}\right) + \frac{1}{2}\left(u_{k,1,j} - u_{k,2,j}\right) \qquad \text{(xxix)}$$

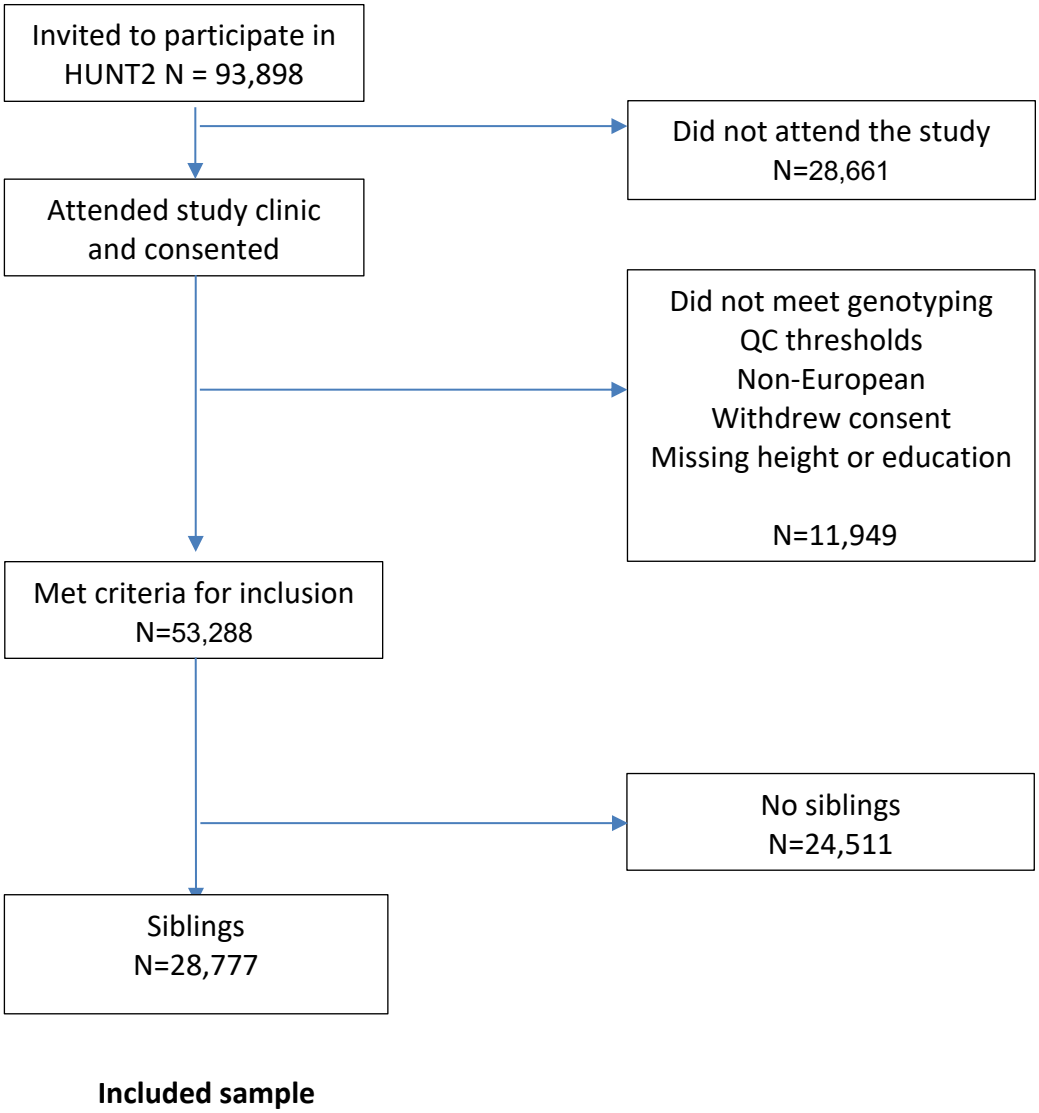
$$\frac{1}{2}x'_k = \gamma_{1,j}\frac{1}{2}g'_{k,j} + \frac{1}{2}u'_{k,j}$$
$$x'_k = \gamma_{1,j}g'_{k,j} + u'_{k,j}$$

and
$$\frac{1}{2}\left(y_{k,1} - y_{k,2}\right) = \beta_1\frac{1}{2}\left(x_{k,1} - x_{k,2}\right) + \frac{1}{2}\left(v_{k,1} - v_{k,2}\right) \quad \text{(xxx)}$$
$$\frac{1}{2}y'_k = \beta_1\frac{1}{2}x'_k + \frac{1}{2}v'_{k,j}$$
$$y'_k = \beta_1 x'_k + v'_{k,j}$$

Therefore, the fixed effects estimator is equal to the first difference estimator with all of the variables included divided by 2. As this transformation is applied to every variable in the model it cancels out across the estimation and the estimated parameters will be the same as in the first difference model.

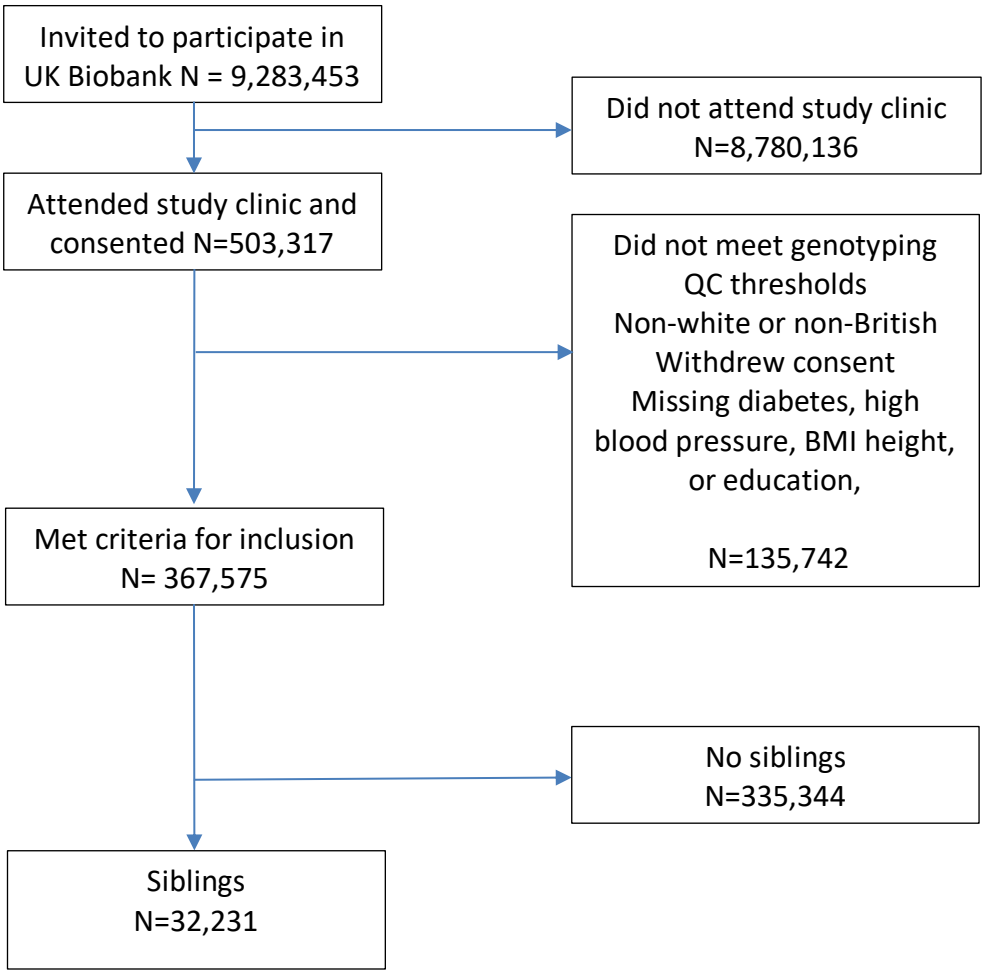
When there are more than two individuals in each family, the difference between these estimators depends on the error terms. If $v_{k,i,j}$ is assumed to be uncorrelated between individuals in a family, then the fixed effects estimator is more efficient. However, if it is assumed to be dependent on the level taken by other family members (i.e.$v_{k,1,j} = v_{k,2,j} + \epsilon_k$ where ϵ_k is a randomly distributed variable) then the first difference estimator is more efficient.[73]

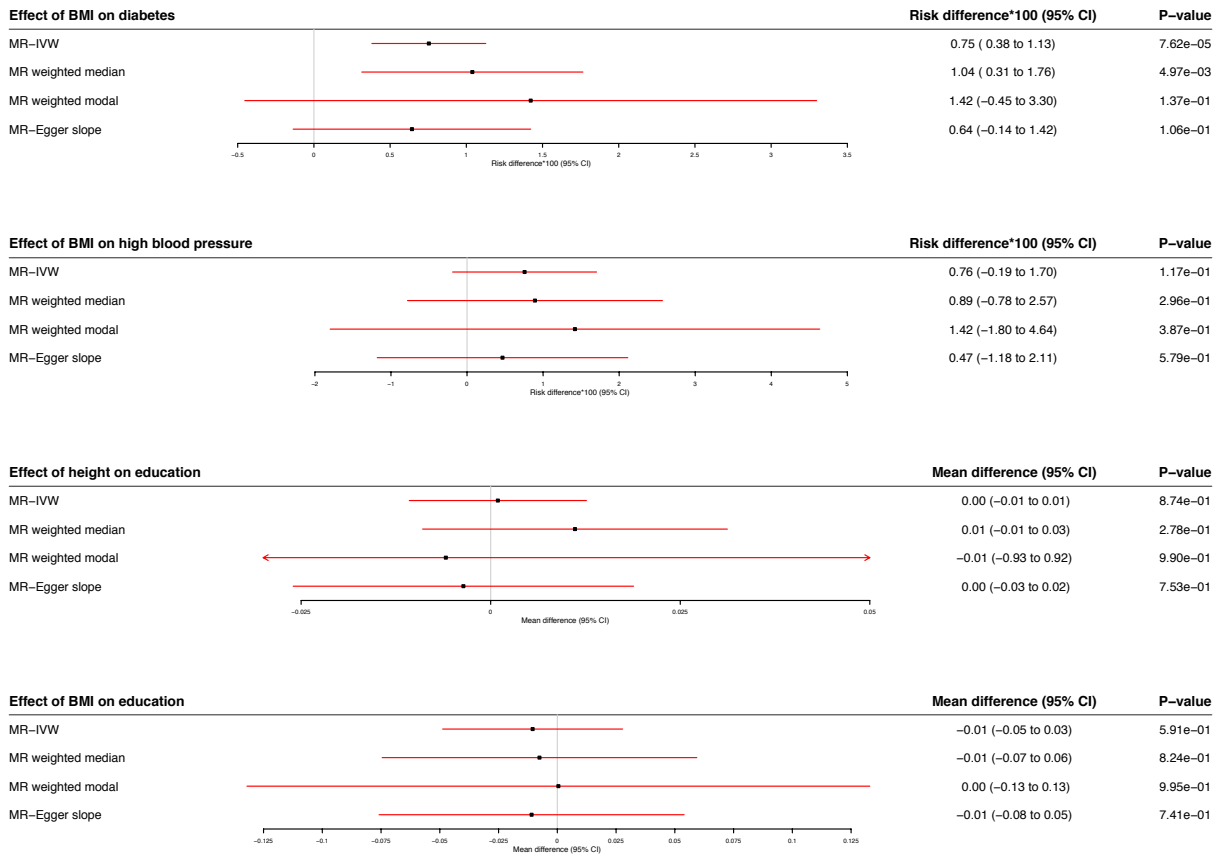
Supplementary Figure 1: STROBE Flow chart representing inclusion into the study sample for HUNT2.



Included sample

Supplementary Figure 2: STROBE Flow chart representing inclusion and inclusion into the study sample for the UK Biobank.



Included sample

Supplementary Figure 3: Estimates of the effect of BMI on self-reported diabetes and high blood pressure and height and BMI on educational attainment using Mendelian randomization in samples of siblings using inverse variance weighted (IVW), weighted median, weighted modal and MR-Egger. This analysis uses a split sample approach, in which the SNP-exposure and SNP-outcome associations are estimated in separate samples and allow for a family fixed effect. The weighted median, weighted modal and MR-Egger estimators are less precise than IVW. We found little evidence of pleiotropy using MR-Egger. There was little evidence of heterogeneity between HUNT and UKBB for any of the estimates. The MR-Egger intercepts for all outcomes found little evidence of directional pleiotropy (p>0.05), however this may be due to lack of power. The total analysed sample size in UK Biobank and HUNT was 61,008.