Human Molecular Genetics, 2019, Vol. 28, No. R2 R170–R179

doi: 10.1093/hmg/ddz204
Advance Access Publication Date: 24 October 2019
4 Invited Review Article

4 INVITED REVIEW ARTICLE

Within family Mendelian randomization studies

Neil M Davies[1,2,*,†], Laurence J Howe[1,2], Ben Brumpton[1,3,4], Alexandra Havdahl[1,5,6], David M Evans[1,7] and George Davey Smith[1,2]

[1]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, BS8 2BN, United Kingdom. [2]Population Health Sciences, Bristol Medical School, University of Bristol, Barley House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom. [3]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway [4]Clinic of Thoracic and Occupational Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway [5]Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway [6]Department of Mental Disorders, Norwegian Institute of Public Health, Oslo, Norway [7]University of Queensland Diamantina Institute, University of Queensland, Brisbane, 4102, Australia

*To whom correspondence should be addressed at: IEU Admin team. Tel: 0117 3310099; Fax: 0117 3314052; Email: neil.davies@bristol.ac.uk

Abstract

Mendelian randomization (MR) is increasingly used to make causal inferences in a wide range of fields, from drug development to etiologic studies. Causal inference in MR is possible because of the process of genetic inheritance from parents to offspring. Specifically, at gamete formation and conception, meiosis ensures random allocation to the offspring of one allele from each parent at each locus, and these are unrelated to most of the other inherited genetic variants. To date, most MR studies have used data from unrelated individuals. These studies assume that genotypes are independent of the environment across a sample of unrelated individuals, conditional on covariates. Here we describe potential sources of bias, such as transmission ratio distortion, selection bias, population stratification, dynastic effects and assortative mating that can induce spurious or biased SNP–phenotype associations. We explain how studies of related individuals such as sibling pairs or parent–offspring trios can be used to overcome some of these sources of bias, to provide potentially more reliable evidence regarding causal processes. The increasing availability of data from related individuals in large cohort studies presents an opportunity to both overcome some of these biases and also to evaluate familial environmental effects.

Mendelian randomization (MR) is an approach that exploits the natural experiment occurring at conception—the random inheritance of germline genetic variation from parents to their offspring (1). MR has transformed our ability to evaluate the causal effects of a wide array of exposures in biomedical research, drug development and social science (2,3). MR is a form of observational study, albeit one which exploits causal genetic anchors (4). Historically, the vast majority of MR studies have used samples of unrelated individuals, with little information on parents or family members. These studies have assumed that what is true at the within-family level—the random inheritance of genetic variants from parents to offspring—is reflected at a population-level in samples of unrelated individuals: i.e. that genetic variants are unlikely to be related to potential confounding factors. In general, at the population level, genetic variants are much less associated with many

[†]Neil M Davies, http://orcid.org/0000-0002-2460-0508
Received: August 13, 2019. **Revised:** August 13, 2019. **Accepted:** August 14, 2019

potential confounders than directly measured exposures of interest (5). However, the random inheritance of genetic variants from parents to offspring does not guarantee that genetic variants and confounders will be independent in samples of unrelated individuals. The initial elaboration of MR stated that it depended on the random allocation of alleles from parent to offspring and that this familial design was closely analogous to a randomized controlled trial (1). However, in 2003 no adequately powered family-based studies were available (or, indeed, imaginable) and thus most applications of MR to date have used population-based association studies of unrelated individuals. This downgraded the robustness of the design, and meant that 'the Mendelian randomization in genetic association studies is approximate, rather than absolute' (1). Here we revisit the situation in light of now available data, review potential sources of bias in MR using unrelated individuals, describe potential sources of bias and current solutions and conclude with suggestions for future research.

Mendel's laws

Causal inference in MR depends on additional assumptions added to what have become known as Mendel's first and second laws of genetic inheritance: the Law of Segregation and the Law of Independent Assortment (6). The Law of Segregation implies that at every point in the genome offspring randomly inherit one of their mother's two alleles and one of their father's two alleles (with the obvious exception of non-pseudoautosomal X loci where the father only has one allele per locus, which is non-randomly transmitted to his offspring). The Law of Independent Assortment states that alleles will segregate to gametes independently of each other, outside of regions of the genome genetically linked in the post-meiotic DNA within the gametes.

As alluded to above, the vast majority of MR studies conducted to date have used samples of unrelated individuals. Indeed, genome-wide association studies (GWAS) of putatively unrelated individuals often remove cryptically related individuals from the analyses. Familial effects can mean that GWAS SNP–phenotype associations from unrelated individuals can differ from the associations of the SNP with the phenotype that would be seen within sibships. Within a population that shares similar family formations this will not influence the average degree of prediction provided by genetic variation. However, between populations and as family formations change over time, these family effects may change, leading to SNP–phenotype associations being unstable. Below we will outline sources of bias and then discuss the implications of these and how they can be overcome using data from related individuals.

Transmission ratio distortion

Transmission ratio distortion (TRD) refers to situations where the transmission of the parents' two alleles to their offspring deviates from the expected 50:50 probability (7). There is an opportunity to empirically evaluate this phenomenon due to the increasing availability of genome-wide data from parent–offspring trios, although statistical power remains an issue even in this era of increasing sample sizes (8). TRD has two broad classes: (1) segregation distortion, in which processes occurring during meiosis ('meiotic drive') or fertilization ('gametic competition'), among others, favor one parental allele over another; (2) viability selection, in which the viability of gametes and zygotes, through to live birth, depends on offspring genotype (7,9). In itself, TRD is not problematic for MR, but it becomes an issue if the environmental factors influence any of the processes. Environmental factors that affect TRD will become associated with genotype and could lead to bias in MR studies. For example, maternal folate status may relate to both meiotic drive and embryo selection (Fig. 1) (10,11). If true, these effects would result in genetic variants becoming associated with folate level, which could bias MR estimates using these variants. Furthermore, if a particular combination of genotype and environment influences survival from birth until study entry, then this can similarly lead to genotypes becoming associated with environmental factors and therefore to biased MR estimates. Thus, MR analyses must assume that the segregation of alleles at germ cell production, their equal representation at zygote formation, the survival of conceptuses carrying different alleles to live birth and then to entry into a study are all independent of the environment. It may be possible to detect such TRD using Hardy–Weinberg tests if the effects are large, although individual SNPs with evidence of violations of Hardy–Weinberg equilibrium expectations are routinely excluded from both GWAS and MR studies (12).

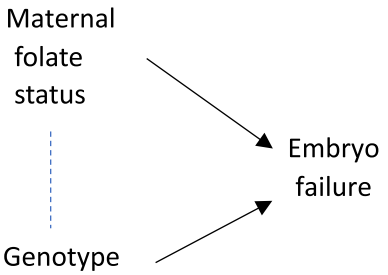

Figure 1. The impact of transmission distortion. If a genotype affects the likelihood of embryo failure, then genotype may no longer be independent of environment. The blue dashed line indicates the induced association.

Selection/collider bias

The bias introduced by forms of non-random segregation and selection through to study entry is a form of selection/collider bias (Fig. 2) (13). If participants are non-randomly selected into studies for any reason, then genotypes associated with selection may become associated with other factors related to selection in ascertained samples *even* if the genotype is independent of the environment in the wider population (14). For example, studies that oversample healthier people may induce associations between SNPs associated with being healthier (e.g. low values of a coronary heart disease polygenic score), and other factors that affect selection into a study (e.g. socioeconomic position or education). Another example is case-only studies of disease progression, where genotypes associated with the incidence of disease may associate with the progression of the disease, even in the absence of a true causal effect on progression (15) Selection of cases may induce associations between a genetic variant that influences disease onset and any other (phenotypic or genotypic) factors associated with incidence of disease. If any of these factors that become associated with the genetic variant themselves influence disease progression, then it will appear that the genetic variant also has such an influence (8). Given certain assumptions, it is possible to perform analyses that are robust to selection bias (16). For example, methodlogists have also proposed methods that may mitigate this bias using inverse probability weights of sampling (14).

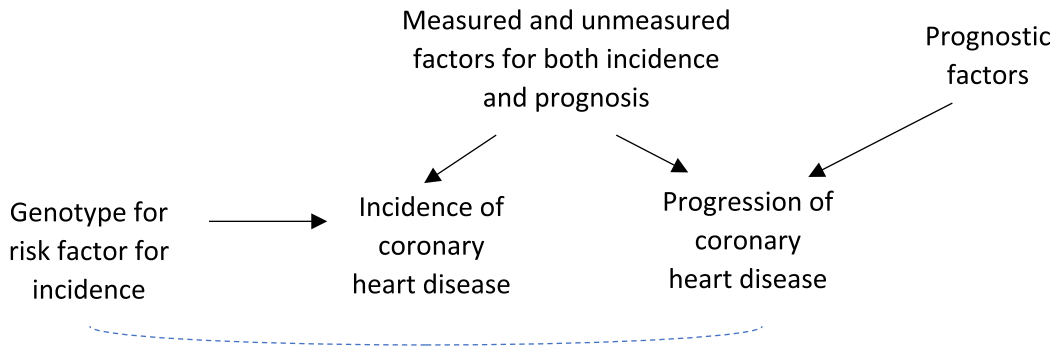

Figure 2. Selection/collider bias can induce associations between genotype and factors related to selection. In a case-only study of disease progression, factors influencing the likelihood of case status may associate with progression of the disease because of collider bias. Figure adapted from Paternoster *et al.* (2017) (15).

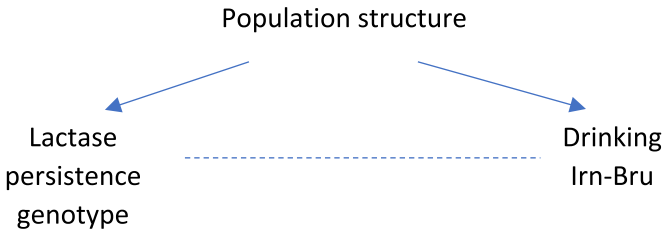

Figure 3. Confounding by population structure can induce associations between genotype and phenotype (dashed blue line). For example, alleles associated with lactase persistence are more common in northern regions of the United Kingdom. On average, these regions also have poorer health; this does not imply that lactase persistence affects health outcomes. (18) Figure adapted from Brumpton *et al.* (2019) (61).

Population stratification

The diverse ancestral origins of many human populations can lead to systematic allele frequency differences across study populations (Fig. 3). Such population structure can result in otherwise unrelated phenotypic differences across the population becoming spuriously associated with genetic variation. For example, genetic variants associated with lactase persistence (the ability to process dairy products after infancy) are stratified across Europe and are more common in northern European populations (17,18). GWAS typically account for these differences using methods that utilize genome-wide SNP data, including principal components, (19) or linear mixed models (20). However, the sample sizes used initially to assess the ability of these methods to control for stratification were relatively limited, (21) and larger biobank scale studies have provided evidence that these commonly used methods to account for population stratification are unlikely to eliminate all residual differences within populations, (22) or differences due to rare variation (23). Within family estimates of SNP–phenotype associations (members of the same sibship having the same ancestry) can correct for this bias (24).

Dynastic effects

Dynastic effects are any indirect effects of the parent's genotype on their offspring that are mediated via the parents' phenotype, i.e. effects that are not mediated via direct inheritance of DNA. As we will discuss later, dynastic effects can be seen most clearly in studies in which the non-transmitted parental alleles at each locus—i.e. the alleles which are not inherited by the offspring—can be shown to relate to offspring phenotype. The intergenerational transmission of education is potentially an example of a dynastic effect. Dynastic effects would induce associations between parents' non-transmitted education-associated alleles and offspring educational attainment (EA) or other outcomes. A more educationally stimulating family environment or higher income of parents carrying such alleles could mediate these effects (24). For transmitted alleles, the offspring-level association will be a combination of the parental (dynastic) effect and the within-individual effect (Fig. 4). Thus in the case of the 1271 SNP–EA associations reported in a GWAS of unrelated individuals (25), these SNPs may associate with offspring EA via a direct effect of the SNP in the offspring (e.g. offspring SNP -> offspring EA), or via a dynastic effect of parental SNPs on parents' phenotypes, which in turn influence offspring EA (e.g. parental SNP -> parental EA -> offspring EA). Therefore the SNP–EA association reflects a combination of the parental effects (genetic nurture effects) and effects of the SNPs in the offspring (24,26). Kong *et al.* estimated the size of dynastic effects using non-transmitted alleles for EA, age at first child, HDL cholesterol, BMI, fasting glucose, height, cigarettes per day and a composite health measure, finding evidence of associations between non-transmitted parental variants and offspring phenotypes of varying magnitude, with the largest effects being for education and age at first child, which may be considered to be the traits with the strongest social component. These findings suggest that SNP–phenotype associations in unrelated individuals may capture broader factors beyond the causal effect of the SNP in the offspring. Dynastic effects that are shared across siblings or that are independent of genotype within families will not bias within-family or within-sibship estimates of the SNP–phenotype association (24).

Assortative mating

Assortative mating has both genotypic and phenotypic consequences for populations (27). Humans do not mate at random; parental pairs are more alike than would be expected for two individuals randomly drawn from the population for many traits,

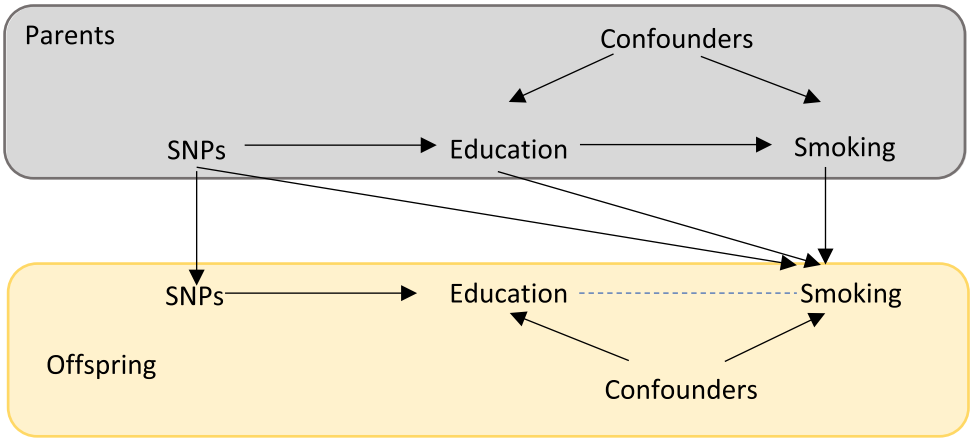

Figure 4. Dynastic effects can induce SNP–phenotype associations that are not due to the effects of the SNP in the individual. For example, suppose parents' smoking affects their offspring's likelihood of developing a wheeze, this would induce associations between SNPs associated with smoking initiation in the offspring and wheezing in the offspring (dashed blue line) even if the offspring's smoking behavior could not cause them to wheeze (i.e. if the offspring were children and had not started to smoke). Figure adapted from Brumpton *et al.* (2019) (61).

including height, education and adiposity (28,29). Assortative mating on a phenotype can be direct (i.e. partners selecting a mate on a specific phenotype), or indirect, such as being a consequence of social homogamy (i.e. partners from a similar background or ancestry being more likely to pair) (30). A further complication is that couples may also influence one another's phenotypes (but not genotype), and become more similar over time after pairing (31). Assortative mating can be on a single trait, e.g. more educated individuals selecting partners who are also more educated, or cross-trait, e.g. more educated people preferring taller partners. Single trait assortative mating can inflate SNP–phenotype associations, (32) but this association remains a valid test of the null hypothesis that the SNP does not affect the phenotype. Single trait assortative mating means that the SNP–exposure associations are likely to overestimate the causal effect of each SNP on the exposure. This overestimation occurs because single trait assortative mating induces associations between SNPs used as instruments and other SNPs that cause the exposure that were not included in the analysis. However, single trait assortative mating on the exposure alone is insufficient to cause bias in MR studies. This is because any change in the magnitude of the SNP–exposure association caused by variants in LD (i.e. induced by assortative mating) will also be reflected by a proportional change in the magnitude of the association between the variant and the outcome. In offspring, cross-trait assortative mating can induce associations between SNPs affecting one phenotype and SNPs affecting another phenotype (32). As a result, cross-trait assortative mating can cause bias in MR estimates and invalidate tests of the null hypothesis (Fig. 5) (32). For example, assortative mating on EA and height will induce associations between SNPs affecting education and measured height, and similarly, it will induce associations between SNPs affecting height and measured EA. The bias caused by assortative mating can be amplified across generations if the same patterns of assortment occur in each generation. The presence of assortative mating can be evaluated by estimating the phenotypic association between spouses, or by estimating the correlation in genetic scores between mates, or from polygenic scores constructed from individual chromosomes within the same individual (30). Genetic variants inherited by offspring are still random within a family. Therefore, studies estimating SNP–phenotype associations within families will be robust to bias from assortative mating.

Impact of these biases for causal inference in genetics

Each of these sources of bias can result in SNP–phenotype associations capturing more than the causal effect of varying the SNP on a phenotype in a single individual. The consequences of this potential misestimation or bias are dependent on the specific research question. If a study is interested in predicting a phenotype within a given population, then predictions using SNP–phenotype associations from unrelated individuals may be valid, as long as family formations are similar in the source GWAS populations. However, misestimation may be more problematic for research questions that involve causal interpretation of results such as MR. Historically, GWAS studies have corrected for population structure by restricting to populations with relatively homogenous ancestry and employing previously described techniques to account for population structure (19). A disadvantage of limiting studies to homogenous populations is that it may reduce the genomic variation and external validity of estimates. Additionally, large biobank studies have suggested that methods generally used to correct for population stratification are unlikely to control for all differences within populations fully (22,33). These results are in contrast to early genetic epidemiological studies on relatively small samples (e.g. the Wellcome Trust Case Control Consortium), which found little evidence of residual confounding after adjusting for PCs (34). The much larger size of recent biobanks increases the ability to detect smaller, but real, biases. Family-based study designs are potentially robust to many of these sources of bias because the random inheritance of DNA from parents to offspring will ensure that conditional on parental genotype, offspring genotype is independent of population structure, dynastic effects and assortative mating. Within family approaches have been widely described and used in genetics, but to date less frequently used in MR studies (35–37). In the following sections, we describe methods for using data from family studies within genetic epidemiology, with a particular focus on MR.

Family-based studies

Siblings

A wide range of genetic epidemiological study designs can use data from siblings. For example, Lionel Penrose used the

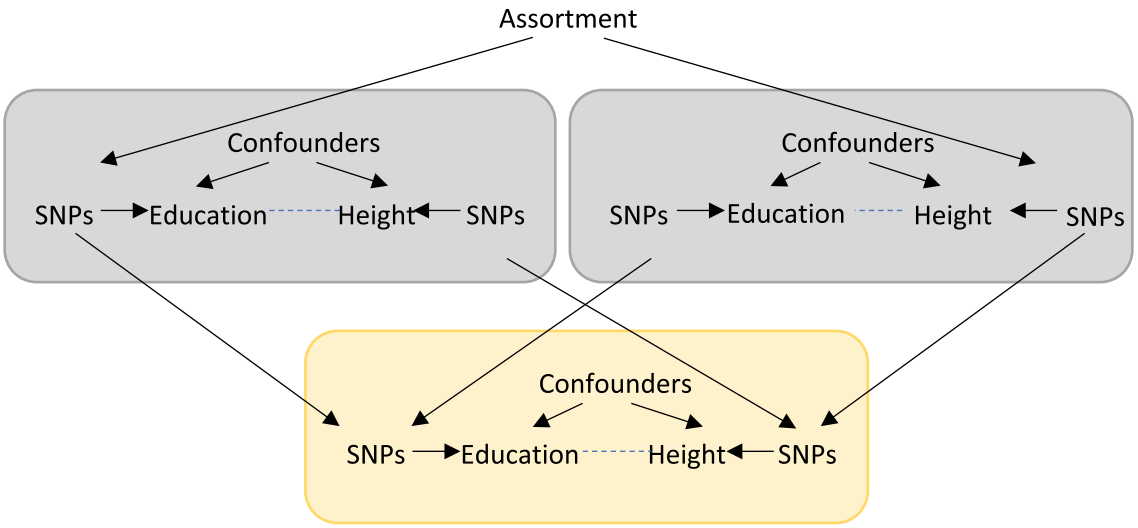

Figure 5. Cross-trait assortative mating can induce associations between SNPs and phenotypes. Figure adapted from Brumpton *et al.* (2019). Cross-trait assortment, for example, if more educated individuals assort with taller people induces associations between genetic variants associated with height and offspring education. The dashed blue line indicates the induced association. Controlling for parental genotype, either directly using genotyped parent–offspring trios or implicitly using siblings (not shown), controls for this bias.

phenotype of siblings of index cases with a range of intellectual disabilities to demonstrate a shift from what would now be called cases with a highly polygenic basis to more severe cases in which there was a single major cause (38). More recent studies have applied this design to similar problems (39). Linkage-based gene mapping studies using affected sib-pair and discordant sib-pair designs have been used to map genetic loci (40,41), and studies using maternal versus paternal half-sib comparisons have investigated both differences in the transmission of effects from mothers and from fathers and shared household effects, given that paternal half-sibs often reside separately (42). Within-sibship analyses can avoid some of the biases above that can arise from dynastic effects, assortative mating and population stratification because each of the meiosis and conception events that determined the sibling's DNA is an independent event conditional on the parental genotypes. The genotypic differences between siblings are random, and these genotypic differences should be independent of environmental confounders (unless TRD related to an environmental factor, as discussed above, is active), and we can obtain estimates that are robust to the biases above. Data from non-identical (dizygotic, DZ) twins can be used in such within-family analyses, but not identical (monozygotic, MZ) twins for obvious reasons. However, if a study also sampled the non-identical siblings of MZ twins, then both the MZ twins and their sibling can be included in within-family analyses. At any given locus the genotype of two siblings may be identical. If this is the case, then at this locus the siblings will not contribute any data to the estimate (which thus reduces statistical power). Analyses using family data should use standard errors that allow for clustering by family.

Numerous estimators can use sibling data, including family fixed effects, difference estimators and within family estimators that control for the mean levels of phenotype and genotype within a family (35,36). Fixed effect estimators can be used to account for all differences between families in both genotype and phenotype. Differences between families in terms of genotype could include differences in allele frequency due to assortative mating, and differences in phenotype that are due to differences in the family environment, such as dynastic effects, and ancestry differences. The fixed effect estimator can be implemented by including binary indicators for each family. However, this approach is relatively computationally inefficient and can be prohibitively time and memory intensive for datasets with a large number of families. An alternative equivalent analytic approach is to use the within-family transformation, which is equivalent to subtracting the family level means from all variables and then running a regression model on the deviations from the family means. This estimator is more computationally efficient (43). A third alternative is to use difference estimators. These take the difference between pairs of siblings in genotype and phenotype and estimate the association between the differences. If there are data for more than one sibling in a family, then all possible sibling pairs can be included. Thus, for the difference estimator, the level of analysis is sibling pair, not the individual. For the difference estimator, standard errors must allow for clustering across all sibling pairs within a family when using more than one pair per family. This approach is simple and computationally efficient. In the special case where there are exactly two siblings per family, the fixed effect, within family and difference estimators are equivalent.

Parent–offspring trios

Studies that sample parent–offspring trios may also assist in controlling for the aforementioned family-level biases. Similar to the logic that applies to analyzing siblings, the meiotic process ensures that (absent of TRD) a zygote's genotype is random conditional on the parents' genotype (i.e. which of the mother or father's allele is passed on at any given locus is entirely stochastic). Again, associations between the offspring genotype and the environment induced by population stratification, dynastic effects or assortative mating may be controlled for after conditioning on parental genotype. If genome-wide data for both parents and offspring are available, it is possible to use inferred patterns of inheritance of haplotypes to identify both the alleles that the offspring inherited and the variants that they did not inherit. The non-inherited variants are a potentially powerful source of inference about the familial effects. Non-inherited variants cannot be biologically expressed in the offspring. Therefore, any association between non-inherited variants and

offspring phenotypes must be via expression in the parents' (or other ancestors') phenotypes. However, the non-inherited variants could associate with the offspring's phenotypes because of population structure, dynastic effects or assortative mating.

There are multiple estimators for use with trio data. The simplest approach is to adjust analyses for parental genotype. Conditioning on parental genotype is likely to account for many forms of familial bias for both SNP–phenotype associations estimated in GWAS or single sample MR analysis. A limitation of this approach is that it only controls for familial effects; it does not estimate the effects of parental phenotypes on their offspring. The effects of parents' phenotype on their offspring's outcomes can also be estimated using multivariable MR in which there are three exposures: the offspring's phenotype, and the mother and father's phenotypes. Estimation of these effects is possible because there are three sets of genetic instruments, the offspring, mother and father's genotypes, one for each phenotype. Therefore, this analysis is exactly identified (one instrument per exposure). However, these models may attribute other familial effects to parents. It is not possible to distinguish whether the source of the familial effects is population stratification, dynastic effects or assortative mating using multivariable MR. It is possible to run some of these models using data from offspring–mother duos. However, these models do not control for familial effects that are mediated via paternal genotype and can suffer from collider bias in situations where there is an open path from paternal genotype to the offspring phenotype of interest (44,45).

Family studies can be used to investigate a range of intergenerational dynastic effects, to test whether the parents' phenotype directly affects their offspring's outcomes. Examples of dynastic effects include intra-uterine effects of maternal characteristics on their offspring, for example, the effect of higher maternal BMI or alcohol consumption during pregnancy on offspring outcomes. Intra-uterine effects can be estimated using parental genetic variation. If there was little evidence that maternal genotypes for the risk factor of interest are associated with offspring outcomes in a well-powered study (1), then it can be inferred that a sizable intra-uterine effect is unlikely. Intra-uterine effects can be estimated using two-sample MR (44,46,47). Studies using MR to assess intra-uterine effects should ideally demonstrate that the genetic variants affect the exposure of interest during pregnancy, and if at all possible, explain why effects at other points in the life-course cannot explain any associations. For example, suppose (in contrast to Richmond *et al.* 2017) (48) maternal genetic variants known to associate with BMI were found to associate with offspring outcomes. Alone these associations would be insufficient to prove that the mechanism was due to an intra-uterine effect. An alternative explanation would be that maternal BMI affects offspring outcomes later in the life course (e.g. via infant and childhood nutrition). A study could identify the timing of these effects if it identified genetic variants that have a differential effect on maternal phenotypes during pregnancy (e.g. genetic variants that influence BMI during pregnancy but not outside of pregnancy). Such variants could be used with multivariable MR to identify the timing of exposure. However, these study designs are often not currently feasible given available data. Paternal samples can be used in negative control analyses to assess the plausibility of MR assumptions. For example, maternal genetic variants in *MTHFR* that are associated with lower folate levels also associate with a higher risk of neural tube defects in offspring, whereas paternal genetic variants in the same gene do not. These associations support the hypothesis of an intrauterine effect of maternal folate levels (1).

Familial effects can confound case-control studies of disease. For example, genetic variants associated with EA positively associate with the risk of autism in case-control studies of unrelated individuals (49). One potential mechanism for generating these associations could be a direct effect of parental education on the likelihood of diagnosis, for example, if more educated individuals were more likely to seek a diagnosis for their offspring and participate in genetic epidemiological studies. Weiner *et al.* (2017) describe a polygenic transmission disequilibrium test that can be used to assess whether variants associated with a given trait were more likely to be inherited by cases (50). In practice, this means calculating the average parental polygenic score of the trait (education) and subtracting the value of the child's polygenic score for this trait. In expectation, across a sample, if there is no effect of the trait on the risk of autism, we would expect the offspring's score to equal the average of their parents' scores. If the offspring score is on average higher than their parents, then this suggests the expression of the trait in the offspring increases risk, whereas if it is on average lower, then this suggests that expression of the trait in offspring decreases risk. Weiner *et al.* used this method to demonstrate that genetic variants in the offspring related to both education and schizophrenia (in addition to genetic variants related to autism) increased risk of an autism diagnosis (50). Therefore, the correlation between education-associated genetic variants and autism is unlikely to be solely attributable to familial factors.

Estimating familial effects

Genetic epidemiology can potentially provide evidence about how individuals within a family affect each other. Below we describe how DNA can provide evidence about specific familial effects.

Assortative mating

MR can potentially be used to investigate the basis of assortative mating, such as the relationship between spousal pairs' alcohol consumption (51). Spousal pairs are more similar in their level of alcohol consumption than expected by chance, but is this due to such similarity increasing the probability of pair formation? Genetic variants associated with alcohol consumption have been used in MR studies (51), with behavior in one spouse considered the exposure and that of the other spouse the outcome. This study found evidence of assortative mating on a genetic predictor of alcohol consumption, and that this association recapitulated the similarity between spouses in their phenotypic alcohol consumption patterns. These results imply that assortative mating occurs by alcohol consumption, which implicitly induces genetic similarity on genotypes related to alcohol consumption. The results for height, another phenotype known to display patterns of assortative mating, were similar. However, more research is needed to understand the specific assumptions required to identify the effects of assortative mating and between spouses.

Familial effects

Individuals within the same family may have effects on each other, for example, dynastic effects, in which the parents' phenotypes directly affect their offspring. Other family members may

affect outcomes via competition and cooperation effects, for example, siblings' phenotypes may influence each other's outcomes. Many studies have described the assumptions required to identify these effects in detail in animal, twin and sibling studies (52–56), but relatively few studies have used molecular genetic data to estimate sibling effects in humans. Samples of siblings and their parents can potentially be used to estimate sibling effects. For example, consider a study estimating the effect of a child's BMI on their siblings' BMI. Conditional on their parents' genotype, the children will have inherited a random set of BMI increasing variants. If the children who inherited more BMI increasing variants than expected have siblings with higher BMI than expected from their genotype, then this would suggest that children's BMI affects their siblings' BMI. Children could also exert effects on their parents, for example, highly educated children may increase their parents' longevity (57). This effect could be investigated through MR in studies with parental and offspring genetic data.

MR-Twin-Direction of Causation

MR-Twin-Direction of Causation combines the logic of MR with classical twin models (58). This approach requires samples of twins with molecular genetic data. The relationships between the twins (specifically the cross-twin cross-trait covariances) can be exploited to identify the causal effects of the exposure and potential pleiotropic effects of the genetic variants by restricting terms in the model. As a result, this model can potentially use polygenic scores constructed using more liberal thresholds that standard MR studies would exclude. A limitation of this approach is that it requires relatively large samples of genotyped twins and cannot use estimates from GWAS of unrelated individuals as in two-sample MR. Also, the method assumes either the absence of unique environmental or common environmental confounding, which may not be realistic. Similar approaches have been proposed for siblings, for example, genetic instrumental variables (59).

Pleiotropy robust within family methods

Methodologists have proposed a large number of pleiotropy robust methods that use summary data from GWAS of unrelated individuals (60). In general, these methods are not robust to familial-level biases described above. Familial effects are likely to induce bias in proportion to the magnitude of each SNP's association with the phenotype of interest. For example, consider dynastic effects due to parental education (and thus parental alleles that influence education). These effects are likely to influence SNP associations with other phenotypes in the offspring in proportion to each SNP's effect on education. Thus, the INSIDE assumption in MR Egger regression is also likely to be violated and hence causal estimates biased. In addition, widely used sensitivity analyses, such as the weighted median and mode, are also likely to be biased, because all SNPs, including the median and modal SNPs, will suffer from bias.

However, if within family summary data of SNP–phenotype associations are available (61), then they can trivially be used with many of the proposed summary data estimators (62–64). Coefficients and standard errors that allow for a familial effect (e.g. a within family estimate from siblings or controlling for parental genotype) can replace estimates from unrelated individuals from GWAS. The precision of these estimates will typically be lower, and hence the power of the MR analysis will be lower. Therefore, in many cases, within-family estimates may be used as a sensitivity analysis, acknowledging the limitations of power. If there is little evidence that the within-family estimates differ from (more precise) estimates in unrelated individuals, then the latter should be preferred.

Triangulation

Multiple sources of evidence about hypotheses can strengthen causal inferences. If the inferences are consistent, then this can increase confidence about results (65,66). Family studies provide an alternative source of evidence about SNP–phenotype associations. If evidence from family studies is consistent with evidence from unrelated individuals, then this may suggest that some of the biases described above are not having a large influence. Alternatively, if there are substantial differences between within family estimates and those from unrelated individuals, then this may provide evidence about the origins of SNP–phenotype associations (e.g. residual population stratification, assortative mating or dynastic effects, etc.). Therefore, family studies can provide a useful additional source of evidence about the effects of SNPs, which are likely to be less precise, but more robust to many sources of bias. Other sources of evidence, including natural experiments, twin, adoption and half-sib studies, can all contribute to the triangulation of evidence in such situations.

Future research

Family-based studies can overcome many of the limitations of MR, such as bias due to assortative mating, dynastic effects, population stratification and horizontal pleiotropy. A limitation of these designs is that while they are more robust to many sources of bias, there are far fewer samples available than for unrelated individuals. However, there are an increasing number of studies with large samples of genetic data from related individuals, such as the Norwegian Mother, Father and Child Cohort Study (MoBa), The Nord-Trøndelag Health Study (HUNT) and Millennium Cohort Study and siblings sampled as part of UK Biobank and a large number of twin studies and registries from around the world (67–70). The increasing availability of data from these studies makes it possible to obtain potentially more reliable estimates of the causal effects of phenotypes on outcomes. However, there are relatively few large case-control studies that have sampled relatives (either siblings or parent–offspring trios). This limitation means it can be very challenging to estimate SNP–disease association, particularly when a condition is rare and/or the effect sizes are small. In sibling studies, this is because only siblings with discordant disease status contribute to the estimates. One approach to overcome this is to use continuous proxies of the condition, either as an underlying causal mechanism (e.g. LDL cholesterol for coronary heart disease), or a symptom score (depression symptoms for depression), or potentially even a survival outcome (time to diagnosis).

Family studies offer the possibility of more reliable causal estimates of the effects of exposures. As a result, family studies could provide new and more robust evidence for a wide range of fields, from social science to drug development. A particularly cost-effective program of research may be to augment existing cohort or clinical studies by recruiting family members.

Author contributions

N.M.D. wrote and revised the paper. G.D.S. conceived and revised the paper. All other authors revised the manuscript.

Acknowledgements

We would like to thank Debbie Lawlor, David Bann, Tim Morris and Amanda Hughes for helpful comments on this manuscript.

Funding

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]. 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 Stiftelsen Kristian Gerhard Jebsen funds the K.G. Jebsen Center for Genetic Epidemiology; 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.

References

1. Davey Smith, G. and Ebrahim, S. (2003) Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.*, **32**, 1–22.
2. Walker, V.M., Kehoe, P.G., Martin, R.M., Davies, N.M. *et al.* (2017) Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. *Int. J. Epidemiol.* [Internet]. 2019 Jul 23 [cited 2019 Oct 15]; Available from: https://academic.oup.com/ije/advance-article/doi/10.1093/ije/dyz155/5537361.
3. von Hinke Kessler Scholder, S., Wehby, G.L., Lewis, S., Zuccolo, L. (2014) Alcohol exposure *in utero* and child academic achievement. *Econ. J.*, **124**, 634–667.
4. Davey Smith, G. and Hemani, G. (2014) Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.*, **23**, R89–R98.
5. Davey Smith, G., Lawlor, D.A., Harbord, R., Timpson, N., Day, I., Ebrahim, S. (2007) Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med.*, **4**, e352.
6. Lock, R.H. (1906) Recent progress in the study of variation, heredity, and evolution. In *Recent Progress in the Study of Variation, Heredity, and Evolution*. John Murray, London.
7. Meyer, W.K., Arbeithuber, B., Ober, C., Ebner, T., Tiemann-Boege, I., Hudson, R.R., Przeworski, M. (2012) Evaluating the evidence for transmission distortion in human pedigrees. *Genetics*, **191**, 215–232. https://doi.org/10.1534/genetics.112.139576
8. Evans, D.M., Morris, A.P., Cardon, L.R., Sham, P.C. (2006) A note on the power to detect transmission distortion in parent-child trios via the transmission disequilibrium test. *Behav. Genet.*, **36**, 947–950. https://doi.org/10.1007/s10519-006-9087-2
9. Huang, L.O., Labbe, A. and Infante-Rivard, C. (2013) Transmission ratio distortion: review of concept and implications for genetic association studies. *Hum. Genet.*, **132**, 245–263.
10. Nadeau, J.H. (2017) Do gametes woo? Evidence for their nonrandom union at fertilization. *Genetics*, **207**, 369–387.
11. Bochud, M., Chiolero, A., Elston, R. C., et al. (2008) A cautionary note on the use of Mendelian randomization to infer causation in observational epidemiology. *Int. J. Epidemiol.*, **37**, 414–416; author reply 416–417.
12. Rodriguez, S., Gaunt, T.R. and Day, I.N.M. (2009) Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies. *Am. J. Epidemiol.*, **169**, 505–514.
13. Munafò, M.R., Tilling, K., Taylor, A.E., Evans, D.M., Davey Smith, G. (2018) Collider scope: when selection bias can substantially influence observed associations. *Int. J. Epidemiol.*, **47**, 226–235. https://doi.org/10.1093/ije/dyx206
14. Hughes, R.A., Davies, N.M., Davey Smith, G., Tilling, K. (2019) Selection bias when estimating average treatment effects using one-sample instrumental variable analysis. *Epidemiology*, **30**, 350–357. https://doi.org/10.1097/EDE.0000000000000972
15. Paternoster, L., Tilling, K. and Davey Smith, G. (2017) Genetic epidemiology and Mendelian randomization for informing disease therapeutics: conceptual and methodological challenges. *PLOS Genetics*, **13**, e1006944.
16. Dudbridge, F., Allen, R.J., Sheehan, N.A., Schmidt, A.F., Lee, J.C., Jenkins, R.G., Wain, L.V., Hingorani, A.D., Patel, R.S. (2019) Adjustment for index event bias in genome-wide association studies of subsequent events. *Nat. Commun.*, **10**, 1561. https://doi.org/10.1038/s41467-019-09381-w
17. Ingram, C.J.E., Mulcare, C.A., Itan, Y., Thomas, M.G., Swallow, D.M. (2009) Lactose digestion and the evolutionary genetics of lactase persistence. *Hum. Genet.*, **124**, 579–591.
18. Davey Smith, G., Lawlor, D.A., Timpson, N.J., Baban, J., Kiessling, M., Day, I.N.M., Ebrahim, S. (2009) Lactase persistence-related genetic variant: population substructure and health outcomes. *Eur. J. Hum. Genet.*, **17**, 357–367. https://doi.org/10.1038/ejhg.2008.156
19. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909. https://doi.org/10.1038/ng1847
20. Loh, P.-R., Tucker, G., Bulik-Sullivan, B.K., Vilhjálmsson, B.J., Finucane, H.K., Salem, R.M., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, , Patterson, N., Price, A.L. (2015) Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.*, **47**, 284–290. https://doi.org/10.1038/ng.3190
21. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.
22. Haworth, S., Mitchell, R., Corbin, L., Wade, K.H., Dudding, T., Budu-Aggrey, A., Carslake, D., Hemani, G., Paternoster, L., Smith, G.D., Davies, N., Lawson, D.J., Timpson, N., (2019) Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat. Commun.*, **10**. https://doi.org/10.1038/s41467-018-08219-1
23. Mathieson, I. and McVean, G. (2012) Differential confounding of rare and common variants in spatially structured populations. *Nat. Genet.*, **44**, 243–246.
24. Kong, A., Thorleifsson, G., Frigge, M.L., Vilhjalmsson, B.J., Young, A.I., Thorgeirsson, T.E., Benonisdottir, S., Oddsson, A.,

Halldorsson, B.V., Masson, G., Gudbjartsson, D.F., Helgason, A., Bjornsdottir, G., Thorsteinsdottir, U., Stefansson, K. (2018) The nature of nurture: effects of parental genotypes. *Science*, **359**, 424–428. https://doi.org/10.1126/science.aan6877

25. Lee, J.J., 23andMe Research Team, COGENT (Cognitive Genomics Consortium), Social Science Genetic Association Consortium , Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A. *et al.* (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.*, **50**, 1112–1121. https://doi.org/10.1038/s41588-018-0147-3
26. Plomin, R. and Bergeman, C.S. (1991) The nature of nurture: genetic influence on "environmental" measures. *Behav. Brain Sci.*, **14**, 373–386.
27. Wright, S. (1921) Systems of Mating. III. Assortative mating based on somatic resemblance. *Genetics*, **6**, 144–161.
28. Silventoinen, K., Kaprio, J., Lahelma, E., Viken, R.J., Rose, R.J. (2003) Assortative mating by body height and BMI: Finnish twins and their spouses. *Am. J. Hum. Biol.*, **15**, 620–627. https://doi.org/10.1002/ajhb.10183
29. Qian, Z. (1998) Changes in assortative mating: the impact of age and education, 1970–1990. *Demography*, **35**, 279.
30. Robinson, M.R., Kleinman, A., Graff, M., Vinkhuyzen, A.A.E., Couper, D., Miller, M.B., Peyrot, W.J., Abdellaoui, A., Zietsch, B.P., Nolte, I.M. *et al.* (2017) Genetic evidence of assortative mating in humans. *Nat. Hum. Behav.*, **1**, 0016. https://doi.org/10.1038/s41562-016-0016
31. Ask, H., Idstad, M., Engdahl, B., Tambs, K. (2013) Non-random mating and convergence over time for mental health, life satisfaction, and personality: the Nord–Trøndelag Health Study. *Behav. Genet.*, **43**, 108–119. https://doi.org/10.1007/s10519-012-9578-2
32. Hartwig, F.P., Davies, N.M. and Davey Smith, G. (2018) Bias in Mendelian randomization due to assortative mating. *Genet. Epidemiol.*, **42**, 608–620.
33. Abdellaoui, A., Hugh-Jones, D., Yengo, L., Kemper, K. E., Nivard, M. G., Veul, L., Holtz, Y., Zietsch, B. P., Frayling, T. M., Wray, N. R., Yang, J., Verweij, K. J. H., Visscher, P. M., (2019) Genetic correlates of social stratification in Great Britain. *Nature Human Behaviour*, doi:10.1038/s41562-019-0757-5.
34. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.
35. Fulker, D.W., Cherny, S.S., Sham, P.C., Hewitt, J.K. (1999) Combined linkage and association sib-pair analysis for quantitative traits. *Am. J. Hum. Genet.*, **64**, 259–267. https://doi.org/10.1086/302193
36. Abecasis, G.R., Cardon, L.R. and Cookson, W.O.C. (2000) A general test of association for quantitative traits in nuclear families. *Am J. Hum. Genet.*, **66**, 279–292.
37. Spielman, R.S. and Ewens, W.J. (1996) The TDT and other family-based tests for linkage disequilibrium and association. *Am. J. Hum. Genet.*, **59**, 983–989.
38. Penrose, L.S. (1949) The biology of mental defect. In *The Biology of Mental Defect*. Sidgwick and Jackson, Ltd., London, UK.
39. Reichenberg, A., Cederlöf, M., McMillan, A., Trzaskowski, M., Kapra, O., Fruchter, E., Ginat, K., Davidson, M., Weiser, M., Larsson, H., Plomin, R., Lichtenstein, P. (2016) Discontinuity in the genetic and environmental causes of the intellectual disability spectrum. *Proc. Natl. Acad. Sci. USA*, **113**, 1098–1103. https://doi.org/10.1073/pnas.1508093112
40. Risch, N. and Zhang, H. (1995) Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science*, **268**, 1584–1589.
41. Risch, N. (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am. J. Hum. Genet.*, **46**, 229–241.
42. Kendler, K.S., Ohlsson, H., Sundquist, J., Sundquist, K. (2019) Maternal half-sibling families with discordant fathers: a contrastive design assessing cross-generational paternal genetic transmission of alcohol use disorder, drug abuse and major depression. *Psychol. Med.*, 1–8. https://doi.org/10.1017/S0033291719000874
43. Greene, W. (2002) *Econometric Analysis*. Prentice Hall, Upper Saddle River, New Jersey.
44. Evans, D.M., Moen, G.-H., Hwang, L.-D., Lawlor, D.A., Warrington, N.M. (2019) Elucidating the role of maternal environmental exposures on offspring health and disease using two-sample Mendelian randomization. *Int. J. Epidemiol.* https://doi.org/10.1093/ije/dyz019
45. Warrington, N.M., Beaumont, R.N., Horikoshi, M., Day, F.R., Helgeland, Ø., Laurin, C., Bacelis, J., Peng, S., Hao, K., Feenstra, B. *et al.* (2019) Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat. Genet.*, **51**, 804–814.
46. Pierce, B.L. and Burgess, S. (2013) Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am. J. Epidemiol.*, **178**, 1177–1184.
47. Hwang, L.-D., Lawlor, D.A., Freathy, R.M., Evans, D.M., Warrington, N.M. (2019) Using a two-sample Mendelian randomization design to investigate a possible causal effect of maternal lipid concentrations on offspring birth weight. *Int. J. Epidemiol.*, dyz160. https://doi.org/10.1093/ije/dyz160
48. Richmond, R.C., Timpson, N.J., Felix, J.F., Palmer, T., Gaillard, R., McMahon, G., Davey Smith, G., Jaddoe, V.W., Lawlor, D.A. (2017) Using genetic variation to explore the causal effect of maternal pregnancy adiposity on future offspring adiposity: a Mendelian randomisation study. *PLoS Med.*, **14**, e1002221. https://doi.org/10.1371/journal.pmed.1002221
49. Hagenaars, S.P., Harris, S.E., Davies, G., Hill, W.D., Liewald, D.C.M., Ritchie, S.J., Marioni, R.E., Fawns-Ritchie, C., Cullen, B., Malik, R. *et al.* (2016) Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Mol. Psychiatry*, **21**, 1624–1632. https://doi.org/10.1038/mp.2015.225
50. Weiner, D.J., Wigdor, E.M., Ripke, S., Walters, R.K., Kosmicki, J.A., Grove, J., Samocha, K.E., Goldstein, J.I., Okbay, A., Bybjerg-Grauholm, J. *et al.* (2017) Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.*, **49**, 978–985. https://doi.org/10.1038/ng.3863
51. Howe, L. J., Lawson, D. J., Davies, N. M., St. Pourcain, B., Lewis, S.J., Davey Smith, G., Hemani, G. (2018) Alcohol consumption and mate choice in UK Biobank: comparing observational and Mendelian randomization estimates. bioRxiv. https://doi.org/10.1101/418269
52. Eaves, L. (1976) A model for sibling effects in man. *Heredity*, **36**, 205.
53. Carey, G. (1986) Sibling imitation and contrast effects. *Behav. Genet.*, **16**, 319–341.
54. Baud, A., Mulligan, M.K., Casale, F.P., Ingels, J.F., Bohl, C.J., Callebert, J., Launay, J.-M., Krohn, J., Legarra, A., Williams, R.W., Stegle, O. (2017) Genetic variation in the social

environment contributes to health and disease. *PLoS Genet.*, **13**, e1006498. https://doi.org/10.1371/journal.pgen.1006498
55. Dolan, C.V., de Kort, J.M., van Beijsterveldt, T.C.E.M., Bartels, M., Boomsma, D.I. (2014) GE covariance through phenotype to environment transmission: an assessment in longitudinal twin data and application to childhood anxiety. *Behav. Genet.*, **44**, 240–253. https://doi.org/10.1007/s10519-014-9659-5
56. Moscati, A., Verhulst, B., McKee, K., Silberg, J., Eaves, L. (2018) Cross-lagged analysis of interplay between differential traits in sibling pairs: validation and application to parenting behavior and ADHD symptomatology. *Behav. Genet.*, **48**, 22–33. https://doi.org/10.1007/s10519-017-9882-y
57. Torssander, J. (2013) From child to parent? The significance of children's education for their parents' longevity. *Demography*, **50**, 637–659.
58. Minică, C.C., Dolan, C.V., Boomsma, D.I., de Geus, E., Neale, M.C. (2018) Extending causality tests with genetic instruments: an integration of Mendelian randomization with the classical twin design. *Behav. Genet.*, **48**, 337–349. https://doi.org/10.1007/s10519-018-9904-4
59. DiPrete, T.A., Burik, C.A.P. and Koellinger, P.D. (2018) Genetic instrumental variable regression: explaining socioeconomic and health outcomes in nonexperimental data. *Proc. Natl. Acad. Sci. U. S. A.*, **115**, E4970–E4979.
60. Hemani, G., Bowden, J. and Davey Smith, G. (2018) Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.*, **27**, R195–R208.
61. Brumpton, B., Sanderson, E., Hartwig, F.P., Harrison, S., Vie, G.Å., Cho, Y., Howe, L.D., Hughes, A., Boomsma, D.I., Havdahl, A., Hopper, J. *et al.* (2019) Within-family studies for Mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases. *Genetics*, (preprint): BioRxiv. https://doi.org/10.1101/602516
62. Bowden, J., Davey Smith G., Burgess, S., et al. (2015) Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.*, **43**, 512–525.
63. Hartwig, F.P., Davey Smith, G. and Bowden, J. (2017) Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.*, **46**, 1985–1998.
64. Bowden, J., Davey Smith, G., Haycock, P.C. , Burgess, S. (2016) Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.*, **40**, 304–314. https://doi.org/10.1002/gepi.21965
65. Lawlor, D.A., Tilling, K. and Davey Smith, G. (2016) Triangulation in aetiological epidemiology. *Int. J. Epidemiol.*, **45**, 1866–1886.
66. Munafò, M.R. and Davey Smith, G. (2018) Robust research needs many lines of evidence. *Nature*, **553**, 399–401.
67. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J. *et al.* (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature*, **562**, 203–209. https://doi.org/10.1038/s41586-018-0579-z
68. Magnus, P., Birke, C., Vejrup, K., Haugan, A., Alsaker, E., Daltveit, A.K., Handal, M., Haugen, M., Høiseth, G., Knudsen, G.P. *et al.* (2016) Cohort profile update: the Norwegian Mother and Child Cohort Study (MoBa). *Int. J. Epidemiol.*, **45**, 382–388. https://doi.org/10.1093/ije/dyw029
69. Krokstad, S., Langhammer, A., Hveem, K., Holmen, T., Midthjell, K., Stene, T., Bratberg, G., Heggland, J., Holmen, J. (2013) Cohort profile: the HUNT study, Norway. *Int. J. Epidemiol.*, **42**, 968–977. https://doi.org/10.1093/ije/dys095
70. Connelly, R. and Platt, L. (2014) Cohort profile: UK Millennium Cohort Study (MCS). *Int. J. Epidemiol.*, **43**, 1719–1725.