Using genetic data to strengthen causal inference in observational research

Jean-Baptiste Pingault[1,2], Paul F. O'Reilly[2], Tabea Schoeler[1], George B. Ploubidis[3], Frühling Rijsdijk[2] and Frank Dudbridge[4]*

[1] Department of Clinical, Educational and Health Psychology, University College London, London, WC1H 0DS, UK.
[2] Social, Genetic, and Developmental Psychiatry, King's College London, De Crespigny Park, London, SE5 8AF, UK.
[3]Centre for Longitudinal Studies, Department of Social Science, UCL Institute of Education, University College London, London, WC1H 0AL, UK.
[4]Department of Health Sciences, University of Leicester, Leicester, LE1 7RH, UK.

***e-mail: j.pingault@ucl.ac.uk**

Abstract | Causal inference, which involves progressing from confounded statistical associations to evidence of causal relationships, is essential across the biomedical, behavioural and social sciences; it can reveal complex pathways underlying diseases and traits and help to prioritize targets for interventions. Recent progress in genetic epidemiology — including statistical innovation, massive genotyped datasets and novel computational tools for deep data mining — has fostered the intense development of methods exploiting genetic data and relatedness to strengthen causal inference in observational research. In this Review, we describe how such genetically informed methods differ in their rationale, applicability and inherent limitations, and outline how they should be integrated in future to offer a rich causal inference toolbox.

Introduction

Identifying causal risk and protective factors for relevant phenotypes constitutes a core objective across the biomedical, behavioural and social sciences. Examples of causal questions — some resolved, some still controversial — include the following. Is vitamin E a protective factor for coronary heart disease (CHD)? Is the same true for high-density-lipoprotein cholesterol (HDL-C)? Does higher income protect against depression? Does maternal smoking during pregnancy lower offspring birth weight? Does cannabis use increase the risk of schizophrenia or is there an effect in the reverse direction? Answering such causal questions can advance fundamental knowledge of complex aetiological pathways and profoundly impact applied settings such as public health and policy[1].

The quest to answer causal questions faces major challenges. A primary challenge is confounding , in which a variable (or set of variables) causally influences both the risk factor and the outcome (e.g. income affecting vitamin E intake and CHD). Confounding can generate associations between risk factors and outcomes in the absence of causal relationships. Genetic confounding occurs when genetic factors generate confounding (e.g. variants associated with HDL-C also directly affect CHD). Challenges to causal inference, detailed in BOX 1, can lead to spurious findings in observational epidemiology, because adjusting for key confounders is typically insufficient. For example, two major observational studies concluded that higher consumption of vitamin E reduces risk for CHD[2,3]. These findings, reported in major media outlets, led to a substantial increase in vitamin E consumption[4]. However, subsequent randomized controlled trials (RCTs) reported null findings[5]. This illustrates the potentially disruptive impact of incorrect inference on our aetiological understanding of diseases and on public health.

RCTs, often regarded as the 'gold standard' for causal inference, suffer from their own methodological shortcomings and may be infeasible and unethical (e.g. random allocation to smoking during pregnancy)[6–8]. RCTs are also inefficient in the absence of reliable evidence to prioritize targets, e.g. low drug development success rates result in $2.6 billion costs per approved drug[9]. To tackle the limitations of RCTs and the challenges of causal inference, methods to strengthen causal inference in observational research have been developed over the past decades. Among causal inference methods , genetically informed methods represent powerful tools to account for genetic and environmental confounding. By genetically informed, we mean methods that exploit genetic information embedded in the study design, including data on familial relationships and/or on genetic variation.

Key features of the genome and its transmission at conception make such genetically informed methods particularly valuable for causal inference. First, the expected degree of genetic similarity is known for different types of relationships, which is exploited in family-based designs to control for genetic and environmental confounding[10]. Second, the genetic sequence is fixed from conception and therefore free from reverse causation (BOX 1)[11]. Third, the genome is randomized at conception, which is critical for the use of genetic variants as instrumental variables to strengthen causal inference, as implemented in Mendelian randomization (MR)[12]. Critical developments in recent years have allowed greatly increased applications of genetically informed methods. First, rapid methodological innovations in the use of genetic variants as instrumental variables has extended the range of phenotypes that can be studied and enabled more robust causal inference[13]. Second, the recent availability of massive genotyped and phenotyped datasets has considerably expanded the applicability of these methods[14,15]. Third, novel informatics tools allow data mining of these resources at phenome-wide scale[16,17]. This has led to converging interests between epidemiologists, primarily concerned with modifiable exposures in the population, and geneticists, concerned with molecular mechanisms underlying diseases, traits and behaviours.

In light of these developments, we provide an integrative review of the current range of genetically informed methods to strengthen causal inference. Considering these methods together allows us to outline a coherent framework to understand their commonalities and differences, and to explain how they should be integrated in the future to offer a rich causal inference toolbox. We start by delineating the counterfactual approach to causal inference, which offers a unifying language to understand current genetically informed methods for

causal inference. We then discuss such methods in the following sections, describing family-based methods and implementations of MR. Finally, we detail emerging methods that move the field forward by embedding genetic instruments within family-based designs and by adopting phenome-wide approaches to causal inference. We will not consider non-genetically informed methods for causal inference (see Refs[1,18,19]) nor the use of family-based and genetic variation data to dissect the genetic architecture of phenotypes (see Refs[20,21]).

[H1] A causal inference framework

The counterfactual (also known as 'potential outcomes') approach offers a unifying framework for causal inference that is relevant to genetically informed designs[22]. In a counterfactual scenario, an individual is simultaneously exposed and non-exposed to a risk factor. In this hypothetical setting in which everything besides the exposure is the same, a causal effect can be defined as a difference in outcomes in the exposed and non-exposed scenarios. For example, if an individual survives after a heart transplant, but the same individual dies without the transplant, we can conclude that the heart transplant caused survival in this individual. Naturally, such a scenario is impossible as an individual cannot be simultaneously exposed and non-exposed to a risk factor. Consequently, strict causal inference cannot be achieved because the counterfactual is missing in reality[22]. All causal inference methods — including RCTs — aim to approximate this ideal scenario by investigating substitutes that enable causal inference under reasonable assumptions.

To attain consistent causal inference, achieving or sufficiently approximating exchangeability is essential[23]. Intuitively, exchangeability occurs when exposed and non-exposed groups are balanced on all confounders. In observational studies, vitamin E consumers were not exchangeable with non-consumers (e.g. because of their income), leading to biased estimates. In subsequent RCTs, randomization ensured exchangeability and their findings suggested no protective effect of vitamin E. Conditional exchangeability — when exchangeability holds in each stratum of a confounder — is sufficient to remove residual confounding and to compute consistent causal estimates if the confounder (or set of confounders) is controlled for.

Directed acyclic graphs (DAGs) provide a formal yet intuitive representation of causal inference[24]. A directed arrow between two variables indicates a causal relationship, i.e. the (counterfactual) values of the variable at the origin cause corresponding (potential) outcomes in the variable at the destination. It can be useful to conceive of the causal effect as the result of an intervention on the variable at the origin, holding all other variables constant. As depicted in BOX 2, figure part **a**, 'blocking' all confounders of an association between a risk factor and an outcome can ensure conditional exchangeability. Genetically informed designs can approximate conditional exchangeability in two main ways. First, following the instrumental variable approach, genetic factors predicting an exposure can be used to estimate the effect of the exposure on an outcome (BOX 2, figure part **b**). Second, designs such as the twin design can be used to control for genetic confounding and, to some extent, environmental confounding (BOX 2 figure part **c**). BOX 2, figure part **d** combines these two approaches and constitutes a general representation of causal inference using genetically informed designs. The designs we present in the following sections can be understood by referring to this general representation. Importantly, genetically informed designs for causal inference do not focus on genetic information as an end objective. Rather, they exploit genetic information as a means to attain reasonable substitutes to the counterfactual situation in order to estimate consistent causal effects.

Family-based designs

Family-based designs have been exploited to strengthen causal inference in observational research for decades and can tackle a wide range of causal questions, from the role of smoking during pregnancy on birth weight[25] to the impact of income on depression[26]. Family-based designs rely on a priori knowledge of genetic relatedness — or absence thereof — between family members (e.g. identical twins versus adopted siblings). As such, genotyping is not necessarily required. Family-based designs for causal inference have in common their ability to control for (some) genetic confounding. They differ with regard to: the extent to which they control for genetic confounding; their ability to control for non-genetic confounding; and their applicability.

Sibling and twin designs. These designs approximate the counterfactual situation because a non-exposed sibling or twin represents a natural match to their exposed co-sibling or twin[10,27]. Siblings and dizygotic (DZ)

twins share 50% of their segregated genetic material on average. Monozygotic twins (MZ) share 100% of their genetic material (with exceptions[28]). By definition, shared environmental factors are all environmental factors that contribute to the similarity of family members, and are 100% in common between the two members of a sibling, DZ, or MZ pair. In the case of a binary exposure, some sibling and twin designs for causal inference therefore compare outcomes in exposed versus non-exposed pair members. Genetic confounding is entirely controlled for only in MZ twins (i.e. blocking all backdoor paths through G in BOX 2, figure part **c**) yielding more accurate causal estimates than siblings or DZ twins. Sibling and twin designs also control for confounding by shared environment. For example, as parental age at birth does not differ between members of a twin pair, confounding effects of parental age are removed. A powerful feature of these designs is that they account for unobserved confounding by unmeasured genetic variation or shared environment. Effect estimation in these models, also called family fixed effects models, is straightforward for discordant designs (binary exposure) and differences designs (continuous exposure)[10,29–31]. Other estimation methods can be used such as structural equation modelling [32].

Sibling and twin designs have been applied to a variety of causal questions in many disciplines, for example: confirming that smoking causes lung cancer[33] but also lowers long-term earnings[34]; and suggesting that higher income and access to green spaces are protective factors for depression[26,35]. Longitudinal extensions of these designs constitute powerful tools to study the duration of effects and reciprocal relationships. For example, evidence from a twin differences design shows that the consequences of exposure to bullying in childhood might be shorter term than suggested by classical longitudinal studies[36]. Using a longitudinal twin differences design, attention-deficit hyperactivity disorder (ADHD) symptoms have been shown to be more predictive of future autistic spectrum disorder (ASD) symptoms than the reverse, i.e., ASD symptoms predicting future ADHD symptoms[37].

Although they control more stringently for confounders than non-genetically informed designs, sibling and twin designs are limited in that they cannot account by design for non-shared environmental confounding. For example, MZ twins discordant for smoking could differ in other lifestyle choices such as alcohol consumption, which may confound the association between smoking and outcomes. Controlling for relevant observed non-shared environmental confounders, such as alcohol consumption, can mitigate this issue (i.e. controlling for the non-shared component of O but not U in BOX 2, figure part **c**; see also Ref[38]). Furthermore, measurement error can be a problem in twin and sibling causal inference designs as different degrees of measurement error between the causal and caused variables can bias inference. This can be addressed by directly modelling measurement error when it can be estimated or by conducting a sensitivity analysis to determine how much difference in measurement error is needed to change the conclusion[38–40]. Another important limitation concerns exposures that do not vary between pair members. For example, twins are perfectly matched for parental age or family income, and such exposures that do not vary within the family cannot be used as predictors in discordant or differences designs. By contrast, parental age can differ between siblings, and the sibling design has been used to demonstrate that paternal age at birth is likely to have widespread effects on offspring psychiatric and academic outcomes that often remain undetected in classical observational studies[41].

Adoption-at-birth and in-vitro fertilization design. These designs compare associations between risk factors and outcomes in genetically related and unrelated parent–child pairs. Adopted children are genetically unrelated to their adoptive parents. *In-vitro* fertilization (IVF) can use either parental gametes (genetically related) or donor gametes (genetically unrelated) for fertilization. Associations in genetically unrelated pairs are free from genetic confounding due to passive gene–environment correlation (BOX 1). These designs are appropriate for examining intergenerational effects. For example, smoking during pregnancy associates with lower birth weight. However, maternal genetic factors contribute to smoking during pregnancy; when transmitted to the offspring, the same genetic factors may influence birth weight, thereby generating an association even in the absence of an effect of smoking. An IVF study demonstrated that smoking during pregnancy was predictive of lower birth weight in both genetically related and unrelated mother–child dyads, ruling out genetic confounding[25]. Similarly, the adoption-at-birth design has been used to investigate the role

of parental psychiatric morbidity in child developmental outcomes[42]. The key limitation of these designs is that, unlike MZ twins, they do not control for environmental confounding. Therefore, it becomes necessary to adjust for observed confounders, with the limitations inherent in that approach.

Direction of causation model. The classical twin design aims to decompose the variance of a phenotype into heritability (additive (A) and dominance (D) effects), and environmental influences (subdivided into shared (C) and non-shared (E) effects). The insight behind the direction of causation (DoC) model is to use these A(D)CE components as instruments to investigate causal relationships (similar to BOX 2, figure part **b**). Interestingly, using A(D)CE components of each phenotype as instruments for the other phenotype enables the investigation of reciprocal causal relationships in cross-sectional designs (similar to bidirectional MR, see below)[43,44]. The DoC model has been implemented for example to investigate the genetic overlap between cognitive functions and schizophrenia. Findings showed that around a quarter of the variance in liability to schizophrenia was explained by variation in cognitive function[45]. However, the scope of application of the DoC model has been limited, as a condition required for its implementation is that the variance components should not be equal for both phenotypes. This condition can be satisfied for example when unequal proportions of variance are explained by A, C and E for each phenotype or when ADE components explain one phenotype and ACE components the other. More-similar components lead to decreasing statistical power.

Mendelian randomization

Over the past decade, MR has become a method of choice to strengthen causal inference in observational research. MR is used to investigate an ever-growing set of causal questions, from the role of molecular biomarkers in CHD to behavioural questions such as possible reciprocal effects between cannabis use and schizophrenia. In contrast with family-based designs described in the previous section, MR exploits genotyping data, most often in unrelated individuals. MR is founded on the realization that a genetic variant associated with an exposure X can be used as an instrumental variable to estimate the causal effect of X on an outcome of interest (BOX 2, figure part **b**)[11,12,46]. Genetic instruments — typically single nucleotide polymorphisms (SNPs), although other sequence variants could be used — can approximate the counterfactual situation. Individuals carrying the risk allele have higher (or lower) levels of X on average than individuals with no risk allele. According to Mendel's laws of segregation and independent assortment, we can assume that the resulting exposed and non-exposed groups satisfy the condition of exchangeability[47,48]. When certain assumptions are satisfied (see below), a difference in the outcome between individuals with and without the risk allele can only be attributed to the causal influence of X. To a certain extent, MR can thus be construed as a natural experiment analogue to RCTs in which participants are allocated to different exposure levels independently of confounding[12,49] (hence the term Mendelian randomization, as 'genetic allocation', similar to randomized allocation, generates variation in the exposure that, under assumptions, should be unaffected by confounding).

A classic example relies on variants in the *CRP* gene to assess the health consequences of elevated circulating C-reactive protein (CRP), a marker of systemic inflammation[50,51]. In an early study, the SNP rs1059 was used as an instrument to investigate whether elevated CRP levels influence blood pressure. The concentration of CRP was 1.81 mg/L (log) in carriers of the GG genotype and 1.39 in non-carriers ($p<0.001$)[51]. Strikingly, although circulating CRP levels were strongly associated with many measured confounders such as low-density-lipoprotein cholesterol (LDL-C) and socioeconomic status, the genetic instrument was independent of all measured confounders. This suggests exchangeability between GG carriers and non-carriers and illustrates the benefits of using genetic instruments rather than observed CRP levels for causal inference. Comparing outcomes between GG carriers and non-carriers suggested no causal relationship, with systolic blood pressure of 147 mmHg in both groups ($p=0.98$). Subsequently, MR analyses have demonstrated that: CRP is likely to be a simple marker rather than a causal risk factor for many phenotypes, including CHD, lung function, and depression, although unexpected suggestive evidence of a protective effect on schizophrenia has recently been reported[52–55]; and similar to vitamin E, vitamin D levels appear unlikely to be causally related to CHD[56] but appear to be causal for multiple sclerosis[57].

To derive reliable causal estimates from MR, genetic instruments must satisfy instrumental variable assumptions. The core assumptions are not fully testable and constitute a serious threat to inference validity (BOX 3). Genetic instruments extracted from a single gene with a well-understood biological function, such as *CRP*, are more likely to meet these assumptions, enabling reliable causal inference and providing targets for pharmacological interventions[13]. However, such monogenic instruments are unavailable for many exposures, leaving only imperfect instruments. Highly polygenic influences on most phenotypes imply small individual SNP effects, which creates potential problems with weak instruments unless large samples are used[58]. Polygenicity also implies that pleiotropy is widespread[59], potentially (but not necessarily) resulting in invalid instruments (detail in BOX 3). Fortunately, polygenicity also provides an antidote, in the form of multiple instruments for any given exposure. In recent years, considerable efforts have been devoted to extensions of MR allowing for multiple imperfect instruments, which we consider in the following section[60–62].

Extensions of Mendelian randomization

[H2] Dealing with imperfect instruments. Modelling multiple imperfect instruments together can substantially increase power[48,63] and mitigate problems due to weak instruments[58,64,65]. Figure 1a illustrates the use of multiple instruments derived from relevant genome-wide association studies (GWAS) to assess the effect of LDL-C on CHD. Estimates of the association between genetic instruments and CHD (β_{ZY}) are regressed on estimates of the association between instruments and LDL-C (β_{ZX}, see BOX 3). We expect that, if LDL-C→CHD is causal, then instruments with larger effects on LDL-C should have proportionally larger effects on CHD. The slope of this regression estimates the causal effect; a flat line implies no causation. The effect estimated from multiple instruments is more precise than the effect based on a single SNP (FIG. 1a).

As illustrated in Figure 1a, most but not all SNPs are aligned with the regression line, resulting in heterogeneity in causal estimates. In the context of MR, heterogeneity occurs when estimates derived from each genetic variant do not all converge to the same causal estimate. Heterogeneity, which can be assessed via graphical inspection and statistical tests, can result in misleading causal conclusions. Heterogeneity may stem in part from pleiotropy (BOX 3): in addition to its effect on CHD through LDL-C, a SNP may have an effect through other pathways, explaining a greater or lower than expected association with CHD. Several methods jointly modelling multiple instruments have been proposed to allow for such invalid instruments (see Table 1; these methods cannot be implemented for instruments using a single genetic variant, for which the validity of the instrument has to be assumed). For example, MR-Egger regression quantifies pleiotropy by estimating an intercept in addition to the slope in the regression shown in Figure 1a, and can yield consistent causal estimates even when all individual instruments are invalid[60]. Compared to the inverse-variance weighted method, which does not account for unbalanced pleiotropy (see Table 1), the MR-Egger regression estimate is reduced for LDL-C and more so for HDL-C (FIG. 1a,b).

Bidirectional MR. This approach investigates possible reciprocal causal relationships between two phenotypes. For example, cannabis use has been implicated in the aetiology of schizophrenia but reverse causation is possible[66]. Bidirectional MR uses genetic instruments for cannabis use to investigate the cannabis→schizophrenia relationship and genetic instruments for schizophrenia to investigate schizophrenia→cannabis (FIG. 1c,d). A first attempt to investigate this question demonstrated that bidirectional causal influences are plausible[66,67]. Importantly, reverse causation between an exposure and an outcome violates an assumption of MR that is explicit in the directed effect of X on Y (BOX 3, figure part a). Therefore, results from bidirectional MR can currently only be regarded as suggestive (see the legend of Figure 1).

Multivariable MR. This method considers several exposures simultaneously and thus allows direct modelling of possible pleiotropic pathways that would violate MR assumptions[68,69]. For example, SNPs associated with either HDL-C, LDL-C or triglycerides are often associated with the other two. Therefore, genetic instruments for HDL-C may affect CHD through pathways other than HDL-C levels, violating a key assumption of MR. Recently, multivariable MR and MR-Egger regression have been combined to further test for pleiotropy[62]. Using multivariable MR-Egger regression, we updated previous findings[65] based on the most

recent GWAS for lipids and CHD[14,70]. Findings confirm the robustness of the effects of LDL-C; however, the ostensibly protective role of HDL-C reported in univariate analyses is not confirmed when using multivariable MR-Egger regression (FIG. 1a,b and legend).

[H2] Intergenerational MR. Similar to the IVF design, intergenerational MR capitalizes on information regarding the mother–child genetic relatedness to account for passive gene–environment correlation. In contrast to the IVF design, intergenerational MR exploits measured genotypes and accounts for environmental confounding. Intergenerational MR was implemented to demonstrate that higher maternal body mass index (BMI) and higher levels of fasting glucose predict larger birth weight in offspring[71]. Importantly, simply deriving instruments from maternal genotypes cannot rule out passive gene–environment correlation. Indeed, the association between maternal genetic instruments and offspring outcomes may arise from the transmission of risk alleles rather than from the causal influence of maternal BMI. Controlling for the genetic instrument in the offspring is a first step to address this issue. However, it creates a collider bias (BOX 2) with the paternal genotype, reintroducing confounding[72]. Two approaches have been proposed to deal with this problem: controlling for paternal genotypes (requiring genotypes on mother, father, and child)[72]; and second, splitting the genetic instrument for the mother into two instruments comprising the non-transmitted and transmitted alleles[73]. The non-transmitted alleles enable causal estimation whereas the transmitted alleles reflect genetic transmission. Notably, splitting the genetic instrument substantially limits power, because power in MR largely depends on how much variance in the predictor is explained by the genetic instrument[63].

[H1] Emerging approaches

MR studies described in the two previous sections typically involve: first, selecting a set of SNPs as instruments; second, using these instruments to investigate one (or a few) risk factor(s) and one outcome; and third, testing whether they are causally related. Here, we describe emerging approaches that go beyond these three features, in particular by exploiting genome-wide and phenome-wide information to delineate complex pathways between multiple phenotypes.

[H2] Polygenic scores. Genetic instruments derived from allelic scores typically use a limited number of SNPs, from a few to a few hundred, thereby leaving out most causal SNPs in the genome and potentially limiting power. The justification for such severe ascertainment is that polygenic scores with many more SNPs are more likely to violate instrumental variable assumptions (see BOX 3). First, polygenic instruments are more likely to correlate with confounders. For example, one study showed that both allelic scores made of known variants and truly polygenic scores using hundreds of thousands of SNPs for BMI, LDL-C and CRP predicted diseases as expected[74]. However, the polygenic scores were less specific, associating with more traits and thus more potential confounders, thereby constituting questionable instruments for their respective exposure (see also Ref[75]). Dynastic effects (BOX 3) are a special case, where a genetic instrument acts as a proxy not only for the exposure (e.g. child BMI) but also an environmental effect (e.g. the obesogenic environment created by parents). Second, polygenic instruments are likely to include many variants with problematic pleiotropic effects, i.e. influencing the outcome not exclusively via the exposure (BOX 3).

A possible strategy for circumventing these issues is to integrate polygenic scores, used as instruments, with family-based designs, either family fixed effects[31,76] or DoC models[77]. In family fixed effects models, differences in the outcome between siblings (or DZ twins) can be explained using differences in sibling's polygenic scores as an instrument[31]. Given the properties of family fixed effects designs (see above) such an instrument is independent of all confounders shared between siblings. Dynastic effects are also controlled for because environmental conditions created by the parents are shared between siblings. Notably, dynastic effects are not controlled for in MR on unrelated individuals, highlighting the benefit of embedding genetic instruments within family-based designs. More generally, MR is only absolute in within-family designs, as the genetic material is randomized in transmission from parent to child, whereas the ‘randomization’ is approximate at a population level[78,79].

A DoC model integrating a polygenic score as an instrument can be identified in several ways, entailing trade-offs between different assumptions[77]. One possibility is to assume the absence of non-shared environmental confounding, similar to the twin differences design (see above). This enables direct estimation of pleiotropic effects, by modelling both directed paths from the instrument to the exposure and from the instrument to the outcome. Releasing the no-pleiotropy assumption in this way provides a promising method to account for the pleiotropy introduced by the use of polygenic scores. Theoretical and empirical research regarding these models is still limited but can develop in diverse ways. For example, an increased number of parameters could be identified by including a wider range of relationships from extended pedigrees or distantly related individuals, potentially reducing the assumptions required[77].

[H2] Phenome-wide approaches and shared aetiology. The wide availability of summary association statistics [80] enables phenome-wide approaches to investigate relationships between thousands of phenotypes. Genetic correlations quantify the magnitude of the shared genetic aetiology between phenotypes[16] (FIG. 2a). Often, the existence of shared genetic aetiology is more relevant than strict causality, such as when an intervention on the exposure cannot be achieved, as for adult height[74] or age at menarche[81]. In these cases, strict adherence to the MR assumption of no pleiotropy is less important than the demonstration that the phenotypes have a common aetiology. Further investigating what gives rise to this shared aetiology can also offer new avenues for interventions if we can identify underlying common causal pathways (illustrated by P in BOX 3, figure part **b**). Genetic correlation estimates are limited in that regard as they do not identify where in the genome shared loci reside, nor do they elucidate mechanisms underlying cross-phenotype relationships[59]. Phenome-wide association studies (PheWAS) or multi-trait GWAS approaches can help in identifying shared loci[82–84], i.e. genetic variants influencing two or more phenotypes. For example, a nonsynonymous variant in the zinc transporter *SLC39A8* associates with schizophrenia, Parkinson disease and height[85]. Identifying such shared loci can be achieved via colocalization methods **[G]**. Two phenotypes colocalize in a genetic region when it contains variants that associate with both phenotypes. This reflects three possible scenarios: first is causality, in which the SNP effect on one phenotype is mediated by its effect on the second phenotype; second is pleiotropy, in which the same SNP independently affects both phenotypes; and third is linkage disequilibrium (LD), in which two or more SNPs in LD affect different phenotypes[86] (BOX 3). Colocalization tests and related methods can provide evidence in favour of the first two scenarios over the third scenario, thereby indicating that at least one causal variant in the genetic region influences the two traits, pointing towards a common causal pathway, which may constitute a target for intervention[85–89].

The mechanisms underlying cross-phenotype relationships can be further elucidated by attempting to distinguish between the first and second scenarios. One approach is to test for asymmetry between two phenotypes, asymmetry being defined as the situation where the SNPs most strongly associated with one phenotype predict the other phenotype, but the reverse is not true[85]. For example, the top SNPs for LDL-C predict CHD, but those for CHD do not predict LDL-C. Such asymmetry is interpreted as more consistent with causal relationships between the two phenotypes (the first scenario) rather than the cross-phenotype association being generated by shared pathways (the second scenario). A study of 42 phenotypes identified five pairs of putative causally related phenotypes, including evidence for higher BMI leading to type 2 diabetes but not the reverse[85]. Notably, this asymmetry analysis would be underpowered to detect cases of true reciprocal causal relationships of similar magnitude. Furthermore, spurious asymmetry patterns can arise in principle through particular algebraic relationships between SNP effects, causal effects and effects of unmeasured confounders.

Such methods can also be applied to probe relationships between phenotypes and biomarkers, such as gene expression. Transcriptome-wide association studies (TWAS) using measured gene expression are susceptible to the same biases as observational studies. Conversely, using summary statistics from expression quantitative trait locus (eQTL) studies and from GWAS enables the detection of genetic variants that affect both expression levels and endpoint phenotypes (the second scenario). Such analyses can help to identify functionally relevant genes, for example by pinpointing TNF receptor associated factor 1 (*TRAF1*) rather than complement C5 (*C5*) as the most functionally relevant gene in the *TRAF1–C5* locus for rheumatoid arthritis. Furthermore, when several eQTLs are detected, asymmetry analysis can be implemented to estimate the

causal effect of gene expression on endpoint phenotypes[59]. The same remarks apply to epigenome-wide association studies (EWAS) to assess the aetiological role of DNA methylation levels.

To summarize, genetically informed phenome-wide approaches can help in: better understanding the shared aetiology between phenotypes; prioritizing putative causal relationships; and refining functional knowledge.

[H2] Dissecting exposures and delineating pathways. Causal questions typically lead to testing whether one exposure causes one outcome. However, exposures are often heterogeneous, conflating distinct sub-components. Such heterogeneous exposures can lead to heterogeneity in causal estimates even in the absence of pleiotropy. BMI is an example of a heterogeneous exposure that could be refined into appetite, adipogenesis, and cardiopulmonary fitness subcomponents. When available, genetic instruments indexing these different subcomponents may provide more specific causal effects and intervention targets[90]. Furthermore, a complex network of pathways often relates multiple exposures to multiple outcomes. Mapping out pathways from exposures to outcomes may provide additional targets for intervention, for example by determining the mediating role of inflammation markers or hormones. As a result, instead of examining a single arrow from one exposure to one outcome, causal analyses may start to resemble causal maps unravelling networks of relationships between phenotypes, as illustrated in Figure 2b[91]. Figure 2b also illustrates the concept of a network of causal relationships, e.g. smoking is protective for type 2 diabetes via its effect of lowering BMI (see Refs[92,93]). To establish such causal maps, different methods outlined in this Review can be implemented, such as network MR that exploits a different genetic instrument to probe each arrow in the network[94], or longitudinal twin designs with relevant phenotypes and biomarkers.

[H1] Conclusions and future perspectives

In this Review, we have described the common logic underlying the use of genetically informed methods to strengthen causal inference, based on the counterfactual approach. We have shown that such genetically informed methods already form a worthy toolbox for causal inference. Researchers can select appropriate tools depending on the characteristics of their research question and data: if exposure varies within families, twin and sibling designs can be considered; if we can find monogenic or polygenic instruments to adequately proxy the exposure, MR and extensions are available; if reverse causation is suspected, the DoC models and bidirectional MR can be explored; if prior knowledge exists regarding possible pleiotropic pathways, multivariable MR is recommended; to investigate the intergenerational transmission of risk, adoption, IVF and intergenerational MR can be applied; if the aim is to identify causal pathways shared between multiple phenotypes, colocalization methods are appropriate. These methods can be further integrated to develop this toolbox and offer new avenues for research. In particular, emerging approaches embedding polygenic instruments within family-based designs can address certain limitations of both approaches. In addition, integrating MR, colocalization methods, and phenome-wide approaches can allow researchers to identify putative causal relationships and shared causal pathways that are relevant to many phenotypes. In future, methodological advances are likely to enrich this toolbox, and applications across disciplines should expand accordingly.

[H2] Methodological advances. We expect a continued burgeoning of method developments for genetically informed causal inference designs. Progress in the near future should lead to yet more robust MR estimators. Methods to refine genetic instruments by leveraging functional knowledge should yield more insightful inferences (e.g. dissecting the effects of heterogeneous exposures). We have outlined how integrating genetic instruments with family-based designs can mitigate problems of MR, such as dynastic effects. This is reminiscent of family-based genetic association tests developed to control for population stratification. Adapting those approaches to the MR paradigm could prove fruitful, for example by conditioning on parental genotypes or by treating family effects as random[95,96]. A further promising area is the use of genome-wide information. As shown in BOX 3, figure part d, fully capturing genetic factors confounding an association would enable better causal inference. However, although polygenic scores are increasingly powerful, they currently explain only a small amount of the variability in phenotypes[97]. We propose that, similar to multivariable adjustment in non-genetic epidemiology and other disciplines, polygenic scores can still be

used in a sensitivity analysis to assess the likelihood that a relationship of interest results entirely from genetic confounding (BOX 4).

[H2] Rapid expansion of applications across disciplines. New, more powerful GWAS, multi-trait GWAS, PheWAS, TWAS and EWAS will considerably increase the scope of applications and the reliability of the methods described in this Review. Inexpensive microarrays also enable genotyping on specific samples such as twins (e.g. TwinsUK[98], Twins Early Development Study (TEDS)[99]), or birth cohort studies (e.g. the Norwegian Mother and Child Cohort Study (MoBa)[100] and the Avon Longitudinal Study of Parents and Children (ALSPAC)[101]); data from such samples can be combined, encouraging a wider application of the aforementioned methods combining family-based designs with genome-wide data. Applications of genetically informed methods for causal inference in medicine already provide evidence of palpable benefits, with pharmaceutical companies implementing these methods for: first, validating (or invalidating) existing drug targets (e.g. discarding CRP or HDL-C for CHD prevention); second, identifying possible off-target effects; third, repurposing existing drugs; and fourth, discovering new targets[13,102,103]. Disciplines that traditionally have largely ignored the role of genetics can no longer justify doing so, such as social sciences and economics[104]. Genetically informed causal inference methods should become routine wherever possible, at the very least to consider the possibility of genetic confounding.

[H2] Pitfalls of causal inference. Conclusions drawn from causal inference methods are only as good as the modelling decisions made and to the extent that assumptions are credible[105]. Assessing credibility requires in-depth knowledge of the question, which, for example, is unlikely in massive hypothesis-free causal inference exercises, such as phenome-wide approaches[13]. The causal map in Figure 2b shows examples of implausible cases resulting from hypothesis-free approaches. Furthermore, each method makes a different set of assumptions, which cannot always be appropriately evaluated. Therefore, triangulation — when conclusions from several study designs converge — will play an increasingly important role in strengthening evidence for causality[106,107 98]. Overall, one should not expect that a single existing or future method for causal inference in observational settings will provide a definitive answer to a causal question. Rather, such methods can substantially improve the strength of evidence on a continuum from mere association to established causality.

In summary, causal inference using genetically informed designs has a long history but has undergone rapid and exciting developments in recent years, with research already reaping valuable benefits. A rich and growing toolbox of genetically informed methods to strengthen causal inference is becoming available, with applications across the biomedical and behavioural sciences and in new areas including social sciences and behavioural economics.

Box 1 | **Challenges to causal inference**

Even the poster child of causal relationships — smoking cigarettes causes lung cancer — was once controversial. In 1957, Ronald Fisher, a founding father of modern statistics and statistical genetics, and himself a smoker, qualified smoking as "possibly an entirely imaginary cause" for lung cancer[108]. He argued that the observed association was due to genetic confounding, in his words "a common cause, in this case the individual genotype". Most putative causal relationships are much harder to establish than this one and confounding is the major challenge for causal inference. Confounding occurs when a third variable causes both the risk factor and the outcome (Fisher's 'common cause'), generating a spurious association. Genetic confounding occurs when 'individual genotype' is the third variable, in other words, when genetic factors affecting the environmental exposure also directly affect the outcome (e.g., genetic factors affecting both cigarette smoking and lung cancer in the figure)[7]. Pleiotropy, a concept related to genetic confounding, is detailed in BOX 3.

Gene–environment correlation[109], can generate genetic confounding. Active or evocative gene–environment correlations occur when the environment experienced by an individual is partly influenced by their genotypes. Such gene–environment correlation explains why even 'environmental variables' such as educational attainment (see the figure) or bullying victimization are partly heritable[110–112]. Similarly, genetic variants in the CHRNA5–A3–B4 nicotinic receptor subunit gene cluster reliably predict smoking heaviness in smokers[113,114]. An exposure such as smoking can thus be genetically influenced. Importantly, gene–environment correlations do not always generate confounding. This is because genetic variants may be associated with the exposure (here, smoking) but only indirectly associated with outcomes through that exposure. In such cases, these genetic proxies for exposures can be used to probe the causal role of these exposures on diverse outcomes (see the Mendelian randomization section in the main text). Passive gene–environment correlation occurs when children inherit parental genetic variants that contribute to the environment that parents create[109]. For example, smoking during pregnancy is genetically influenced and the offspring can receive both the genetic variants associated with smoking and the smoking environment. Such passive gene–environment correlation can confound observed associations between smoking during pregnancy and offspring's outcomes (see the figure).

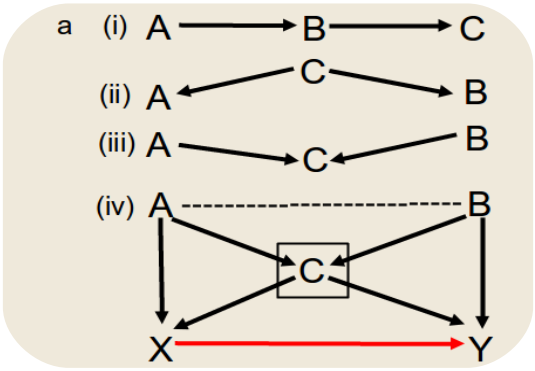
Reverse causation constitutes another major issue. Even if causal relationships are established between risk factors and outcomes, the direction may remain unclear. Reverse causation is relevant to many causal questions, e.g., does alcohol abuse cause depression or does depression lead to alcohol abuse (see the figure). No reverse causation exists between germline genetic variants and phenotypes. For example, alcohol abuse may cause a disease or alcohol abuse may increase in response to disease onset. But germline genetic variants associated with alcohol abuse will not be modified by disease onset[47], which is advantageous when using genetic variants for causal inference (BOX 3).

Measurement error in the exposure or the outcome can hinder the detection of causal effects, e.g., in the figure, imprecise measures of alcohol abuse may prevent the detection of its effect on depression. Conversely, even slight measurement error in confounders can lead to biased estimates as confounders are not appropriately controlled for[115]. Genetic proxies of exposures can be less susceptible to measurement error and reporting bias (e.g. variants in the nicotinic receptor gene cluster predict objective measures of tobacco exposure better than they predict self-reported smoking[113]).

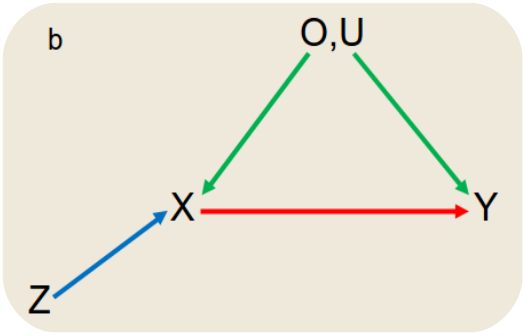
Misidentification occurs when the putative causal risk factor is only correlated with the true causal risk factor, e.g., in the figure, the tobacco inside the cannabis joint causes cancer, rather than the cannabis per se[116]. Misidentification may also happen when a genetic proxy for a given exposure is not entirely relevant to that exposure (BOX 3), yielding causal estimates that do not accurately reflect the effect of the exposure under scrutiny.

Box 2 | **Directed acyclic graphs for causal inference exploiting genetics**

Directed acyclic graphs (DAGs) provide a useful graphical language for causal inference in general and for genetically informed causal inference methods in particular. DAGs displayed here provide a conceptual framework to understand methods presented in this Review: part **a** of the figure illustrates key DAG concepts; parts **b** and **c** represent two main ways in which genetically informed designs can strengthen causal inference; and part **d** shows how these two approaches can be merged into a single representation.

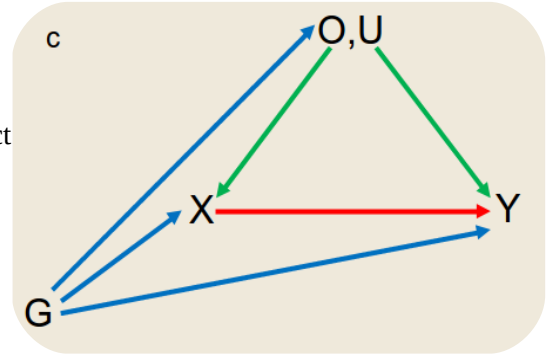


The figure, part **a**, illustrates four alternative DAG structures. Solid arrows represent a directed causal path. No arrow means no directed path. In (i), A→B→C is a directed path: all arrows point forwards. In (ii), A←C→B is a backdoor or unblocked path: A is associated with B through C. Such an unblocked path creates an observed association between A and B, even in the absence of a causal path (no A→B). C is thus a confounder, generating a correlation between A and B, despite neither causing the other. In 'potential outcomes' terminology, exposed and non-exposed participants on factor A are not exchangeable. Formally, exchangeability requires that potential outcomes are independent of the observed exposure[22]. For example, we assume that, in a randomized controlled trial (RCT), observed levels of B in the treatment group (A_1) would have been the same in the control group (A_0) had control participants been exposed to the treatment. Here, C prevents exchangeability, leading to a biased estimate of A→B. In (iii), C is a collider: arrows 'collide' at C. The path A→C←B is blocked: A is not associated with B through C. Controlling for a collider (C) creates a spurious correlation between A and B. In (iv), the exposure–outcome path X→Y (red) is confounded. Controlling for C (the square around C) blocks the backdoor path X←C→Y. However, controlling for the collider C unblocks the path X←A--B→Y, which confounds X→Y. Controlling for C alone is therefore not sufficient, but controlling in addition for either A or B solves this problem, by blocking this newly created path (see Ref[117]). To ensure conditional exchangeability of exposed and non-exposed individuals, all backdoor paths between X and Y should be blocked. When this is achieved (here by controlling for C and A or B), then X is 'd-separated ' from Y, which provides an unconfounded causal estimate of X→Y.

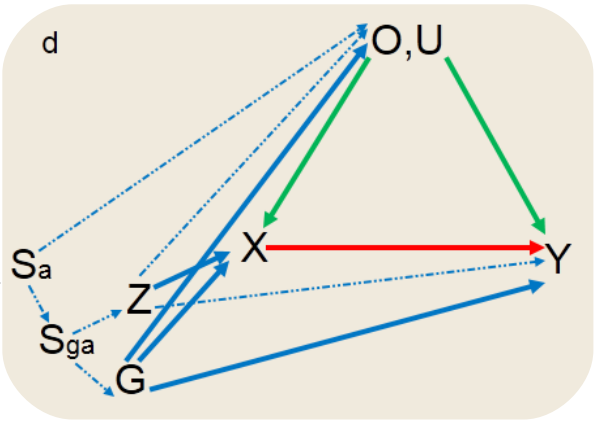


The figure, part **b** illustrates an instrumental variable analysis, using an instrument Z to estimate X→Y. To conclude that X is a causal risk factor for Y, three assumptions must be satisfied: relevance , exchangeability, and exclusion restriction **[G]**. Relevance implies that the chosen instrument Z reliably predicts the risk factor of interest (solid Z→X arrow). Second, the instrument must be independent of all observed (O) and unobserved (U) confounders to ensure exchangeability between exposed and non-exposed individuals (no Z→O,U). Third, exclusion restriction means that, conditional on exposure and confounders, the instrument is independent of the outcome. More intuitively, exclusion restriction signifies that the genetic instrument must affect the outcome exclusively through its effect on X (i.e. solid path Z→X→Y but no other path from Z to Y)[118].

In the figure, part **c**, G represents a latent variable capturing all genetic influences on X, Y and O,U. Note that the previous conditions for an instrumental approach are not satisfied: G directly influences both Y and O,U. To estimate the causal effect of X on Y — i.e. to d-separate X from Y — it is necessary to adjust for G and O,U. Naturally, d-separation is challenging because all relevant genetic variants and environmental confounders must have been identified and measured without error, which highlights the difficulty of causal inference in observational research.

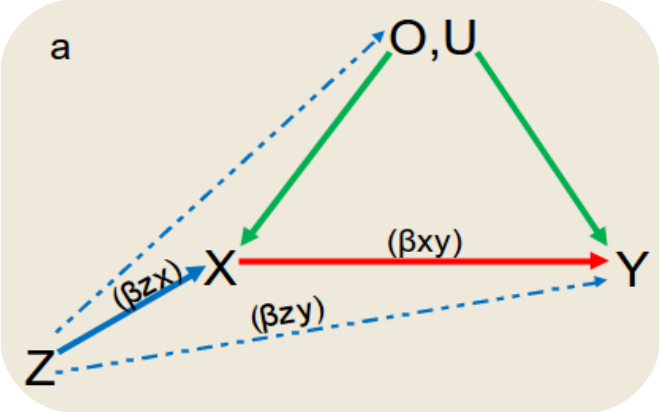


The figure, part **d** provides a general representation of approaches for causal inference using genetic data, combining the instrumental and the direct control for confounders. Note that the dashed lines represent violations of instrumental variable assumptions: instrument Z is related to confounders; Z is directly related to Y; Z is associated with Y through its association with other genetic factors (G), due to shared genetic ancestry (S_{ga})[119]; Z is associated with Y via S_a→O,U (e.g. shared cultural ancestry affecting social factors). Z–G interactions and Z–O,U interactions, not represented here, can also generate assumption violations.

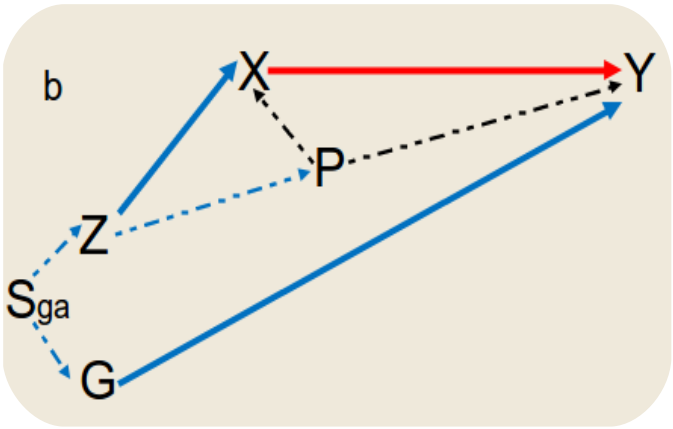


Box 3 | **Mendelian randomization: principles and assumptions**

A genetic instrument Z must satisfy the assumptions of relevance, exchangeability, and exclusion restriction (BOX 2). Relevance can be tested by using genetic association studies for X, which provide an effect size estimate (βzx) and a test of significance[120] (see the figure, part **a**). Exchangeability can be tested by verifying whether Z predicts observed confounders (dashed Z→O representing associations that should not be present). For example, a genetic instrument for C-reactive protein (CRP) was found to be independent of 21 potential observed confounders of the association between CRP and coronary heart disease (CHD)[120]. Generally, genetic instruments are likely to provide more reliable causal estimates of X→Y (red) than direct estimates using observed X and Y, even when controlling for observed risk factors. One study[121] demonstrated that 96 behavioural, socioeconomic and physiological characteristics were strongly interrelated. By contrast, genetic variants showed no more associations with these potential confounders than expected by chance. Despite these encouraging findings, exchangeability cannot be proven because some relevant confounders may be unobserved. To fulfil the exclusion restriction assumption, there should be no direct causal path from Z to Y. Note that the observed βzy is not null. However, the exclusion restriction assumption implies that the observed βzy results only from the indirect effect through X, i.e. βzy = βzx*βxy. Based on observed βzx and βzy, we can therefore estimate the causal estimate (βxy), using the ratio βxy = βzy/βzx. A relevant instrument can be weak (significant but small βzx). A weak instrument leads to a small denominator (βzx), which results in imprecise estimates, and biases the estimated causal effect towards the observational association when βzx and βzy are estimated in the same sample, or towards the null when they are estimated in independent samples[122]. Notably, if the three aforementioned assumptions are satisfied, we can conclude that X causes Y, but additional parametric assumptions (e.g. linearity) are required for the ratio to be reliable[120].



The notion of pleiotropy[123] — when a genetic locus affects more than one phenotype — is key when assessing exclusion restriction. Mediated pleiotropy[124] (also called vertical pleiotropy[125], or causality[86]) occurs when Z and Y associate because Z affects Y through X. This fulfils the exclusion restriction assumption and is consistent with causality. Unmediated or biological pleiotropy[124] (horizontal pleiotropy[125], or simply pleiotropy[86]) is when Z affects both X and Y but through different pathways. Such pleiotropy can be: direct, as in the path from Z to Y in the figure, part **a**, or indirect either via O,U or via intermediate pathways P in the figure, part **b**. This type of pleiotropy is informative about shared aetiology (X and Y are both caused by P) but the instrument will yield biased βxy. Finally, spurious pleiotropy[124] is when two (rather than one) causal variants explain Z and Y but the variants are in linkage disequilibrium — i.e. associated because of shared ancestry (S_{ga} in the figure, part **b**)[119]. Exclusion restriction is violated as the observed βzy not only reflects Z→X→Y but also the association via other variant(s) in G. Note that Z need not be a causal variant for X. Z can be tagging a causal variant affecting X if both tagging and causal variants fulfil exchangeability and exclusion restriction assumptions[118]. Finally, dynastic effects also violate the exclusion restriction assumption[76]. Dynastic effects occur when parental genotypes affect the child via the environment that

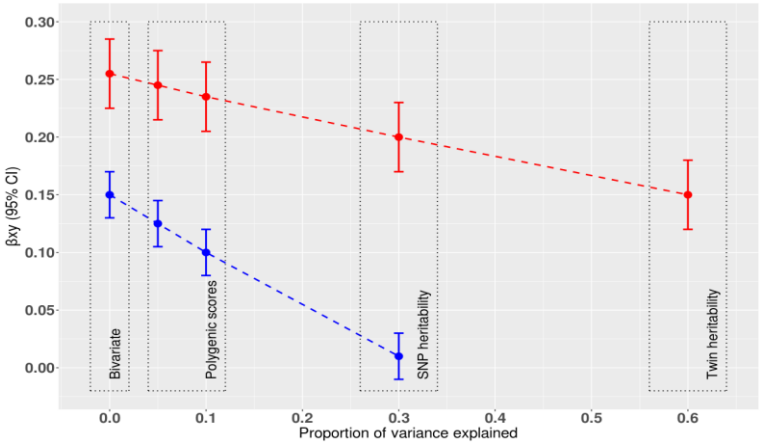


parents create for their child by affecting the parental phenotype accordingly (also called ‘genetic nurture’ as genotypes affect the nurturing environment[126]). As a result, the genetic instrument in the child is correlated with the environment created by the parents. Dynastic effects therefore open a backdoor path between instruments and outcomes via parental environments.

Genetic instruments have additional advantages (e.g. reducing reverse causation, reporting bias, and measurement error) and limitations (e.g. limited power, population stratification, and developmental compensation) that are summarized elsewhere[12,48,49,120].

Box 4 | **Sensitivity analyses using genome-wide data**

We outline here possible avenues to further utilize genome-wide data for causal inference. Unlike Mendelian randomization (MR), the approach suggested here does not follow instrumental variable principles. Instead, it builds on a classical multivariable framework in epidemiology, with the additional advantage of directly modelling genetic confounding (see part **c** versus part **b** of the figure in Box 2). A drawback of classic multivariable approaches (e.g. multivariable regression) is their inability to control for unobserved confounding. Here, we propose a sensitivity analysis to assess the extent to which estimates might result from unobserved confounding, in particular genetic confounding. Such a method could represent an alternative when no appropriate genetic instrument is available, which may frequently happen for complex phenotypes of interest to social scientists (e.g., income).



Sensitivity analyses constitute common epidemiological tools to probe the robustness of findings. One such sensitivity analysis aims to assess to what extent the estimate of the effect of an exposure X on an outcome Y (βxy) would change if additional confounders were observed. In other words, how large should unobserved confounding be for the observed association to become null? A similar approach, schematically represented in the figure, can be adopted using polygenic scores and heritability estimates to test the likelihood that the association partly or fully results from genetic confounding.

Two cases corresponding to two outcomes are represented, one outcome being more strongly influenced by X as shown by the standardized bivariate estimates of βxy on the Y-axis: 0.25 (red) and 0.15 (blue). The first step is to compute polygenic scores corresponding to each outcome using increasing *p*-value thresholds, which leads to more single nucleotide polymorphisms (SNPs) being used to generate the polygenic scores[127]. This first step results in several polygenic scores predicting increasing levels of variance in each outcome (represented on the X-axis in the figure with 5% and 9% of the variance explained in the two outcomes by the resulting polygenic scores). We then regress Y on X to estimate βxy while controlling for the polygenic score explaining 5% of the variance, and then repeat the operation with the polygenic score explaining 9% of the variance. This should lead to a progressive decrease in βxy, proportional to the amount of genetic confounding, as represented in the polygenic scores section of the figure. However, this would still only capture a small fraction of genetic confounding because even genome-wide polygenic scores may not capture all genetic influences on Y. Available heritability estimates for Y, based on both SNP and twin data, provide useful benchmarks for the sensitivity analysis. We can estimate βxy in an ideal scenario where available polygenic scores capture the entire heritability of the outcome, thereby estimating the full impact of genetic confounding on βxy. As shown in the figure, these scenarios can be based on available estimates of SNP-heritability (i.e. heritability explained collectively by common SNPs) or twin-heritability. Lines in the figure therefore represent the decrease in βxy as a function of the variance explained in Y by genetic factors. The following estimates of βxy are represented: (i) bivariate estimate; (ii) estimates when controlling for observed polygenic scores (here two); (iii) estimate under the SNP-heritability scenario (here 30% of variance explained); (iv) estimate under the twin-heritability scenario (here 60% of variance explained). Two possibilities are represented: βxy is still significant even under the twin-heritability scenario (red); βxy is already non-significant under the SNP-heritability scenario (blue). The sensitivity analysis therefore allows us to assess how likely it is that a given effect is entirely genetically confounded. It can be expanded in at least two ways: (i) by including polygenic scores for X and Y; and (ii) by integrating known environmental confounders.

Figure 1 | **Including multiple instruments in Mendelian randomization.** Each dot corresponds to one genetic variant, with 95% confidence interval (CI) of its association with the risk factor (horizontal) and the outcome (vertical). Regression lines correspond to different estimators (listed in the legend and explained in Table 1); numerical results are given for the inverse variance weighted (IVW) and MR-Egger regression methods. **a** | Association between low-density lipoprotein cholesterol (LDL-C) and coronary heart disease (CHD) (138 single nucleotide polymorphisms (SNPs)). Causal estimate derived from a single SNP (rs11591147) is 0.53 (95%CI: 0.30–0.77), which is less precise than the estimate derived from all SNPs (0.47; 95%CI: 0.40–0.54); Multivariable MR-Egger estimate (386 SNPs) is 0.41 (95%CI: 0.34–0.48); all estimates are consistent with causality. **b** | Association between high-density lipoprotein cholesterol HDL-C and CHD (183 SNPs). Multivariable MR-Egger estimate is –0.03 (95%CI: –0.10–0.03), which is not consistent with causality. **c** | Association between initiation of cannabis use and schizophrenia (21 SNPs); the IVW estimate is consistent with causality. **d** | Association between schizophrenia and initiation of cannabis use (107 SNPs); the IVW estimate is consistent with causality. Bidirectional MR (**c** and **d**) requires that the instrumental variable assumptions hold in both directions. Instruments with direct effects on both exposure and outcome are not informative on the direction of causality. Additional details on the data sources and analysis methods to generate this figure are provided in Supplementary information S1 (box).

Figure 2 | **Causal mapping.** Phenotypes of interest for various fields were selected to illustrate the possibilities and the pitfalls of a phenome-wide causal map. Estimates (see below) were computed based on association summary statistics for each phenotype. Only significant estimates at $p<0.001$ are shown. **a** | Genetic correlations were estimated between all phenotypes using linkage disequilibrium (LD) score regression[16] implemented in LD Hub (link in further information) **b** | Mendelian randomization (MR) causal effects were estimated in both directions for all phenotypes using an inverse variance weighted estimator, implemented in MR-base (link in further information). The map shows causal relationships in expected directions, such as low-density lipoprotein cholesterol (LDL-C) to coronary heart disease (CHD) and not the reverse, or from body mass index (BMI) to type 2 diabetes (T2D). However, some relationships are also not plausible, such as years of education determining childhood IQ. Overall, therefore, genetic correlations indicate shared genetic aetiology between phenotypes, which can be dissected in MR analyses to better assess whether they arise from pleiotropic effects and/or from causal effects in either or both directions. Phenome-wide analyses help in prioritizing plausible causal relationships and should be considered as an invitation to further probe the causal nature of detected relationships; however, they do not provide a definitive answer, as illustrated by the output of some implausible causal relationships. Upstream filtering based on a priori knowledge (e.g. temporality precludes a causal relationship from years of education to childhood IQ) or evidence from other designs can further increase causal evidence. Additional details on the data sources and analysis methods to generate this figure are provided in Supplementary information S1 (box).

Table 1. **| Estimators for Mendelian randomization using summary statistics**

Implementation	Limitations	Refs
Inverse variance weighted (IVW)-based methods		
The IVW method involves a weighted linear regression of SNP effects on the outcome on SNP effects on the risk factor, without an intercept term. The regression slope is equivalent to a weighted average of the ratio estimates (BOX 3), based on the precision of the causal estimate for each SNP used as an instrument[a,b]. IVW methods are more powerful than other methods (e.g. MR-Egger).	Unlike the other methods described below, IVW cannot account for directional (unbalanced) pleiotropy[c]. Balanced pleiotropic effects[d] can be accounted for in random effects IVW models (by allowing for heterogeneity) if the InSIDE assumption[e] holds true.	128
Methods based on Egger regression		
Linear regression with an intercept term using inverse variance weights[a,b]. MR-Egger regression provides consistent estimates even if all genetic instrumental variables are invalid under the InSIDE assumption[c]. This analysis is robust to directional (unbalanced) pleiotropy[c]. The intercept can be interpreted as the average pleiotropic effect across the genetic instrumental variables. Significance of the intercept term indicates the presence of unbalanced pleiotropy or violation of the InSIDE assumption[e].	Egger regression is less efficient and powerful than other methods because it allows for heterogeneity due to pleiotropy. It requires the InSIDE assumption[e].	60
Median-based methods		
Median-based methods allow some (but not all) instrumental variables to be invalid instruments. The median estimate is obtained by first calculating the ratio causal estimate for each instrumental variable and then taking their median. In the unweighted version, each genetic instrumental variable receives equal weight in the analysis. In the weighted version, the median is calculated using the inverse variance weights[b]. Median-based methods are more robust to directional pleiotropy than IVW and are more robust to individual genetic variants with outlying causal estimates than IVW and MR-Egger regression.	These methods assume that at least 50% of the instrumental variables are valid instruments (unweighted median estimates) or that the instrumental variables that represent 50% of the weight in the analysis are valid instruments (weighted median estimates).	61
Mode-based methods		
These methods allow the majority of the genetic instrumental variables to be invalid instruments under the ZEMPA assumption[f]. In the unweighted version of the mode estimate, each genetic instrumental variable receives equal weight in the analysis. In the weighted version, the mode is calculated using the inverse variance weights[b]. Mode-based methods are more robust to directional pleiotropy than IVW and more powerful than MR-Egger regression.	The methods assume that the largest number of instrumental variable estimates comes from valid instruments (ZEMPA assumption[f]), i.e. that the invalid instrumental variables have heterogeneous effect estimates. They have less power than IVW and median methods.	129,130
Multiple methods		
In practice, it is recommended to apply each of these methods to assess the robustness of the assumptions relevant for the different estimators, including the IVW estimator (all instruments are valid), the Egger estimator (all instruments		131

may be invalid if the 'InSIDE' assumption[e] is verified) and the median and modal estimators (a subset of genetic variants are valid instruments).

[a]Can also be calculated using robust estimates, which downweights the contribution of IVs with outlying ratio estimates. This can reduce bias and imprecision due to the influence of outlying variants. [b]Can also be calculated using penalized estimates, which downweights/penalizes the contribution of IVs with heterogeneous ratio estimates and gives more weight to genetic variants with homogeneous ratio estimates. This can reduce bias and imprecision if a small number of candidate instruments have heterogeneous or outlying causal estimates. [c]Directional (unbalanced) pleiotropy. Pleiotropic effects are more (or less) likely to be positive than negative, resulting in an average pleiotropic effect that is different to zero (significant intercept in MR-Egger regression under the InSIDE assumption). [d]Balanced pleiotropy. Pleiotropic effects are equally likely to be positive as negative (i.e. ratio estimates for individual SNPs above or below the true causal value), resulting in a null average pleiotropic effect. [e]The InSIDE assumption. The instrument strength independent of direct effects (InSIDE) is the assumption that pleiotropic effects are independent of the effects on the exposure, which is untestable and is violated when the pleiotropic effects act via confounders of the exposure and outcome. [f]ZEMPA assumption. The zero modal pleiotropy assumption (ZEMPA) states that the largest subset of genetic instrumental variables with the same ratio estimate comprises the valid instruments. SNP, single nucleotide polymorphism.

GLOSSARY

Causal risk and protective factors
Factors whose different values predict different risks of the outcome (either an elevated risk or a protective effect), all other factors being held constant.

Phenotypes
Measurable individual characteristics.

Confounding
A phenomenon whereby a variable (the confounder) has a causal effect on both the risk factor and the outcome, generating a spurious association between the two.

Genetic confounding
Confounding created by genetic factors influencing both the risk factor and the outcome.

Causal inference methods
Methods that aim to clarify the causal status of a risk factor, either by providing a direct estimate of the causal effect or by ruling out possible sources of confounding (e.g. removing the possibility of genetic confounding).

Genetically informed methods
Methods that use genetic information, such as known genetic relationships (e.g. twins) or genetic variation data.

Instrumental variables
Variables that are used as a proxy for an exposure X to estimate the causal effect of X on an outcome. This variable must be robustly associated with X, independent of all confounders of the effect of X on an outcome Y, and its effect on Y must be entirely mediated by X.

Mendelian randomization
A method that uses single nucleotide polymorphisms associated with an exposure, as instruments to probe the causal nature of the relationship between this exposure and an outcome of interest.

Counterfactual
(Also known as potential outcomes). The counterfactual is a treatment (or value of a risk factor) that an individual is not exposed to. The potential outcome is the outcome that would obtain under this counterfactual treatment.

Exchangeability
Verified when the expected outcome in the non-treated group would have been the same as the outcome in the treated group, if subjects in the non-treated group had received the treatment. Conditional exchangeability occurs when exchangeability is verified in each stratum of a confounder, after conditioning (adjusting) for the confounder.

Genetic relatedness
Occurs when two individuals share a proportion of their genome identical by descent, as a result of inheritance from a recent common ancestor.

Backdoor paths
(Also known as unlocked paths). A path between an exposure X and an outcome Y through a confounder, which biases the estimation of the causal effect of X on Y.

Structural equation modelling
Multivariate statistical technique combining factor analysis and regression analysis to estimate networks of relationships between latent and observed variables.

Sensitivity analysis
An analysis conducted to assess how robust an association of interest is to potential unobserved confounding or other sources of bias

Heritability
The proportion of variance in a phenotype that can be attributed to genetic differences among individuals in a given population. Narrow-sense heritability estimates additive genetic effects. Broad-sense heritability includes both additive and dominance effects.

Environmental influences
Environmental influences that contribute to make two individuals (e.g. twins) similar to each other (shared environmental influences) or dissimilar (non-shared environmental influences).

Single nucleotide polymorphisms
(SNPs). DNA sequence variation arising from differences in a single nucleotide: adenine (A), thymine (T), cytosine (C) or guanine (G).

Polygenic
Influenced by variants in many genes.

Pleiotropy
Occurs when a genetic locus (e.g. a single nucleotide polymorphism (SNP)) affects more than one trait.

Genome-wide association studies
(GWAS). Studies in which each of hundreds of thousands to millions of genetic variants is tested for an association with a phenotype.

Heterogeneity
Whether in meta-analyses or in Mendelian randomization analyses using many genetic instruments, heterogeneity occurs when several estimates of the same effect do not converge towards the same value.

Collider bias
When a variable (the collider) is independently caused by the exposure and outcome of interest, controlling for it creates an association between exposure and outcome.

Allelic scores
Computed as a polygenic score, but summarizes genetic information derived from a few to a few hundred single nucleotide polymorphisms (SNPs) as opposed to polygenic scores, which rely on thousands up to all SNPs in the genome.

Polygenic scores
Individual-level scores that summarize genetic risk (or protection) for a given phenotype. For each single nucleotide polymorphism (SNP), a score is computed by counting effect alleles in an individual and weighting them by the effect size of this SNP. A polygenic score is computed summing scores from a large number, potentially all, of the SNPs in the genome.

Dynastic effects
Occur when genetic variants in parents are transmitted to the offspring but also contribute to parental phenotype and in turn to the environment experienced by the child. This induces a correlation between offspring genotypes and offspring's environment.

Summary association statistics

Effect sizes and standard errors derived from a genome-wide association study for each single nucleotide polymorphism. They may include other summary statistics (e.g. allele frequency, imputation accuracy).

Genetic correlations
The correlation between causal effect sizes for two phenotypes across single nucleotide polymorphisms (SNPs). Typically reported as the correlation across the whole genome, and will differ when restricted to pleiotropic SNPs only.

Phenome-wide association studies
(PheWAS). These studies estimate the association of one or a few genetic variants of particular interest against many phenotypes, i.e. a selection of all possible phenotypes or phenome.

Colocalization methods
When a genetic region contains variants associated with more than one phenotype, colocalization methods aim to determine whether this is due to shared or distinct causal variants.

Linkage disequilibrium
Non-random associations between alleles at different loci.

D-separated
An exposure X and an outcome Y are d-separated through the process of d-separation, in which all backdoor paths between X and Y are blocked, to estimate the unconfounded effect of X on Y.

Relevance
A core assumption of instrumental variable estimation, whereby the instrument used must be robustly associated with the exposure of interest.

Exclusion restriction
A core assumption of instrumental variable estimation, whereby the effect of the instrument on the outcome must act entirely through its effect on the exposure (i.e. not directly and not via confounders or other mediators).

References

1. Glass, T. A., Goodman, S. N., Hernán, M. A. & Samet, J. M. Causal inference in public health. *Annu Rev Public Heal.* **34,** 61–75 (2013).
2. Rimm, E. B. *et al.* Vitamin E consumption and the risk of coronary heart disease in men. *N Engl J Med.* **328,** 1450–1456 (1993).
3. Stampfer, M. J. *et al.* Vitamin E consumption and the risk of coronary disease in women. *N Engl J Med.* **328,** 1444–1449 (1993).
4. Millen, A. E., Dodd, K. W. & Subar, A. F. Use of vitamin, mineral, nonvitamin, and nonmineral supplements in the United States: the 1987, 1992, and 2000 National Health Interview Survey results. *J Am Diet Assoc.* **104,** 942–950 (2004).
5. Eidelman, R. S., Hollar, D., Hebert, P. R., Lamas, G. A. & Hennekens, C. H. Randomized trials of vitamin E in the treatment and prevention of cardiovascular disease. *Arch Intern Med.* **164,** 1552–1556 (2004).
6. Imai, K., King, G. & Stuart, E. A. Misunderstandings between experimentalists and observationalists about causal inference. *J R Stat Soc Ser. A Stat Methodol.* **171,** 481–502 (2008).
7. Jaffee, S. R. & Price, T. S. The implications of genotype-environment correlation for establishing causal processes in psychopathology. *Dev Psychopathol.* **24,** 1253–1264 (2012).
8. Deaton, A. & Cartwright, N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med.* (2017). doi:10.1016/j.socscimed.2017.12.005

9. DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Heal. Econ.* **47,** 20–33 (2016).
10. McGue, M., Osler, M. & Christensen, K. Causal inference and observational research: the utility of twins. *Perspect Psychol Sci.* **5,** 546–556 (2010).

An introduction to the twin model from a causal inference perspective. It includes a discussion of concepts, estimation and limitations.

11. Davey Smith, G. & Ebrahim, S. What can Mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ* **330,** 1076–1079 (2005).
12. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* **23,** R89-98 (2014).
13. Burgess, S., Timpson, N. J., Ebrahim, S. & Davey Smith, G. Mendelian randomization: where are we now and where are we going? *Int J Epidemiol.* **44,** 379–388 (2015).
14. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* **47,** 1121–1130 (2015).
15. 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 Med.* **12,** e1001779 (2015).
16. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet.* **47,** 1236–1241 (2015).
17. Hemani, G. *et al.* MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. *bioRxiv* 78972 (2016). doi:10.1101/078972
18. Stuart, E. A. Matching methods for causal inference: a review and a look forward. *Stat. Sci. a Rev. J. Inst. Math. Stat.* **25,** 1 (2010).
19. Angrist, J. D., Imbens, G. W. & Rubin, D. B. Identification of causal effects using instrumental variables. *J Am Stat Assoc.* **91,** 444–455 (1996).
20. Tenesa, A. & Haley, C. S. The heritability of human disease: estimation, uses and abuses. *Nat Rev Genet.* **14,** 139–149 (2013).
21. Speed, D., Cai, N., Johnson, M. R., Nejentsev, S. & Balding, D. J. Reevaluation of SNP heritability in complex human traits. *Nat Genet.* **49,** 986–992 (2017).
22. Hernán, M. A. A definition of causal effect for epidemiological research. *J Epidemiol Community Heal.* **58,** 265–271 (2004).

A pedagogical introduction to the counterfactual or potential outcomes framework for causal inference. It includes mathematical notations and a discussion of key concepts such as association, causation and exchangeability.

23. Imbens, G. W. & Rubin, D. B. *Causal inference for statistics, social, and biomedical sciences*. (Cambridge Univ. Press, Cambridge, 2015).
24. Pearl, J. *Causality*. (Cambridge University Press, 2009).
25. Rice, F. *et al.* Disentangling prenatal and inherited influences in humans with an experimental design. *Proc Natl Acad Sci.* **106,** 2464–2467 (2009).

Example of the application of the in vitro fertilization design to examine the effect of smoking during pregnancy on birth weight.

26. Mezuk, B., Myers, J. M. & Kendler, K. S. Integrating social science and behavioral genetics: testing the origin of socioeconomic disparities in depression using a genetically informed design. *Am J Public Heal.* **103 Suppl,** S145-151 (2013).
27. Kendler, K. S. & Gardner, C. O. Dependent stressful life events and prior depressive episodes in the prediction of major depression: the problem of causal inference in psychiatric epidemiology. *Arch Gen Psychiatry* **67,** 1120–1127 (2010).
28. Bruder, C. E. G. *et al.* Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am J Hum Genet.* **82,** 763–771 (2008).
29. Carlin, J. B., Gurrin, L. C., Sterne, J. A., Morley, R. & Dwyer, T. Regression models for twin studies: a critical review. *Int J Epidemiol.* **34,** 1089–1099 (2005).
30. Vitaro, F., Brendgen, M. & Arseneault, L. The discordant MZ-twin method: one step closer to the holy grail of causality. *Int J Behav Dev.* **33,** 376–382 (2009).
31. Fletcher, J. M. & Lehrer, S. F. Genetic lotteries within families. *J Heal. Econ.* **30,** 647–659 (2011).

This paper provides a model combining family fixed-effects and genetic instruments, with a discussion of important concepts such as dynastic effects.

32. Kohler, H.-P., Behrman, J. R. & Schnittker, J. Social science methods for twins data: integrating causality, endowments, and heritability. *Biodemography Soc Biol.* **57,** 88–141 (2011).
33. Hjelmborg, J. *et al.* Lung cancer, genetic predisposition and smoking: the Nordic Twin Study of Cancer. *Thorax* **72,** 1021–1027 (2017).
34. Bröckerman, P., Hyytinen, A. & Kaprio, J. Smoking and long-term labour market outcomes. *Tob Control.* **24,** 348–353 (2015).
35. Cohen-Cline, H., Turkheimer, E. & Duncan, G. E. Access to green space, physical activity and mental health: a twin study. *J Epidemiol Community Heal.* **69,** 523–529 (2015).
36. Singham, T. *et al.* Concurrent and longitudinal contribution of exposure to bullying in childhood to mental health: the role of vulnerability and resilience. *JAMA Psychiatry* **74,** 1112–1119. (2017).
37. Taylor, M. J. *et al.* Developmental associations between traits of autism spectrum disorder and attention deficit hyperactivity disorder: a genetically informative, longitudinal twin study. *Psychol Med.* **43,** 1735–1746 (2013).
38. Frisell, T., Öberg, S., Kuja-Halkola, R. & Sjölander, A. Sibling comparison designs: bias from non-shared confounders and measurement error. *Epidemiology* **23,** 713–720 (2012).
39. Heath, A. C. *et al.* Testing hypotheses about direction of causation using cross-sectional family data. *Behav. Genet.* **23,** 29–50 (1993).
40. Neale, M. C. & Cardon, L. R. *Methodology for genetic studies of twins and families*. (Kluwer Academic, 1992).
41. D'Onofrio, B. M. *et al.* Paternal age at childbearing and offspring psychiatric and academic morbidity. *JAMA Psychiatry* **71,** 432–438 (2014).
42. Tully, E. C., Iacono, W. G. & McGue, M. An adoption study of parental depression as an environmental liability for adolescent depression and childhood disruptive disorders. *Am J Psychiatry* **165,** 1148 (2008).
43. Duffy, D. L. & Martin, N. G. Inferring the direction of causation in cross-sectional twin data: theoretical and empirical considerations. *Genet Epidemiol.* **11,** 483–502 (1994).
44. Wood, A. C., Rijsdijk, F., Asherson, P. & Kuntsi, J. Inferring causation from cross-sectional data: examination of the causal relationship between hyperactivity-impulsivity and novelty seeking. *Front Genet.* **2,** 6 (2011).
45. Toulopoulou, T. *et al.* Reciprocal causation models of cognitive vs volumetric cerebral intermediate phenotypes for schizophrenia in a pan-European twin cohort. *Mol Psychiatry* **20,** 1386 (2015).
46. Katan, M. B. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet* **1,** 507–508 (1986).
47. Davey Smith, G. Mendelian randomization for strengthening causal inference in observational studies: application to gene x environment interactions. *Perspect Psychol Sci.* **5,** 527–545 (2010).
48. Brion, M.-J. A., Benyamin, B., Visscher, P. M. & Smith, G. D. Beyond the single SNP: emerging developments in Mendelian randomization in the 'Omics' era. *Curr Epidemiol Rep.* **1,** 228–236 (2014).
49. Nitsch, D. *et al.* Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. *Am J Epidemiol.* **163,** 397–403 (2006).
50. Davey Smith, G. *et al.* Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet* **366,** 1484–1498 (2005).
51. Davey Smith, G. *et al.* Association of C-reactive protein with blood pressure and hypertension: life course confounding and Mendelian randomization tests of causality. *Arter. Thromb Vasc Biol.* **25,** 1051–1056 (2005).
52. Hartwig, F. P., Borges, M. C., Horta, B. L., Bowden, J. & Davey Smith, G. Inflammatory biomarkers and risk of schizophrenia: a 2-sample Mendelian randomization study. *JAMA Psychiatry* **74,** 1226 (2017).
53. Wensley, F. *et al.* Association between C reactive protein and coronary heart disease:

mendelian randomisation analysis based on individual participant data. *BMJ* **342,** d548 (2011).
54. Bolton, C. E. *et al.* The CRP genotype, serum levels and lung function in men: the Caerphilly Prospective Study. *Clin Sci (Lond).* **120,** 347–355 (2011).
55. Pingault, J.-B., Cecil, C. a. M., Murray, J., Munafo, M. & Viding, E. Causal inference in psychopathology: a systematic review of Mendelian randomisation studies aiming to identify environmental risk factors for psychopathology. *Psychopathol Rev.* **4,** 4–25 (2017).
56. Manousaki, D., Mokry, L. E., Ross, S., Goltzman, D. & Richards, J. B. Mendelian randomization studies do not support a role for vitamin D in coronary artery disease. *Circ Cardiovasc Genet.* **9,** 349–356 (2016).
57. Mokry, L. E. *et al.* Vitamin D and risk of multiple sclerosis: a Mendelian randomization study. *PLoS Med.* **12,** e1001866 (2015).
58. Sheehan, N. A. & Didelez, V. Commentary: Can 'many weak' instruments ever be 'strong'? *Int J Epidemiol.* **40,** 752–754 (2011).
59. Visscher, P. M. & Yang, J. A plethora of pleiotropy across complex traits. *Nat Genet.* **48,** 707 (2016).
60. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol.* **44,** 512–525 (2015).

This study introduces the use of a meta-analytical method known as Egger regression to Mendelian randomization analysis. Under certain assumptions, this approach enables causal estimation even when all instruments are invalid.

61. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol.* **40,** 304–314 (2016).
62. Rees, J. M. B., Wood, A. M. & Burgess, S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat Med.* **36,** 4705–4718 (2017).

This study provides the analytical framework to combine multivariable-MR and MR-Egger methods, which yields causal estimates robust to invalid genetic instruments.

63. Brion, M.-J. A., Shakhbazov, K. & Visscher, P. M. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol.* **42,** 1497–1501 (2013).
64. Burgess, S. & Thompson, S. G. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med.* **30,** 1312–1323 (2011).
65. Burgess, S. & Thompson, S. G. Improving bias and coverage in instrumental variable analysis with weak instruments for continuous and binary outcomes. *Stat Med.* **31,** 1582–1600 (2012).
66. Gage, S. H. *et al.* Assessing causality in associations between cannabis use and schizophrenia risk: a two-sample Mendelian randomization study. *Psychol Med.* **47,** 971–980 (2017).
67. Stringer, S. *et al.* Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium. *Transl Psychiatry* **6,** e769 (2016).
68. Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol.* **181,** 251–260 (2015).
69. Burgess, S., Freitag, D. F., Khan, H., Gorman, D. N. & Thompson, S. G. Using multivariable Mendelian randomization to disentangle the causal effects of lipid fractions. *PLoS One* **9,** e108891 (2014).
70. Liu, D. J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet.* **49,** 1758 (2017).
71. Tyrrell, J. *et al.* Genetic evidence for causal relationships between maternal obesity-related traits and birth weight. *JAMA* **315,** 1129–1140 (2016).
72. Richmond, R. C. *et al.* Using genetic variation to explore the causal effect of maternal pregnancy adiposity on future offspring adiposity: A Mendelian randomisation study. *PLoS*

Med. **14,** e1002221 (2017).
73. Zhang, G. *et al.* Assessing the causal relationship of maternal height on birth size and gestational age at birth: a Mendelian randomization analysis. *PLoS Med.* **12,** e1001865 (2015).

This study introduces intergenerational MR by computing allelic scores in the mother containing variants either transmitted or non-transmitted to the offspring. The method enables the estimation of the effect of maternal risk factors on the offspring free from passive gene–environment correlation.

74. Evans, D. M. *et al.* Mining the human phenome using allelic scores that index biological intermediates. *PLoS Genet.* **9,** e1003919 (2013).
75. Krapohl, E. *et al.* Widespread covariation of early environmental exposures and trait-associated polygenic variation. *Proc Natl Acad Sci.* **114,** 11727–11732 (2017).
76. Fletcher, J. M. The promise and pitfalls of combining genetic and economic research. *Heal. Econ.* **20,** 889–892 (2011).
77. Minica, C. C., Dolan, C. V., Boomsma, D. I., Geus, E. de & Neale, M. C. Extending causality tests with genetic instruments: an integration of Mendelian randomization and the classical twin design. *bioRxiv* 134585 (2017). doi:10.1101/134585
78. Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* **32,** 1–22 (2003).
79. Davey Smith, G. Capitalizing on Mendelian randomization to assess the effects of treatments. *J R Soc Med.* **100,** 432–435 (2007).
80. Pasaniuc, B. & Price, A. L. Dissecting the genetics of complex traits using summary association statistics. *Nat Rev Genet.* **18,** 117–127 (2017).
81. Gill, D. *et al.* Age at menarche and lung function: a Mendelian randomization study. *Eur J Epidemiol.* **32,** 701–710 (2017).
82. Bush, W. S., Oetjens, M. T. & Crawford, D. C. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nat Rev Genet.* **17,** 129 (2016).
83. O'Reilly, P. F. *et al.* MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS one.* **7,** e34861 (2012).
84. Porter, H. F. & O'Reilly, P. F. Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Sci Rep.* **7,** 38837 (2017).
85. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet.* **48,** 709 (2016).

This study introduces a method to detect shared genetic influences on multiple traits. It includes a test of asymmetry which helps to identify pairs of phenotypes that are causally related and which phenotype influences the other (i.e. direction of causation).

86. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* **48,** 481–487 (2016).

Applies Summary Mendelian randomization (SMR) methods to expression data, and enables the distinction between shared aetiology between expression and phenotypes due to shared causal variants or distinct variants in linkage disequilibrium.

87. Richardson, T. G. *et al.* Mendelian Randomization analysis identifies CpG sites as putative mediators for genetic influences on cardiovascular disease risk. *Am J Hum Genet.* **101,** 590–602 (2017).
88. Wallace, C. Statistical testing of shared genetic control for potentially related traits. *Genet Epidemiol.* **37,** 802–813 (2013).
89. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10,** e1004383 (2014).

This paper introduces a Bayesian colocalization method to identify shared causal variants between phenotypes.

90. Walter, S. *et al.* Revisiting mendelian randomization studies of the effect of body mass index on depression. *Am J Med Genet B Neuropsychiatr Genet.* **168B,** 108–115 (2015).
91. Hemani, G. *et al.* Automating Mendelian randomization through machine learning to

construct a putative causal map of the human phenome. *bioRxiv* 173682 (2017). doi:10.1101/173682

92. Davey Smith, G. *et al.* Incidence of type 2 diabetes in the randomized multiple risk factor intervention trial. *Ann Intern Med.* **142,** 313–322 (2005).
93. Åsvold, B. O. *et al.* Causal associations of tobacco smoking with cardiovascular risk factors: a Mendelian randomization analysis of the HUNT Study in Norway. *Int J Epidemiol.* **43,** 1458–1470 (2014).
94. Burgess, S., Daniel, R. M., Butterworth, A. S. & Thompson, S. G. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. *Int J Epidemiol.* **44,** 484–495 (2015).
95. Chen, W.-M. & Abecasis, G. R. Family-based association tests for genomewide association scans. *Am J Hum Genet.* **81,** 913–926 (2007).
96. Dudbridge, F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum Hered.* **66,** 87–98 (2008).
97. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9,** e1003348 (2013).
98. Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort Profile: TwinsUK and healthy ageing twin study. *Int J Epidemiol.* **42,** 76–85 (2013).
99. Haworth, C. M. A., Davis, O. S. P. & Plomin, R. Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Res Hum Genet.* **16,** 117–125 (2013).
100. Magnus, P. *et al.* Cohort profile update: The Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol.* **45,** 382–388 (2016).
101. Fraser, A. *et al.* Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Clin Pr.* **42,** 97–110 (2013).
102. Walker, V. M., Davey Smith, G., Davies, N. M. & Martin, R. M. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. *Int J Epidemiol.* **46,** 2078–2089 (2017).
103. Scott, R. A. *et al.* A genomic approach to therapeutic target validation identifies a glucose-lowering GLP1R variant protective for coronary heart disease. *Sci Transl Med.* **8,** 341ra76-341ra76 (2016).
104. Lehrer, S. F. & Ding, W. Are genetic markers of interest for economic research? *IZA J Labor Policy.* **6,** 2 (2017).
105. Glymour, M. M., Tchetgen Tchetgen, E. J. & Robins, J. M. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol.* **175,** 332–339 (2012).
106. Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in aetiological epidemiology. *Int J Epidemiol.* **45,** 1866–1886 (2016).
107. Munafò, M. R. & Davey Smith, G. Robust research needs many lines of evidence. *Nature* **553,** 399–401 (2018).
108. Fisher, R. A. Alleged dangers of cigarette-smoking. *BMJ* **2,** 4 & 297–298 (1957).
109. Knopik, V. S., Neiderhiser, J. M., DeFries, J. C. & Plomin, R. *Behavioral Genetics*. (Worth Publishers, New York, 2016).
110. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533,** 539–542 (2016).
111. Kendler, K. S. & Baker, J. H. Genetic influences on measures of the environment: a systematic review. *Psychol Med.* **37,** 615–626 (2007).
112. Krapohl, E. & Plomin, R. Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Mol Psychiatry.* (2015). doi:10.1038/mp.2015.2
113. Munafò, M. R. *et al.* Association between genetic variants on chromosome 15q25 locus and objective measures of tobacco exposure. *J Natl Cancer Inst.* **104,** 740–748 (2012).
114. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet.* **42,** 441–447 (2010).
115. Morral, A. R., McCaffrey, D. F. & Paddock, S. M. Reassessing the marijuana gateway effect.

Addiction **97,** 1493–1504 (2002).
116. Rutter, M. Proceeding from observed correlation to causal inference: the use of natural experiments. *Perspect Psychol Sci.* **2,** 377–395 (2007).
117. Greenland, S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* **14,** 300–306 (2003).
118. Sheehan, N. A., Didelez, V., Burton, P. R. & Tobin, M. D. Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med.* **5,** e177 (2008).
119. Didelez, V. & Sheehan, N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res.* **16,** 309–330 (2007).
120. Burgess, S. & Thompson, S. G. *Mendelian Randomization: methods for using genetic variants in causal estimation*. (CRC Press, Boca Raton, 2015).
121. Davey Smith, G. *et al.* Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med.* **4,** e352 (2007).
122. Pierce, B. L. & Burgess, S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol.* **178,** 1177–1184 (2013).
123. Hu, J. X., Thomas, C. E. & Brunak, S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet.* **17,** 615–629 (2016).
124. Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet.* **14,** 483–495 (2013).
125. Paaby, A. B. & Rockman, M. V. The many faces of pleiotropy. *Trends Genet.* **29,** 66–73 (2013).
126. Kong, A. *et al.* The nature of nurture: effects of parental genotypes. *Science (80-.).* **359,** 424–428 (2018).
127. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31,** 1466–1468 (2015).
128. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol.* **37,** 658–665 (2013).
129. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol.* **46,** 1985–1998 (2017).
130. Burgess, S., Zuber, V., Gkatzionis, A., Rees, J. M. B. & Foley, C. Improving on a modal-based estimation method: model averaging for consistent and efficient estimation in Mendelian randomization when a plurality of candidate instruments are valid. *bioRxiv* 175372 (2017). doi:10.1101/175372
131. Bowden, J. *et al.* A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med.* **36,** 1783–1802 (2017).

Acknowledgements
The authors thank S. Gage and J.M. Vink for cannabis initiation summary statistics and syntax and J. Rees for multivariable MR syntax. JBP is a fellow of MQ: Transforming Mental Health (MQ16IP16) and affiliated to the CESP, INSERM, Univ. Paris-Sud, UVSQ, Université Paris-Saclay, Paris, France. PFO receives funding from the UK Medical Research Council (MR/N015746/1) and the Wellcome Trust (109863/Z/15/Z). This report represents independent research (part)-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

Competing interests
The authors declare no competing interests.

FURTHER INFORMATION
Twin Registries: http://www.twinstudies.org/information/twinregisters/
DAGitty: http://www.dagitty.net/
LD Hub: http://ldsc.broadinstitute.org/
MR-base: http://www.mrbase.org/
Access to this interactive links box is free online.

Author contributions:

J-B.P. and T.S. researched data for the article. J-B.P and F.D. wrote the manuscript. All authors contributed substantially to discussion of content and reviewed/edited the manuscript before submission.