# Mendelian randomization: genetic anchors for causal inference in epidemiological studies

George Davey Smith

Gibran Hemani

MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, School of Social and Community Medicine, Bristol, UK

Many examples exist of apparently robust observational associations between behavioural, pharmacological or physiological measures and disease risk which, when subjected to randomised controlled trials, do not deliver the anticipated health benefits because the causal effect has been incorrectly inferred. These include many nutritional factors (e.g. several vitamins), pharmacological agents (e.g. hormone replacement therapy) and circulating biomarkers (e.g. HDL cholesterol). Confounding, reverse causation and various biases can generate the associations, and even with careful study design and statistical adjustment incorrect causal inference is possible.[[1]](#endnote-1) [[2]](#endnote-2) The recognition of these problematic aspects of epidemiological investigation has led to the application of a series of methods aimed at improving causal inference[[3]](#endnote-3) [[4]](#endnote-4). A successful approach is to use genetic variants as indicators of exposure that are not subject to the influences that vitiate conventional study designs, an approach known as Mendelian randomization (MR)[[5]](#endnote-5) [[6]](#endnote-6). We will not repeat the many detailed reviews that now exist of MR5 [[7]](#endnote-7) [[8]](#endnote-8) [[9]](#endnote-9) [[10]](#endnote-10) [[11]](#endnote-11) [[12]](#endnote-12) [[13]](#endnote-13) , nor summarise the hundreds of empirical studies applying the technique to a wide range of exposures and disease outcomes, rather, after a brief summary of the foundational principles, we will outline recent developments and potential future directions of the field.

## Mendelian randomization: basic principles

Inferring the causal direction between correlated variables is a pervasive issue in biology. But attempting to answer this question can actually be achieved by asking a more fundamental question: what is the effect on the dependent variable that specifically acts through the hypothesised causal variable? Simple regression analysis alone is not equipped to answer this. The putative association certainly could be a result of the hypothesised causal relationship. However it is also possible that the causal effect is in the opposite direction (reverse causality), or crucially, there are likely many unobserved factors that could be causing both variables and leading to their association (confounding) (Figure 1). For both latter scenarios the effect on the dependent variable specific to the independent variable is zero. Even if the hypothesised causal direction were true, in the event that the independent variable is correlated with some unobserved confounders then the estimate of its causal effect could be biased. Mendelian randomisation (MR) is a statistical technique that is designed to make an unbiased estimate of the causal effect acting on the dependent variable specifically through the independent variable, and in doing so make inference as to whether or not the hypothesised causal relationship is true.

Suppose that trait A and trait B are correlated, it follows that if this correlation arises because A is causing B, then any variable that influences trait A should also influence trait B. So the key to inferring a causal relationship between A and B is to identify an ‘instrument’ which is known to be reliably associated with A. Biologists are in a privileged position in this regard because it is no exaggeration to say that virtually all traits of interest are at least partially influenced by genetic effects, and genetic effects can serve as excellent instruments for a number of reasons. First, and most crucially, in a genetic association the direction of causation is from the genetic polymorphism to the trait of interest, and not vice versa. Second, conventionally measured environmental exposures are often associated with a wide range of behavioural, social, and physiological factors that confound associations with outcomes[[14]](#endnote-14). Genetic variants on the other hand can serve as unconfounded indicators of a particular trait. Third, genetic variants and their effects are subject to relatively little measurement error or bias. Fourth, the exact causal variant for the trait is not required, a marker in linkage disequilibrium (LD) with the causal variant will satisfy the conditions for MR. Finally, in the era of genome wide association studies (GWAS) and high throughput genomic technologies, genetic instruments are routinely obtainable for huge sample sizes for a vast number of traits.

### Analogy between Mendelian randomization and randomised controlled trials

An intuitive way to understand how MR can be used to infer causality is by analogy with randomised controlled trials (RCTs). In RCTs the study participants are randomly allocated to one or another treatment, thus any potential confounding between treatment and outcome is avoided, and if the treatment truly has an effect then causal inference is unambiguous. MR creates a similar scenario for us. Supposing a particular allele is robustly related to trait A, and trait A causes trait B. From Mendel’s first law of inheritance, we know that the alleles passed from parents to offspring are randomly selected, and subjects who inherit the allele are in effect being randomly allocated a high dosage of trait A, while those who don’t inherit the allele are being randomly allocated a low dosage of trait A. Thus, by stratifying by genotype Mendel’s second law (independent segregation) ensures we are severing any potential associations between confounders that may underlie correlations between trait A and trait B. If the random allocation to genotype class has an association with trait B then trait A can be deemed causal. Empirical evidence that there is lack of confounding of genetic variants with factors that confound exposures in conventional observational epidemiological studies comes from several sources[[15]](#endnote-15). We outline the conditions for MR and a simple method in which it is performed in Box 1.

To date MR has been successfully applied to a wide range of observational associations, covering the range of applications from the causal effects of biomarkers, understanding the correlation between physiological measures, and estimating the causal effects of various behaviours and maternal influences (Table 1). But, inevitably, there are a number of limitations to MR that should be considered when using this tool (Table 1). The remainder of this paper will outline recent advances in MR that seeks to address some of these limitations.

## Recent extensions to basic Mendelian Randomization

### Use of multiple variants to increase power and test assumptions

Ideally MR is performed using a single variant whose biological effect is understood on the trait for which it is an instrument. But this comes with two potential limitations that can be partially mitigated by increasing the number variants used as instruments.

First, the genetic effect may not be particularly large, resulting in a weak instrument. By increasing the number of variants the proportion of variance explained by the instrument increases, thus improving precision in 2SLS[[16]](#endnote-16).

Second, the variant could be pleiotropic or in LD with a variant that affects the outcome, thus violating conditions (2) or (3) for being a valid instrument. One can interrogate this potential caveat by using multiple instruments. For example it would be increasingly improbably that two, three, or more independent instruments all result in the same conclusion due to the colocalisation of genetic factors for both traits. For a convincing example demonstrating the causal influence of low-density lipoprotein cholesterol (LDL-C) on coronary heart disease (CHD) see Figure 2 where 9 polymorphisms from 6 genes independently lead to very similar predicted causal effects using instrumental variables analyses[[17]](#endnote-17).

Typically genetic variants are only used as instruments if they are reliably detected and replicated in GWAS. But predictive power is often improved when SNPs that do not reach significance thresholds are also included, the rationale being that they may be false negatives due to small effect size[[18]](#endnote-18). Using this approach has the potential to dramatically improve the power of MR, but caution should be made due to the increased chance of including pleiotropic effects (Box 2).[[19]](#endnote-19)

### Two sample Mendelian randomization

It is often the case that an observational association between two variables exists, but high measurement costs or lack of appropriate biospecimens leads to prohibitively small datasets with both measurements and instruments. Methods have been developed to perform MR when the exposure variable and the outcome variable are measured in two independent datasets[[20]](#endnote-20).

Another scenario in which two sample MR can be used is if the data in which MR is being performed is the same as is being used to identify instruments. GWAS is known to lead to overestimation of genetic effect sizes due to the phenomenon of winner’s curse, and this can lead to bias in MR. Subsetting the data into two (or more) samples for estimation and testing can mitigate this problem. This method has been applied in a study that demonstrates that childhood adiposity causes a reduction in physical activity[[21]](#endnote-21).

### Bidirectional Mendelian randomization

A major limitation of the MR is that using a single genetic variant it is difficult to distinguish between an exposure causing an outcome and an outcome causing a trait because the genetic variant could be biologically causing either variable. For example, atheroma and BMI influence C-reactive protein (CRP) levels, and apparent misleading causal effects can be generated if the genetic variant is mistaken as being a variant for the exposure (e.g. an atheroma or obesity variant taken to be one for CRP)[[22]](#endnote-22).

Aside from focusing on instruments for which there exists some degree of biological understanding, another method is to use bi-directional MR. Here, instruments are required for both variables and MR is performed in both directions (Figure 1). If trait A causes trait B, then the instrument will be significantly associated with both A and B. However, a second instrument specific to trait B, will be associated with trait B, and not with trait A. This method was used in a recent paper that demonstrated that obesity is a causal factor in 25(OH)D levels, concluding that population level interventions to reduce BMI are expected to reduce instances of vitamin D deficiency[[23]](#endnote-23).

### Factorial Mendelian randomization

The manner by which causes of disease act together to increase disease risk can have important public health implications, as above additive effects lead to the clustering of risk factors generating a greater burden of disease in the population. For example it has been suggested that the risk of liver disease associated with the combination of obesity and heavy alcohol consumption is greater than multiplicative [[24]](#endnote-24), and adverse trends for these two risk factors would be predicted to generate considerable increases in liver disease. It is difficult to estimate such effects, however, as confounding can be magnified when examining two already confounded risk factors. By analogy with factorial randomised controlled trials, where separate randomisation of different treatments allows characterization of interactions between them[[25]](#endnote-25), Mendelian randomization studies can investigate disease risk associated with combinations of genetic variants and through this obtain unconfounded estimates of the effect of co-occurrence of the two risk factors that the genetic variants are taken to be indicators of.

### Multiphenotype Mendelian randomization

In some situations genetic variants tend to be associated with multiple intermediate phenotype, and estimating the causal effect of one particular intermediate phenotype is problematic. For example, categories of blood lipids associated with coronary heart disease, but HDL cholesterol and triglycerides, which are highly inversely correlated cannot be reliably separated using observational studies[[26]](#endnote-26). The large number of genetic variants related to HDL-C and triglycerides generally associate with both[[27]](#endnote-27) and thus genetic instruments created with multiple variants do not purely relate to one of the two lipids. In this context regression methods can be applied to attempt to separate the effects; two independent studies utilizing this approach suggested that the causal influence of triglycerides was robust, whereas the apparent protective effect of HDL-C was not[[28]](#endnote-28) [[29]](#endnote-29). Unlike in the situation with factorial Mendelian randomization there is a dependence on attempting to statistically separate effects, which reintroduces problems in conventional observational studies. Both the appropriateness of different statistical approaches and whether reliable answers can be obtained in the multiphenotype context remain areas of active investigation.

### Hypothesis free Mendelian randomization

The majority of MR studies have been focused on testing hypotheses that arose from established associations between traits from observational studies. But is this only the tip of the iceberg? In a study of unprecedented size, mining the records of 110 million patients uncovered 2909 associations between Mendelian diseases and complex traits, the majority of which were previously unreported[[30]](#endnote-30). As high-throughput ‘omics technologies continue to reduce in price, datasets with comprehensive phenotyping are destined to grow, and increasingly comprehensive catalogues of genetic variants that can be used as instruments will emerge. Indeed, there already exist reports for methylome-wide searches for causal mediators for various traits, for example increased folate in red blood cells in cord blood leads to a (mostly) positive influence on methylation at seven sites across the genome[[31]](#endnote-31).

## Conclusions

Deconstructing observational correlations into causal relationships is an elusive problem at the heart of biological understanding, pharmaceutical development, and medical practice. MR is a statistically robust method for this endeavour, whose scope for application widens as the cost of data generation continues to reduce.

## Acknowledgements

Thanks to Professor Sheila Bird who (in 2002) suggested the term “factorial Mendelian randomization” and to Dr Tom Palmer who suggested the term “multiphenotype Mendelian randomization”

Table 1: Examples of Mendelian randomisation

|  |  |  |  |
| --- | --- | --- | --- |
| Type | Exposure / trait | Disease / outcome | Example |
| Biomarkers | CRP | Coronary heart disease | Observational association between CRP and coronary heart disease is a result of confounding with BMI[[32]](#endnote-32) |
|  | Uric acid | Coronary heart disease | Observational association between uric acid and coronary heart disease is due to confounding by BMI[[33]](#endnote-33) |
|  | Homocysteine | Stroke | MR suggests a causal influence of homocysteine on risk of stroke[[34]](#endnote-34) |
|  | Macrophage migration inhibitory factor (MIF) | Type 2 diabetes | Elevated MIF, instrumented by a genetic marker, is associated with higher risk of type 2 diabetes, suggesting a causal role[[35]](#endnote-35) |
| Behaviours | Smoking | Anxiety/depression | Anxiety and depression amongst smokers is unlikely to be a consequence of the activity[[36]](#endnote-36) |
|  | Alcohol consumption | Blood pressure | Genetic variants for alcohol metabolism are associated with increase blood pressure amongst individuals who regularly consume alcohol[[37]](#endnote-37) |
| Physiological measures | BMI | Symptomatic gallstone disease | Three variants associated with BMI are associated with elevated risk of symptomatic gallstone disease, suggesting a causal role[[38]](#endnote-38) |
| Maternal influences | Alcohol consumption | Childhood IQ | Genetic variants for alcohol metabolism in children were associated with lower IQ at age 8 only if mothers were moderate drinkers during pregnancy[[39]](#endnote-39) |
|  | Maternal BMI | Fat mass | Fat mass in children aged 9-11 is not influenced by BMI of mothers during pregnancy[[40]](#endnote-40) |

Table 2: Limitations of Mendelian randomisation

|  |  |  |
| --- | --- | --- |
| Limitation | Role in Mendelian randomization (MR) studies | Approaches to evaluating or avoiding the limitation |
| Low statistical power | MR studies are often of low power and effect estimates are imprecise because of this | Increase sample size and/or combine genetic variants so they explain more of the variance of the intermediate phenotype |
| Reverse causation | A genetic variant may be causing the disease outcome which in turn causes the biomarker, or the causal direction could be in the opposite direction. 2SLS will not distinguish between these cases | Bi-directional MR can be used to distinguish between the two causal models. |
| Population stratification | Spurious associations used as instruments can lead to faulty causal inference | Restrict analyses to ethnically homogeneous groups, families and/or apply correction methods using ancestrally informative markers or principal components from genome wide data |
| Reintroduced confounding though pleiotropy | A genetic variant may directly influence more than one post-transcriptional process. Known to be the case for some genetic variants | When possible utilise cis variants with respect to the intermediate phenotype under study, as these may be less likely to have pleiotropic effects. Apply multiple instrument approaches with more than one independent genetic variant as unlikely pleiotropy will generate the same associations for different instruments |
| Linkage disequilibrium (LD) induced confounding | LD is crucial in genetic association studies as it allows marker SNPs to proxy for un-genotyped causal SNPs. However this can reintroduce confounding if LD leads to the association of SNPs related to more than one post-transcriptional process. This case will be similar to the pleiotropy situation | Studies can be carried out in populations with different LD structures. Approaches to avoiding distortion by pleiotropy will also counter problems due to LD |
| Canalization / developmental compensation | During development compensatory processes may be generated that counter the phenotypic perturbation consequent on the genetic variant utilized as an instrument | No general approach developed, although context–specific biological knowledge can be appealed to. Period of lifecourse when influence of genetic variation on IPs emerge can indicate whether canalization could, in principle, be an issue |
| Lack of genetic variants to proxy for modifiable exposure of interest | No reliable genetic variant associations for many intermediate phenotypes of interest, although an increasing number of these now identified | Continued genome wide and sequencing based studies |
| Complexity of associations | Without adequate biological knowledge misleading inferences regarding intermediate phenotypes and disease may be drawn | Increased biological understanding of genotype – phenotype links |

Figure 1: Directed acyclic graphs (DAGs) depicting MR and bi-directional MR

A. MR can be used to test the hypothesis that trait A causes trait B, provided that conditions (1), (2), and (3) are met adequately, governing that is a valid instrument. B. In bi-directional MR prior knowledge of the causal direction between traits A and B (if any) can be elucidated if valid instruments are present for each trait.

Macintosh HD:Users:explodecomputer:repo:mr_review_hmg:dag.pdf

Figure 2: Effect of lower LDL-C on risk of CHD (taken from Ference *et al.* (2012))

Boxes represent the proportion risk reduction (1-OR) of CHD for each exposure allele plotted against the absolute magnitude of lower LDL-C associated with that allele (measured in mg/dl). Vertical lines represent 1 SE above and below the point estimate of proportional risk reduction. SNPs are plotted in order of increasing absolute magnitude of associations with lower LDL-C. The line (forced to pass through the origin) represents the increase in proportional risk reduction of CHD per unit lower long-term exposure to LDL-C.

Macintosh HD:Users:explodecomputer:repo:mr_review_hmg:ference_fig3.pdf

Box 1: Performing Mendelian randomisation

Conventional instrumental variable (IV) analysis requires that the instruments are valid, and in order to be valid they must meet three conditions. An instrument for trait A must be

1. reliably associated with trait A;
2. associated with the outcome (trait B) only through trait A; and
3. independent of unobserved confounders that influence traits A and B after conditioning on observed confounders.

In MR, condition (1) is straightforward to test, but (2) and (3) cannot be directly proven. For example, if the variant is pleiotropic (see Box 2), or if it is in linkage disequilibrium (LD) with a genetic variant that influences the outcome through a different mechanism, this can lead to erroneous causal estimation. We discuss ways in which to address this issue later. If the above conditions are met, then the unbiased estimate of the effect of trait A on the outcome, B, can be made using two-stage least squares (2SLS) regression.

In stage 1 a predictor for A is constructed from its instrument, and in stage 2 the effect of the predictor for A acting on B is estimated. The intuition here is that A is potentially associated with B due to many confounding effects, and we wish to estimate the effect of A on B that occurs only via the instrument, which we know is in the correct causal direction for A. Thus, if the predictor for A is associated with B in the estimate from stage 2 then this is only occurring through a path which is has no confounding.

Stage 1’s predictor for A, , is obtained from , where is the instrument for A. In stage 2 the regression coefficient is obtained from . thus represents an asymptotically unbiased estimate of the effect of A on B that occurs due to alone. In order to transform this back into the full effect of A on B we calculate , where the potentially biased is obtained from . There exist several implementations of 2SLS in open source software.

Box 2: Consequences of pleiotropy on the interpretation of MR

Pleiotropy is the phenomenon by which a single locus influences multiple phenotypes[[41]](#endnote-41). Depending on the form it takes, pleiotropy is a potential limitation to interpretation of MR4, so distinguishing between its different types is important. In the context of MR there are two mechanisms by which pleiotropy occurs: a single process leading to a cascade of events (e.g. a locus influences one particular protein product, and this causes perturbations in many other phenotypes); or a single locus directly influencing multiple phenotypes[[42]](#endnote-45) [[43]](#endnote-46). Amongst its many names, the former has been termed “spurious pleiotropy” [[44]](#endnote-47), or “type II pleiotropy” [[45]](#endnote-49); the latter “biological pleiotropy” [[46]](#endnote-51), “type I pleiotropy” [[47]](#endnote-53). Type II pleiotropy is not only unproblematic for Mendelian randomization, it is the very essence of the approach, in which the downstream effects of a perturbed phenotype are estimated through the causal genetic 4. Type I pleiotropy, is problematic for the interpretation of MR.

As may be intuited, estimates of the degree of pleiotropy suggest that type II pleiotropy is the more pervasive form[[48]](#endnote-54), with type I pleiotropy being more pronounced at the level of the gene than at the level of single SNPs1 [[49]](#endnote-55). Greater pleiotropic effects are seen for mutations with larger effects on the primary trait[[50]](#endnote-56) [[51]](#endnote-57), as would be anticipated for type II pleiotropic influences. Consider the instrument of common variation in *FTO*, known to influence body mass index (BMI)[[52]](#endnote-59), probably through influencing caloric intake[[53]](#endnote-60) [[54]](#endnote-61). Higher BMI is expected to lead to a wide range of downstream phenotypes; thus it is a more parsimonious explanation that type II pleiotropy underlies the observation that BMI index and *FTO* relate to blood pressure and hypertension[[55]](#endnote-62), coronary heart disease [[56]](#endnote-63), fasting insulin, glucose, HDL cholesterol and trigylcerides[[57]](#endnote-64), bone mineral density[[58]](#endnote-65), chronic renal disease [[59]](#endnote-66), and diabetes[[60]](#endnote-67).

## References

1. Davey Smith G, Ebrahim S. Epidemiology - is it time to call it a day? Int J Epidemiol 2001;30:1-11. [↑](#endnote-ref-1)
2. Fewell Z, Davey Smith G, Sterne JAC. The impact of residual and unmeasured confounding in epidemiological studies; a simulation study. Am J Epidemiol 2007;166:646-655. [↑](#endnote-ref-2)
3. Lipsitch M, Tchetgen ET, Cohen T. [Negative Control Exposures in Epidemiologic Studies](http://journals.lww.com/epidem/Citation/2012/03000/Negative_Control_Exposures_in_Epidemiologic.28.aspx" \o "Negative Control Exposures in Epidemiologic Studies). Epidemiology. 2012;23:351-352 [↑](#endnote-ref-3)
4. Davey Smith G. Assessing intrauterine influences on offspring health outcomes: can epidemiological findings yield robust results? Basic and Clinical Pharmacology and Toxicology 2008;102:245-256. [↑](#endnote-ref-4)
5. Davey Smith G, Ebrahim S. ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiology 2003;32:1-22. [↑](#endnote-ref-5)
6. Timpson NJ, Wade KH, Davey Smith G. Mendelian Randomization: Application to Cardiovascular Disease. [Current Hypertension Reports](http://www.springerlink.com/content/1522-6417/) 2012;14:29-37. [↑](#endnote-ref-6)
7. Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. Int J Epidemiol 2004;33:30-42. [↑](#endnote-ref-7)
8. Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian Randomisation and Causal Inference in Observational Epidemiology. PLoS Med 2008;5: e177 [↑](#endnote-ref-8)
9. Lawlor DA, Harbord RM, Sterne JAC, Timpson NJ, Davey Smith G. Mendelian randomization : Using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine 2008; 27:1133-1163 [↑](#endnote-ref-9)
10. Bochud M, Rousson V. Usefulness of Mendelian Randomization in Observational Epidemiology. International Journal of Environmental Research and Public Health. 2010; 7(3):711-728. [↑](#endnote-ref-10)
11. VanderWeele TJ.; Tchetgen Tchetgen, EJ.; Cornelis, M. [Methodological Challenges in Mendelian Randomization](http://journals.lww.com/epidem/Abstract/2014/05000/Methodological_Challenges_in_Mendelian.14.aspx). Epidemiology. 2014;25:427-435 [↑](#endnote-ref-11)
12. Richmond RC, Davey Smith G, Ness AR, den Hoed M, McMahon G, Timpson NJ. Assessing causality in the association between child adiposity and physical activity levels: A Mendelian randomization analysis. PLoS Medicine 2014; doi:10.1371/journal.pmed.1001618 [↑](#endnote-ref-12)
13. Davey Smith G. Use of genetic markers and gene-diet interactions for interrogating population-level causal influences of diet on health. Genes & Nutrition 2011;6:27–43. [↑](#endnote-ref-13)
14. Davey Smith G, Lawlor DA, Harbord R, Timpson NJ, Day I, Ebrahim S. Clustered Environments and Randomized Genes: a fundamental distinction between conventional and genetic epidemiology. PLoS Medicine 2007;4:1985-1992. PMCID: PMC2121108 [↑](#endnote-ref-14)
15. Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 447:661–678. [↑](#endnote-ref-15)
16. Marie-Jo A Brion, Konstantin Shakhbazov, Peter M Visscher. Calculating statistical power in Mendelian randomization studies. Int. J. Epidemiol. (2013) 42 (5): 1497-1501 doi:10.1093/ije/dyt179 [↑](#endnote-ref-16)
17. Ference BA, Yoo W, Alesh I, Mahajan N, Mirowska KK, Mewada A, Kahn J, Afonso L, Williams KA Sr, Flack JM. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. J Am Coll Cardiol. 2012 Dec 25;60(25):2631-9 [↑](#endnote-ref-17)
18. International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009 Aug 6;460(7256):748-52. [↑](#endnote-ref-18)
19. Evans DM, Brion MJ, Paternoster L, Kemp JP, McMahon G, Munafò M, Whitfield JB, Medland SE, Montgomery GW; GIANT Consortium; CRP Consortium; TAG Consortium, Timpson NJ, St Pourcain B, Lawlor DA, Martin NG, Dehghan A, Hirschhorn J, Davey Smith G. Mining the human phenome using allelic scores that index biological intermediates. PLoS Genet. 2013 Oct;9(10):e1003919. [↑](#endnote-ref-19)
20. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. Am J Epidemiol. 2013 Oct 1;178(7):1177-84 [↑](#endnote-ref-20)
21. Richmond RC, Davey Smith G, Ness AR, den Hoed M, McMahon G, Timpson NJ. Assessing causality in the association between child adiposity and physical activity levels: a Mendelian randomization analysis. PLoS Med. 2014 Mar 18;11(3):e1001618 [↑](#endnote-ref-21)
22. Bowden J, Vansteelandt S. [Mendelian randomization analysis of case-control data using structural mean models](http://onlinelibrary.wiley.com/doi/10.1002/sim.4138/abstract). Statistics in Medicine 2011;30:678-694 [↑](#endnote-ref-22)
23. Vimaleswaran KS, Berry DJ, Lu C, Tikkanen E, Pilz S, Hiraki LT, Cooper JD, Dastani Z, Li R, Houston DK, Wood AR, Michaëlsson K, Vandenput L, Zgaga L, Yerges-Armstrong LM, McCarthy MI, Dupuis J, Kaakinen M, Kleber ME, Jameson K, Arden N, Raitakari O, Viikari J, Lohman KK, Ferrucci L, Melhus H, Ingelsson E, Byberg L, Lind L, Lorentzon M, Salomaa V, Campbell H, Dunlop M, Mitchell BD, Herzig KH, Pouta A, Hartikainen AL; Genetic Investigation of Anthropometric Traits-GIANT Consortium, Streeten EA, Theodoratou E, Jula A, Wareham NJ, Ohlsson C, Frayling TM, Kritchevsky SB, Spector TD, Richards JB, Lehtimäki T, Ouwehand WH, Kraft P, Cooper C, März W, Power C, Loos RJ, Wang TJ, Järvelin MR, Whittaker JC, Hingorani AD, Hyppönen E. Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. PLoS Med. 2013;10(2):e1001383 [↑](#endnote-ref-23)
24. Hart C, Morrison DS, Batty GD, Mitchell RJ, Davey Smith G. Effect of body mass index and alcohol consumption on liver disease: analysis of data from two prospective cohort studies. BMJ 2010;34:c1240. PMCID: PMC2837144 [↑](#endnote-ref-24)
25. Montgomery A, Peters TJ, Little P. Design, analysis and presentation of factorial randomised controlled trials. BMC Medical Research Methodology 2003, 3:26

    doi:10.1186/1471-2288-3-26 [↑](#endnote-ref-25)
26. Phillips A, Davey Smith G. How independent are "independent" effects? Relative risk estimation when correlated exposures are measured imprecisely. J Clin Epidemiol 1991;44:1223-31. [↑](#endnote-ref-26)
27. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. Nat. Genet.2013;45:1274–1283 [↑](#endnote-ref-27)
28. Do R, Willer CJ, Schmidy EM et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. Nature Genetics 2013;45, 1345-1352 doi:10.1038/ng.2795 [↑](#endnote-ref-28)
29. Holmes MV, Asselbergs FW, Palmer TM, et al. Mendelian randomization of blood lipids for coronary heart disease. Eur Heart J 2014, doi:10.1093/eurheartj/eht571 [↑](#endnote-ref-29)
30. Blair DR, Lyttle CS, Mortensen JM, Bearden CF, Jensen AB, Khiabanian H, Melamed R, Rabadan R, Bernstam EV, Brunak S, Jensen LJ, Nicolae D, Shah NH, Grossman RL, Cox NJ, White KP, Rzhetsky A. A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. Cell. 2013 Sep 26;155(1):70-80 [↑](#endnote-ref-30)
31. Binder AM, Michels KB. The causal effect of red blood cell folate on genome-wide methylation in cord blood: a Mendelian randomization approach. BMC Bioinformatics. 2013 Dec 4;14:353 [↑](#endnote-ref-31)
32. C Reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC), Wensley F, Gao P, Burgess S, Kaptoge S, Di Angelantonio E, Shah T, Engert JC, Clarke R, Davey-Smith G, Nordestgaard BG, Saleheen D, Samani NJ, Sandhu M, Anand S, Pepys MB, Smeeth L, Whittaker J, Casas JP, Thompson SG, Hingorani AD, Danesh J. Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. BMJ. 2011 Feb 15;342:d548 [↑](#endnote-ref-32)
33. Palmer TM, Nordestgaard BG, Benn M, Tybjærg-Hansen A, Davey Smith G, Lawlor DA, Timpson NJ. Association of plasma uric acid with ischaemic heart disease and blood pressure: mendelian randomisation analysis of two large cohorts. BMJ. 2013 Jul 18;347:f4262 [↑](#endnote-ref-33)
34. Casas JP, Bautista LE, Smeeth L, Sharma P, Hingorani AD (2005) Homocysteine and stroke: Evidence on a causal link from Mendelian randomisation. Lancet 365: 224–232 [↑](#endnote-ref-34)
35. Herder C, Klopp N, Baumert J, Muller M, Khuseyinova N, et al. (2008) Effect of macrophage migration inhibitory factor (MIF) gene variants and MIF serum concentrations on the risk of type 2 diabetes: Results from the MONICA/KORA Augsburg Case-Cohort Study, 1984–2002. Diabetologia 51: 276–284 [↑](#endnote-ref-35)
36. Bjørngaard JH, Gunnell D, Elvestad MB, Davey Smith G, Skorpen F, Krokan H, Vatten L, Romundstad P. The causal role of smoking in anxiety and depression: a Mendelian randomization analysis of the HUNT study. Psychol Med. 2013 Apr;43(4):711-9 [↑](#endnote-ref-36)
37. Chen L, Davey Smith G, Harbord R, Lewis S (2008) Alcohol intake and blood pressure: a systematic review implementing Mendelian randomization approach. PLoS Med 5:461 [↑](#endnote-ref-37)
38. Stender, S., Nordestgaard, B. G. and Tybjærg-Hansen, A. (2013), Elevated body mass index as a causal risk factor for symptomatic gallstone disease: A Mendelian randomization study. Hepatology, 58: 2133–2141 [↑](#endnote-ref-38)
39. Lewis SJ, Zuccolo L, Davey Smith G, Macleod J, Rodriguez S, et al. (2012) Fetal Alcohol Exposure and IQ at Age 8: Evidence from a Population-Based Birth-Cohort Study. PLoS ONE 7(11): e49407 [↑](#endnote-ref-39)
40. Lawlor DA, Timpson NJ, Harbord RM, Leary S, Ness A, et al. (2008) Exploring the developmental overnutrition hypothesis using parental–offspring associations and *FTO* as an instrumental variable. PLoS Med 5: e33 [↑](#endnote-ref-40)
41. Stearns FW. One hundred years of pleiotropy: a retrospective. Genetics 2010; 186: 767-773. [↑](#endnote-ref-41)
42. Hodgkin J. Seven types of pleiotropy. Int. J. Dev. Biol 1998; 42: 501-505. [↑](#endnote-ref-45)
43. Pyeritz RE. Pleiotropy revisited: molecular explanations of a classic concept. American Journal of Medical Genetics 1989; 34: 124-134. [↑](#endnote-ref-46)
44. Gruneberg H. An analysis of the “pleiotropic” effects of a lethal mutation in the rat. Proc. R. Soc. Lond. B. 1938; 125: 123-144. [↑](#endnote-ref-47)
45. Wagner GP, Zhang J. The pleiotropic structure of the genotype – phenotype map: the evolvability of complex organisms. Nature Reviews Genetics 2011; 12: 204-213. [↑](#endnote-ref-49)
46. Stearns FW. One hundred years of pleiotropy: a retrospective. Genetics 2010; 186: 767-773. [↑](#endnote-ref-51)
47. Gruneberg H. An analysis of the “pleiotropic” effects of a lethal mutation in the rat. Proc. R. Soc. Lond. B. 1938; 125: 123-144. [↑](#endnote-ref-53)
48. He X, Zhang J. Toward a molecular understanding of pleiotropy. Genetics 2006; 173: 1885-1891. [↑](#endnote-ref-54)
49. Stern DL. Evolutionary developmental biology and the problem of variation. Evolution 2000; 54: 1079-1091. [↑](#endnote-ref-55)
50. Wang Z, Liao B-Y, Zhang J. Genomic patterns of pleiotropy and the evolution of complexity. PNAS 2010; doi: 10.7073/pnas.1004666107. [↑](#endnote-ref-56)
51. Wagner GP. Pleiotropic scaling of gene effects and the “cost of complexity”. Nature 2008 452; 470-472. [↑](#endnote-ref-57)
52. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Davey Smith G; The Wellcome Trust Case Control Consortium; Hattersley AT, McCarthy MI. A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. Science 2007;316:889-894. [↑](#endnote-ref-59)
53. Timpson NJ, Emmett PM, Frayling TM, Rogers I, Hattersley AT, McCarthy MI, Davey Smith G. The fat mass- and obesity- associated locus and dietary intake in children. Am J Clin Nutr 2008;88:971-978. [↑](#endnote-ref-60)
54. Richmond RC, Timpson NJ. Recent Findings on the Genetics of Obesity: Is there Public Health Relevance? Current Nutrition Reports. 2012;DOI 10.1007/s13668-012-0027-x [↑](#endnote-ref-61)
55. Timpson N, Harbord R, Davey Smith G, Zacho J, Tybaerg-Hansen A, Nordestgaard BG. Does Greater Adiposity Increase Blood Pressure And Hypertension Risk? Mendelian Randomization Using Fto/Mc4r Genotype. Hypertension 2009;54:84-90. [↑](#endnote-ref-62)
56. Nordestgaard BG, Palmer TM, Benn M, Zacho J, Tybjærg-Hansen A, Davey Smith G, Timspon NJ. The Effect of Elevated Body Mass Index on Ischemic Heart Disease Risk: Causal Estimates from a Mendelian Randomisation Approach. PLoS Med 2012;9(5): e1001212. doi:10.1371/journal.pmed.1001212 [↑](#endnote-ref-63)
57. Freathy RM, Timpson NJ, Lawlor DA, Pouta A, Ben-Shlomo Y, Ruokonen A, Ebrahim S, Shields B, Zeggini E, Weedon WM, Lindgren CM, Lango H, Melzer D, Ferrucci L, Paolisso G, Neville MJ, Karpe F, Palmer CAN, Morris AD, Elliott P, Jarvelin MR, Davey Smith G, McCarthy MI, Hattersley AT, Frayling TM. Common variation in the FTO gene alters diabetes-related metabolic traits to extent expected, given its effect on BMI. Diabetes 2008;57;1419-1426. [↑](#endnote-ref-64)
58. Timpson NJ, Tobias JH, Richards JB, Soranzo N, Duncan EL, Sims AM, Whittaker P, Kumanduri V, Zhai G, Glaser B, Eisman J, Jones G, Nicholson G, Prince R, Seeman E, Spector T, Brown MA, Peltonen L, Davey Smith G, Deloukas P, Evans DM. Common variants in the region around Osterix are associated with bone mineral density and growth in childhood. Hum Mol Genet. 2009;18:1510-1517. [↑](#endnote-ref-65)
59. Hubacek, JA and Viklicky, O and Dlouha, D and Bloudickova, S and Kubinova, R and Peasey, A and Pikhart, H and Adamkova, V and Brabcova, I and Pokorna, E, Bobak, M. The FTO gene polymorphism is associated with end-stage renal disease: two large independent case-control studies in a general population. **Nephrol Dial Transplant** 2012;27:1030 - 1035. [↑](#endnote-ref-66)
60. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Davey Smith G; The Wellcome Trust Case Control Consortium; Hattersley AT, McCarthy MI. A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. Science 2007;316:889-894. [↑](#endnote-ref-67)