International Journal of Epidemiology, 2015, 379–388
doi: 10.1093/ije/dyv108
EDITORIAL

EDITORIAL

Mendelian randomization: where are we now and where are we going?

The methodology and application of Mendelian randomization to study causal mechanisms in health and disease has developed dramatically over the past decade. New methods, large-scale genome-wide analyses, molecular epigenetics and other new -omics technologies are all providing exceptional opportunities for the exploitation of Mendelian randomization approaches to understand causes of complex traits and disease outcomes. This research has the potential to identify new approaches for the prevention and treatment of common conditions.

The origins of what is now termed 'Mendelian randomization' (Figure 1, see caption for assumptions) can be traced back over half a century,[1] although the first extended presentation of the principles was in this journal just over a decade ago,[2] Since then it has become a widely utilized methodology, with publications covering many branches of biomarker,[3–12] behavioural[13–16] and infectious disease[17,18] epidemiology. Mendelian randomization studies with clear implications for pharmacotherapeutics are also becoming commonplace,[19–21] and applications to social science and to economics (the field in which the statistical technique of instrumental variables analysis central to Mendelian randomization was initially conceived[22]) are being developed.[23,24]

Methodological advances

Over the past few years, several methodological advances have been made. The basic assumption—that genetic variants which can proxy for a potentially modifiable exposure are essentially unrelated to confounding factors—has been demonstrated to have widespread plausibility.[25] The connection between the standard Mendelian randomization experiment and the theory of instrumental variables has been elaborated upon.[26,27] Extensions to use multiple genetic variants for increasing power and investigating the influence of pleiotropy have been theorized[28] and implemented.[29–32] Bidirectional Mendelian randomization for informing the direction of causal effects has been exemplified[33,34] and extended to consider more complex networks.[35] Methods for the estimation of non-linear causal effects have been proposed.[36,37] Causal effects of related phenotypes with common genetic predictors in a multivariable analysis framework have been estimated.[38,39] Factorial Mendelian randomization to predict the separate and combined effect of treatments using different genetic proxies has been undertaken.[40] Sensitivity analyses for investigating the biasing effects of pleiotropy have been developed.[41,42] Extensions to consider gene-by-environment interactions have been outlined and applied.[43–45] The integration of epigenetic profiles as an intermediate phenotype has been proposed[46,47] and implemented.[48,49] The development of Mendelian randomization into the hypothesis-free resolution of causal directions in correlated networks has been outlined.[50] In summary, methodological development has been undertaken in response to the challenges of new substantive applied questions and increasingly detailed genetic data. This development has enabled (and continues to enable) more sophisticated questions to be answered using the framework of Mendelian randomization.

Mendelian randomization in the post genome-wide association study era

Initial applications of Mendelian randomization generally incorporated a single genetic variant, and assessed the causal relationship of the modifiable intermediate phenotype on the outcome in a single sample. The proliferation of genome-wide association study (GWAS) data, and in particular publicly available GWAS data[51] (such as summary genetic associations with coronary artery disease in over 60 000 cases and 130 000 controls from the CARDIoGRAMplusC4D consortium[52]) provides opportunities to extend this via the use of the following.

i. Increased sample sizes. Consortia with GWAS data on large sample sizes are available for many phenotypic traits and disease outcomes. This increases the power of Mendelian randomization investigations.[53]
ii. Multiple genetic variants. For many intermediate phenotypes investigated in Mendelian randomization

studies, GWAS investigations have been able to identify multiple genetic variants contributing to variation in the phenotype. Again, this increases the power of Mendelian randomization investigations.[54]

iii. Two-sample Mendelian randomization. The ideal context for the precise estimation of genetic associations with modifiable intermediate phenotypes is population-based cohort studies. In contrast, the ideal context for the precise estimation of genetic associations with disease outcomes is case-control studies. Two-sample Mendelian randomization is a design strategy whereby genetic associations with the phenotype and with the outcome are taken from separate samples.[55] Provided that the samples come from the same underlying population (for example, the same ethnicity), valid causal estimates can be obtained even if concomitant data on the genetic variants, intermediate phenotype and outcome are not available for any individuals. Moreover, such estimates can be obtained from summarized data rather than individual-level data.[56,57] This allows the efficient evaluation of causal effects in large sample sizes without requiring sharing of individual-level data.

As it is not required for the phenotype and outcome in two-sample Mendelian randomization to be estimated on the same individuals, genetic associations with the phenotype and outcome can be taken from large consortia, thus potentially greatly increasing power compared with a one-sample Mendelian randomization analysis.[51]

Over the past decade, the heritability of many complex traits has been explored using GWAS. In general, common genetic variants have small effects on complex traits. In the recently completed UK10K study [www.uk10k.org], novel genetic variants with relatively large phenotypic effects were observed.[58] However, large effect sizes seemed to be confined to the rarest detectable signals and, for the most part, effects attributable to common genetic variants were small. This is rather disappointing from the viewpoint of developing predictive tools for even highly heritable traits. Studies like UK10K assessing the genetic architecture of complex traits more thoroughly through sequencing suggest that, for complex traits, this picture is unlikely to change. But even variants with modest effect sizes provide opportunities for the investigation of potential novel causal pathways using Mendelian randomization, particularly given the development of novel statistical tools for detecting and adjusting for pleiotropy from multiple genetic variants.[41]

The promise of -omics

Mendelian randomization studies have generally focused on a limited number of intermediate phenotypes, but recent applications of -omic technologies into large-scale population-based studies present new opportunities for identifying novel predictive biomarkers and causal links between established phenotypes and disease outcomes.[47,59–63] Both metabolomic and DNA methylation data are increasingly being exploited.[49,64]

Metabolomic data, representing multiple metabolic pathways in systemic metabolism, can be quantified by targeted mass spectroscopy or by proton nuclear magnetic resonance spectroscopy. With this, it has been possible to examine the causal role of risk factors such as body mass index (BMI) in the formation of metabolomic profiles and thus to consider the finer aetiology of possible disease effects.[65] Furthermore, many metabolites have substantial heritability and robust genetic variant associations have already been identified.[66,67] Metabolite profiles have proved useful in the prediction of cardiometabolic disease,[68,69] although their role as modifiable targets for intervention or causal mediators of disease risk is unclear. The availability of genetic instruments for many metabolites provides opportunities to assess the causal effects of metabolites on disease risk. Both bi-directional (see above) and hypothesis-generating (see below) applications of Mendelian randomization are likely to be useful in exploiting these data.

Methylation of DNA is a partially stable mechanism for gene regulation, occurring from the earliest stages of development onwards, under genetic, environmental and stochastic influences.[70] In a similar way to metabolomic data, the availability of large collections of genome-wide epigenetic data marks presents a valuable opportunity to consider the role of gene regulation in the aetiology of complex disease. In this case, methylation-related genetic variants (mQTLs) are used as proxy markers of DNA regulatory variation, which maybe causally implicated in diseases. A theoretical framework for this work has been developed[46,47,71] and applied.[48,49] (Figure 2).

As well as being potential targets for intervention, both metabolomic[72,73] and methylation data may serve as indicators of exposure to difficult-to-measure intermediate phenotypes. In the case of DNA methylation data in particular, these could provide proxy measures of long-term[74] or critical period exposure[75,76] that could otherwise not be assessed on large population samples.

Taxonomy of Mendelian randomization investigations

Limitations in our understanding of genetic variants used in Mendelian randomization has led to suggestions that evidence from Mendelian randomization studies in informal evidence synthesis should be down-weighted.[77–79] However, not all applications of Mendelian randomization

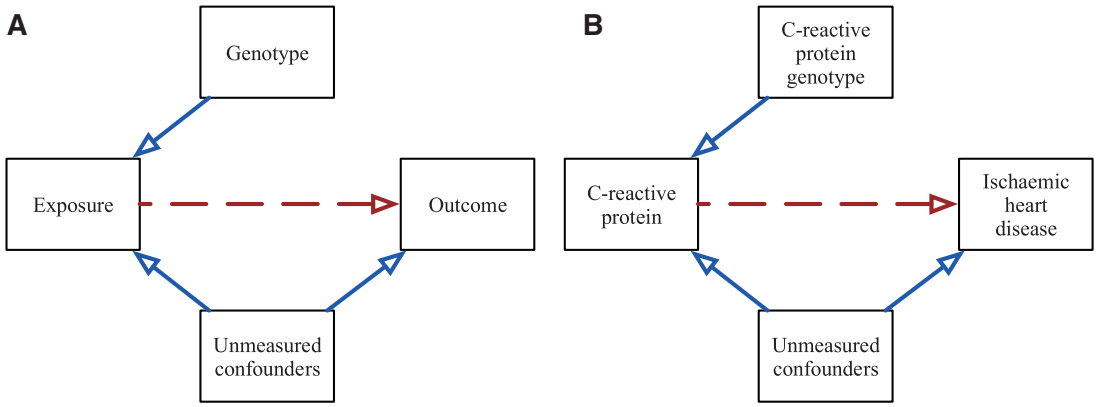

Figure 1. Mendelian randomization: using genetic variants as instrumental variables to establish whether an exposure is causally related to a disease or trait. (A) The genotype acts as an instrumental variable if: (i) it is associated with the exposure; (ii) it is independent of measured or unmeasured confounders; and (iii) it can only influence that outcome via the causal effect of the exposure. (B) Under the instrumental variable assumptions, the lack of association between the C-reactive protein genotype and disease risk indicates that C-reactive protein is not a causal risk factor for ischaemic heart disease. An association between the genotype and disease outcome would indicate a causal relationship of the exposure on the outcome.

are the same in terms of their aims, procedures and quality of evidence generated. We provide a taxonomy of Mendelian randomization investigations into three broad categories, based largely on the nature of the intermediate phenotype evaluated and the biological plausibility of the genetic variants for use in assessing causal effects. These categories are presented separately but form a spectrum of evidence quality, as some investigations will not fall neatly into a single category.

Validation of potential drug targets

Some phenotypes have a genetic aetiology dominated by a relatively small number of key coding or functionally relevant loci (such as C-reactive protein,[3] interleukin-6,[19,20] lipoprotein-associated phospholipase A2,[80] or secretory phospholipase A2,[81] bilirubin,[82] uric acid[83]). Mendelian randomization investigations conducted using a small number of genetic variants in a single gene region having clear biological links to the intermediate phenotype provide the closest parallels to a randomized trial.[84] These are the most plausible Mendelian randomization investigations, in terms of the validity of the instrumental variable assumptions that the variants are specific proxies for the phenotype, as well as providing evidence to aid the prioritization and development of pharmacological interventions which have a reasonable likelihood of producing health benefits.[85] This type of Mendelian randomization experiment mirrors the potential effects of a drug acting on the same pathway. Such applications have advantages for pharmaceutical companies in prioritizing drugs for clinical trials, and for investigating unintended consequences of drugs (both for drug repositioning and for investigating safety signals).

There are several examples of Mendelian randomization investigations relevant to pharmacological investigations. Drugs to inhibit C-reactive protein were not developed further after Mendelian randomization experiments demonstrated no causal role of C-reactive protein in cardiovascular disease.[3,86] In contrast, the interleukin-6 receptor can be blocked by a monoclonal antibody (tocilizumab) which was developed for the treatment of rheumatoid arthritis. A variant in the *IL6R* gene region shows an association with coronary heart disease risk,[87,88] so consequently tocilizumab would be worthwhile taking forward into trials for cardiovascular risk prevention.[89] As another example, statins are associated with an increased risk of type 2 diabetes. A Mendelian randomization study using genetic variants coding for HMGCoA reductase (the protein target that is inhibited by statins) demonstrated that these variants were associated with an increase in type 2 diabetes.[90,91] The inference from these findings is that attempts to make statins more specific and thereby reduce off -target effects will not avoid the increased risk of the diabetes. Genetic variants in the *CETP* gene region have been used as proxies for cholesterylester transfer protein (CETP) inhibitors, such as dalcetrapib.[92] These drugs are developed to raise high-density lipoprotein cholesterol levels. Variants in the *CETP* region have shown null associations with coronary artery disease risk,[21] although null associations with blood pressure suggest that the blood pressure-increasing effect of torcetrapib[93] is an off-target effect rather than a downstream consequence of CETP inhibition.[94]

A recent investigation to assess the impact ofinterleukin-1 inhibition (e.g. by use of the drug anakinra, which is beneficial in rheumatoid arthritis) on cardiometabolic disorders found that genetic variants which proxy the effects of sustained dual interleukin-1α/β inhibition were associated with an increased risk of cardiovascular diseases.[95]

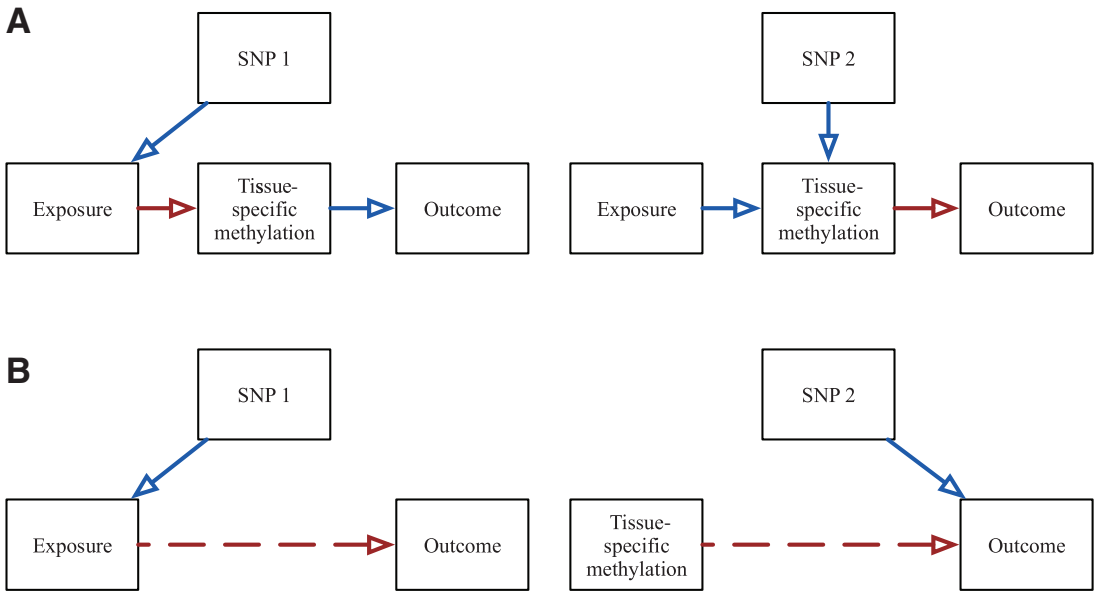

Figure 2. Two-step and integrated two-step/two-sample approaches in the application of Mendelian randomization to methylation data. (A) Two-step Mendelian randomization: genetic variants can be used as instrumental variables in a two-step framework to establish whether methylation is on the causal pathway between exposure and disease. First, a genetic variant (SNP 1) associated with the exposure of interest to assess the causal impact of the exposure on an intermediate trait (in this case tissue-specific methylation). Second, a different genetic variant (SNP 2) associated with the intermediate trait (and not associated with the exposure) is used to assess the causal impact of the intermediate trait on the outcome. (B) Two-sample/two-step Mendelian randomization: we consider tissue-specific DNA methylation as a potentially causal intermediate phenotype. In a potentially smaller first sample, the association of the exposure to tissue-specific DNA methylation is established using a Mendelian randomization approach (with the exposure-related SNP 1). A genetic variant associated with the same methylation difference but not related to the exposure is identified (SNP 2). In a potentially larger second sample, the exposure is shown to influence the outcome through the use of SNP 1, and the exposure-related methylation is shown to influence the outcome through the use of SNP 2. (Adapted with permission.[71]).

Two notable aspects of this investigation are the use of positive control variables (variables that should be affected by the phenotype according to biological knowledge) and the consideration of multiple outcomes. Clinical trials of anakinra show decreases in C-reactive protein and interleukin-6 levels that are also predicted by the associations of the genetic variants. The concordant associations with these positive controls increase the plausibility that the genetic variants are good proxies for the pharmacological intervention. The investigation of large numbers of outcomes, made practical by publicly available GWAS data, enables both the search for potential causal mediators of disease risk (in this case, proatherogenic lipids) and drug repositioning. Here, rather than finding another disease outcome that may be beneficially treated by anakinra, an important safety signal was discovered.

Investigation of complex intermediate phenotypes

Many intermediate phenotypes are not regulated by single metabolic pathways but are influenced by multiple genetic variants. Examples of multifactorial and polygenic risk factors include body mass index,[96] height[97] and blood pressure.[56] In these situations, Mendelian randomization investigations often proceed in a different manner, and on the basis of a large number of genetic variants in different gene regions. These variants may be discovered in GWAS investigations and the biological pathways linking each variant to the intermediate phenotype may be unknown. Clearly, the formal instrumental variable assumptions that the only causal pathway from the genetic variants to the outcome passes via the phenotype of interest are rarely satisfied.[98] Plausibility of a causal effect can be increased by empirical evidence that the genetic variants are not associated with measured confounders, as well as by demonstrating consistency and directional concordance of the causal estimate across genetic variants in multiple gene regions with different biological effects on the same phenotype. If many different independent genetic variants all suggest the same direction of causal effect, and if the overall statistical result is not dependent on one or two variants, then a causal conclusion is most plausible.[50] However, the associations of genetic variants with unmeasured or unknown confounders cannot be assessed, and so the instrumental variable assumptions are not fully testable. Additionally, even if a genetic variant is associated with a measured covariate, it is not possible to tell empirically whether this association is a (horizontally) pleiotropic effect of the genetic variant (hence a violation of the assumptions), or an

effect of the intermediate phenotype (a mediated, or vertically pleiotropic effect). In the latter case, provided that the only causal pathway from the genetic variant to the outcome is via the intermediate phenotype, the instrumental variable assumptions are not violated.

In these cases, the aim of a Mendelian randomization investigation is not only to give a more definitive answer as to whether the intermediate phenotype is causal or not, but also to investigate mechanisms linking the phenotype to the outcome. Particularly for phenotypes such as adult height, which is not readily modifiable, the findings of the analysis usually go beyond a simple instrumental variable analysis and investigate potential causal pathways.

Hypothesis-generating investigations

A final category of analyses (which some may feel are not true Mendelian randomization analyses) are termed 'hypothesis-generating investigations'. As with GWAS studies, these are undertaken particularly for intermediate phenotypes that do not have strong known genetic determinants, such as educational attainment.[99,100] Automated analyses of associations between a range of risk factors and outcomes have been undertaken using whole-genome scores[101] and summarized data from across the whole genome,[102] to investigate whether common genetic predictors correlate with phenotypic and outcome traits. Such investigations have given mixed results, and should be regarded as hypothesis-generating rather than assessments of causation. Nonetheless, they represent a natural extension to the methods of Mendelian randomization. Findings will be more speculative, but the statistical power to detect a causal effect may be greater. To this end, the application of automated two-sample Mendelian randomization in a hypothesis-generating approach is likely to expand rapidly the capacity of conventional epidemiology to generate plausible hypotheses. In this case, derived genetic instruments may be exported to existing large GWAS collections of any disease or outcome and employed to give estimates of the causal implications of exposure to novel modifiable risk factors. This would yield a potential return on the large collections of genetic variant data in the GWAS community which are, as yet, underutilised.

In this themed issue of the journal we have published both methodological developments and substantive findings from many research groups. Methodology for improving quality of reporting,[103] bias detection due to invalid instruments[41] and mediation in causal pathways[35] are covered. The effects of a wide range of intermediate phenotypes on disease outcomes using genetic instruments are also examined. These range from sex-hormone binding globulin[64], tobacco (smoking does lower body weight),[104] coffee,[105] milk[106] and alcohol[107] intakes, obesity,[108,109] vitamin D[110] and testosterone.[111,112] These analyses using genetic instruments provide a means of interrogating potential causal associations, particularly in circumstances where associations are likely to be heavily confounded and randomized experiments are not feasible.

Caution and conclusion

Potential limitations of the Mendelian randomization strategy were discussed extensively in its initial formal presentation[2] and have been reiterated elsewhere.[113–116] Largely as a function of the potentially overwhelming collection of genetic variants available to the epidemiologist looking to practise Mendelian randomization, the potential to fall into one of a series of analytical traps has been increased. Power, linkage disequilibrium, pleiotropy, canalization and population stratification have all been recognized as potential flaws in the Mendelian randomization approach as methods have been developed. While avoidance strategies for these limitations are now really beginning to appear, further limitations are being realized. In circumstances where we are less likely to have well-characterized and biologically understood genetic variants as instruments, it is tempting to use the totality of available variants in an analysis, for example in a genetic risk score approach.[117] Although it is attractive at the outset to amalgamate genetic variants into comprehensive genetic scores which have the potential to increase variance in the phenotype explained (and thus increase power),[118] it is increasingly clear that where these scores are not understood completely, the potential for inferential complication is greater now than ever.

Using the example of educational attainment, large-scale GWAS meta-analysis has successfully identified genetic variants reliably correlated with education.[99] However, these signals represent a small fraction of the total variability in educational attainment.[100] Genome-wide predictors will enhance the power of a Mendelian randomization analysis, with genetic scores including all variants (even those not associated at a conventional level of significance) explaining around 3% of the variance (see Figure 3). However, as a result of the combined impact of linkage disequilibrium, genetic contributions from many different biological pathways and the possible biasing effects of pleiotropy, the use of such a genome-wide estimator may sadly produce effect estimates which suffer the similar limitations as a more conventional, observational estimates.

The next decade will see a deeper understanding of the properties of genetic variants which will be crucial to the appropriate implementation and interpretation of Mendelian

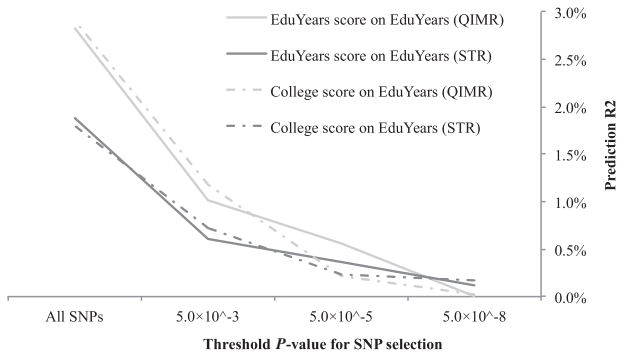

Figure 3. Estimates of variance explained by common genetic variants for educational attainment for differing portions of the genetic association spectrum. (Image adapted from Rietveld *et al.*[99]) Light grey lines (solid and mixed) show results from regressions of the number of years in education (Edu Years) on linear polygenic scores in a set of unrelated individuals from two independent epidemiological studies; Queensland Institute of Medical Research (QIMR) ($n = 3526$, light grey) and Swedish Twin Registry (STR) ($n = 6770$, dark grey). Solid lines indicate results from regressions of EduYears on linear polygenic derived from the genome-wide association analysis of years in education; and dashed lines indicate the genome-wide association analysis of the presence or absence of college education (Adapted from the original[99]).

randomization analyses. Over the past decade, Mendelian randomization has provided a novel and flexible paradigm to understand the causal nature of associations between modifiable risk factors and common diseases. Mendelian randomization has made use of the massive investment in human genetic research, focusing on causal mechanisms that have the promise of identifying worthwhile targets for pharmacological research and for preventive public health interventions that are already making a difference and will continue to do so in the coming decade.

Stephen Burgess[1,2,3,†]
[1]*Department of Public Health and Primary Care, University of Cambridge*
[2]*MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol*
[3]*Homerton College, University of Cambridge*

Nicholas J Timpson[2,†]
[2]*MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol*

Shah Ebrahim[4]
[4]*Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine*

George Davey Smith [2,*]
[2]*MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol*

**Corresponding author. MRC Integrative Epidemiology Unit, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom. E-mail: julia.mackay@bristol.ac.uk*

[†]These authors contributed equally to this paper.

References

1. Davey Smith G. Capitalizing on Mendelian randomization to assess the effects of treatments. *J R Soc Med* 2007;**100:** 432–35.
2. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32:**1–22.
3. C-Reactive Protein Coronary Heart Disease Genetics Collaboration (CCGC). Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. *BMJ* 2011;**342:**d548.
4. Clarke R, Bennett DA, Parish S *et al*. Homocysteine and coronary heart disease: meta-analysis of MTHFR case-control studies, avoiding publication bias. *PLoS Med* 2012;**9:** e1001177.
5. Benn M, Tybjærg-Hansen A, McCarthy MI, Jensen GB, Grande P, Nordestgaard BG. Nonfasting glucose, ischemic heart disease, and myocardial infarction: a mendelian randomization study. *J Am Coll Cardiol* 2012;**59:**2356–65.
6. McArdle PF, Whitcomb BW, Tanner K, Mitchell BD, Shuldiner AR, Parsa A. Association between bilirubin and cardiovascular disease risk factors: using Mendelian randomization to assess causal inference. *BMC Cardiovasc Disord* 2012;**12**:16.
7. Benn M, Tybjærg-Hansen A, Stender S, Frikke-Schmidt R, Nordestgaard BG. Low-density lipoprotein cholesterol and the risk of cancer: a mendelian randomization study. *J Natl Cancer Inst* 2011;**103:**508–19.
8. Trummer O, Pilz S, Hoffmann M *et al*. Vitamin D levels, vitamin D insufficiency genotypes and mortality: A Mendelian randomization study. *Bone* 2012;**50:**S108–09.
9. Kamstrup PR, Tybjærg-Hansen A, Nordestgaard BG. Genetic evidence that lipoprotein(a) associates with atherosclerotic stenosis rather than venous thrombosis. *Arterioscler Thromb Vasc Biol* 2012;**32:**1732–41.
10. Pfister R, Sharp S, Luben R *et al*. Mendelian randomization study of B-type natriuretic peptide and type 2 diabetes: evidence of causal association from population studies. *PLoS Med* 2011; **8:**e1001112.
11. Nordestgaard BG, Palmer TM, Benn M *et al*. The effect of elevated body mass index on ischemic heart disease risk: causal estimates from a mendelian randomisation approach. *PLoS Med* 2012;**9:**e1001212.
12. Pfister R, Barnes D, Luben R *et al*. No evidence for a causal link between uric acid and type 2 diabetes: a Mendelian randomisation approach. *Diabetologia* 2011;**54:**2561–69.
13. Wang Y, Zhang Y, Zhang J *et al*. Association of a functional single-nucleotide polymorphism in the ALDH2 gene with essential hypertension depends on drinking behavior in a Chinese Han population. *J Hum Hypertens* 2013;**27:**181–86.
14. Bjørngaard JH, Gunnell D, Elvestad MB *et al*. The causal role of smoking on anxiety and depression – A Mendelian Randomization analysis of the HUNT study. *Psychol Med* 2013;**43:**711–19.

15. Lewis SJ, Araya R, Davey Smith G *et al*. Smoking is associated with, but does not cause, depressed mood in pregnancy – a Mendelian randomization study. *PLoS One* 2011;**6:**e21689.
16. Drogan D, Sheldrick AJ, Schütze M *et al*. Alcohol consumption, genetic variants in alcohol deydrogenases, and risk of cardiovascular diseases: a prospective study and meta-analysis. *PLoS One* 2012;**7:**e32176.
17. Scott JAG, Berkley JA, Mwangi I *et al*. Relation between falciparum malaria and bacteraemia in Kenyan children: a population-based, case-control study and a longitudinal study. *Lancet* 2011;**378:**1316–23.
18. Kang H, Kreuels B, Adjei O, Krumkamp R, May J, Small DS. The causal effect of malaria on stunting: a Mendelian randomization and matching approach. *Int J Epidemiol* 2013;**42:** 1390–98.
19. The Interleukin-6 Receptor Mendelian Randomisation Analysis (IL6R Mendelian randomization) Consortium. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet* 2012;**379:**1214–24.
20. IL6R Genetics Consortium Emerging Risk Factors Collaboration. Interleukin-6 receptor pathways in coronary heart disease: a collaborative meta-analysis of 82 studies. *Lancet* 2012;**379:**1205–13.
21. Voight BF, Peloso GM, Orho-Melander M *et al*. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 2012;**380:**572–80.
22. Angrist JD, Krueger AB. Instrumental variables and the search for identification: from supply and demand to natural experiments. *J Econ Perspect* 2001;**15:**69–85.
23. von Hinke Kessler Scholder S, Davey Smith G, Lawlor DA, Propper C, Windmeijer F. Mendelian randomization: the use of genes in instrumental variable analyses. *Health Econ* 2011;**20:** 893–96.
24. von Hinke Kessler Scholder S, Davey Smith G, Lawlor DA, Propper C, Windmeijer F. The effect of fat mass on educational attainment: examining the sensitivity to different identification strategies. *Econ Hum Biol* 2012;**10:**405–18.
25. 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 Med* 2008;**4:**1985–92.
26. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 2007;**16:**309–30.
27. Lawlor DA, Harbord RM, Sterne JAC, Timpson NJ, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008; **27:**1133–63.
28. Davey Smith G, Timpson N, Ebrahim S. Strengthening causal inference in cardiovascular epidemiology through Mendelian randomization. *Ann Med* 2008;**12:**524–41.
29. Palmer TM, Lawlor DA, Harbord RM *et al*. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res* 2011;**21:** 223–42.
30. Palmer TM, Sterne JAC, Harbord RM *et al*. Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses. *Am J Epidemiol* 2011;**173:** 1392–403.
31. Timpson NJ, Sayers A, Davey Smith G, Tobias JH. How does body fat influence bone mass in childhood? a mendelian randomisation approach. *J Bone Miner Res* 2009:**24:** 522–33.
32. 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.
33. Timpson NJ, Nordestgaard BG, Harbord RM *et al*. C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization. *Int J Obes* 2011;**35:**300–08.
34. Welsh P, Polisecki E, Robertson M *et al*. Unravelling the directional link between adiposity and inflammation: a bidirectional Mendelian randomisation approach. *J Clin Endocrinol Metab* 2009;**95:**93–99.
35. Burgess S, Daniel RM, Butterworth AS, Thompson SG. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. *Int J Epidemiol* 2015;**44**:484–95.
36. Burgess S, Davies NM, Thompson SG; EPIC-InterAct Consortium. Instrumental variable analysis with a nonlinear exposure-outcome relationship. *Epidemiology* 2014;**25:** 877–85.
37. Silverwood RJ, Holmes MV, Dale CE *et al*. Testing for nonlinear causal effects using a binary genotype in a Mendelian randomization study: application to alcohol and cardiovascular traits. *Int J Epidemiol* 2014;**43:**1781–90.
38. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 2015;**181:**251–60.
39. Burgess S, Freitag DF, Khan H, Gorman DN, Thompson SG. Using multivariable mendelian randomization to disentangle the causal effects of lipid fractions. *PLoS One* 2014; **9:**e108891.
40. Ference BA, Majeed F, Penumetcha R, Flack JM, Brook RD. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: A 2×2 factorial mendelian randomization study. *J Am Coll Cardiol* 2015;**65:**1552–61.
41. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;**44:** 512–25.
42. Smith JG, Luk K, Schulz CA *et al*. Association of low-density lipoprotein cholesterol–related genetic variants with aortic valve calcium and incident aortic stenosis. *JAMA* 2014;**312:** 1764–71.
43. Freathy RM, Kazeem GR, Morris RW *et al*. Genetic variation at CHRNA5-CHRNA3-CHRNB4 interacts with smoking status to influence body mass index. *Int J Epidemiol* 2011;**40:** 1617–28.
44. Davey Smith G. Use of genetic markers and gene-diet interactions for interrogating population-level causal influences of diet on health. *Genes Nutr* 2011;**6:**27–43.
45. Davey Smith G. Mendelian randomization for strengthening causal inference in observational studies: application to gene by environment interaction. *Perspect Psychol Sci* 2010;**5:** 527–45.

46. Relton CL, Davey Smith G. Epigenetic epidemiology of common complex disease: prospects for prediction, prevention and treatment. *PLoS Med* 2010;**7:**e1000356.
47. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol* 2012; **41:**161–76.
48. Groom A, Potter C, Swan DC *et al*. Postnatal growth and DNA methylation are associated with differential gene expression of the TACSTD2 gene and childhood fat mass. *Diabetes* 2012;**61:** 391–400.
49. Liang L, Willis-Owen SA, Laprise C *et al*. An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature* 2015;**520:**670–74.
50. Davey Smith G. Random allocation in observational data: how small but robust effects could facilitate hypothesis-free causal inference. *Epidemiology* 2011;**22:**460–63.
51. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG; EPIC-InterAct Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol* 2015, Mar 15. [Epub ahead of print.] PMID: 25773750.
52. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, Willenborg C *et al*. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* 2013;**45:** 25–33.
53. Freeman G, Schooling CM, Cowling BJ. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *Int J Epidemiol* 2013;**42:**1157–63.
54. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol* 2011;**40:** 740–52.
55. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol* 2013;**178:**1177–84.
56. International Consortium for Blood Pressure Genome-Wide Association Studies; Ehret GB, Munroe PB, Rice KM *et al*. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011;**478:**103–9.
57. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;**37:**658–65.
58. Timpson NJ, Walter K, Min JL *et al*. A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in Europeans. *Nature Commun* 2014;**5:**4871.
59. Ala-Korpela M, Kangas AJ, Soininen P. Quantitative high-throughput metabolomics: a new era in epidemiology and genetics. *Genome Med* 2012;**4:**36.
60. Nordstrom A, Lewensohn R. Metabolomics: moving to the clinic. *J Neuroiuume Pharmacol* 2010;**5:**4–17.
61. Zhang GF, Sudhukhan S, Tochtrop GP, Brunengraber H. Metabolomics, pathway regulation and pathway discovery. *J Biol Chem* 2011;**286:**23631–5.
62. Patti GJ, Yanes O, Siuzdak G.. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell* 2012;**13:** 263–69.
63. Davies MN, Volta M, Pidsley R *et al*. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol* 2012;**13:**R43.
64. Wang Q, Kangas AJ, Soininen P *et al*. Sex hormone-binding globulin associations with circulating lipids and metabolites and the risk for type 2 diabetes: observational and causal effect estimates. *Int J Epidemiol* 2015;**44:**623–37.
65. Würtz P, Wang Q, Kangas AJ *et al*. Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change. *PLoS Med* 2014;**11:**e1001765.
66. Shin SY, Petersen AK, Wahl S *et al*. Interrogating causal pathways linking genetic variants, small molecule metabolites and circulating lipids. *Genome Med* 2014;**6:**25.
67. Kettunen J, Tukiainen T, Sarin A-P *et al*. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet* 2012;**44:**269–76.
68. Wang TJ, Larson MG, Vasan RS *et al*. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011;**17:**448–53.
69. Wang Z, Klipfell E, Bennett BJ *et al*. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 2011;**472:**57–63.
70. Relton CL, Davey Smith G. Is epidemiology ready for epigenetics? *Int J Epidemiol* 2012;**41:**5–9.
71. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;**23:**R89–98.
72. Chadeau-Hyam M, Athersuch TJ, Keun HC *et al*. Meeting-in-the-middle using metabolic profiling – a strategy for the identification of intermediate biomarkers in cohort studies. *Biomarkers*. 2011;**16:**83–88.
73. Holmes E, Loo RL, Stamler J *et al*. Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 2008;**453:**396–400.
74. Shenker NS, Ueland PM, Polidoro S *et al*. DNA methylation as a long-term biomarker of exposure to tobacco smoke. *Epidemiology* 2013;**24:**712–16.
75. Tobi EW, Goeman JJ, Monajemi R *et al*. DNA methylation signatures link prenatal famine exposure to growth and metabolism. *Nat Commun* 2014;**5:**5592.
76. Hernandez-Vargas H, Castelino J *et al*. Exposure to aflatoxin B1 *in utero* is associated with DNA methylation in white blood cells of infants in The Gambia. *Int J Epidemiol* 2015;doi:10.1093/ije/dyv027.
77. Glynn RJ. Promises and limitations of mendelian randomization for evaluation of biomarkers. *Clin Chem* 2010;**56:**388–90.
78. Ridker PM, Paynter NP, Danik JS, Glynn RJ. Interpretation of Mendelian randomization studies and the search for causal pathways in atherothrombosis: the need for caution. 2010;**8:** 465–69.
79. Johansen CT, Hegele RA. Mendelian Randomization. Using Mendelian randomization to determine causative factors in cardiovascular disease. *J Intern Med* 2013;**273:**44–47.
80. Casas JP, Ninio E, Panayiotou A, Palmen J, Cooper JA, Ricketts SL. PLA2G7 genotype, lipoprotein-associated phospholipase A2 activity, and coronary heart disease risk in 10 494 cases and 15 624 controls of European ancestry. *Circulation* 2010;**121:**2284–93.
81. Holmes MV, Simon T, Exeter HJ *et al*. Secretory phospholipase A(2)-IIA and cardiovascular disease: a mendelian randomization study. *J Am Coll Cardiol* 2013;**62:**1966–76.
82. Stender S, Frikke-Schmidt R, Nordestgaard BG, Grande P, Tybjaerg-Hansen A. Genetically elevated bilirubin and risk of ischaemic heart disease: three Mendelian randomization studies and a meta-analysis. *J Intern Med* 2013;**273:**59–68.

83. Palmer TM, Nordestgaard BG, Benn M *et al*. Association of plasma uric acid with ischaemic heart disease and blood pressure: Mendelian randomisation analysis of two large cohorts. *BMJ* 2013;**347:**f4262.
84. Hingorani A, Humphries S. Nature's randomised trials. *Lancet* 2005;**366:**1906–08.
85. Mokry JE, Ahmad O,Forgetta V, Thanassoulis G, Richards JB. Mendelian randomisation applied to drug development in cardiovascular disease: a review. *J Med Genet* 2015;**52:** 71–79.
86. Zacho J, Tybjaerg A, Jensen JS, Grande P, Sillesen H, Nordestgaard BG. Genetically elevated C-reactive protein and ischemic vascular disease. *N Engl J Med* 2008;**359:** 1897–908.
87. Interleukin-6 Receptor Mendelian Randomisation Analysis (IL6R MR) Consortium. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet* 2012;**379:**1214–24.
88. IL6R Genetics Consortium. Interleukin-6 receptor pathways in coronary heart disease: a collaborative meta-analysis of 82 studies. *Lancet* 2012;**379:**1205–13.
89. Ridker PM, Lüscher TF. Anti-inflammatory therapies for cardiovascular disease. *Eur Heart J* 2014;**35:**1782–91.
90. Swerdlow DI, Preiss D, Kuchenbaecker KB *et al*. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. *Lancet* 2015;**385:**351–61.
91. Frayling TM. Statins and type 2 diabetes: genetic studies on target. *Lancet* 2015;**385:**310–12.
92. Schwartz GG, Olsson AG, Abt M *et al*. Effects of dalcetrapib in patients with a recent acute coronary syndrome. *N Engl J Med* 2012;**367:**2089–99.
93. Bots ML, Visseren FL, Evans GW *et al*. Torcetrapib and carotid intima-media thickness in mixed dyslipidaemia (RADIANCE 2 study): a randomised, double-blind trial. *Lancet* 2007;**370:** 153–60.
94. Sofat R, Hingorani AD, Smeeth L *et al*. Separating the mechanism-based and off-target actions of cholesteryl ester transfer protein inhibitors with CETP gene polymorphisms. *Circulation* 2010;**121:**52–62.
95. Freitag DF, Butterworth AS, Willeit P *et al*. for the Interleukin 1 Genetics Consortium. Cardiometabolic effects of genetic upregulation of the interleukin 1 receptor antagonist: a Mendelian randomisation analysis. *Lancet Diabet Endocrinol* 2015;**3:** 243–53.
96. Holmes MV, Lange LA, Palmer T, Lanktree MB, North KE, Almoguera B. Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis. *Am J Hum Genet* 2014;**94:**198–208.
97. Nelson CP, Hamby SE, Saleheen D *et al*. Genetically determined height and coronary artery disease. *N Engl J Med* 2015; **372:**1608–61.
98. Glymour MM, Tchetgen Tchetgen EJ, Robins JM. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol* 2012; **175:**332–39.
99. Rietveld CA, Medland SE, Derringer J *et al*. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 2013;**340:**1467–71.
100. Ward ME, McMahon G, St Pourcain B *et al*. Genetic variation associated with differential educational attainment in adults has anticipated associations with school performance in children. *PLoS One* 2014;**9:**e100248.
101. Evans DM, Brion MJ, Paternoster L *et al*. Mining the human phenome using allelic scores that index biological intermediates. *PLoS Genet* 2013;**9:**e1003919.
102. Bulik-Sullivan B, Finucane HK, Anttila V *et al*. An atlas of genetic correlations across human diseases and traits. *Bioarxiv* 2015. doi: http://dx.doi.org/10.1101/014498.
103. Boef AGD, Dekkers OM, le Cessie S. Mendelian randomisation studies: a review of the approaches used and the quality of reporting. *Int J Epidemiol* 2015;**44**:496–511.
104. Winsløw UC, Rode L, Nordestgaard B. High tobacco consumption lowers body weight: a Mendelian randomization study of the Copenhagen General Population study. *Int J Epidemiol* 2015;**44**:540–50.
105. Nordestgaard AT, Thomsen M, Nordestgaard BG. Coffee intake and risk of obesity, metabolic syndrome and type 2 diabetes: a Mendelian randomization study. *Int J Epidemiol* 2015; **44**:551–65.
106. Bergholdt HKM, Nordestgaard BG, Varbo A, Ellervik C. Milk intake is not associated with ischemic heart disease in observational of Mendelian randomization analyses in 98 529 Danish adults. *Int J Epidemiol* 2015;**44**: 587–603.
107. Wium-Andersen MK, Ørsted DD, Schurmann Tolstrup J, Nordestgaard BG. Increased alcohol consumption as a cause of alcoholism, without similar evidence for depression: A Mendelian randomization study. *Int J Epidemiol* 2015;**44**: 526–39.
108. Hägg S, Fall T, Ploner A *et al*. Adiposity as a cause of cardiovascular disease: a Mendelian randomization study. *Int J Epidemiol* 2015;**44**:578–86.
109. Sungin D, Cornelis MC, Divaris K *et al*. Using genetics to test the causal relationship of total adiposity and periodontitis: Mendelian 5 randomization analyses in the Gene-Lifestyle Interactions and Dental Endpoints (GLIDE) Consortium. *Int J Epidemiol* 2015;**44**:638–50.
110. Brøndum-Jacobsen P, Benn M, Afzal S, Nordestgaard BG. No evidence that genetically reduced 25-hydroxyvitamin D is associated with increased risk of ischemic heart disease or myocardial infarction. A Mendelian randomization study. *Int J Epidemiol* 2015;**44**:651–61.
111. Zhao J, Jiang, C, Lam TH, Liu B, Cheng KK, Xu L. Genetically predicted testosterone and electrocardiographic QT interval duration in Chinese: a Mendelian randomization analysis in the Guangzhou Biobank Cohort Study. *Int J Epidemiol* 2015;**44**: 613–20.
112. Gong J, Zubair N. Commentary: Mendelian randomization, testosterone, and cardiovascular disease. *Int J Epidemiol* 2015; **44**:621–22.
113. Nitsch D, Molokhia M, Smeeth L, DeStavola BL, Whittaker JC, Leon DA. Limits to causal inference based on mendelian randomization: a comparison with randomized controlled trials. *Am J Epidemiol* 2006;**163:**397–403.
114. Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med* 2008;**5:**e177.
115. Glymour MM, Tchetgen EJ, Robins JM. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol* 2012; **175:**332–39.

116. VanderWeele T, Tchetgen Tchetgen E, Cornelis M, Kraft P. Methodological challenges in Mendelian randomization. *Epidemiology* 2014;**25:**427–35.
117. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol* 2013; **42:**1134–44.
118. Richmond RC, Davey Smith G, Ness AR, den Hoed M, MacMahon G, Timpson NJ. Assessing causality in the association between child adiposity and physical activity levels: A Mendelian randomization analysis. *PLoS Med* 2014;**11:** e1001618.