Editorial

Mendelian randomization: the challenge of unobserved environmental confounds

Philipp D Koellinger[1]*[†] and Ronald de Vlaming[1,2†]

[1]Department of Economics, School of Business and Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands and [2]Department of Applied Economics, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands

*Corresponding author. Department of Economics, School of Business and Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. E-mail: p.d.koellinger@vu.nl

[†]These authors contributed equally to this work

The standard approach for scientists to demonstrate causal relationships is to conduct experiments in which everything is held constant except for the element that is suspected to induce a change in the outcome of interest.[1,2] In epidemiology, however, there are limits to the types of experiments we can run. For example, we cannot randomly assign body height, trauma or college degrees to people to study their causal effects on health outcomes. If we want to move beyond observing co-occurrences in such situations and get closer to understanding causal mechanisms, the best we can often do is to search for naturally occurring experiments that induce random variation in the exposure of interest and use such exogenous shocks to estimate the causal effects.

One type of such exogenous shocks is the specific genetic variants a person is born with. The specific combination of genetic variants we inherit are effectively a random draw from the genotypes of our parents, thanks to the naturally occurring experiment of meiosis.[3] Because these randomly assigned genes tend to influence virtually all dimensions of diversity among people,[4] they are a window that allows us to look at the causal relationships between many phenomena that we cannot or should not manipulate experimentally. This is the basic idea behind Mendelian randomization (MR), a technique that is sometimes also referred to as using genes as instrumental variables.[5]

MR has rapidly gained popularity in recent years, thanks to the growing availability of genetic data and well-powered genome-wide association studies (GWASs) on an ever growing number of traits,[6,7] as well as continuing improvements in statistical methods.[8–14] According to the Web of Science, there were 492 publications mentioning MR in 2018, compared with just 56 a decade earlier in 2008.[15] Thus, scientists are increasingly relying on MR as a way to establish causality and to draw conclusions about policies, health advice, interventions and treatments.

Key assumptions of Mendelian Randomization

Several reviews of MR have been published that discuss the available methods and assumptions in some detail.[5,9,10,16–18] In essence, the key assumptions behind classic MR studies can be paraphrased as follows:

(Assumption 1) genes are randomly assigned among people;
(Assumption 2) some genes influence the exposure of interest;
(Assumption 3) the genes that influence the exposure do not influence the outcome via any other channel than the exposure.

Large-scale GWASs have uncovered many genetic loci for many exposures of interest,[6,7,19] thus addressing Assumption 2 and enabling a rising number of MR studies. However, the growing number of GWASs also reveal that most genetic loci tend to be associated with different

outcomes, sometimes even traits that seem unrelated at first glance—a phenomenon referred to as pleiotropy.[12,19] Thus, GWASs show that Assumption 3 is potentially problematic.

However, not all forms of pleiotropy constitute a violation of Assumption 3.[20] Specifically, if a gene shows pleiotropic effects because it influences the outcome via the exposure, this is not a violation of Assumption 3. Instead, this type of cascade effect that is often referred to as 'vertical pleiotropy' would be observed for any valid genetic instrument. In contrast, so-called horizontal pleiotropic effects that resemble 'parallel' processes (e.g. where a gene affects both exposure and outcome directly or via some unobserved mechanism that is relevant for both) constitute a challenge for classic MR studies. Unfortunately, it is difficult to know for certain whether genetic associations with both exposure and outcome are due to horizontal or vertical pleiotropy or a mixture of the two. Nevertheless, many methodological developments in the MR literature in the past few years have tried to tackle the potential violation of Assumption 3 by proposing approaches that are more robust to potential horizontal pleiotropic effects of genes.

However, little attention has been paid to Assumption 1. We believe that this deserves more consideration because genes are anything but randomly assigned within a population, and both GWASs and MR analyses are typically conducted in population samples. Specifically, genetic variants tend to be clustered within families, and—by extension—groups of people that share common ancestors. Family members tend to live in geographical proximity to each other, which induces correlations of genes with geography, as well as with cultural, economic, social, political and other environmental factors.[21–23] Thus, if we want to use genes as instrumental variables in a population sample, Assumption 1 is almost certainly violated and can therefore introduce bias. If we would know and observe all the relevant environmental factors, we could control for them and circumvent this problem. Unfortunately, the relevant environmental factors are often unknown or unobserved, leaving MR analyses that are purely based on population samples with an Achilles' heel.

Figure 1 illustrates the basic setup of an MR study.[8,17] *X* is the exposure of interest (e.g. educational attainment)
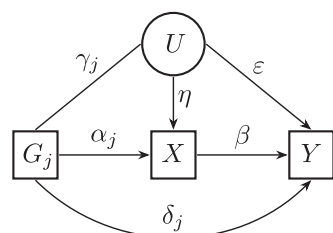
and *Y* is the health outcome of interest [e.g. body mass index (BMI)]. G_j ($j = 1,.., M$) are genetic markers that are influencing *X*, and *U* is any confounder that may influence both *X* and *Y* (e.g. another phenotype such as height or an environmental factor, such as the parenting style, that influences children's attitude towards school and their propensity to engage in sport). Importantly, *U* can also be correlated with G_j. In this diagram, α_j denotes the causal effect of genetic marker *j* on the exposure, *X*. The parameter of interest is the causal effect of *X* on *Y* (i.e. β). In our example, by estimating the causal parameter, β, we are trying to find out by how much BMI (*Y*) is expected to change on average if we would change only educational attainment (*X*), holding everything else constant, including *U*.

If the genes that influence educational attainment do not influence BMI in any other way than via educational attainment (i.e. $\delta_j = 0$ and $\gamma_j = 0$), we could use those genes as a naturally occurring experiment that induces exogenous variation in education. In turn, this variation could be used to obtain an unbiased estimate of the causal effect, β.[24,25] However if genes, for whatever reason, co-occur with any unobserved confound, *U* (i.e. $\gamma_j \neq 0$; e.g. as a result of parenting style or geographical differences in school quality, or the presence of some unobserved phenotype that is partly influenced by G_j and that affects both *X* and *Y*) or if genes still influence BMI even when controlling for educational attainment and *U* (i.e. $\delta_j \neq 0$; e.g. pleiotropic effects on the thyroid), classic MR will yield biased results. In fact, the bias induced by weak instruments in combination with even just slight violations of $\gamma_j = 0$ and $\delta_j = 0$ may be much worse than the bias in a naive ordinary regression of *Y* on *X*.[26,27]

Recent developments

Last year, several promising new MR-like methods (MR-PRESSO, GSMR, LCV and GIV) have been introduced that leverage GWAS summary statistics for identification.[11–14] Since most exposures of interest to epidemiologists are genetically complex, using many or even all genetic variants in MR analyses can boost statistical power and reduce the risk of weak instrument bias and false-positive findings.

All four methods try to address pleiotropy, although using different strategies. MR-PRESSO, GSMR, and LCV rely primarily on GWAS summary statistics and do not require access to individual-level data to examine the relationship between exposure and outcome. In contrast, GIV regression requires access to a hold-out sample with individual-level genetic data that were not included in the GWAS on the outcome of interest. Furthermore, GIV regression requires that the GWAS samples can be split

Figure 1. Structural model with confounder.

to obtain GWAS estimates from two non-overlapping samples. LCV and GIV regression make use of all single nucleotide polymorphisms (SNPs) included in a GWAS, whereas MR-PRESSO and GSMR are restricted to a subset of SNPs.

MR-PRESSO[12] and GSMR[11] both try to account for pleiotropy using a two-step approach. In the first stage, SNPs that are likely to have direct pleiotropic effects (i.e. not mediated by the exposure) are detected and removed. MR-PRESSO and GSMR use different tests that rely on different assumptions for this first stage. For example, MR-PRESSO assumes that more than 50% of SNPs are non-pleiotropic. If this assumption is violated, MR-PRESSO may yield biased results. In the second stage, the causal effect β is estimated by comparing the GWAS estimates of the effects of selected SNPs on X and Y. Intuitively, when X has a causal effect on Y, any SNP that causes X should also be associated with Y. If X has a very strong causal effect on Y, SNPs that cause X should have almost equally strong associations with Y. However, if the causal relationship of X on Y is weak, SNPs causing variation in X would have an attenuated association with Y, allowing estimation of the strength of the causal relationship.

Both MR-PRESSO and GSMR require the so-called InSIDE assumption[8] to hold. In addition, both methods require there to be no confounders that are associated with the SNPs under consideration. These two assumptions jointly can be considered as a relaxation of Assumption 3—for these two assumptions to hold, the effects of SNPs on the exposure should be uncorrelated with a specific part of the effects of those SNPs on the outcome (i.e. the part that is not mediated by the exposure). That is, in terms of the coefficients in Figure 1, $\alpha_j + \gamma_j\eta$ should be uncorrelated with $\delta_j + \gamma_j\varepsilon$ for $j = 1,.., M$.

Thus, both GSMR and MR-PRESSO assume that there is no additional phenotype that (i) is associated with the genes that are used for inferring the causal effect of interest and (ii) has a causal bearing on both exposure and outcome which has not been fully controlled for. Whether this assumption is reasonable depends on the variables being studied and how well population structure has been controlled for in the GWASs. It is interesting to note that the presence of this 'class' of confounders causes $\alpha_j + \gamma_j\eta$ to become correlated with $\delta_j + \gamma_j\varepsilon$, yielding biased estimates.

A further implication of these assumptions is that methods like MR-PRESSO and GSMR effectively also require the absence of unobserved environmental confounds that are correlated with the SNPs under consideration. This tacit assumption is easy to overlook, but could induce substantial bias if it is violated. After all, genetic variants are not random draws with respect to environmental factors in population samples: they are correlated across the genome with the family environment one is born into, which, in turn, can affect the outcome.[23] Again, the presence of unobserved environmental confounds that are, for whatever reason, correlated with the SNPs under consideration leads to bias because $\alpha_j + \gamma_j\eta$ and $\delta_j + \gamma_j\varepsilon$ become correlated.

Similar to MR-PRESSO and GSMR, LCV[14] only requires GWAS summary statistics rather than individual-level data. However, LCV uses a different approach to deal with pleiotropy. In contrast to MR-PRESSO and GSMR, LCV does not rely on the InSIDE assumption, yet still allows some or even all SNPs to have pleiotropic effects. Thus, it is not necessary for LCV to identify specific SNPs that influence X without pleiotropic effects on Y. Instead, LCV assumes that exactly one unobserved, latent variable mediates the genetic correlation between X and Y, and estimates a so-called genetic causality proportion. This parameter ranges between zero (no genetic causality) and one (full genetic causality, i.e. the entire genetic architecture of X is causal for Y). LCV is more powerful than traditional MR methods as well as MR-PRESSO and GSMR, as it uses the full set of GWAS results rather than just a subset of supposedly non-pleiotropic, top SNPs. However, violations of the LCV assumption that just one latent variable mediates the genetic correlation between X and Y can induce bias, and it is arguable whether this assumption is strictly weaker than the InSIDE assumption.

GIV regression[13] takes yet another approach. In contrast to MR-PRESSO, GSMR and traditional MR-methods, GIV regression does not use genetic variables as instruments for the exposure. Instead, the method only attempts to correct for all pleiotropic genes influencing both X and Y. GIV does so by constructing two polygenic scores,[28,29] based on two non-overlapping GWASs on the outcome of interest, in a single hold-out sample. The estimated effect in GIV regression can be thought of as the association between X and Y when controlling for the pleiotropic effects of genes on Y, which are not mediated by X. For example, a GIV regression of BMI on educational attainment aims to correct for the part of the phenotypic correlation that is due to pleiotropic effects of genes on both traits (e.g. due to genes that regulate cell growth and cell metabolism). This estimate may still deviate from the true causal effect of education on BMI because environmental confounds that are unrelated with genetic influences (e.g. war or economic crises influencing both BMI and education for non-genetic reasons) are not eliminated by taking pleiotropic effects into account. Furthermore, GIV regression does not resolve bias due to potential reverse causality (e.g. from BMI during childhood on educational

attainment due to discrimination against children with a high BMI).

The challenge of unobserved environmental confounds

Thus, whereas all four methods explicitly try to address the challenge of pleiotropy, they remain vulnerable to environmental confounds that are correlated with genes ($\gamma_j \neq 0$) to various degrees. Disconcertingly for all MR-like methods, correlations between genes and environments are pervasive and come in various forms. To make matters worse, these gene-associated environments also tend to be associated with broad ranges of phenotypes.

For example, Abdellaoui et al.[21] and Haworth et al.[22] find evidence for geographical clustering of polygenic score values for various traits in the UK Biobank which cannot be accounted for by the standard practice of adding principal components of the genetic data as control variables.[30] More generally, subgroups of a population that share common ancestors tend to differ in their allele frequencies. If these subgroups also differ in their environments, this can induce spurious genetic associations with outcomes that are influenced by these environments (e.g. the 'chopstick gene effect').[31] This type of confound in genetic association studies is often referred to as population stratification. Many statistical methods have been developed to address this issue in a GWAS and to test for its presence.[30,32–37] Although these methods go a long way in addressing population stratification, none of them guarantees the complete absence of the issue from genetic association studies in population samples, as the results from Abdellaoui et al.[21] and Haworth et al.[22] demonstrate.

Furthermore, Kong et al.[23] show that the non-transmitted genetic alleles of parents can still influence their children via environmental channels that are influenced by the parents' genes (e.g. via parenting style, role modelling, stability of the family and the socioeconomic conditions that parents pass on to their children). Thus, there is evidence that parental genotypes, to which the genotypes of the offspring are correlated, affect relevant environmental factors. Therefore, genetic nurture is a major challenge to MR methods. Arguably, genetic nurture may play a more important role for behavioural and socioeconomic traits (e.g. smoking, drinking, diet, educational attainment, aggression) than for biologically more proximate outcomes (e.g. height, bone density, macular degeneration). However, even if that is the case, most physical and medical outcomes are correlated or even partially caused by behaviour and socioeconomic status. For example, unbalanced diet, lack of physical activity, substance use, stress and lack of access to education and health care are risk factors for a broad range of medical outcomes including cardiovascular, metabolic and infectious diseases, cancer, mental health and risk of injury.[38–43] Therefore, we cannot easily dismiss the potential relevance of genetic nurture for medical outcomes. As a corollary, MR methods that are vulnerable to genetic nurture may yield biased and misleading results for medical outcomes. To our knowledge, GIV regression is currently the only MR-like method that seems to be robust to genetic nurture even if both the GWASs and polygenic prediction are carried out in population samples.[13]

Table 1 summarizes the main assumptions, properties and sources of biases of the four methods we discussed, when applied to (results from) population samples. Importantly, MR-PRESSO, GSMR and LCV are all biased in the presence of the gene–environment correlations that may be induced by genetic nurture and population stratification. GIV regression, on the other hand, seems to be more robust when gene–environment correlations induced by genetic nurture are at play. When it comes to population stratification, however, it is currently unclear whether and, if so, to what extent GIV regression is biased.

Within-family analyses can help to address the challenge of unobserved environmental confounds in all kinds of MR analyses, exploiting the truly random genetic differences between non-identical twins or siblings with the same parents (as pointed out by Davey Smith and Ebrahim[44]). For example, if these random genetic differences between offspring of the same parents would be used in a GWAS, the resulting estimates would be completely immune to population structure. Similarly, if samples of trios (father-mother-child) would be available, it would be possible to use only the deviations of the child's genotype from the average genotype of the parents to obtain GWAS results that are unaffected by population structure. However, since most traits of interest are genetically complex and the effect sizes of each genetic variant tend to be tiny, very large samples of siblings or trios would be required for well-powered within-family GWAS. Fortunately, efforts to pool the available genotyped family samples are already on their way, and first results using within-family MR suggest differences in the estimated causal effects for some pairs of traits compared with classic MR (e.g. BMI and educational attainment).[45] Thus, we assert that well-powered GWAS results that are unaffected by population structure would benefit any type of MR analysis, but they are particularly important as input for methods such as MR-PRESSO, GSMR and LCV—these methods all rely exclusively on GWAS summary statistics to make causal inferences.

Even if one could obtain GWAS summary statistics from a population sample which are completely unaffected

Table 1. Key assumptions, properties and sources of bias that apply to MR-PRESSO, GSMR, LCV and GIV

	Method			
	MR-PRESSO	GSMR	LCV	GIV
Panel A. Assumptions and properties				
1. Requires InSIDE for SNPs under consideration	Yes	Yes[a]	No	No
2. Removal of problematic pleiotropic SNPs	Yes	Yes	No	No
3. Returns estimate of causal effect of X on Y	Yes	Yes	No[b]	No[c]
Panel B. Inferences when using population samples under presence of:				
1. Environmental confounders associated with many SNPs				
A. Genetic nurture effects	Biased[d]	Biased[d]	Likely biased	Unbiased[e]
B. Population stratification not controlled for in GWAS	Biased[d]	Biased[d]	Biased	Unclear
2. Confounders not associated with SNPs[f]	Unbiased	Unbiased	Unbiased	Biased

[a]The InSIDE assumption is not explicitly mentioned by Zhu *et al.* (2018). However, correlation between SNP-exposure associations and SNP-outcome associations when controlling for exposure leads to a bias in the method.

[b]Estimates so-called genetic causality proportion, a parameter between zero (no genetic causality) and one (full genetic causality, i.e. the entire genetic architecture of X is causal for Y).

[c]Estimates association between X and Y, controlling for pleiotropic effects of genes on X and Y that are not mediated by X.

[d]Provided confounder is associated with most SNPs considered.

[e]Provided genetic nurture effects are identical across different GWAS samples.

[f]For example, purely environmental shocks.

by population structure, this still would not solve the challenge of genetic nurture—the estimated effects of causal genetic variants could still be conflated by the effect of the genotypes of close relatives at those loci, via environmental channels that are affected for example by parental genotypes. Thus, MR-PRESSO, GSMR and LCV remain vulnerable to this issue unless, again, GWAS results from appropriate within-family analyses are used.

By contrast, traditional MR analyses as well as GIV regression, which both use individual-level data for their inferences, can be subjected to a within-family design more readily (e.g. for GIV regression, the underlying GWAS estimates used to construct polygenic scores do not need to be based on GWASs using family data—one only needs a hold-out sample that consists of family data).

In such within-family designs, Assumption 1 of MR would actually be true—within each family, conditional on parental genotypes, genes are assigned randomly. Indeed, the authors of GIV regression specifically recommend such within-family analyses for this very reason. Furthermore, the recently started consortium that pools genotyped family samples will enable such analyses for a broad number of traits.[45]

By using a within-family design, GIV regression is also able to address the presence of environmental confounders that are not correlated in any way with genes—any environmental shock that is shared between family members is effectively controlled when using a within-family design.

It is an interesting and open question to what extent the current MR literature is biased (and potentially misleading) due to challenges we discussed above. Within-family analyses in large samples will help to answer this question. One

particular issue that cannot be addressed by methods such as MR-PRESSO and GSMR, even when using a within-family design, is the presence of one or more phenotypes that are not controlled for in the GWAS, but do affect both exposure and outcome and are associated with the relevant SNPs. LCV faces the issue that it may suffer from bias in the presence of multiple latent factors, even when applied to GWAS results based on a within-family design. Finally, under a within-family design, GIV is unable to deal with environmental confounders that are neither shared between family members nor correlated with genes.

Conclusion

We believe that none of the currently available MR-like methods alone will be able to tackle all the challenges discussed above. However, converging evidence from a combination of various MR methods that rely on different sets of assumptions, together with data from large, genotyped family samples, will allow us to make progress with the important objective of identifying causal effects in non-experimental data.

Conflict of interest: None declared.

References

1. Popper K. *The Logic of Scientific Discovery*. Abingdon, UK: Routledge, 2005.
2. Pearl J. *Causality*. New York, NY: Cambridge University Press, 2009.
3. John B. *Meiosis*. New York, NY: Cambridge University Press, 2005.
4. Polderman TJC, Benyamin B, Leeuw CD *et al*. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 2015;**47**:702–09.
5. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008;**27**:1133–63.
6. Visscher PM, Wray NR, Zhang Q *et al*. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* 2017;**101**:5–22.
7. Mills MC, Rahal C. A scientometric review of genome-wide association studies. *Commun Biol* 2019;**2**:9.
8. 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.
9. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res* 2017;**26**:2333–55.
10. Zheng J, Baird D, Borges M-C *et al*. Recent developments in Mendelian Randomization studies. *Curr Epidemiol Rep* 2017;**4**: 330–45.
11. Zhu Z, Zheng Z, Zhang F *et al*. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 2018;**9**:224.
12. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 2018;**50**:693–98.
13. DiPrete TA, Burik C, Koellinger P. Genetic instrumental variable regression: explaining socioeconomic and health outcomes in non-experimental data. *Proc Natl Acad Sci U S A* 2018;**115**:E4970–79.
14. O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet* 2018;**50**:1728–34.
15. Web of Science. https://www.webofknowledge.com/ (26 April 2019, date last accessed).
16. Davey Smith G, Hemani G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;**23**:R89–98.
17. Slob EAW, Burgess S. A comparison of robust Mendelian randomization methods using summary data. *bioRxiv* 2019. https://doi.org/10.1101/577940.
18. Hinke Kessler Scholder SV, 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.
19. Welter D, MacArthur J, Morales J *et al*. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucl Acids Res* 2014;**42**: D1001–06.
20. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet* 2018;**27**:R195–208.
21. Abdellaoui A, Hugh-Jones D, Kemper KE *et al*. Genetic consequences of social stratification in Great Britain. *bioRxiv* 2018. https://doi.org/10.1101/457515.
22. Haworth S, Mitchell R, Corbin L *et al*. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat Commun* 2019;**10**:333.
23. Kong A, Thorleifsson G, Frigge ML *et al*. The nature of nurture: effects of parental genotypes. *Science* 2018;**359**:424–28.
24. Roger J, Bowden DAT. *Instrumental Variables*. Cambridge, UK: Cambridge University Press, 2008.
25. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: Massachusetts Institute of Technology, 2002.
26. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997;**65**:557.
27. Palmer TM, Lawlor DA, Harbord RM *et al*. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res* 2012;**21**:223–42.
28. Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 2008;**3**:e3395.
29. Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM. Research review: polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry* 2014;**55**:1068–87.
30. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;**38**:904–09.
31. Hamer DH, Sirota L. Beware the chopsticks gene. *Mol Psychiatry* 2000;**5**:11–13.
32. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* 2014;**46**:100–06.
33. Loh P-R, Tucker G, Bulik-Sullivan BK *et al*. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015;**47**:284–90.
34. Bulik-Sullivan BK, Loh P-R, Finucane HK *et al*. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015;**47**:291–95.
35. Yang J, Weedon MN, Purcell S *et al*. Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* 2011;**19**: 807–12.
36. Lee JJ, Wedow R, Okbay A *et al*. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 2018;**50**: 1112–21.
37. Karlsson Linnér R, Biroli P, Kong E *et al*. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat Genet* 2019;**51**:245–57.
38. Willett WC. Diet and health: what should we eat? *Science* 1994; **264**:532–37.
39. World Health Organization. WHO Healthy Cities Project. *Social Determinants of Health: The Solid Facts*. Geneva: World Health Organization, 2003.
40. Stringhini S, Carmeli C, Jokela M *et al*. Socioeconomic status and the 25×25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1.7 million men and women. *Lancet* 2017;**389**:1229–37.

41. Warburton DER, Nicol CW, Bredin S. Health benefits of physical activity: the evidence. *CMAJ* 2006;**174**:801–09.
42. Hsia J, Kemper E, Kiefe C *et al*. The importance of health insurance as a determinant of cancer screening: evidence from the Women's Health Initiative. *Prev Med* 2000;**31**:261–70.
43. Movig KLL, Mathijssen MPM, Nagel PHA *et al*. Psychoactive substance use and the risk of motor vehicle accidents. *Accid Anal Prev* 2004;**36**:631–36.
44. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32**:1–22.
45. Brumpton B, Sanderson E, Hartwig FP *et al*. Within-family studies for Mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases. *bioRxiv* 2019. https://doi.org/10.1101/602516v1.