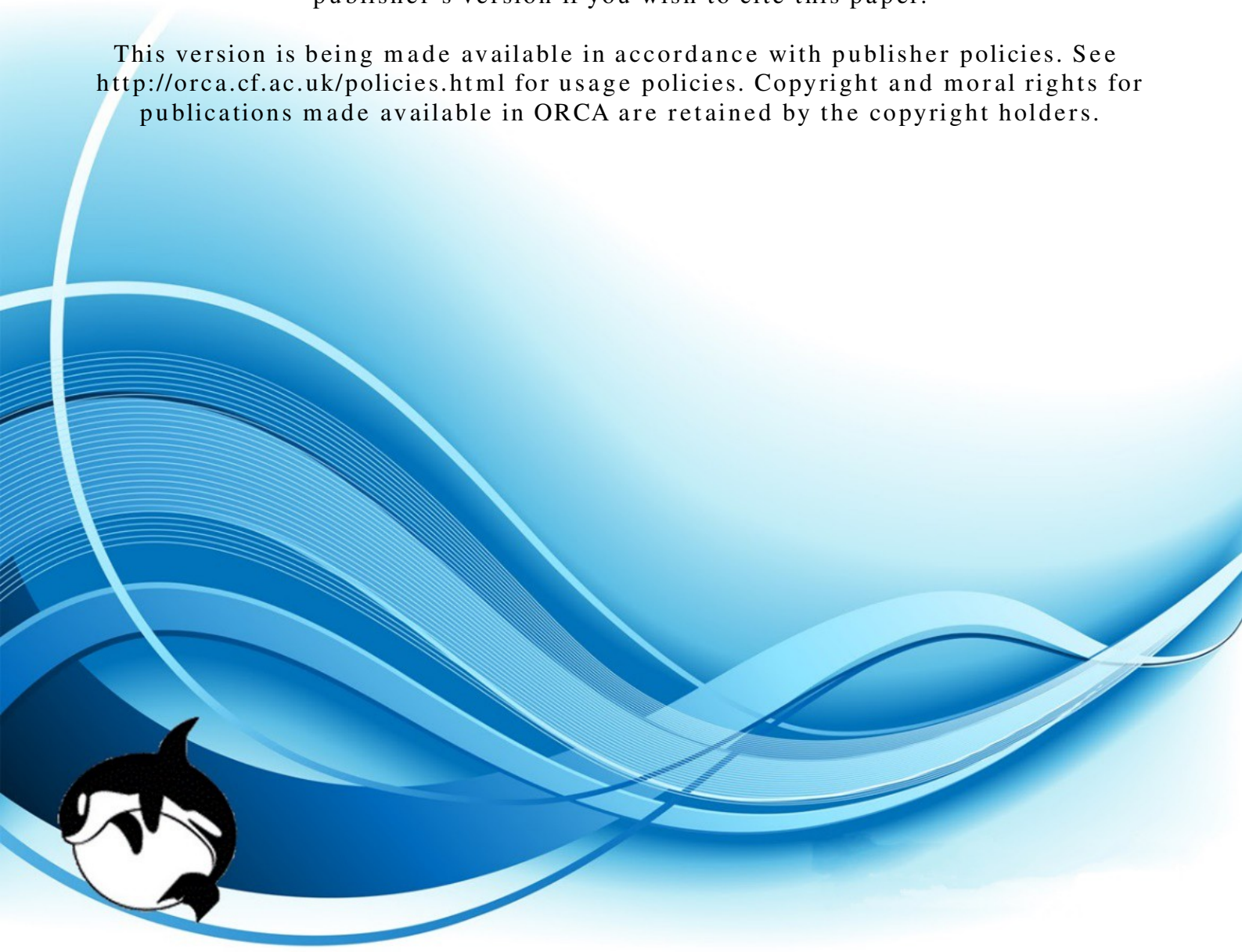

TH
OP

Mendelian Randomization and the Goal of Inferring Causation from Observational Studies in the Vision Sciences

Journal:	*Ophthalmic and Physiological Optics*
Manuscript ID	OPO-IR-2499.R2
Manuscript Type:	Invited Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Plotnikov, Denis; Cardiff University, School of Optometry & Vision Sciences Guggenheim, Jeremy; Cardiff University, School of Optometry & Vision Sciences
Keywords:	Mendelian Randomization, Epidemiology, Randomized Controlled Trial, Observational study

Title:

Mendelian Randomization and the Goal of Inferring Causation from Observational Studies in the Vision Sciences

Running title:
Mendelian Randomization Studies in the Vision Sciences

Authors:
Denis Plotnikov and Jeremy A. Guggenheim

Author affiliation:
School of Optometry & Vision Sciences, Cardiff University, Cardiff, UK

Corresponding author:
Professor Jeremy A. Guggenheim
School of Optometry & Vision Sciences
Cardiff University
Maindy Road,Cardiff, CF24 4HQ, UK
Tel +44 (0) 29 2087 4904
Email. GuggenheimJ1@cardiff.ac.uk

Keywords:
Mendelian Randomization; Epidemiology; Randomized Controlled Trial; Observational study

Conflicts of interest statement:
The authors report no conflicts of interest and have no proprietary interest in any of the materials mentioned in this article.

Acknowledgements:
This work was supported by funding from Cardiff University and the Global Education Program of the Russian Federation government.

Abstract

Purpose: Randomized controlled trials (RCTs) allow reliable causal inferences to be drawn regarding the effectiveness of specific interventions. However, they are expensive to carry out, and not all exposure-outcome relationships can be tested in an RCT framework: for example, it would be unethical to deliberately expose participants to a putative risk factor, or the time-scale involved may be prohibitive. Mendelian randomization (MR) has been proposed as an alternative approach for drawing causal inferences, with the major advantage that the method can often be applied to existing, cross-sectional study datasets. Therefore, results from an MR study can be obtained much more quickly and cheaply than through an RCT.

Recent findings: The validity of causal inferences from an MR study are dependent on two key assumptions, neither of which can be tested fully. Nevertheless, several approaches have been proposed in the last three years that either highlight questionable results, or provide valid causal inference if the necessary assumptions are met only in part. Compared to certain other areas of clinical practice, the ophthalmic research community has been slow to adopt MR.

Summary: An MR study cannot match an RCT in its strength of evidence for a claim of causality. However, MR still has much to offer. In some circumstances, an MR study can provide causal insight into research questions that cannot be addressed by an RCT, while more generally, an MR study can be used to evaluate the supporting evidence before deciding to embark on a lengthy and costly RCT.

INTRODUCTION

Terminology

The glossary section explains the meaning of technical terms frequently encountered in the Mendelian randomization literature, including in this review: assortative mating; collider bias; directional pleiotropy; funnel plot; genetic variant; genome-wide association study (GWAS); horizontal and vertical pleiotropy; instrument strength independent of direct effect (InSIDE) assumption; instrumental variable; weak instrument bias.

What is Mendelian Randomization?

Mendelian randomization (MR) is an epidemiology research method designed to estimate the causal effect of exposure to a putative risk factor on an outcome.[1,2] An example would be a study designed to test whether, and to what degree, additional dietary intake of carotenoids reduces the risk of age-related macular degeneration (AMD). In this example, the exposure of interest is 'additional dietary intake of carotenoids' and the outcome of interest is AMD. In contrast to a randomized controlled trial (RCT), which is generally regarded as the gold-standard research method for drawing causal inferences in epidemiology, MR can be applied to cross-sectional data obtained from observational studies.[3] Thus, whereas addressing a research question by running an RCT requires a considerable investment in time and resources, running an MR study potentially offers a fast and cost-efficient alternative approach that can utilize existing, large-scale cross-sectional datasets.

Standard cross-sectional analyses of observational data have a poor track record of successfully identifying modifiable risk factors,[1] as exemplified by the caveat, '*association does not imply causation*'. A key limitation of standard cross-sectional methods is bias from confounders (a confounder is defined as a variable with causal effects on both the exposure and the outcome). In the carotenoids-AMD example, the list of potential confounders would include factors such as socioeconomic position, level of education, and ethnicity. For instance, wealthier, health-conscious individuals might choose to eat a diet rich in carotenoids while also engaging in other behaviours that reduced their risk of AMD independently of dietary carotenoid intake.[4] In this scenario, an association between (reduced) carotenoid intake and AMD could arise in the absence of a causal relationship. Standard cross-sectional analyses are also susceptible to 'reverse causation', the situation in which an outcome has a causal influence on the exposure of interest. For example,

reverse causation could result in a non-causal association between AMD and dietary carotenoid intake if patients diagnosed with AMD were recommended by their eye-care provider to eat a diet rich in carotenoids.[4] MR is free from bias due to reverse causation, and – as discussed below – is generally less prone to bias from confounders such as socioeconomic position than standard observational analyses.

MR is an example of an 'instrumental variable' analysis method.[5,6] An instrumental variable, also known as an 'instrument', is a variable that meets the following 3 criteria: (1) it is robustly associated with the exposure of interest, (2) it is not associated with confounders of the exposure-outcome relationship, and (3) it is not associated with the outcome except via the exposure. These criteria are most readily understood with reference to a pathway diagram, such as Figure 1, which illustrates causal relationships between variables using arrows (where A $\rightarrow$ B is interpreted as, 'Variable A is a cause of variable B'). To be a valid instrumental variable, coefficient β_1 must be non-zero (criterion #1), coefficient β_5 must be zero (criterion #2), and coefficient β_6 must be zero (criterion #3).

Instrumental variables enable the causal effects of an exposure to be assessed by supplying a 'causal handle' for the exposure of interest that is unrelated to, i.e. statistically independent of, the confounders.[7,8] The effects on the outcome resulting from a change in the level of the exposure variable can therefore be assessed free from the influence of the confounders, and free from the effects of reverse causation. In MR, the instrumental variable is a genetic variant associated with the exposure variable, or a collection of such genetic variants. This concept of using genetic variants to obtain evidence for a causal effect free from reverse causation and confounder bias is generally attributed to Katan[9] (although the necessary data were not available for Katan to test the specific hypothesis he had in mind: that the link between low serum cholesterol and the risk of cancer was non-causal). The term 'Mendelian randomization' was first used[10] by Gray and Wheatley in 1991 when describing the advantages of MR over an RCT study design to assess the efficacy of allogenic bone marrow transplantation vs. 'conventional' therapy. They advocated, and carried out a pilot study, comparing outcomes in children with leukaemia whose sibling were vs. were not compatible (genetically matched) for a bone marrow transplant.

Assumptions inherent to Mendelian Randomization

In order for a genetic variant to be a valid instrumental variable, it must meet all 3 of the criteria listed above. It is straightforward to choose a genetic variant that meets criterion #1 (i.e. that the genetic variant is robustly associated with the exposure) and to test that this assumption is met. Typically, genetic variants identified in a genome-wide association study (GWAS) for the exposure trait that exceed the genome-wide statistical significance threshold (typically $P < 5 \times 10^{-8}$) are chosen for use in MR analyses. Their validity can be confirmed by testing for an association with the exposure in an independent dataset (but see the discussion of 'weak instruments' in the '*Few vs. many genetic variants*' section). By contrast, it is not possible to test fully that criteria #2 and #3 are met. Lack of an association between the genetic variant and *known* confounders can be confirmed, but clearly not all confounders will be known or measurable for many exposure-outcome relationships of interest. Genetic variants with pleiotropic effects (defined as effects on more than one trait) are therefore potentially problematic, since this could mean the variant is not a valid instrumental variable.

Of the various ways of classifying pleiotropy, two are central to the validity of the MR assumptions: horizontal pleiotropy and vertical pleiotropy. A genetic variant that exhibits vertical pleiotropy has a causal relationship with the exposure via a path that is indirect, i.e. a relationship with the exposure variable that is mediated by one or more intermediate trait(s). This genetic variant does satisfy the 3 instrumental variable criteria and therefore can be used to draw valid causal inferences in an MR analysis. In contrast, a genetic variant displaying horizontal pleiotropy exerts effects on the outcome via two or more causal pathways: a pathway via the exposure of interest and at least one pathway acting via another route. Such a variant does not satisfy instrumental variable criterion #3, and therefore would not be valid for use in an MR study.

While much attention has been focused on the issue of pleiotropy in MR studies, the potential for 'collider bias' has rarely been raised[11] (a collider is defined as a variable influenced independently by two or more other variables; collider bias is defined as bias in an exposure-outcome relationship induced by 'conditioning on', or stratifying the sample by, a collider[12]). Commonly encountered reasons for collider bias to occur are selection bias[12] and survivor bias.[13] Regarding the AMD example discussed above, health-conscious individuals may choose to participate in a research study more often than less health-conscious individuals. Likewise, participants with AMD might choose to participate more often than those not affected by AMD. In this scenario, in which both being health-conscious and having AMD are associated with participation, participation is a

collider: therefore, genetic variants associated with being health conscious could produce a biased causal effect estimate in an MR study investigating if a healthy lifestyle influences the risk of AMD. Similarly, the fact that health-conscious individuals will on average live longer than less health-conscious individuals – and will therefore be more likely to suffer AMD during their lifetime – could also potentially introduce bias (for example, if a genetic variant used as an instrumental variable in MR was associated with AMD via an effect on mortality).[13]

Relationship between Mendelian Randomization and Randomized Controlled Trials

In an RCT, random assignment to the intervention or control group has the dual role of modifying the level of the exposure in the intervention group whilst ensuring that levels of confounder variables are balanced between the 2 groups (panel A of Figure 2). The random assortment of alleles during meiosis (Mendel's second law), which holds true for the vast majority of genetic loci, provides an analogy between an RCT and MR.[14,15] In an MR analysis, the assumption is made that the assortment of alleles is independent of levels of the confounder variables, i.e. that assortment is indeed random (panel B of Figure 2). This assumption seems highly plausible; for instance, socioeconomic position would be very unlikely to sway the inheritance of one allele over another. An important exception to this rule is ethnicity: allele frequencies vary widely between populations of differing ancestry or demographic history, therefore alleles associated with an exposure may also be associated with levels of confounder variables – a phenomenon termed 'population stratification'. As an example, individuals from one ethnic group may choose to eat a vegetarian diet that is not only rich in carotenoids but also in a number of other dietary components that may influence the risk of AMD. For this reason, it is essential for MR studies to account for ethnic background in their design. Typically this is done by restricting the analysis to individuals of a single, homogenous, genetically-inferred ancestry group. The results of an MR study will only be relevant to the chosen study population, and hence may not necessarily be applicable more widely. Another potential exception to the random inheritance of specific alleles is assortative mating.[16] For example, if (i) taller individuals tend to choose each other as spouses (single trait assortative mating) and height, education and refractive error have genetic determinants in common, or (ii) myopic individuals are more likely to choose better educated spouses (so-called cross-trait assortative mating), then a Mendelian randomization analysis testing for a causal effect of education on myopia could produce biased results.

There are other important differences between MR and an RCT. The alleles used as instrumental variables in MR usually produce very small changes in the level of the exposure variable, whereas in RCTs the intervention typically has a much larger effect. In order to gauge whether MR results would be clinically meaningful, the results are generally assumed to scale linearly. For example, if a genetic variant imparts a change in exposure level of δx and this is associated with a change of δy in the outcome, then it is assumed that a change in exposure of $100 \times \delta x$ will cause a change in the outcome of $100 \times \delta y$. Another fundamental difference between RCTs and MR is that in an RCT, the intervention is introduced at a specific point during the lifecourse, while in MR the change in exposure imparted via inheritance will have been present from conception. For this reason it can be argued that an MR study can never provide proof that an intervention will succeed in the clinical environment, even if all MR assumptions are fully met.[14,15,17] Thus, it has been suggested that MR studies are well-suited as rapid, inexpensive preliminary tests of novel interventions that can be used to prioritize investment in RCTs.

FUNDAMENTAL METHODOLOGICAL CONSIDERATIONS

Few vs. many genetic variants

MR studies can be performed with just a single genetic variant, with multiple variants, or with a 'genetic risk score' (also known as an 'allele score') calculated by summing the effects of multiple variants. In early MR investigations, the genetic variants chosen as instrumental variables were typically few in number and had known functional relevance to the exposure of interest. For example, to examine the relationship between serum complement factor-H (CFH) levels and AMD, Sharma et al.[18] tested a single genetic variant (rs1061170) within the *CFH* gene coding region, which they suspected to lower serum CFH levels. The rs1061170 variant's alleles, T and C, code for a tyrosine or histidine (amino acid symbol Y and H), respectively, at amino acid 402 of the CFH protein; hence, termed the Y402H polymorphism. Sharma et al.'s MR analysis was carried out under the assumption that the C allele reduced serum CFH levels,[18] however other work suggests this not to be the case.[19,20] By contrast, Cuellar-Partida et al.[21] created a genetic risk score by combining 17,749 genetic variants associated with educational attainment in order to study the causal impact of education on myopia.

It is rarely possible to find more than a handful of genetic variants associated with an exposure whose mechanisms of action have been established. Therefore, a disadvantage of using multiple

genetic variants for an MR analysis (whether combined into a genetic risk score or not) is that the molecular/physiological pathway between the variant and the exposure is typically unknown. This risks at least some of the variants having a horizontally pleiotropic relationship with the outcome and thus biasing the MR causal effect estimate. Balancing this risk is the potentially greater precision that can be obtained from using multiple variants (so long as each variant is robustly associated with the exposure; otherwise weak instrument bias may actually worsen precision). Moreover, using multiple variants provides an opportunity to test for pleiotropic effects (see the '*Sensitivity Analyses and New Directions*' section). Hence, there has been a tendency for recent MR studies to use tens or hundreds of variants.

Combining genetic variants into a genetic risk score[22] protects against 'weak instrument bias'. The latter phenomenon occurs when an MR analysis has insufficient statistical power, i.e. the genetic effect of the instrument variable is too small, given the sample size of the study, to adequately gauge the true causal effect. Crucially, rather than biasing the causal effect estimate towards zero, weak instrument bias in the '1-sample' setting (see the '*One-sample vs. two-sample Mendelian Randomization*' section) biases the causal effect estimate towards that estimated in a standard cross-section analysis. In this situation, an MR result may be given undeserved credence when in reality it is no better than that obtained from a standard, ordinary least squares analysis. The disadvantage of combining genetic variants into a genetic risk score is that they can no longer be used to test for pleiotropy (see the '*Sensitivity Analyses and New Directions*' section). Also, in order to combine information into a genetic risk score the researcher must have access to 'individual level' genetic data (the genotypes of each participant in the sample). Frequently, only 'summary level' data are available for reasons of privacy, which thus rules out the option of conducting a genetic risk score MR analysis. If individual level data are available, there is nothing to stop the investigator performing a genetic risk score MR analysis followed by a multiple variant MR sensitivity analysis.

One-sample vs. two-sample Mendelian Randomization

In 1-sample MR, the association of the genetic variants with both the exposure and the outcome is estimated in a single sample of participants. In a 2-sample MR, the degree of association with the exposure and with the outcome are estimated in different samples.[23] As mentioned above, a key advantage of the 2-sample MR study design is protection against 'weak instrument bias', since in the 2-sample setting lack of statistical power will bias the MR causal estimate towards zero

whereas in the 1-sample setting the causal effect estimate is biased towards the estimate from a standard cross-section analysis. Sample overlap in the 2-sample MR setting provides an intermediate level of protection against weak instrument bias proportional to the degree of overlap.[24]

Another attractive feature of 2-sample MR is that the analysis can be carried out using summary statistics (summary level data) from a GWAS for the exposure of interest and summary statistics from a GWAS for the outcome of interest. These summary statistics datasets, which include regression coefficients and associated standard errors, are often made publicly available by large research consortia who have accrued very large sample sizes. Platforms such as MR-Base[25] facilitate access to these datasets and their integration with state-of-the-art analysis tools.

Sample size and statistical power

Most genetic variants associated with exposure variables have very small effect sizes. This imposes a requirement for extremely large sample sizes in order to gauge the impact of the variants – and thus the exposure – with a trait or disease outcome. Insufficient power will either lead to biased inference of the causal effect, or failure to identify a modestly-sized causal effect (as discussed above). With the advent of large-scale GWAS analyses from samples of hundreds of thousands of participants, lack of statistical power is becoming less of a limitation than in the past. It could be argued that performing studies using very, very large sample sizes will lead to the discovery of statistically significant but biologically meaningless findings. Nevertheless, a counter-argument is that so long as the *effect sizes* of risk factors are reported, not just their associated *P*-values, then the greater precision offered by a very large sample size will be generally be an advantage. Formulae for performing statistical power calculations for MR have been published.[26-28]

SENSITIVITY ANALYSES AND NEW DIRECTIONS

Tests for markers exhibiting horizontal pleiotropy

A number of tests have been proposed for detecting genetic variants with horizontally pleiotropic effects,[29-32] which work under the assumption that variants with unusual variant-exposure and variant-outcome relationships are likely to be pleiotropic. A sensitivity analysis can be performed with these 'outlier' variants excluded. An interesting alternative is Steiger filtering,[33] which identifies (and removes) variants that explain more of the variance in the outcome than the

exposure, under the assumption such variants may have reverse-causal relationships with the outcome and exposure (namely, genetic variant → outcome → exposure).

Care is needed when interpreting the findings from all of the available outlier detection methods, and the related methods described below; for instance, an apparent outlier variant could be the only reliable instrumental variable if in fact all of the remaining variants have pleiotropic effects. Alternatively, even if a full set of genetic variants are valid instrumental variables, a variant with an unusually strong effect could still act as an outlier. See Hemani et al.[34] for an in-depth discussion of these issues.

MR-Egger

The terms 'directional pleiotropy' and 'balanced pleiotropy' refer, respectively, to multiple variant MR analyses in which the weaker variants do or do not have effects biased in one direction. Directional pleiotropy can be visualized in a funnel plot of the causal effect estimate vs. instrumental variable 'strength' relationships[35] or a scatter plot of the variant-outcome vs. variant-exposure regression coefficients[36] (Figure 3). In general, it is difficult to distinguish between bias arising from directional pleiotropy and bias arising from variants with pleiotropic effects on the outcome variable acting through a confounder, i.e. failure of the so-called InSIDE (Instrument Strength Independent of Direct Effect) assumption.

MR-Egger applies the principle of Egger-regression meta-analysis to multiple-variant MR.[35] Specifically, an intercept term is included in the model used to combine and weight the causal effect estimates from the genetic variants. Directional pleiotropy will shift the intercept away from zero while still providing a valid causal effect estimate.[35] This is an informative and commonly-used sensitivity analysis, however the statistical power to detect a causal effect is reduced with MR-Egger compared to a standard, inverse variance weighted (IVW) meta-analysis model for combining information from multiple MR variants.[37]

Median and Mode-based Mendelian Randomization estimates

Following the widespread adoption of MR-Egger, several alternative methods have been proposed for combining information in a multiple variant MR framework in order to reduce the influence of pleiotropy. Bowden et al.[38] introduced the weighted median causal effect estimate, which is valid even if up to 50% of the information in the analysis is from genetic variants with horizontally

pleiotropic effects. Loosely, this can be interpreted as suggesting that a weighted median-based MR causal estimate will be reliable so long as at least half of the variants are valid instrumental variables. Along similar lines, Hartwig et al.[39] proposed a mode-based estimator (MBE), which can potentially provide a reliable causal effect estimate even if the majority of instrumental variables are invalid because of pleiotropy. Both approaches are useful sensitivity analyses: caution is needed when interpreting findings if the IVW, weighted-median, and MBE estimates differ widely.

Multivariable Mendelian Randomization

Distinct from the use of multiple genetic variants to gauge the effect of a *single* exposure, multivariable MR employs multiple genetic variants to gauge the effects of an exposure *while accounting for pleiotropic effects on one or more additional, specified exposures*. To date, multivariable MR has been adopted most often in studies examining the risks conferred by different lipid traits.[30,40-42] High-density lipoprotein (HDL), low-density lipoprotein (LDL) and triglyceride levels in the blood are influenced by many genetic variants, some of which have pleiotropic effects on more than one lipid fraction. This makes single-exposure (univariable) MR studies of lipid traits difficult to interpret. However, by accounting for the effects of genetic variants on all three lipid traits simultaneously, multivariable MR has been used to disentangle the causal effect of triglycerides, HDL and LDL on AMD.[40,43] MR-Egger can also be applied in the multivariable setting.[44]

New directions

The increasing popularity of MR has been accompanied by several innovative developments in recent months (reviewed by Zheng et al.[45]). MR is being applied on a genome-wide scale to leverage causal information from transcriptomics and epigenomics datasets.[29,46-49] These approaches can help determine the genes and pathways through which GWAS variants exert their effects, which is an important goal in genetics.

Methodological advances such as genetic instrumental variable (GIV) regression[50] and mixture-of-experts (MR-MoE) machine learning[33] offer improved frameworks for drawing causal inferences. GIV regression utilizes summary statistics from two independent GWAS analyses for the outcome variable (or a split-sample GWAS for the outcome) so that genetic risk scores obtained from one dataset can be used as an instrumental variable for an MR analysis in the other dataset,[50] and vice versa. This idea builds on an existing approach used to correct for measurement error.[5] As a result, GIV regression has the potential to provide causal estimates free from bias due to horizontal

pleiotropy. MR-MoE provides a standardized approach for choosing from the myriad of available MR analysis methods (IVW, MR-Egger, weighted-median, MBE, etc.) the one that is most appropriate for a given situation.[33] The approach works by categorizing features of the summary statistics of 2-sample MR analyses best suited to, say, an IVW MR analysis, and then applying the IVW method to subsequent datasets that match these features. To 'train' the MoE model, the authors simulated 2-sample MR datasets and used a random forest classifier to select the analysis method that provided the optimum trade-off between statistical power and bias from pleiotropic instrumental variables.

Finally, Staley and Burgess[51] have described two MR methods for assessing the 'shape' of the exposure-outcome causal relationship. Both approaches require access to individual level data. In an applied example, the methods were used to provide evidence of non-linearity in the causal relationship between body mass index (BMI) and blood pressure. A progressively higher BMI was found to cause progressively higher blood pressure across most of the BMI distribution, yet the relationship plateaued or reversed in hyper-obese individuals.[51]

Further guidance on the interpretation of Mendelian Randomization studies

We highly recommend a recent BMJ article by Davies et al.[52] for clinicians interested in learning more about how to interpret the strengths and weaknesses of published MR studies. This article provides guidance on how the plausibility of the MR assumptions in a published study can be gauged, since this is a key determinant of the weight of evidence of an MR study compared to other epidemiological approaches. In keeping with moves to standardise the reporting and interpretation of RCTs (e.g. CONSORT[53], CASP), the authors provide a 'critical appraisal checklist for evaluating MR studies'.

Review of Mendelian Randomization studies in the vision sciences

We conducted a literature search of PubMed and Web of Science to identify studies applying Mendelian randomization to study risk factors for eye disorders. The search was restricted to articles written in English and published in peer-reviewed journals between 2008 and 2018. The search strategy is described in the Appendix. Only 8 studies were identified (Table 1).

Three of the ophthalmic MR studies we found addressed research questions relating to myopia.[21,54,55] In the most recent of these, a UK research team tested the hypothesis that education

has a causal effect on myopia development. The results supported the hypothesis, confirming a similar conclusion from a smaller scale study carried out 2 years earlier.[21] The other MR study examining risk factors for myopia[55] provided evidence refuting a causal role for (low) serum vitamin D level in myopia development. This result implied that the association between serum vitamin D and refractive error observed in several cross-sectional epidemiology studies[56-63] is non-causal, most likely mediated by the time individuals spend outdoors.

There were also 3 ophthalmic MR studies addressing research questions related to AMD.[18,40,43] Two of the publications estimated the effect of plasma lipid levels on AMD, with both finding evidence of an effect of HDL cholesterol, but not for LDL cholesterol or triglycerides.[40,43] As mentioned above, the other AMD-related study used MR to assess whether a low serum complement factor H (CFH) level predisposes individuals to AMD. The result was inconclusive, perhaps due to the use of only a single genetic variant as an instrumental variable.[18] Notably, the latter study was published in 2013, whereas the remaining studies were all published in during the period 2016-2018.

A single study investigated primary open-angle glaucoma (POAG) as an outcome.[64] Shen et al. found strong evidence from their MR analyses to support observational evidence that individuals with type-2 diabetes (T2D) are at an increased risk of glaucoma. Notably, Shen et al. carried out a series of separate MR analyses using allele scores designed to investigate the causal effects of specific mechanisms implicated in T2D pathogenesis (adiposity, β-cell function, insulin regulation, and other metabolic processes) as well as a non-mechanism-specific, T2D allele score analysis. One reason why pathway-specific MR analyses such as this are not common in the literature is that they can be difficult to interpret, e.g. if genetic variants have pleiotropic effects on more than one disease mechanism; a problem analogous to the difficulty of inferring the causal effects of individual lipid traits using univariable rather than multivariable MR.

The final ophthalmic MR study that we identified evaluated the risk associated with plasma HDL cholesterol, LDL cholesterol and triglycerides, on the incidence of diabetic retinopathy (DR).[65] None of the 3 lipid fractions was found to be causally associated with the risk of DR, either when the outcome was 'any DR' or 'severe DR'. However, the authors were careful to point out that the study had limited statistical power to detect subtle causal risks, since the GWAS sample size used to obtain genetic effect estimates for association with DR was relatively small (2,969 cases and 4,096 controls).

In summary, despite the increasing adoption of MR in fields of health research such as cardiology and rheumatology, the number of MR studies applying this approach to identify and to estimate the causal effect of risk factors for eye diseases remains limited. To date, the main ophthalmic-related outcomes of interest for researchers are myopia and AMD (Table 1).

CONCLUSIONS

GWAS summary statistics for a wide range of potential risk factors, analysed in samples of tens or hundreds of thousands of participants, are publically available. These summary statistics provide an excellent resource for identifying instrumental variables for use in MR. GWAS summary statistics are also available for several ophthalmic traits, including refractive error, diabetic retinopathy, intra-ocular pressure, glaucoma and cataract. Together, these resources can be harnessed to carry out 2-sample MR analyses for addressing a wide range of epidemiological research questions, facilitated by platforms such as MR-Base. Although the ophthalmic research community has been relatively slow to adopt MR compared to some disciplines, the approach offers significant potential for independently supporting and clarifying causal relationships inferred from observational studies, and for prioritizing investment in RCTs.

References

1. Ebrahim S & Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet* 2008; 123: 15-33.

2. Lawlor DA, Harbord RM, Sterne JAC, Timpson N & Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stats Med* 2008; 27: 1133-1163.

3. Davey Smith G & Ebrahim S. What can Mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ* 2005; 330: 1076-1079.

4. Meyers KJ, Mares JA, Igo JRP, *et al.* Genetic evidence for role of carotenoids in age-related macular degeneration in the carotenoids in age-related eye disease study (CAREDS). *Invest Ophthalmol Vis Sci* 2014; 55: 587-599.

5. Angrist JD & Pischke J-S. Mostly harmless econometrics: an empiricist's companion. Princeton: Princeton University Press; 2009.

6. Hernan MA & Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiol* 2006; 17: 360-372.

7. Davey Smith G & Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014; 23: R89-R98.

8. Hemani G, Tilling K & Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* 2017; 13: e1007081.

9. Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet* 1986; 1: 507-508.

10. Gray R & Wheatley K. How to avoid bias when comparing bone marrow transplantation with chemotherapy. *Bone Marrow Transplant* 1991; 7 (Suppl 3): 9-12.

11. Munafo MR, Tilling K, Taylor AE, Evans DM & Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol* 2017; 47: 226–235.

12. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiol* 2003; 14: 300-306.

13. Vansteelandt S, Dukes O & Martinussen T. Survivor bias in Mendelian randomization analysis. *Biostatistics* 2017: kxx050.

14. 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.

15. Swanson SA, Tiemeier H, Ikram MA & Hernan MA. Nature as a trialist? Deconstructing the analogy between Mendelian randomization and randomized trials. *Epidemiol* 2017; 28: 653-659.

16. Hartwig FP, Davies NM & Davey Smith G. Bias in Mendelian randomization due to assortative mating. *Genet Epidemiol* 2018: 42; 608-620.

17. Burgess S & Malarstig A. Using Mendelian randomization to assess and develop clinical interventions: Limitations and benefits. *J Comp Eff Res* 2013; 2: 209-212.

18. Sharma NK, Gupta A, Prabhakar S*, et al.* Association between CFH Y402H polymorphism and age related macular degeneration in north Indian cohort. *PLoS ONE* 2013; 8: e70193.

19. Smailhodzic D, Klaver CCW, Klevering BJ*, et al.* Risk alleles in CFH and ARMS2 are independently associated with systemic complement activation in age-related macular degeneration. *Ophthalmology* 2012; 119: 339-346.

20. Silva AS, Teixeira AG, Bavia L*, et al.* Plasma levels of complement proteins from the alternative pathway in patients with age-related macular degeneration are independent of complement factor H Tyr(402)His polymorphism. *Mol Vision* 2012; 18: 2288-2299.

21. Cuellar-Partida G, Lu Y, Kho PF*, et al.* Assessing the genetic predisposition of education on myopia: a Mendelian randomization study. *Genet Epidemiol* 2016; 40: 66-72.

22. Burgess S, Butterworth A & Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013; 37: 658-665.

23. Pierce BL & Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol* 2013; 178: 1177-1184.

24. Burgess S, Davies NM & Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* 2016; 40: 597-608.

25. Hemani G, Zheng J, Elsworth B*, et al.* The MR-base platform supports systematic causal inference across the human phenome. *eLife* 2018; 7: e34408.

26. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int J Epidemiol* 2014; 43: 922-929.

27. Freeman G, Cowling BJ & Schooling CM. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *Int J Epidemiol* 2013; 42: 1157-1163.

28. Brion M-JA, Shakhbazov K & Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol* 2013; 42: 1497-1501.

29. Zhu Z, Zhang F, Hu H*, et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 2016; 48: 481-487.

30. 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.

31. 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: 693-698.

32. Corbin LJ, Richmond RC, Wade KH*, et al.* Body mass index as a modifiable risk factor for type 2 diabetes: refining and understanding causal estimates using Mendelian randomisation. *Diabetes* 2016; 65; 3002–3007.

33. Hemani G, Bowden J, Haycock PC*, et al.* Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. *bioRxiv* 2017: 173682.

34. Hemani G, Bowden J & Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet* 2018; 27(R2): R195-R208.

35. Bowden J, Smith GD & Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015; 44: 512-525.

36. Burgess S & Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol* 2017; 32: 377-389.

37. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N & Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stats Med* 2017; 36: 1783-1802.

38. Bowden J, Davey Smith G, Haycock PC & Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016; 40: 304-314.

39. Hartwig FP, Davey Smith G & Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017; 46: 1985-1998.

40. Burgess S & Davey Smith G. Mendelian randomization implicates high-density lipoprotein cholesterol-associated mechanisms in etiology of age-related macular degeneration. *Ophthalmology* 2017; 124: 1165-1174.

41. Burgess S & Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 2015; 181: 251-260.

42. 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.

43. Fan Q, Maranville JC, Fritsche L, *et al.* HDL-cholesterol levels and risk of age-related macular degeneration: a multiethnic genetic study using Mendelian randomization. *Int J Epidemiol* 2017: dyx189-dyx189.

44. Rees JMB, Wood AM & Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stats Med* 2017; 36: 4705-4718.

45. Zheng J, Baird D, Borges MC, *et al.* Recent developments in Mendelian randomization studies. *Curr Epidemiol Rep* 2017; 4: 330-345.

46. 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: 161-176.

47. Richmond RC, Hemani G, Tilling K, Davey Smith G & Relton CL. Challenges and novel approaches for investigating molecular mediation. *Hum Mol Genet* 2016; 25: R149-R156.

48. Richardson TG, Zheng J, Davey Smith G, *et al.* Mendelian randomization analysis identifies CpG sites as putative mediators for genetic influences on cardiovascular disease risk. *Am J Hum Genet* 2017; 101: 590-602.

49. Richardson TG, Haycock PC, Zheng J, *et al.* Systematic Mendelian randomization framework elucidates hundreds of CpG sites which may mediate the influence of genetic variants on disease. *Hum Mol Genet* 2018: 27; 3293-3304.

50. DiPrete TA, Burik CAP & Koellinger PD. Genetic instrumental variable regression: explaining socioeconomic and health outcomes in nonexperimental data. *Proc Natl Acad Sci USA* 2018; 115: E4970-E4979.

51. Staley JR & Burgess S. Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. *Genet Epidemiol* 2017; 41: 341-352.

52. Davies NM, Holmes MV & Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 2018; 362: k601.

53. Moher D, Schulz KF & Altman DG. The consort statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Medical Research Methodology* 2001; 1: 2.

54. Mountjoy E, Davies NM, Plotnikov D*, et al.* Education and myopia: assessing the direction of causality by Mendelian randomisation. *BMJ* 2018; 361: k2022.

55. Cuellar-Partida G, Williams KM, Yazar S*, et al.* Genetically low vitamin D concentrations and myopic refractive error: a Mendelian randomization study. *Int J Epidemiol* 2017; 46: 1882–1890.

56. Choi JA, Han K, Park YM & La TY. Low serum 25-hydroxyvitamin D is associated with myopia in Korean adolescents. *Invest Ophthalmol Vis Sci* 2014; 55: 2041-2047.

57. Kwon JW, Choi JA & La TY. Serum 25-hydroxyvitamin D level is associated with myopia in the Korea National Health and Nutrition Examination Survey. *Medicine* 2016; 95: e5012.

58. Guggenheim JA, Williams C, Northstone K*, et al.* Does vitamin D mediate the protective effects of time outdoors on myopia? Findings from a prospective birth cohort. *Invest Ophthalmol Vis Sci* 2014; 55: 8550–8558.

59. Mutti DO & Marks AR. Blood levels of vitamin D in teens and young adults with myopia. *Optom Vis Sci* 2011; 88: 377-382.

60. Pan CW, Qian DJ & Saw SM. Time outdoors, blood vitamin D status and myopia: a review. *Photochem Photobiol Sci* 2016; 16: 426-432.

61. Tideman JW, Polling JR, Voortman T*, et al.* Low serum vitamin D is associated with axial length and risk of myopia in young children. *Eur J Epidemiol* 2016; 31: 491-499.

62. Williams KM, Bentham GC, Young IS*, et al.* Association between myopia, ultraviolet B radiation exposure, serum vitamin D concentrations, and genetic polymorphisms in vitamin D metabolic pathways in a multicountry European study. *JAMA Ophthalmol* 2017; 135: 47-53.

63. Yazar S, Hewitt AW, Black LJ*, et al.* Myopia is associated with lower vitamin D status in young adults. *Invest Ophthalmol Vis Sci* 2014; 55: 4552-4559.

64. Shen L, Walter S, Melles RB, Glymour MM & Jorgenson E. Diabetes pathology and risk of primary open-angle glaucoma: evaluating causal mechanisms by using genetic information. *Am J Epidemiol* 2016; 183: 147-155.

65. Sobrin L, Chong YH, Fan Q, *et al.* Genetically determined plasma lipid levels and risk of diabetic retinopathy: a Mendelian randomization study. *Diabetes* 2017; 66: 3130-3141.

66. Lambert JC, Ibrahim-Verbaas CA, Harold D, *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; 45: 1452-1458.

67. Laeng B, Mathisen R & Johnsen J-A. Why do blue-eyed men prefer women with the same eye color? *Behavioral Ecol Sociobiol* 2007; 61: 371-384.

68. Nordsletten AE, Larsson H, Crowley JJ, Almqvist C, Lichtenstein P & Mataix-Cols D. Patterns of nonrandom mating within and across 11 major psychiatric disorders. *JAMA Psychiatry* 2016; 73: 354-361.

69. Chakravarthy U, Wong TY, Fletcher A, *et al.* Clinical risk factors for age-related macular degeneration: a systematic review and meta-analysis. *BMC Ophthalmol* 2010; 10: 31.

70. Burgess S, Bowden J, Fall T, Ingelsson E & Thompson SG. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. *Epidemiol* 2017; 28: 30-42.

71. Angrist JD. Lifetime earnings and the vietnam era draft lottery: evidence from social security administrative records. *Am Econ Rev* 1990; 80: 313-336.

72. Davies NM, Scholder SvHK, Farbmacher H, Burgess S, Windmeijer F & Davey Smith G. The many weak instruments problem and Mendelian randomization. *Stats Med* 2015; 34: 454-468.

73. Burgess S, Thompson SG & CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* 2011; 40: 755-764.

Table 1. Mendelian randomization studies examining the effects of specific exposures on ophthalmic traits.

Study	Exposure/ risk factor	Outcome	Instrumental variable(s)	Findings
Sharma et al. (2013) [18]	Complement factor H level	AMD	1 SNP	Causal relationship
Burgess et al. (2017) [40]	Plasma lipid levels	AMD	185 SNPs	Causal effect of HDL Cholesterol
Fan et al. (2017) [43]	Plasma lipid levels	AMD	185 SNPs	Causal effect of HDL Cholesterol
Cuellar-Partida et al. (2016) [21]	Educational attainment	Myopia	Allele score	Causal relationship
Cuellar-Partida et al. (2017) [55]	Serum vitamin D level	Myopia	6 SNPs	No causal relationship
Mountjoy et al. (2018) [54]	Educational attainment	Myopia	Allele score	Causal relationship
Shen et al. (2016) [64]	Type 2 diabetes	Glaucoma	Allele score	Causal relationship
Sobrin et al. (2017) [65]	Plasma lipid levels	Diabetic retinopathy	157 SNPs	No causal relationship

Figure 1. Properties of an instrumental variable. Arrows depict causal relationships amongst variables, with solid arrows denoting known or strongly-suspected relationships and dashed arrows indicating putative relationships. Beta coefficients represent the strength of the causal relationships. The parameter of primary research interest is coefficient β2, which gauges the causal effect of the exposure on the outcome.

Figure 2. Analogy between a randomized controlled trial (RCT) and a Mendelian randomization (MR) analysis. Panel A: In an RCT, randomization serves, firstly, to cause an increase in the level of the exposure in the intervention group relative to the control group. Secondly, randomization serves to balance the levels of both known and unknown confounders between the intervention and control groups. Panel B: In an MR analysis, random assortment of alleles at meiosis creates the setting for a 'natural experiment' in which some individuals are genetically-predisposed to a higher level of the exposure than others. If the assortment of alleles during meiosis is not influenced by known or unknown confounders of the exposure-outcome relationship, then levels of these confounders will be balanced between the 2 groups (i.e. those with and without a genetic predisposition due to the genetic variant of interest).

Figure 3. MR sensitivity analyses. Panel A: Scatter plot of single nucleotide polymorphism (SNP) genetic variant regression coefficients quantifying the level of association with the exposure (Alzheimer's disease; x-axis) and with the outcome (self-reported glaucoma; y-axis) in an MR analysis. The solid blue line represents the Inverse Variance Weighted (IVW) and the dashed green line the MR-Egger methods of combining information across variants. A possible outlier variant is shown in red. Error bars indicate 1 standard error (SE). Panel B: Funnel plot for the same MR analysis shown in A. Each data point represents a genetic variant. The possible outlier variant plotted in red in panel A is also plotted in red in panel B. Data for these plots were obtained from MR-Base,[25] for the traits "UKB-a:79" (self-reported glaucoma in UK Biobank) and "#298" (Alzheimer's disease[66]). The MR-Egger analysis suggests minimal evidence of directional pleiotropy, and both the IVW and MR-Egger analyses suggest negligible causal impact of Alzheimer's disease on self-reported glaucoma.

Appendix: Literature search methodology

PubMed (https://www.ncbi.nlm.nih.gov/pubmed)

Publication date: from 01/01/2008 to 15/07/2018

Search query:

("Mendelian Randomization Analysis"[Mesh] OR "Mendelian randomisation"[all fields] OR "Mendelian randomization"[all fields] OR (Mendelian[all fields] AND (("Mendelian Randomization Analysis"[Mesh] OR "Mendelian randomisation"[all fields] OR "Mendelian randomization"[all fields] OR (Mendelian[all fields] AND randomi*[all fields]) OR "genetic instrumental variable"[all fields] OR "genetic instrumental variables"[all fields] OR "genetic instrument"[all fields] OR "genetic instruments"[all fields] OR "genes as instruments"[all fields] OR "gene as instrument"[all fields] OR "genes as instrument"[all fields] OR "gene as instruments"[all fields] OR (instrument*[ti] AND (gene[ti] OR genes[ti] OR genetic*[ti] OR mendel*[ti])) OR (("instrumental variable"[all fields] OR "instrumental variables"[all fields] OR "instrumented analysis"[all fields] OR "instrumented analyses"[all fields] OR "instrumental variable analysis"[all fields] OR "instrumental variable analyses"[all fields] OR "instrumental variables analysis"[all fields] OR "instrumental variables analyses"[all fields]) AND (gene OR genes OR genetics OR mendel OR mendelian)) OR ("mendelian"[all fields] AND ("randomisation"[all fields] OR "randomization"[all fields] OR "randomising"[all fields] OR "randomizing"[all fields])))) AND ("myopi*" OR "eye" OR "ophthalm*" OR "AMD" OR"macula*" OR"retin* " OR"glauc*" OR"refract*")

Web of Science (http://apps.webofknowledge.com)

Publication date: from 2008 to 2018

Language: English

Document types: Article

Search query:

TI=((("Mendelian randomisation" OR "Mendelian randomization" OR "genetic instrumental variable" OR "genetic instrumental variables" OR "genetic instrument" OR "genetic instruments" OR "mendel randomise" OR "mendel randomize" OR "mendel randomization" OR "mendel randomisation" OR "random Mendelian" OR "genes as instruments" OR "gene as instrument" OR "genes as instrument" OR "gene as instruments" OR "instrumental genetic variable" OR "instrumental genetic variable")) AND("myopi*" OR "eye" OR "ophthalm*" OR "AMD" OR"macula*" OR"retin*" OR"glauc*" OR"refract*")))

Glossary: Terminology in Mendelian randomization studies

Assortative mating

Definition. *Assortative mating refers to an individual's choice of mate (spouse) being non-random. Positive and negative assortative mating refers to mate selection on the basis of similarity or dissimilarity for particular trait(s) of interest. Single trait assortative mating describes mate choice on the basis of just one trait, while cross-trait assortative mating occurs when individuals with a certain level of one trait choose mates with a certain level of another trait. Assortative mating has the potential to bias the causal estimate from a MR analysis even if the exposure and outcome are not directly subject to assortative mating.[16] Hartwig et al.[16] outline a method for correcting bias due to assortative mating using data from family members.*

Example. It has been suggested that assortative mating occurs for eye colour (an example of single trait assortative mating)[67] and that cross-trait assortative mating is common across a range of psychiatric conditions.[68]

Collider bias (also known as **collider stratification bias**)

Definition. *A collider is a variable that is affected by two or more downstream variables. Stratifying an analysis on the basis of a collider (or adjusting for a collider in a regression analysis) can introduce an entirely spurious association (or create a systematically over-estimated or under-estimated degree of association) between the downstream causal variables.*

Example. UV exposure and hyperopia are both risk factors for AMD.[69] Hence, an analysis of patients with AMD (i.e. stratifying on AMS status) would risk identifying a purely spurious association between UV exposure and hyperopia.

[Figure 4 about here]

Directional pleiotropy (also known as **unbalanced pleiotropy**)

Definition. *The occurrence of horizontal pleiotropy in which the effects of genetic variants acting via confounding trait(s) are <u>not</u> balanced with respect to size and direction, i.e. either outcome-increasing or outcome-decreasing horizontally pleiotropic effects predominate. The MR-EGGER intercept test can be used to test for directional pleiotropy: under the null hypothesis of balanced pleiotropy, the intercept from an MR-EGGER analysis will be zero. The slope from an MR-EGGER analysis provides a valid causal effect estimate in the presence of directional pleiotropy, whereas a standard (inverse variance-weighted) causal effect estimate will be biased.*

Example. The scatterplots show simulated data for a Mendelian randomization analysis, with SNP-exposure and SNP-outcome effect sizes plotted on the x-axis and y-axis, respectively. In the plot on the left, the data were fitted using an inverse variance-weighted Mendelian randomization model, i.e. with the intercept constrained to zero (red dashed line). The steep slope of this line suggests a large causal effect estimate. In the plot on the right, the data were fitted using MR-EGGER. The black dotted line indicates the MR-EGGER intercept (the weighted mean SNP-outcome effect size). The shallow slope of the MR-EGGER regression line suggests a small causal effect estimate. A parsimonious interpretation is that the non-zero MR-EGGER intercept results from directional pleiotropy, and that the small causal effect estimate from the MR-EGGER analysis is better supported than the large causal effect estimate from the inverse variance-weighted analysis.

[Figure 5 about here]

Funnel plot

Definition. *A funnel plot is a scatterplot of effect size (x-axis) versus precision (y-axis). If the distribution of points is asymmetric with respect to the average effect size, this may indicate a source of bias. Funnel plots are commonly used to test for publication bias (in which an asymmetric distribution may indicate bias towards publishing positive findings while not publishing negative findings). In Mendelian randomization, the data points of a funnel plot correspond to the causal effect estimate (x-axis) versus a measure of the genetic variant's expected precision, e.g. the reciprocal of a genetic variant's standard error for association with the outcome.*[70] *Asymmetry in a Mendelian randomization funnel plot may indicate a departure from instrumental variable criteria #2 or #3, most likely due to horizontal pleiotropy, and thus suggest that the causal effect estimate is biased..*

Example. Mendelian randomization funnel plots with symmetric (left) and asymmetric (right) profiles.

[Figure 6 about here]

Genetic variant (also known as a **DNA sequence polymorphism**)

Definition. *A genetic variant is a difference in DNA sequence between individuals in a population at a specific position in the genome. The most common type is a single base difference, called a single nucleotide polymorphism (SNP). Other types of genetic variant include 'indels' (the insertion or deletion of one or more bases), microsatellite repeat polymorphisms (differences in the number of a repeating series of bases) and large structural rearrangements. The vast majority of genetic variants used in Mendelian randomization studies are SNPs, since they are common in the population, and inexpensive and accurate to determine (a process known as, 'genotyping').*

Example. Schematic diagram of a region of a chromosome containing a genetic variant. Individuals each carry two copies of the chromosome. Individual #1 is homozygous for the C nucleotide while individual #2 is heterozygous. Hydrogen bonds between bases of the two strands of the DNA double helix are indicated (= and ≡).

[Figure 7 about here]

GWAS (Genome-wide association study)

Definition. *A GWAS is a systematic search through the genome for genetic variants associated with a trait of interest. Each genetic variant is tested in turn, typically using logistic regression for case/control traits and using linear regression for quantitative traits. Because several million genetic variants are tested in a GWAS, the threshold chosen for declaring 'genome-wide statistical significance' is very stringent, e.g. $P < 5 \times 10^{-8}$. The full GWAS results (so called 'summary statistics') for a wide variety of potential exposure and outcome traits have been made freely available for download. Genetic variants identified in GWAS analyses are a source of potential instrumental variables for MR studies. Furthermore, in '2-sample MR' study designs (in which separate samples of participants are used to quantify the genetic variant-exposure and the genetic variant-outcome relationships) all of the information required for the MR analysis can be obtained from GWAS summary statistics. The MR Base website[25], has collected together information from available GWAS summary statistics to facilitate 2-sample MR analyses.*

Horizontal pleiotropy and vertical pleiotropy

Definition. *A genetic variant that has effects on more than one trait is said to exhibit pleiotropy. Of the various types of pleiotropy, horizontal and vertical pleiotropy are the forms most relevant to Mendelian randomization. In this context, a horizontally pleiotropic genetic variant has <u>independent</u> effects on both the exposure and at least one other trait that directly or indirectly influences the outcome. This invalidates a key instrumental variable requirement, namely, that the genetic variant*

influences the outcome only via the exposure (criterion #3). In the context of Mendelian randomization, a vertically pleiotropic genetic variant has non-independent effects on both the exposure and at least one other trait that directly or indirectly influences the outcome. Instrumental variable criterion #3 still holds for such a genetic variant.

Example. In a Mendelian randomization analysis designed to test for a causal role of carotenoid levels in protecting against AMD, a genetic variant that exerts an effect on carotenoid levels via education – an example of vertically pleiotropy – would be a valid instrumental variable (pathway diagram A). In contrast, a genetic variant with independent effects on carotenoid levels and education – an example of horizontal pleiotropy – would not be a valid instrumental variable since it will influence AMD risk via both a change in education and a change in carotenoid level, making it impossible to determine the role of carotenoids alone.

[Figure 8 about here]

InSIDE (Instrument Strength Independent of Direct Effect) assumption

Definition. *The InSIDE assumption posits that, for the set of genetic variants used in a Mendelian randomization analysis, the variants' effects on the exposure are not correlated with their effects on other horizontally pleiotropic trait(s). The InSIDE assumption must be satisfied in order for the MR-EGGER test to be a valid test for directional pleiotropy. Burgess et al.[70] reason that the InSIDE assumption is more likely to be violated if the set of genetic variants' pleiotropic effects act via a single confounder. With reference to Figure 1, the InSIDE assumption defines that the $\beta 1$ coefficients for a set of variants are uncorrelated with their $\beta 5$ and $\beta 6$ coefficients.*

Instrumental variable

Definition. *An instrumental variable is a variable that meets the following three criteria: (1) it is robustly associated with the exposure of interest, (2) it is not associated with confounders of the exposure-outcome relationship, and (3) it is not associated with the outcome except via the exposure (see Figure 1). Since an instrumental variable is not associated with confounders (criterion #2) it can be used to gauge the impact of an exposure free from the confounder bias typically present in*

observational studies. Furthermore, when genetic variants are used as instrumental variables, the risk of reverse causation is usually negligible, since it is much more likely that a genetic variant will influence an outcome via its effects on the exposure, than that an outcome will have altered an individual's genotype.

Example. Instrumental variables are widely used in econometrics. For example, Angrist[71] used assignment into the United States armed forces by the Vietnam-era draft lottery as an instrumental variable to estimate the effects of military service on earnings in later civilian life. The draft lottery assigned individuals into military service at random and therefore would have been free from the influence of the usual confounders (socio-economic position, parental military service, etc.) that would otherwise bias estimates of the effect of military service on earnings.

Weak instrument bias

Definition. *Instrumental variables that are only weakly associated with the exposure (i.e. not satisfying instrumental variable criteria #1) will bias causal effect estimates.[72] In a 1-sample Mendelian randomization analysis (i.e. the same sample of participants is used to determine both the genetic variant-exposure and genetic variant-outcome effects) weak instrument bias will be in the direction of the observational association between exposure and outcome. In a 2-sample Mendelian randomization analysis (i.e. different samples of participants are used to determine the genetic variant-exposure and genetic variant-outcome effects) weak instrument bias will be towards the null. Selecting genetic variants that attain genome-wide significance in a GWAS for the exposure as instrumental variables and performing the Mendelian randomization in a sufficiently large sample[26-28] will minimize the risk of weak instrument bias. A commonly used approach to examine the strength of an instrumental variable is to confirm that the F-statistic (Cragg-Donald F-statistic) from a variant-exposure regression model is at least 10, although such an approach does not guarantee against weak instrument bias.[73]*

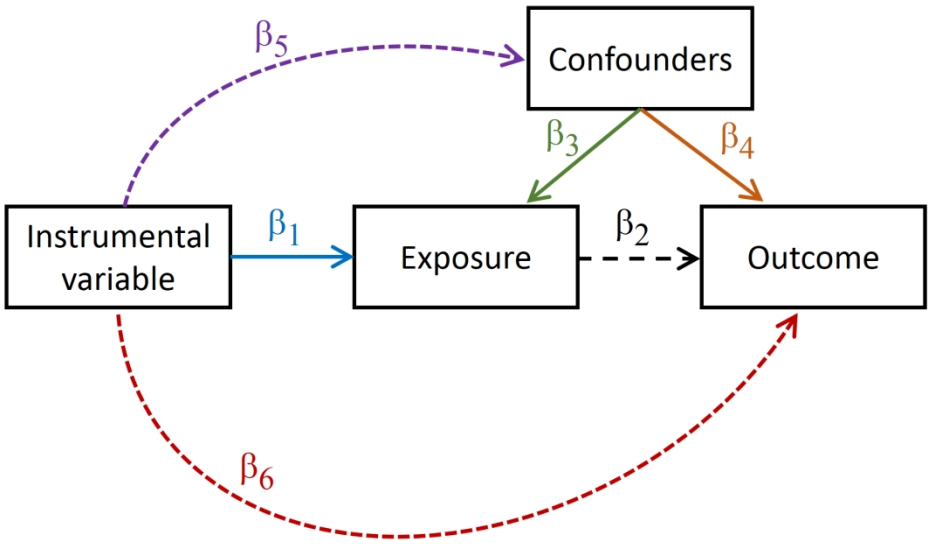

Figure 1

178x107mm (300 x 300 DPI)

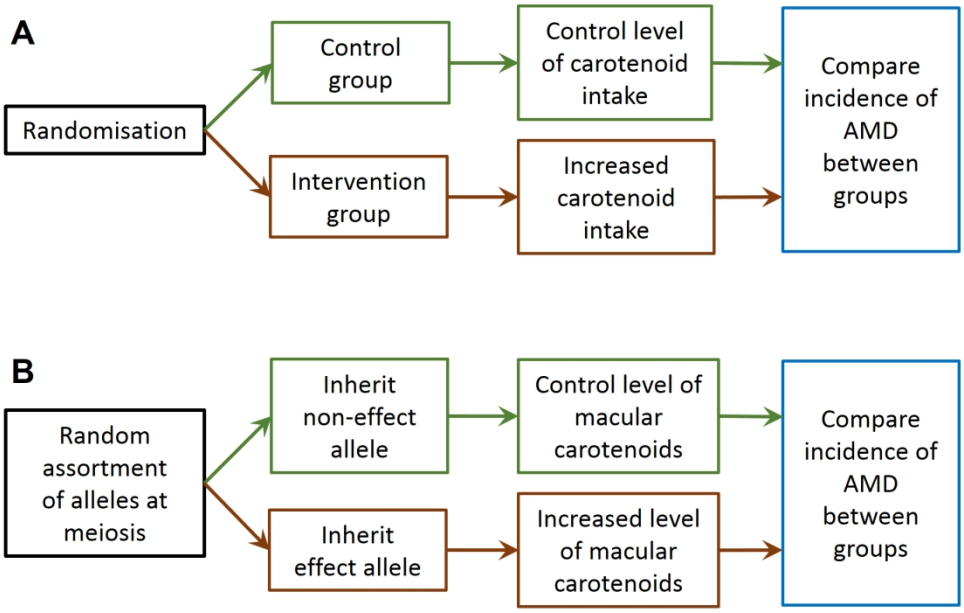

Figure 2

170x110mm (300 x 300 DPI)

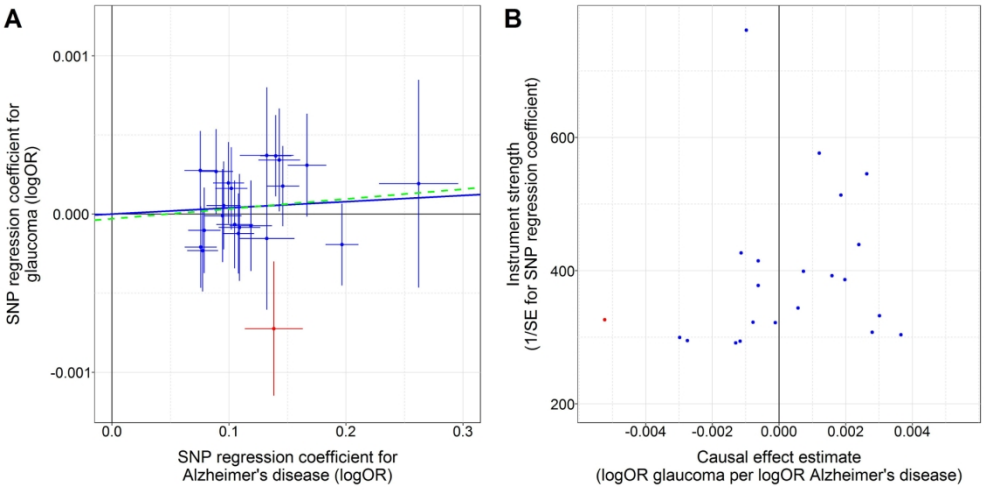

Figure 3

170x85mm (300 x 300 DPI)

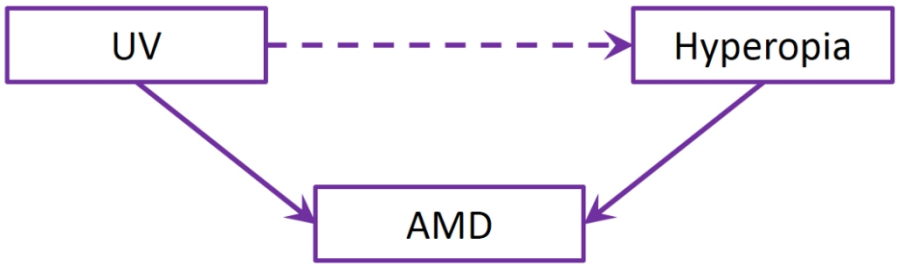

Figure 4

99x35mm (300 x 300 DPI)

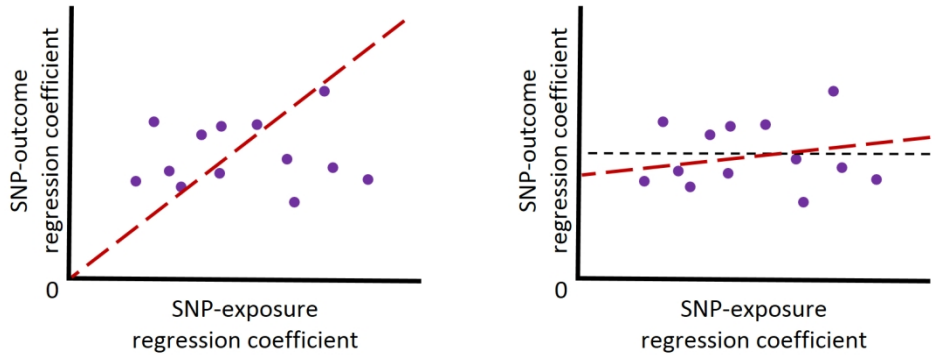

Figure 5

149x61mm (300 x 300 DPI)

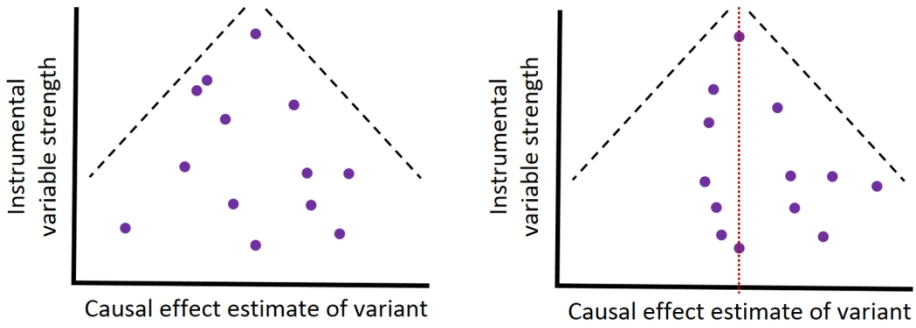

Figure 6

150x56mm (300 x 300 DPI)

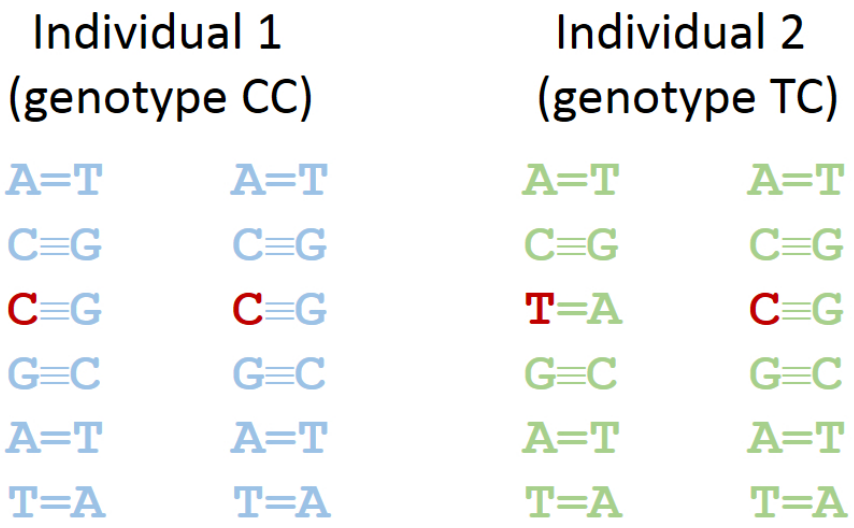

Figure 7

78x48mm (300 x 300 DPI)

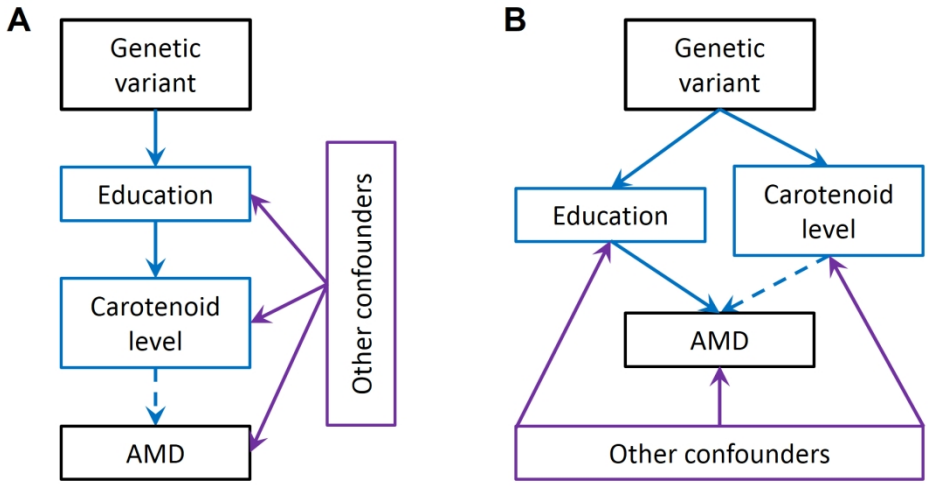

Figure 8

170x91mm (300 x 300 DPI)