JAMA Guide to Statistics and Methods

Mendelian Randomization

Connor A. Emdin, DPhil; Amit V. Khera, MD; Sekar Kathiresan, MD

Mendelian randomization uses genetic variants to determine whether an observational association between a risk factor and an outcome is consistent with a causal effect.[1] Mendelian randomization relies on the natural, random assortment of genetic variants during meiosis yielding a random distribution of genetic variants in a population.[1] Individuals are naturally assigned at birth to inherit a genetic variant that affects a risk factor (eg, a gene variant that raises low-density lipoprotein [LDL] cholesterol levels) or not inherit such a variant. Individuals who carry the variant and those who do not are then followed up for the development of an outcome of interest. Because these genetic variants are typically unassociated with confounders, differences in the outcome between those who carry the variant and those who do not can be attributed to the difference in the risk factor. For example, a genetic variant associated with higher LDL cholesterol levels that also is associated with a higher risk of coronary heart disease would provide supportive evidence for a causal effect of LDL cholesterol on coronary heart disease.

Author Audio Interview

CME Quiz

One way to explain the principles of mendelian randomization is through an example: the study of the relationship of high-density lipoprotein (HDL) cholesterol and triglycerides with coronary heart disease. Increased HDL cholesterol levels are associated with a lower risk of coronary heart disease, an association that remains significant even after multivariable adjustment.[2] By contrast, an association between increased triglyceride levels and coronary risk is no longer significant following multivariable analyses. These observations have been interpreted as HDL cholesterol being a causal driver of coronary heart disease, whereas triglyceride level is a correlated bystander.[2] To better understand these relationships, researchers have used mendelian randomization to test whether the observational associations between HDL cholesterol or triglyceride levels and coronary heart disease risk are consistent with causal relationships.[3-5]

Use of the Method

Why Is Mendelian Randomization Used?

Basic principles of mendelian randomization can be understood through comparison with a randomized clinical trial. To answer the question of whether raising HDL cholesterol levels with a treatment will reduce the risk of coronary heart disease, individuals might be randomized to receive a treatment that raises HDL cholesterol levels and a placebo that does not have this effect. If there is a causal effect of HDL cholesterol on coronary heart disease, a drug that raises HDL cholesterol levels should eventually reduce the risk of coronary heart disease. However, randomized trials are costly, take a great deal of time, and may be impractical to carry out, or there may not be an intervention to test a certain hypothesis, limiting the number of clinical questions that can be answered by randomized trials.

What Are the Limitations of Mendelian Randomization?

Mendelian randomization rests on 3 assumptions: (1) the genetic variant is associated with the risk factor; (2) the genetic variant is not associated with confounders; and (3) the genetic variant influences the outcome only through the risk factor. The second and third assumptions are collectively known as independence from pleiotropy. *Pleiotropy* refers to a genetic variant influencing the outcome through pathways independent of the risk factor. The first assumption can be evaluated directly by examining the strength of association of the genetic variant with the risk factor. The second and third assumptions, however, cannot be empirically proven and require both judgment by the investigators and the performance of various sensitivity analyses.

If genetic variants are pleiotropic, mendelian randomization studies may be biased. For example, if genetic variants that increase HDL cholesterol levels also affect the risk of coronary heart disease through an independent pathway (eg, by decreasing inflammation), a causal effect of HDL cholesterol on coronary heart disease may be claimed when the true causal effect is due to the alternate pathway.

Another limitation is statistical power. Determinants of statistical power in a mendelian randomization study include the frequency of the genetic variant(s) used, the effect size of the variant on the risk factor, and study sample size. Because any given genetic variant typically explains only a small proportion of the variance in the risk factor, multiple variants are often combined into a polygenic risk score to increase statistical power.

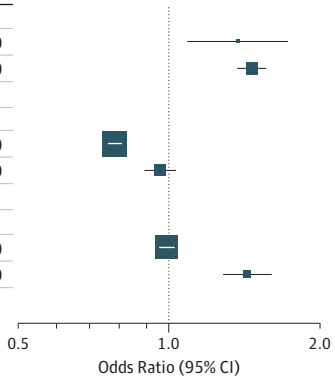
How Did the Authors Use Mendelian Randomization?

In a previous report in *JAMA*, Frikke-Schmidt et al[4] initially applied mendelian randomization to HDL cholesterol and coronary heart disease using gene variants in the *ABCA1* gene. When compared with noncarriers, carriers of loss-of-function variants in the *ABCA1* gene displayed a 17-mg/dL lower HDL cholesterol level but did not have an increased risk of coronary heart disease (odds ratio, 0.93; 95% CI, 0.53-1.62). The observed 17-mg/dL decrease in HDL cholesterol level is expected to increase coronary heart disease by 70% and this study had more than 80% power to detect such a difference; thus, the lack of a genetic association of *ABCA1* gene variants and coronary heart disease was unlikely to be due to low statistical power. These data were among the first to cast doubt on the causal role of HDL cholesterol for coronary heart disease. In other mendelian randomization studies, genetic variants that raised HDL cholesterol levels were not associated with reduced risk of coronary heart disease, a result consistent with HDL cholesterol as a noncausal factor.[5]

Low HDL cholesterol levels track with high plasma triglyceride levels, and triglyceride levels reflect the concentration of triglyceride-rich lipoproteins in blood. Using multivariable mendelian randomization, Do et al[3] examined the relationship among correlated risk factors such as HDL cholesterol and triglyceride levels. In an

Figure. Comparison of Observational Estimates and Mendelian Randomization Estimates of the Association of Low-Density Lipoprotein (LDL) Cholesterol, High-Density Lipoprotein (HDL) Cholesterol, and Triglycerides With Coronary Heart Disease

Analysis	Source	Odds Ratio (95% CI)
LDL cholesterol		
Observational	ERFC[2]	1.37 (1.09-1.73)
Mendelian randomization	Do et al[3]	1.46 (1.37-1.56)
Test for heterogeneity: $P = .60$		
HDL cholesterol		
Observational	ERFC[2]	0.78 (0.76-0.81)
Mendelian randomization	Do et al[3]	0.96 (0.89-1.03)
Test for heterogeneity: $P < .01$		
Triglycerides		
Observational	ERFC[2]	0.99 (0.96-1.03)
Mendelian randomization	Do et al[3]	1.43 (1.28-1.60)
Test for heterogeneity: $P < .01$		



Observational estimates are derived from the Emerging Risk Factors Collaboration (ERFC).[2] Mendelian randomization estimates are derived from Do et al[3] based on an analysis of 185 genetic variants that alter plasma lipids and mutually adjusted for other lipid fractions (eg HDL cholesterol and triglycerides for LDL cholesterol). A formal test of heterogeneity (Cochran Q test) shows that the observational and mendelian randomization causal estimates are consistent for LDL cholesterol but not so for HDL cholesterol or triglycerides.

analysis of 185 polymorphisms that altered plasma lipids, a 1-SD increase in HDL cholesterol level (approximately 14 mg/dL) due to genetic variants was not associated with risk of coronary heart disease (odds ratio, 0.96; 95% CI, 0.89-1.03; **Figure**). In contrast, a 1-SD increase in triglyceride level (approximately 89 mg/dL) was associated with an elevated risk of coronary heart disease (odds ratio, 1.43; 95% CI, 1.28-1.60). LDL cholesterol and triglyceride-rich lipoprotein levels, but not HDL cholesterol level, may be the causal drivers of coronary heart disease risk as demonstrated by these mendelian randomization studies.

Caveats to Consider When Evaluating Mendelian Randomization Studies

The primary concern when evaluating mendelian randomization studies is whether genetic variants used in the study are likely to be pleiotropic. Variants in a single gene that affects an individual risk factor are most likely to affect the outcome only through the risk factor and not have pleiotropic effects. For example, variants in *CRP*, the gene encoding C-reactive protein, have been used in a mendelian randomization study to exclude a direct causal effect of C-reactive protein on coronary heart disease.[6] However, variants in single genes that encode a risk factor of interest are often not available. In these cases, pleiotropy can be examined by testing whether the gene variants used are associated with known confounders such as diet, smoking, and lifestyle factors.[7] More advanced statistical techniques, including median regression[8] and use of population-specific instruments,[7] have recently been proposed to protect against pleiotropic variants biasing results.

A second concern relates to whether the mendelian randomization study has adequate statistical power to detect an association. Consequently, an estimate from a mendelian randomization study that is nonsignificant should be accompanied by a power analysis based on the strength of the genetic instrument and the size of the study. Furthermore, mendelian randomization estimates should be compared with results from traditional observational analyses using a formal test for heterogeneity.

ARTICLE INFORMATION

Author Affiliations: Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston (Emdin, Khera, Kathiresan); Cardiovascular Disease Initiative, Broad Institute, Cambridge, Massachusetts (Emdin, Khera, Kathiresan).

Corresponding Author: Sekar Kathiresan, MD, Center for Genomic Medicine, Massachusetts General Hospital, 185 Cambridge St, CPZN 5.830, Boston, MA 02114 (skathiresan1@mgh.harvard.edu).

Section Editors: Roger J. Lewis, MD, PhD, Department of Emergency Medicine, Harbor-UCLA Medical Center and David Geffen School of Medicine at UCLA; and Edward H. Livingston, MD, Deputy Editor, *JAMA*.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

REFERENCES

1. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1-22.

2. Di Angelantonio E, Sarwar N, Perry P, et al; Emerging Risk Factors Collaboration. Major lipids, apolipoproteins, and risk of vascular disease. *JAMA*. 2009;302(18):1993-2000.

3. Do R, Willer CJ, Schmidt EM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet*. 2013;45(11):1345-1352.

4. Frikke-Schmidt R, Nordestgaard BG, Stene MCA, et al. Association of loss-of-function mutations in the *ABCA1* gene with high-density lipoprotein cholesterol levels and risk of ischemic heart disease. *JAMA*. 2008;299(21):2524-2532.

5. Voight BF, Peloso GM, Orho-Melander M, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet*. 2012;380(9841):572-580.

6. Zacho J, Tybjaerg-Hansen A, Jensen JS, Grande P, Sillesen H, Nordestgaard BG. Genetically elevated C-reactive protein and ischemic vascular disease. *N Engl J Med*. 2008;359(18):1897-1908.

7. Emdin CA, Khera AV, Natarajan P, et al. Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. *JAMA*. 2017;317(6):626-634.

8. 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(4):304-314.