Advanced Gene Mapping Course: Mendelian Randomization ANSWERS

Andrew DeWan, PhD, MPH
January 27, 2021

This exercise is designed to give you practical experience conducting a two-sample Mendelian randomization study using the online version of MR-base. You will be conducting an analysis to investigate the causal relationship between low density lipoprotein (LDL) and coronary artery disease (CHD) based on summary statistics from previously published GWAS data.

Exposure: Fasting LDL measurements from in 173,082 subjects and 2,437,752 genetic variants. Subjects are of European, East and South Asian and African ancestry.

Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013 Nov;45(11):1274-1283. doi: 10.1038/ng.2797. Epub 2013 Oct 6. PMID: 24097068; PMCID: PMC3838666.

Outcome: CHD (e.g. myocardial infarction (MI), acute coronary syndrome, chronic stable angina, or coronary stenosis >50%) in 184,305 subjects (60,801 cases and 123,504 controls) and 9,455,779 genetic variants. Subjects are of European, East and South Asian, Hispanic and African ancestry.

Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015 Oct;47(10):1121-1130. doi: 10.1038/ng.3396. Epub 2015 Sep 7. PMID: 26343387; PMCID: PMC4589895.

1) Conduct an MR analysis of LDL and CAD. Studies can be search by PubmedID in MR-base (make sure PubmedID is checked), however, please note the following:

    A. For the exposure for this publication, use the larger set of subjects for this first analysis (N=173,082)
    B. For the exposure, use a p-value threshold of 5e-8, LD Rsq = 0.001 and clumping distance of 1000kb. Also make sure “Perform Clumping” is checked.
    C. For the outcome for this publication, use the trait denoted “Coronary heart disease”
    D. When running the MR analysis you will want to allow LD proxies to be selected for the outcome using a minimum Rsq of 0.8 and also allow for palindromic SNPs with a MAF threshold of 0.3. Make sure you set “Allele harmonization” to “Attempt to align strands for palindromic SNPs”
    E. Select the following methods:

a. Inverse variance weighted (NOTE: this is a random effects model)
b. MR Egger
c. Weighted Median

Questions:

1. How many variants are included in your genetic instrument for the exposure and how many are included in the outcome analysis? Of these, how many are proxies?

   There are 79 variants that surpass the p<5e-8 threshold for LDL in the exposure GWAS. Of these 77 are identified in the CHD outcome GWAS, 1 of which is a proxy.

2. Based on the descriptions above, is the study used to define the IV appropriate for the outcome population?

   They are fairly well matched in terms of the population ancestries in the two studies, however, the outcome GWAS has subjects of Hispanic ancestry which could be a minor issue. This would be something to mention in the Discussion section of a manuscript. There is a subset of only European subjects for LDL but not for CHD, however, if you had access to the original data you could subset the subjects by ancestry to better match the exposure and outcome groups.

3. Is there evidence of an association between LDL and CHD?

   Yes, the IVW yields a beta = 0.4114 (p=1.626e-15) which corresponds to an OR of 1.51 (95% CI: 1.36 – 1.67) per SD increase in LDL.

4. Is there evidence of heterogeneity in the genetic effects?

   Yes, there is significant heterogeneity across effects of each SNP on CHD (p=2.822e-40) indicating that horizontal pleiotropy may be an issue.

5. Is there evidence of pleiotropy?

   From the MR-Egger regression there is no significant evidence of pleiotropy as the regression intercept is not significantly different from zero (p=0.0046).

6. How would you interpret the results of the three analyses together (i.e. IVW, MR Egger and Weighted Median)?

The IVW method (OR = 1.51, 95% CI: 1.36 – 1.67, p=1.626e-16), MR-Egger (OR = 1.66, 95% CI: 1.42 – 1.93, p=1.086e-8) and Weighed median (OR = 1.49, 95% CI: 1.36 – 1.63, p=1.962e-19) are relatively consistent meaning the causal effect estimate is likely to be between 1.49 and 1.66. There is no evidence that this estimate is influence by horizontal pleiotropy as the MR-Egger intercept is not significant.

2) Re-run the analysis but for myocardial infarction (MI) using outcome data from the same publication.

   Questions:

   1. Is there evidence of an association between LDL and MI?

      Yes, IVW method provides significant evidence of an association between LDL and MI (OR = 1.48, 95% CI: 1.33 – 1.66, p=1.42e-12). The other MR measures of association are consistent with this estimate and there is again no evidence of horizontal pleiotropy.

   2. Can the association between LDL and CHD be explained by MI?

      We would need to test the other traits included in the CHD definition to see if they were associated with LDL or not and test for heterogeneity of the effects. However it is reassuring that the effect estimates are consistent between the larger CHD group and the smaller subgroup of subjects with MI.

3) Feel free to explore associations with additional exposures such as HDL, BMI (you can use the Yengo et al. SNPs) or other exposures/outcomes of interest to you.

   I'm more than happy to discuss additional results one-on-one or during the time we discuss the answers to this exercise.