**Lab 1: Worked example: Conduct a two sample MR analysis**

Programs to use “leadIHD” and “leaddiab”

A recent highly publicized paper using NHANES <https://edition.cnn.com/2018/03/12/health/lead-exposure-cardiovascular-disease-study/index.html> suggested that lead is an important cause of cardiovascular disease. However, this observation could be open to confounding. Here, you will use Mendelian Randomization to estimate the effect of lead on ischemic heart disease (IHD) and on a control outcome, i.e., diabetes which is thought to be unrelated to lead.

**MR estimate for lead on IHD using R**

1. **Install packages in R** (using leadIHD-mez)
2. “MendelianRandomization”
3. **SNPs to exposure**
4. Create a CSV file of strong genetic predictors of lead from Supplementary Table 1 of this paper <https://www.ncbi.nlm.nih.gov/pubmed/25820613>. Only use the rows with the rsid highlighted in yellow. Use Table 1 of the same paper to identify the effect allele. Use the following self-explanatory column headings (chrpos, other\_allele, effect\_allele, eaf, beta, se, gene, SNP). Make sure the letters in the effect allele column are in upper case. Upload this file to RStudio Cloud as lead.txt, i.e., click on Upload at the lower left.
5. **SNPs to IHD**
6. The largest publicly available files giving SNP-specific genetic associations with IHD is CARDIoGRAMplusC4D 1000 Genomes-based GWAS - Additive from the Cardiogram website <http://www.cardiogramplusc4d.org/data-downloads/>. The associated publication is https://pubmed.ncbi.nlm.nih.gov/26343387/
7. We have created an extract of this file for you to use, which is called cad-extract.csv. This file is already uploaded ready for you.
8. **Obtain estimates**

Continue using program leadIHD-mez to

1. Read your lead.txt file into R called “vtas” and format it
2. Read information about the associations of your genetic instruments with the outcome CAD from the file” cad-extract.csv”, we prepared from Cardiogram
3. Rename the columns in the outcome file to be the same for every file
4. Merge SNP to exposure and SNP to outcome files
5. Align the SNPs on the same effect allele for exposure and outcome and have a look at the alignment.
6. Generate a forest plot of the SNP-specific Wald estimates using fixed effects inverse variance weighting, note the estimate is presented as an odds ratio (OR) because the betas provided as logodds were exponentiated when plotted
7. Obtain a fixed effects inverse variance weighting estimate from the MendelianRandomization package, note the estimates are in OR because they were exponentiated by the program
8. Obtain a random effects inverse variance weighting estimate from the MendelianRandomization package, note the estimates are in OR because they were exponentiated by the program
9. Save the merged file for future use as leadihd21.csv

**MR estimate for lead on IHD using MRBase**

1. **SNPs to exposure**

Create a CSV file of strong genetic predictors of lead from Supplementary Table 1 of this paper <https://www.ncbi.nlm.nih.gov/pubmed/25820613>. Only use the rows with the rsid highlighted in yellow. Use Table 1 of the same paper to identify the effect allele. Use the following self-explanatory column headings (SNP, beta, se, other\_allele effect\_allele).

1. **Use MRBase to get estimate**
2. Go into MRBase
3. Click on “Perform MR analysis”
4. Click on “Choose exposures”
5. Under “Choose exposures” select manual file upload and upload your text file, check the delimiter is space, so the entries appear neatly in each column
6. Click on “Choose outcome”
7. Search for “cardiogram”
8. Highlight the row for the largest and most recent entry where the author is Nikpay
9. Click on “Run MR”
10. Click on “Perform MR analysis” (using the defaults)
11. The estimates are presented as extracted from Nikpay. In the Nikpay publication it says that the results from logistic regression of SNPs on IHD are provided as logodds. Ideally for presentation these should be converted to odds ratios (OR), where the estimate (b) is simply exponentiated (exp(b)), and the confidence interval (CI) for the IVW estimate is the estimate (b) plus or minus 1.96 times the standard error (se), i.e., lower CI is exp(b-1.96 times se) and upper CI is exp(b+1.96 times se)
12. **Use MRBase code to get estimate.**

Use the code from the “Analysis R code” section of the screen, which has been placed in the program leadIHD-mez for you, at the end

**Now answer the following questions**

1. Do you think all the SNPs predicting lead are valid instruments? Please explain your answer.
2. Based on your findings do you think lead causes IHD?
3. Is there any other analysis you would like to do?

**MR estimate for lead on diabetes using MRBase**

1. Go into MRBase
2. Click on “Perform MR Analysis”
3. Click on “Choose exposures”
4. Under “Choose Instruments” select “Manual file upload” and upload your file of lead instruments
5. Click on “Choose outcomes”
6. Search for “Type 2 diabetes”
7. Highlight the row of your choice, where the author is Xue A
8. Click on “Run MR”
9. Press “Perform MR Analysis”
10. Review the estimates

**MR estimate for lead on diabetes**

1. The largest available publicly available file giving SNP-specific genetic associations with Type 2 diabetes in Europeans is from a publication by Xue A https://pubmed.ncbi.nlm.nih.gov/30054458/
2. Now use the program leaddiab-mez
3. Leaddiab-mez simply uses MR-Base code to get the estimates

**Now answer the following questions**

1. Do you think all the SNPs predicting lead are valid instruments for diabetes? Please explain your answer.
2. Based on your findings do you think lead causes diabetes?
3. Is there any other analysis you would like to do?