**Two-sample Mendelian randomization using summary genetic data - practical**

In this practical we are going to use two-sample Mendelian randomization to appraise the causal relevance of body mass index (BMI) for risk of coronary heart disease;

**Files you will need:**

1. R script: MRB-practice-2020.R

**ANALYSIS 1: Two-sample Mendelian randomization of one exposure on one outcome**

Coronary heart disease is one of the leading causes of death and morbidity in the world. Of the known risk factors, body mass index (BMI) is one of the most important. In prospective observational studies (Wormser et al Lancet 2011 377:1085), each standard deviation (SD) (~4.56 kg/m2) increase in BMI is associated with a relative risk of 1.23 (95% CI: 1.17–1.29) for coronary heart disease, after adjustment for conventional vascular risk factors, including smoking, cholesterol, diabetes, and blood pressure. However, whether the association reflects a causal effect of BMI on disease.

**A. Estimate the effect of genetically elevated BMI on CHD**

1. Estimate the effect of BMI on CHD using inverse-variance weighted (IVW) linear regression, Weighted median (WM) and MR Egger regression (Egger).
   1. How do the IVW, WM and Egger results compare?
2. Estimate Wald ratios for each SNP and their standard errors
   1. The Wald ratio corresponds to the log odds ratio for coronary heart disease per unit change in BMI due to the SNP

Note: notice that the Wald ratio is simply the gene-outcome effect scaled to reflect a unit change in the gene-exposure effect, e.g. if G-O = 0.5 & G-E = 0.5, then E-O effect due to G is 1

**B. Sensitivity analyses and evaluating assumptions**

* + - 1. Create a scatter plot of the SNP-CHD and SNP-BMI associations. Does the SNP-CHD association increase linearly as the SNP-BMI association increases? What could deviations from linearity mean? Are there any unusual data points?
      2. Create a funnel plot of the results. Does the funnel plot look symmetric? What could asymmetry mean? Are there any outliers?
      3. Create a forest plot of the results. Is there heterogeneity in the effects amongst SNPs? What could that indicate?

**C. Interpret the results**

1. Do you think BMI causes CHD? Consider these questions in your discussion:

* 1. What is the odds ratio for coronary heart disease per unit increase in genetically elevated BMI?
  2. Is there evidence for pleiotropy or violations of MR assumptions?

Also consider these assumptions of Mendelian randomization in your results:

**Core assumptions of Mendelian randomization:**

1) The instrument is associated with the exposure

2) The instrument is not associated with confounders of the exposure-outcome association

3) The instrument is associated with the outcome exclusively via its effect on the exposure (also known as the exclusion restriction assumption)

**Other assumptions and practical limitations**

1. Samples should be independent
2. Two samples should be from the same population
3. Potential impact of weak instruments bias
4. Potential impact of winner’s curse, whereby the effects of the SNPs on BMI may be overestimated in the discovery stage of the GWAS
5. Strand ambiguity (when two GWAS studies use different reference strands)
6. SNPs should be independent