**Lab 2: Assumption testing**

Programs to use: fcalc

Utilities to use: MRBase

Now we have got an estimate for lead on IHD and diabetes we want to test carefully the assumptions of MR.

1. **SNPs are not weak instruments**
2. Calculate the F-statistic
3. Use the program “fcalc” to read in the genetic associations with the exposure and hence estimate the F-statistics

Now answer the following questions

1. Did any of the SNPs have F-statistic <10?
2. What would you do if the standard error for SNP on exposure was missing?
3. **Is the instrument for iron associated with potential confounders?**
4. Look up associations with potential confounders (education, smoking, alcohol use, physical activity) in the UKBiobank using MR-Base. UK Biobank is the largest and most extensive cohort study currently available. Use the file of lead SNPs you created in Lab 1 as the exposure, make sure the column names are the following (SNP, eaf, beta, se, SNP, other\_allele, effect\_allele). Select suitable UK Biobank outcomes that give potential confounders. Specifically,
   * 1. Go to MR Base <http://app.mrbase.org>
     2. Click “Chose exposures”
     3. Under Choose instruments, select “Manual file upload”
     4. Upload the file of instruments you created for Lab 1 with the relevant column names, make sure you are using the correct Separator for your file
     5. Click “Chose outcomes”
     6. Select, by highlighting” relevant UK Biobank variables of interest as the outcomes, making sure that the analysis pertains to most UK Biobank participants, possible outcomes could be
        1. “age completed full time education”
        2. “Smoking status: Never”
        3. “Alcohol intake frequency”
        4. “Number of days/week of vigorous physical activity 10+ minutes
     7. Click “Run MR”
     8. Click on “Perform MR analysis”
     9. Click on “Download harmonized summary statistics”.
     10. Open the file you have downloaded

More information about UK Biobank data is available on the UK Biobank data showcase https://biobank.ndph.ox.ac.uk/showcase/search.cgi

If MR-Base keeps timing out on you, then you can do this a different way

* 1. Find the codes for the potential confounders you want using a program to search MR-Base, it is called “lab2 find codes”
  2. Get the effect of lead on each of these potential confounders using a program called “lab2 lead on confounders”

Now answer the following questions

1. Was iron associated with any of these potential confounders?
2. What would you if iron was associated with these potential confounders?
3. Were all the SNPs available in UKBiobank?
4. What should you do for the unavailable SNPs?
5. **Is the instrument for iron associated with survival?**
   1. Go through the same steps as before to use MR-Base to look at the association of lead with survival. In brief use your lead file of instruments as the exposure and use survival as the outcome. Note the best measure of survival in MR-Base is parental attained age, so search for that

If MR-Base keeps timing out on you, then you can do this a different way

1. Find the codes for parental attained age using a program to search MR-Base, it is called “lab23 find codes”
2. Get the effect of lead on parental attained age using a program called “lab2 lead on lifespan”
3. Be aware that any exposure associated with parental attained age is open to selection bias due to survival if the outcome also affects survival

Now answer the following questions

1. Was iron associated with parental attained age?
2. What would you if iron was associated with parental attained age?
3. Were all the SNPs available in UKBiobank?
4. What should you do for the unavailable SNPs?
5. **Is the instrument for iron pleiotropic**
   1. Search for potential effects of SNP on outcome other than via the exposure (lead) using
      1. GWAS catalog, https://www.ebi.ac.uk/gwas/
      2. Phenoscanner, http://www.phenoscanner.medschl.cam.ac.uk
      3. MR-Base Phewas, https://gwas.mrcieu.ac.uk/phewas/
   2. Check whether any of the SNPs are in genes known to be highly pleiotropic

Now answer the following questions

1. Did you find any potentially pleiotropic SNPs for lead on IHD or lead on diabetes?
2. If so what would you do?
3. Does this help interpret your forest plots for Lead from Lab 1?
4. Do you think the GWAS for all the outcomes considered are free from bias?