A practical on Mendelian Randomization (MR)

A study used Mendelian Randomization with GRS for BMI as an instrument to estimate causal effect of BMI on the risk of Coronary Heart Disease:

http://jamanetwork.com/journals/jamacardiology/article-abstract/2635826


They found: “Mendelian randomization analysis showed significant positive associations between genetically instrumented higher BMI and risk of hypertension (odds ratio [OR] per 1-SD higher BMI, 1.64; 95% CI, 1.48-1.83; \( P = 1.1 \times 10^{-19} \)), coronary heart disease (OR, 1.35; 95% CI, 1.09-1.69; \( P = .007 \)) and type 2 diabetes (OR, 2.53; 95% CI, 2.04-3.13; \( P = 1.5 \times 10^{-17} \)), ...”

What does the result mean (in their data mean BMI is 27.53, SD=4.83)?

They also reported an observational association: 1.43 (1.40-1.46).
Is the MR estimate different?

Now let’s repeat a similar analysis in the Estonian Biobank data (mean BMI=26.5, SD=5.5)

A regression analysis of scaled BMI on scaled GRS results in:

```r
> summary(glm(scale(bmi)~scale(GRS)+age+sex, data=f1g))
```

Coefficients:

| Estimate  | Std. Error | t value | Pr(>|t|) |
|-----------|------------|---------|----------|
| (Intercept) | -0.5582564 | 0.0324872 | -17.184 | <2e-16 *** |
| scale(GRS) | 0.2306108 | 0.0078299 | 29.453 | <2e-16 *** |
| age | 0.0169582 | 0.0004053 | 41.839 | <2e-16 *** |
| sex | -0.1368019 | 0.0160076 | -8.546 | <2e-16 *** |

A logistic regression analysis of CAD on scaled GRS results in:

```r
> summary(glm(cad~scale(GRS)+age+sex, family=binomial, data=f1g))
```

Coefficients:

| Estimate  | Std. Error | z value | Pr(>|z|) |
|-----------|------------|---------|----------|
| (Intercept) | -6.654535 | 0.187950 | -35.406 | <2e-16 *** |
| scale(GRS) | 0.076317 | 0.036136 | 2.112 | 0.0347 * |
| age | 0.091733 | 0.002424 | 37.851 | <2e-16 *** |
| sex | -0.646091 | 0.072745 | -8.882 | <2e-16 *** |

Try to get the Mendelian Randomization estimate based on this output (you have to divide two coefficients and then take an exponential for an Odds Ratio)
To get standard errors as well, we use the Mendelian Randomization package “MendelianRandomization” (Stephen Burgess). The simplest analysis can be ran by:

```r
library(MendelianRandomization)
mrdat<-mr_input(bx = 0.231, bxse=0.00783, by = 0.0763, byse=0.0361)
mr_ivw(mrdat)
```

Make sure you understand where the coefficients in `mr_input` come from!
Is the result similar to what was seen in the UK Biobank?
(Remember, to take the exponential, if you want to see the OR)

When a logistic regression model for CAD was fitted with scaled BMI as the only covariate, the ln(OR) estimate was 0.354, SE=0.034. Is it similar to the MR estimate?

What is the benefit of using Mendelian Randomization here instead of just regressing CAD on BMI?

What do you think, is the Mendelian Randomization estimate valid? Do you see any possibility for pleiotropy?

Optional:
To conduct the Egger analysis, you can look at the example using the inbuilt dataset in the package:
```r
ldldat <- mr_input(bx = ldlc, bxse = ldlcse,by = chdlodds, byse = chdloddsse)
ldldat
mr_allmethods(ldldat, method="egger")
```

plots:
```r
mr_plot(mr_allmethods(ldldat, method="egger"))
```