Advanced MR Short Course - Practical 3

Within Families MR

Like all research approaches, Mendelian randomization depends on assumptions. Specifically that the genetic variants used as an instrumental variable for the exposure of interest i) are associated with the exposure (relevance), ii) have no common cause with the outcome (independence) and iii) only affect the outcome via the exposure (exclusion). Here we will focus on violations of the second assumption – independence. What can we do if we think there are instrument-outcome confounders?

Conventional Mendelian randomization studies use samples of unrelated individuals from the population. This depends on the assumption that what is true within a family (independent and random transmission of genetic variation from parents to their children) is true across the population. For many outcomes and exposures this may not be true. Three potential sources of bias have been widely discussed in the literature, 1) population structure, 2) dynastic effects, and 3) assortative mating. Each of these can induce spurious associations between instruments and outcomes. For example, in the population, genetic variants for height and BMI associate with educational outcomes (Brumpton et al. 2020, and Howe et al. 2022). However, using family based methods and samples of siblings, that control for these biases, these effects are entirely attenuated.

In this practical, we will apply a within family Mendelian randomization methods to overcome these biases. WF MR methods always require data on related individuals, e.g. siblings or parent-offspring trios. We will control for most forms of confounding using three different methods, i) controlling for parental genotype, ii) within sibling fixed effects, and iii) using the within transformation. We will apply this to the example of the effects of BMI on educational attainment. We want to know if the effects detected of BMI on educational attainment by classic MR are due to individual level causal effects, or some form of bias.

Within the script, there are notes (indicated by green writing), which will help navigate through the practical. Good luck!

**Datasets & scripts:**

***Scripts***

Practical3\_wf\_code.R > code used for the practical

generate\_sim.R > code used to generate the simulated data (we don’t need this for the practical but it is provided for reference)

***Data***

family\_sims.RData > simulated data used in the practical

***Part 1 – Control for confounding using parental genotypes***

1. Load the data into R (family\_sims.RData). Summarise and explore the data.
   * 1. How many genetic variants does each parent have?

1

* + 1. How many Siblings are there?

2

1. How can we test whether the inheritance from parent to offspring is random?

Using a linear regression of offspring genotype on parents genotype.

What are the correlations of the maternal and paternal and offspring 1’s genotype:

|  |  |  |
| --- | --- | --- |
| **Outcome** | **Exposure** | **Coefficient and SE** |
| Offspring allele 1 | Maternal allele 1 | 0.5067129 0.0035282 |
| Offspring allele 1 | Maternal allele 2 | 0.4917998 0.0035288 |
| Offspring allele 1 | Paternal allele 1 | -0.0049777 0.0035228 |
| Offspring allele 1 | Paternal allele 2 | 0.0009532 0.0035308 |

1. Repeat these regressions for offspring 2. What do these results demonstrate about the inheritance of genetic variants from parents to offspring?

|  |  |  |
| --- | --- | --- |
| **Outcome** | **Exposure** | **Coefficient and SE** |
| Offspring allele 2 | Maternal allele 1 | -0.0004353 0.0035431 |
| Offspring allele 2 | Maternal allele 2 | -0.0005959 0.0035438 |
| Offspring allele 2 | Paternal allele 1 | 0.4929283 0.0035377 |
| Offspring allele 2 | Paternal allele 2 | 0.5069566 0.0035457 |

Inheritance is a random function of chance, and parents’ genotype.

1. Estimate the association of offspring 1 and offspring 2’s genotype (using the additive measure).
   1. What does this demonstrate about the correlation between offspring’s genotypes?

Sibling’s genotypes are correlated, with an R^2 =0.25.

1. Estimate the phenotypic association between offspring BMI and educational attainment.

|  |  |  |  |
| --- | --- | --- | --- |
| **Outcome** | **Exposure** | **Method** | **Coefficient and SE** |
| Educational attainment | BMI | OLS | 0.32895 0.00795 |

* 1. What does this suggest about the effect of BMI on educational attainment?

BMI is associated with educational attainment. It is unclear if this is causal.

1. Repeat the regression, but control for additive parental genotype.

|  |  |  |  |
| --- | --- | --- | --- |
| **Outcome** | **Exposure** | **Method** | **Coefficient and SE** |
| Educational attainment | BMI | OLS controlling for parental genotype | 0.011331 0.007708 |

* 1. What does this suggest about the association of BMI and educational attainment?

It is confounded by genotype.

* 1. Would we find a similar change in the estimates after controlling for genotype using non-simulated data?

No, genetic variants in real data are likely to have much smaller effects and are unlikely to fully attenuate the estimates as in this simulated example.

1. We can use parental genotype to control for biases in MR models as well. Estimate the effect of BMI on education using the offspring genotype and the ivreg function.

|  |  |  |  |
| --- | --- | --- | --- |
| **Outcome** | **Exposure** | **Method** | **Coefficient and SE** |
| Educational attainment | BMI | Classic MR | 0.98569 0.01506 |

* 1. What does this suggest about the effect of BMI on educational attainment?

BMI affects educational attainment.

1. Include the parents’ genotype as addition covariate (i.e. not as an instrument, but included in both parts of the IV regression.

|  |  |  |  |
| --- | --- | --- | --- |
| **Outcome** | **Exposure** | **Method** | **Coefficient and SE** |
| Educational attainment | BMI | Within family MR | 0.01194 0.01732 |

* 1. What does this suggest about the effect of BMI on educational attainment?

After controlling for parental genotype, there is little evidence that BMI affects educational attainment.

***Part 2 – Control for confounding using sibling fixed effects***

1. The data has two siblings. Reshape the data to long format, so that you have one sibling per row (the code for this is given in lines 120-146)
2. Use the plm package to estimate the effects of BMI on education using the pooled estimator (i.e. model=”pooling”)

|  |  |  |  |
| --- | --- | --- | --- |
| **Outcome** | **Exposure** | **Method** | **Coefficient and SE** |
| Educational attainment | BMI | Classic MR | 0.993815 0.010744 |

1. As the siblings are not independent adjust the standard errors for this non-random sampling using the coeftest function.

|  |  |  |  |
| --- | --- | --- | --- |
| **Outcome** | **Exposure** | **Method** | **Coefficient and SE** |
| Educational attainment | BMI | Classic MR | 0.993815 0.012190 |

* 1. How have the point estimates and the standard errors changed?

The point estimates have not changed, but the standard error has slightly increased.

1. Include the additive parental genotype as a covariate

|  |  |  |  |
| --- | --- | --- | --- |
| **Outcome** | **Exposure** | **Method** | **Coefficient and SE** |
| Educational attainment | BMI | Within family MR | -0.0061713 0.0123816 |

* 1. How have the estimates changed?

Attenuated to the null.

1. Adjust the standard errors to account for non-independence of samples

|  |  |  |  |
| --- | --- | --- | --- |
| **Outcome** | **Exposure** | **Method** | **Coefficient and SE** |
| Educational attainment | BMI | Within family MR | -0.0061713 0.0121903 |

1. How do these estimates compare to the WF MR estimate above using the ivreg package?

Attenuated to the null.

1. Now use the plm package to estimate the effects of BMI on education using the within estimator (i.e. model=”within” without adjusting for the parental genotypes

|  |  |  |  |
| --- | --- | --- | --- |
| **Outcome** | **Exposure** | **Method** | **Coefficient and SE** |
| Educational attainment | BMI | Within family MR | -0.0076306 0.0100510 |

* 1. How do the estimates compare to the WF estimate above?

Very similar.

1. Adjust the standard errors to account for non-independence of samples

|  |  |  |  |
| --- | --- | --- | --- |
| **Outcome** | **Exposure** | **Method** | **Coefficient and SE** |
| Educational attainment | BMI |  | -0.0076306 0.0121903 |

Point estimate the same, standard errors slightly larger.

***Part 3 – Control for confounding using the within transformation***

1. Calculate the mean sibling genotype (lines 212-217)
2. Estimate the effects of BMI on educational attainment using instrumental variable regression and the plm function, without adjusting for parental genotype, but including the mean sibling genotype.

|  |  |  |  |
| --- | --- | --- | --- |
| **Outcome** | **Exposure** | **Method** | **Coefficient and SE** |
| Educational attainment | BMI | Within transformation | -0.0076306 0.0183318 |

1. Estimated standard errors clustered by families

|  |  |  |  |
| --- | --- | --- | --- |
| **Outcome** | **Exposure** | **Method** | **Coefficient and SE** |
| Educational attainment | BMI | Within transformation | -0.0076306 0.0121903 |

1. How do these estimates compare to the parentally adjusted estimates and the sibling fixed effects estimates?

They are very similar. The sibling based methods are considerably more precise because we are using more data. If we reran, but using both siblings and clustering the SEs by family we would get almost identical results because the fixed effects and within transformation are identical.