mrrobust: A Stata package implementing MR-Egger regression type analyses

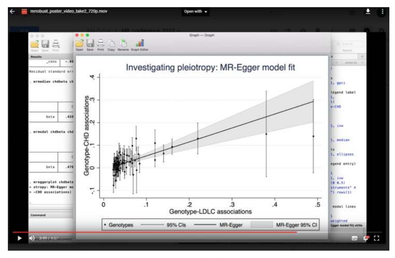
Tom Palmer[1] **Wesley Spiller**[2] **Neil Davies**[2]

1. Department of Mathematics and Statistics, Lancaster University, Lancaster, UK.
2. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

tom.palmer@lancaster.ac.uk

Summary

- **Scan QR code for a short video explaining the package!**



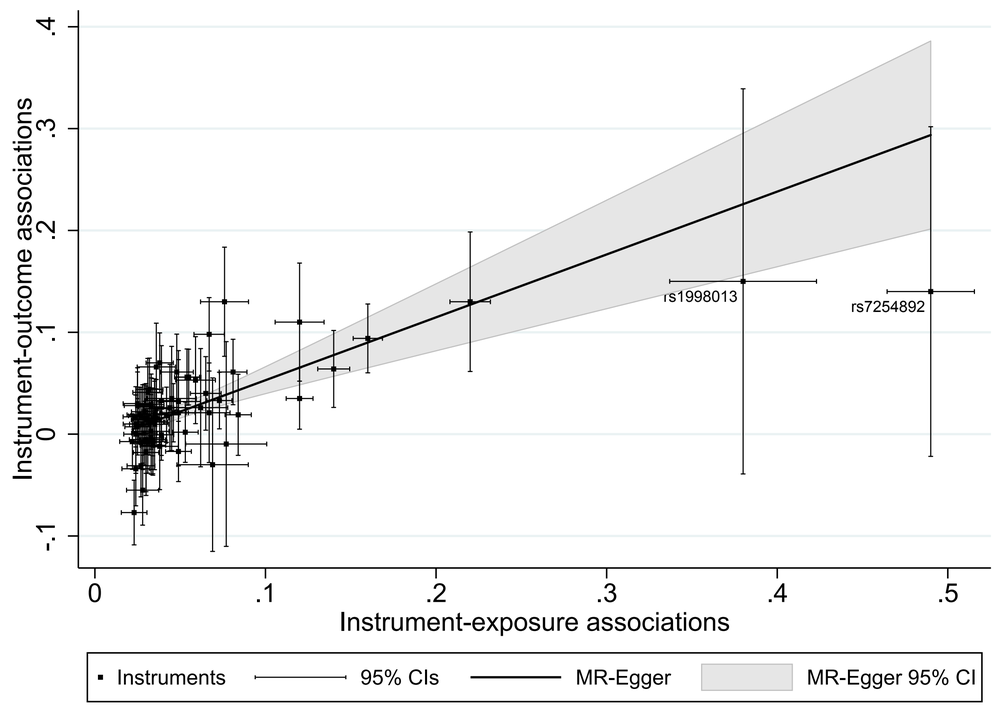
- Mendelian randomization studies using summary data from genome-wide association studies are becoming increasingly common.
- `mrrobust` is a Stata package implementing several of the latest methods.
- It is a free download from `https://raw.github.com/remlapmot/mrrobust`
- See Spiller *et al.* (2017) for further details.

Introduction

- The `mrrobust` package includes the following commands:
  - `mrratio`: ratio (Wald) estimator for a single genotype/instrumental variable (IV);
  - `mrivests`: generate ratio estimates in current dataset;
  - `mregger`: inverse-variance weighted (IVW) and MR-Egger estimators, and I^2_{GX} statistic (Bowden *et al.*, 2015, 2016a);
  - `mrmedian`: median estimators (Bowden *et al.*, 2016b);
  - `mrmodal`: zero modal estimator (Hartwig *et al.*, 2017);
  - `mreggerplot`: Egger regression type plot;
  - `mrforest`: Forest plot of IV estimates.

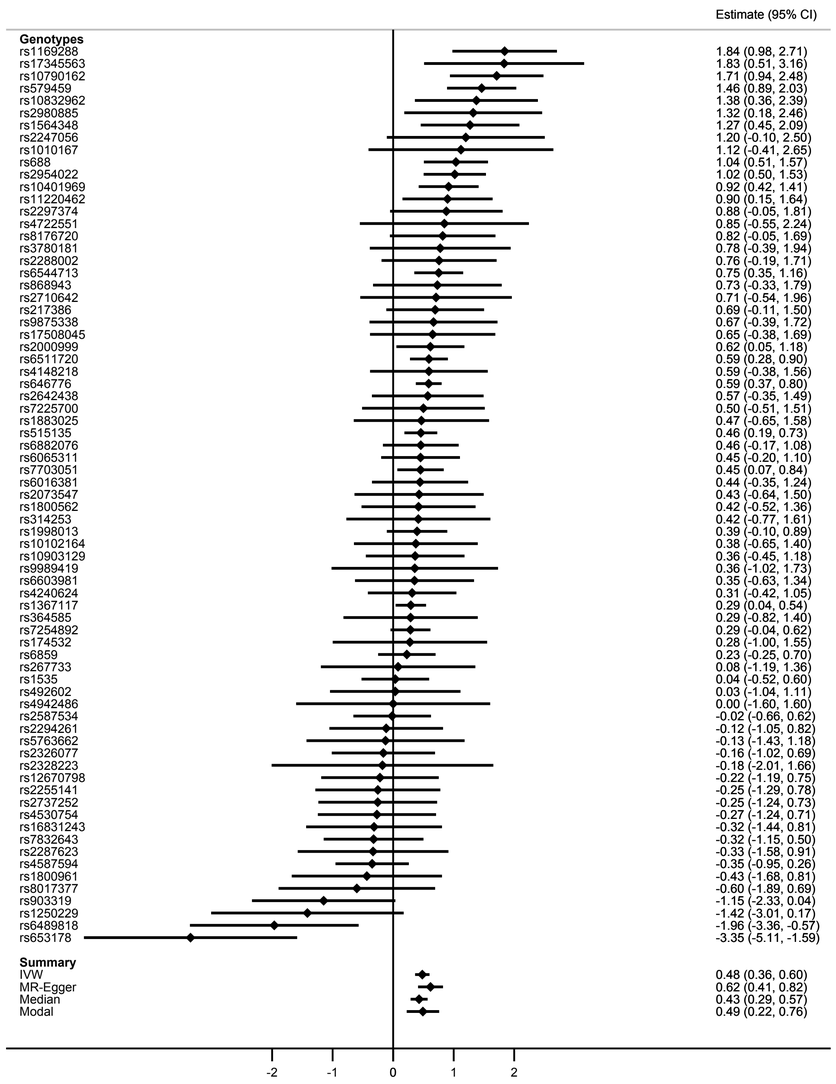
Example analysis

- The package assumes that you have imported summary data, possibly from a repository such as MR-Base `http://www.mrbase.org` (Hemani *et al.*, 2016a).
- The package also assumes that you have appropriately harmonised your data (Hartwig *et al.*, 2016).
- We use summary data provided by Do & et al. (2013) to investigate the causal effect of low-density lipoprotein cholesterol (LDL-C) on the risk of coronary heart disease (CHD).
- 73 genotypes achieved genome-wide statistical significance ($p < 1 \times 10^{-8}$) for their association with LDL-C.
- Plot of the individual IV estimates and MR-Egger fitted line with 95% CI:



- The modal estimate is similar to the IVW estimator.
- The MR-Egger estimate is the largest and the median estimate the smallest.
- The I^2_{GX} statistic of 98.5% shows that there should be less that 1.5% bias in the MR-Egger estimate due to regression dilution bias.
- The MR-Egger intercept of -0.009 (95% CI -0.020, 0.002) provides no strong evidence against the null hypothesis of no pleiotropy.

- Forest plot of genotype specific and summary IV estimates:



I_{GX}^2=98.5%

Discussion

- The `TwoSampleMR` package (Hemani *et al.*, 2016b) and the `MendelianRandomization` package (Yavorska & Burgess, 2016, 2017) provide similar functionality in R.

Acknowledgements: The Medical Research Council (MRC) and the University of Bristol fund the MRC Integrative Epidemiology Unit [MC UU 12013/1, MC UU 12013/9].

The authors would like to thank Michael Holmes, Caroline Dale, Amy Taylor, Rebecca Richmond, Judith Brand, Yanchun Bao, Kawthar Al-Dabhani, Michalis Katsoulis, and Ghazaleh Fatemifar for helpful feedback and suggestions.

References

Bowden, J., Davey Smith, G., & Burgess, S. 2015. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *International Journal of Epidemiology*, **44**(2), 512–525.

Bowden, J., Del Greco M, F., Minelli, C., Davey Smith, G., Sheehan, N. A., & Thompson, J. R. 2016a. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I^2 statistic. *International Journal of Epidemiology*, **45**(6), 1961–1974.

Bowden, J, Davey Smith, G., Haycock, P. C., & Burgess, S. 2016b. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiology*, **40**(4), 304–314.

Do, R., & et al. 2013. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nature Genetics*, **45**, 1345–1352.

Hartwig, F. P., Davies, N. M., Hemani, G., & Davey Smith, G. 2016. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *International Journal of Epidemiology*, **45**(6), 1717–1726.

Hartwig, F. P., Davey Smith, G., & Bowden, J. 2017. Robust inference in summary data Mendelian randomisation via the zero modal pleiotropy assumption. *bioRxiv*.

Hemani, G., Zheng, J., Wade, K. H., Laurin, C., Elsworth, B., Burgess, S., Bowden, J., Langdon, R., Tan, V., Yarmolinsky, J., Shihab, H. A., Timpson, N., Evans, D. M., Relton, C., Martin, R. M., Davey Smith, G., Gaunt, T. R., & Haycock, P. C. 2016a. MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. *bioRxiv*.

Hemani, G., Haycock, P., & Zheng, J. 2016b. *TwoSampleMR: Two Sample MR functions and interface to MR Base database.* R package version 0.2.0.

Spiller, W., Davies, N. M., & Palmer, T. M. 2017. Software Application Profile: mrrobust - A Tool For Performing Two-Sample Summary Mendelian Randomization Analyses. *bioRxiv*.

Yavorska, O. O., & Burgess, S. 2016. *MendelianRandomization: Mendelian Randomization Package.* R package version 0.2.0.

Yavorska, O. O., & Burgess, S. 2017. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *International Journal of Epidemiology*.

Feedback zone

- Vote on these potential great new features!

Feature	Tally marks
Extract data from MR-Base	
Better forest-type plot with lots of genotypes	
SIMEX for MR-Egger	

- Any other comments: