

# Recent challenges for Mendelian randomisation analyses

Vanessa Didelez\*

[didelez@leibniz-bips.de](mailto:didelez@leibniz-bips.de); [vanessa.didelez@bristol.ac.uk](mailto:vanessa.didelez@bristol.ac.uk)

---

Mendelian randomisation (MR) refers to situations where a genetic predisposition can be exploited as an instrumental variable (IV) to estimate the causal effect of a modifiable risk factor or exposure on an outcome of interest. For example, the ALDH2 gene is associated with alcohol consumption, and has therefore successfully been used as an IV to estimate the causal effect of alcohol on outcomes related to coronary heart disease. MR analyses have become very popular especially recently with the increased availability of GWAS data. This gives rise to the following challenges:

- (1) It is common that several SNPs are found to be associated with an exposure of interest, i.e. there are potentially numerous IVs; if these are all valid IVs, methods for multiple instruments are called for.
- (2) It is also common that many of these numerous potential IVs are only weakly associated with the exposure of interest; the phenomenon of weak IV bias is well-known for the simple case, and of course it also affects the multiple IV case; hence methods for multiple weak IVs are needed; it has been proposed to combine SNPs into an allele score, i.e. a single hopefully stronger IV, but this can lead to bias if done in a data-driven manner.
- (3) Further it is unlikely that all such SNPs are actually valid instruments for the causal effect of interest; they could for instance have pleiotropic effects or violated the IV conditions in other ways. Some first proposals to deal with such violations of assumptions suggest methods that do not require knowledge of which IVs are valid and which aren't, e.g. similar to Egger regression in meta analyses.

---

\*Leibniz Institute and Department of Mathematics, University of Bremen, Achterstr. 30, Bremen, 28359, GERMANY

- (4a) Data is often only available from different sources, one with instrument-exposure data, and a different source with instrument-outcome data. This means inference has to be based on two bivariate samples (possibly with additional covariates), instead of a joint sample. Two-stage-least-squares (TSLS) can be adapted to this case as "two-sample TSLS", but more robust methods would be desirable.
- (4b) Even less information is available if an MR analysis has to make do with summary data (often when based on case-control studies, but also otherwise). This means that from a number of primary analyses we only have measures of the instrument-exposure and (possibly from different studies) measures of the instrument-outcome associations.
- (5) Moreover, typical data available for MR analyses often comes from case-control studies, i.e. we have a binary outcome and sampling is conditional on case or control status; linear models are not appropriate in this case and if the retrospective nature of the sampling is ignored this can induce selection bias; even if data was sampled prospectively, selection bias can occur if e.g. volunteering is related to exposure / outcome status. In this context it is particularly important to ensure that the chosen methods for analysis have the null-preservation property, i.e. are consistent under the null-hypothesis of no causal effect.

In this presentation, I will give an overview over the above challenges as well as existing approaches to tackle them, their strengths and limitations.