Matching approaches for efficient and robust causal inferences in Mendelian randomization

Supervisors: Stephen Burgess and Brian Tom

Mendelian randomization is the use of genetic variants as proxy measurements for a modifiable exposure to judge whether interventions on the exposure are likely to reduce disease risk. Mendelian randomization has a history of correctly predicting the results of randomized trials of pharmacological interventions, and has wide applicability in a broad range of scientific fields for making the crucial distinction between causation and correlation. An association between the genetic variant that is a proxy for the exposure and the outcome variable (usually disease) implies that the exposure is a causal risk factor for the disease.

The aim of this project is to investigate the use of traditional approaches for causal inference in Mendelian randomization, in particular: i) matching on covariates, and ii) matching by design. The use of covariates for matching should lead to more efficient instrumental variable estimators, and may result in more robust estimators, as covariates are similar within the matched pairs. Such approaches have not been previously considered in detail in the context of Mendelian randomization, but should be feasible in the UK Biobank dataset. Another reason for using matching in this context is to divide the UK Biobank dataset into a discovery cohort (comprising non-diseased individuals) and a validation dataset (comprising diseased individuals and matched controls). This enables the dataset used to choose which genetic variants to include in the analysis to be separate from the dataset in which the causal hypothesis is tested.

Alternatively, paired designs, such as the comparison of sibling pairs, may be worthwhile in Mendelian randomization, particularly for rare genetic variants that naturally cluster within families, and hence are unlikely to satisfy the Mendelian randomization assumptions. While population-based studies have overtaken family-based studies in the current GWAS era, recall-by-genotype experiments (where additional carriers of rare variants are found amongst the relatives of those who carry the variants) will lead to matched analyses becoming increasingly important. Methods will be developed using data from the Swedish Twin Registry.

Start date: Easter Term (April) or Michaelmas Term (October) 2019

All application queries regarding eligibility should be directed to phdstudy@mrc-bsu.cam.ac.uk

How to Apply: Applications should be made on-line via www.graduate.study.cam.ac.uk/applicant-portal selecting course details MDBI22 PhD in Biostatistics

Deadline for applications: 3rd January 2019