Improving power and methods to systematically detect conflict in evidence synthesis

Supervisors: Anne Presanis and Daniela De Angelis

Location: MRC Biostatistics Unit, University of Cambridge

Detailed description: Observational data from multiple diverse sources, such as electronic health records, public health surveillance, and different types of surveys, are increasingly available. A much deeper understanding of important public health questions than ever before can be achieved through the synthesis of these diverse sources of information. Example public health questions include the estimation of hidden characteristics of epidemics, such as HIV prevalence or influenza incidence. However, synthesising such diverse data is challenging, due to each data source providing perhaps only indirect, incomplete evidence on the quantities of interest; unaccounted selection biases inherent to observational data; and varying relevance, quality, quantity and availability of the data (De Angelis et al. 2015).

Models relating the available data to the quantities to be estimated can be formulated, but an important part of the model development and criticism is to understand whether the model is consistent with each data source, and importantly, whether each data source gives consistent or conflicting evidence about the characteristics to be estimated. Traditional model criticism approaches, such as the examination of residuals, can only give a partial view of the adequacy of the complex probabilistic models resulting from the combination of multiple sources of evidence. "Conflict p-values", based on cross-validation approaches separating evidence into different partitions, have been developed to detect and measure inconsistency in particular parts of a model. However, targeted conflict assessment may not be sufficient when it is unclear which aspects of a model to target. Systematic conflict diagnostics checking for conflict throughout a model is possible but challenging: it involves simultaneous hypothesis testing for every parameter in a model, for which there might not be enough power; and is computationally expensive.

Addressing these challenges, using examples from HIV and influenza evidence syntheses, is the subject of this project. Possible directions include adapting approximate computational approaches or efficient model-building methods to the problem of systematic conflict assessment; or investigating the relationship between power to detect conflict and methods measuring the value of collecting more information.

This project would suit someone interested in developing general methods that are important in current biostatistical/public health applications, and for the more computational aspects, an interest in developing and coding relevant algorithms would be necessary.

Start date: Michaelmas Term 2018

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: 4th January 2018