Dynamic prediction of in-patient mortality based on electronic health record data: a comparison of landmarking and machine learning approaches

Supervisors: Steven Kiddle and Jessica Barrett

Background:
There is great potential to use electronic health record (EHR) datasets to improve care of patients, as EHR are typically bigger, longer and more representative of the healthcare population than traditional research cohorts. As EHRs contain routinely collected data, any model trained on them can be readily implemented into clinical care. However, despite their wealth of longitudinal data most EHR risk prediction studies have not made the most of longitudinal data to improve prediction accuracy [1]. Two promising approaches are the use of landmarking [2] and the use of machine/deep learning [3].

The aim of this project to generate models to accurately predict in-patient mortality in a range of clinical settings. To achieve this we will compare existing cross-sectional prediction methods with landmarking and machine/deep learning, within a rigorous cross-validation scheme.

Details of the project:
The project will begin with a review of methods for longitudinal data in risk prediction, and which have been applied successfully to EHR data. Then the student will familiarise themselves with the methodological approaches, clinically important variables and with the EHR datasets: MIMIC-III [4] and the EPIC health records used by Addenbrooke’s hospital.

MIMIC-III contains data on critical care, which is particularly rich in vital signs and other continuous monitoring data. Within EPIC we will focus on in-patient medicine for the elderly.

A rigorous cross-validation scheme will be set-up, and then the most promising methods will be implemented and compared. As MIMIC-III is open access, by releasing our code we will allow others to repeat our analysis and match our cross-validation splits to compare their own methods. Limitations of existing methods could be used to inspire methodological development. The student will work with clinicians to assess the potential clinical utility of the model, and how it could be implemented in practice.

We will follow the guidelines from a review of this topic [1] and reporting guidelines for multivariable prognosis models [5].

References:


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