Prediction of In-patient Mortality using Advanced Statistical Modeling and Deep Learning on Electronic Health Record Data

EBI PI: Ewan Birney

The EMBL-EBI,
Wellcome Trust Genome Campus,
Hinxton, Cambridge,
CB10 1SD
birney@ebi.ac.uk

BRC PI: Steven Kiddle

MRC Biostatistics Unit,
University of Cambridge,
Forvie site, Robinson Way,
Cambridge, CB2 0SR
steven.kiddle@mrc-bsu.cam.ac.uk

Background
There is a strong interest in building predictive models tailored to individuals for personalised treatment, care plans and enhanced patient safety. Electronic health records (EHRs) are a large and rich source of information on typical patients, despite this EHRs remain underused for research [1]. Where used in prediction models they typically crudely model co-morbidity (patients with multiple diseases) and ignore patient history (i.e. longitudinal information) [2]. More broadly, the complex and detailed information in EHRs needs to be cast into more tractable models for downstream use.

Since 2014, Addenbrooke’s Hospital in Cambridge has used an advanced electronic health records system (EPIC) that can generate anonymised datasets for service evaluation and research under the appropriate clinical governance and ethical regulations. Our collaborator Dr Romero-Ortuno was able to use EPIC electronic health records to conduct an internal service evaluation of the association of frailty and co-morbidities on older patients' outcomes in a real healthcare setting [3]. We intend to take this to a research level using advanced prediction modelling. This will enable us to generate models to provide clinicians with personalised outcome predictions for their patients, allowing them to design more appropriate care.

The EBPOD candidate will work collaboratively between the Birney Research group at EMBL-EBI and the Kiddle group at MRC Biostatistics Unit, University of Cambridge. The fellow will also closely collaborate with clinician Dr Roman Romero-Ortuno and the Department of Medicine for the Elderly in Addenbrooke's Hospital.

Dataset
Access to EPIC for research purposes is subject to ethics approval, which will be in place by project start. We will use EPIC data for individuals older than 75 admitted non-electively to the Department of Medicine for the Elderly at Addenbrookes hospital. We will mainly focus on the prediction of inpatient mortality, but may also explore other outcomes. As is typical for EHR the data is high-dimensional and contains a large proportion of missingness leading to a need for robust methods.

Aim 1 – Improve prediction of in-patient mortality by taking better account of patients’ co-morbidities
Dr Romero-Ortuno used a model including 10 variables to predict in-patient mortality with an Area Under the Curve (AUC) of 0.80 [3]. One of these variables is
the Charlson Co-morbidity Index (CCI), a score which reflects both the number and life-threatening nature of a patient’s co-morbidities. The limited utility of CCI in this context is highlighted by the benefit of adding a separate dementia variable to Dr Romero-Ortuno’s model. We believe that we can improve upon the 20-year-old CCI approach using our large dataset and modern statistical methods, specifically the use of data on each disease separately, the consideration of pair-wise interactions and elastic net regression. Nested cross-validation will be used to select penalty terms and to compare our models’ performance with existing approaches.

Aim 2 – Develop dynamic prediction models
The models in aim 1 are static in that they make a prediction about in-patient mortality at the start of their stay, but are not updated to ensure they are based on the best currently available information. We will address this by developing dynamic prediction models [4], where we attempt to predict a patient’s probability of survival based on data collected during their stay, e.g. longitudinal lab results and vital signs. We will explore transforming the vital sign data using machine learning. A key aspect will be making the best use of informative presence, the fact that the presence and number of observations for each variable can itself be predictive of mortality.

Aim 3 – Assess the potential of deep learning to improve prediction
Deep learning across large variable sets has been shown to often outperform prediction models limited to pre-defined clinically relevant variables [5]. We will test this in the context of predicting in-patient mortality, comparing the results with those from the previous aims, both in terms of predictive ability and interpretability, which can be a problem for deep learning models.

Partners and training opportunities
The candidate will be embedded in a strong multidisciplinary team. Steven Kiddle is a MRC Career Development Award fellow in statistics, with a group focusing on dynamically predicting mortality using health record data on co-morbidities and biomarkers. Ewan Birney’s research group has a broad range of human patient multidimensional datasets. Both groups already have postdocs working on EHRs. The fellow will be fully embedded in both groups and be able to acquire a unique grounding in health informatics and biostatistics.

References