Bayesian Modelling and Statistical Machine Learning for Morbidity Rate Prediction

Kaitlyn Louth

The University of Edinburgh and Heriot-Watt University
65 Abbeyhill, Edinburgh, EH8 8DY
United Kingdom

Joint work with: Prof George Streftaris and Prof Ruth King

Earlier work [1-3] has shown that population morbidity rates are idiosyncratic to a number of demographic and socioeconomic factors. More recent work [4,5, 6] has investigated morbidity risk for cancer, where it has also suggested differences in related time trends between the general population and an insured population in the UK, and respiratory admission rates in a US insured population. This project builds on this work to identify trends and important morbidity risk factors for cancer and a wider range of illnesses, also relating to healthcare provision and insurance. The principal aim of this research is to develop, evaluate and assess models for morbidity risk and related healthcare rates, under statistical approaches such as GLM-type and hierarchical Bayes models, that allow for uncertainty quantification. Aiming at improved predictions, traditional methods are combined with artificial neural network (ANN) deep learning techniques, by adopting a hybrid approach that embeds regression models in ANN settings. The work addresses the timely and pressing need to develop robust predictive models for rapidly changing morbidity risks and relevant impact on healthcare. Data from various sources are used, including the Office for National Statistics (ONS), NHS Open Data platform and insured population data provided by the Institute and Faculty of Actuaries (IFoA). The project includes the investigation of changes in morbidity rates for different diseases over time, and regional or socioeconomic differences. Further methodological work will be carried out towards assessing the robustness of the developed predictive models under a set of criteria designed to optimise the interpretability, predictive quality and uncertainty quantification. The research will provide robust models for morbidity rates, characterised for specific illnesses (e.g. cancer) and factors such as age, gender, and socio-economic status.