Biostatistics – Methods

Work Package 14b

Dementias Platform UK
Medical Research Council

Objective(s):

To develop and apply state-of-the-art statistical stratification methods to DPUK cohorts. It will also provide exemplar statistical analyses to test and demonstrate the utility of the informatics portal for integrated analyses.

More specifically the objectives were:

1. To develop robust (statistical) strategies/algorithms to identify risk stratified subgroups, which can be recruited either into Experimental Studies (including clinical trials) or followed for disease progression based on cross-sectional information.
2. To develop robust statistical strategies/algorithms for trajectory risk stratification that can identify subgroups of individuals in pre-clinical/prodromal stages of disease who have different progression rates to dementia using longitudinal information.
3. To statistically identify, incorporate and evaluate varying types and complexities of biomarkers (e.g. structural MRI, urine or blood markers) or combinations of biomarkers that associate with brain Aβ pathology or with onset of dementias or predictive of treatment response.

Overview Summary:

More accurate models can predict an individual's risk of rapid cognitive decline, the onset of dementia, contribute to understanding disease processes and offer insight into the effectiveness of potential new therapeutics. These models can also help identify individuals in the early stages of disease who would benefit from help in managing identified risks and may be good candidates for recruitment into clinical trials. This project proposed the development of appropriate statistical methodology for dementia risk stratification and prediction using individual-level data, obtained from DPUK cohorts accessed through the DPUK portal. The work would offer additional

insight as instead of a single summary measure, it would utilise multiple pieces of information related to risk (e.g. cognitive functioning, genetics, MRI volumes, age, sex, education) to provide more accurate and detailed models.

Executive Summary:

The use of a diverse set of clinical and biological markers, alongside genetics and other traditional risk factors can lead to more refined models for predicting an individual's risk of rapid cognitive decline or onset of dementia, increased understanding of disease processes and insight into potential new therapeutic targets. Additionally, these models can help identify individuals in the early stages of disease who are at highest risk and therefore potential candidates for recruitment into clinical trials, as well as help tailor management of subjects based on their estimated individualised risk. WP14B aimed to develop appropriate statistical methodology for dementia risk stratification and prediction using individual-level data, obtained from DPUK cohorts accessed through the DPUK portal. Our Bayesian semi-supervised mixture modelling methodology focussed on integrating data from different modalities, some of which are collected longitudinally, to identify subgroups of individuals who have particular profiles of cognitive, imaging and genetic profiles that are indicative of higher risk of disease progression. Due to various issues concerning timely access to data from the targeted DPUK cohorts through the DPUK portal and availability of genetic and biological data, through the portal, we developed our models primarily using ADNI and CFAS I. Analyses using linked ELSA cohort and ELSA METADAC genetic data obtained in 2020 are ongoing.

Summary of Outputs: (as per Researchfish categories)

Publications:

Rouanet, A, Richardson, SR, and Tom, BD (2020). Benefit of Bayesian clustering of longitudinal data: study of cognitive decline for precision medicine. Book chapter in "Bayesian Methods in Pharmaceutical Research" edited by E. Lesaffre, G. Baio and B. Boulanger

For book overview see- https://www.crcpress.com/Bayesian-Methods-in-Pharmaceutical-Research/Lesaffre-Baio-Boulanger/p/book/9781138748484

A Bayesian Dirichlet Mixture model with Gaussian Process priors for identifying subpopulations of patients with different covariate profiles which are linked to different cognitive functioning trajectories was produced. Four subpopulations with differing longitudinal cognitive trajectories linked to profiles described by 6 MRI volumetric imaging biomarkers, gender, APOE4 carrier status and educational attainment were described. One subpopulation is associated with steep cognitive decline and characterised by low levels of hippocampal and entorhinal cortex volume and high prevalence of APOE4 carriers and low proportion with 16 or more years of education. Persons identified as belonging to this cluster earlier on in their cognitive decline can be managed more intensively or be recruited into clinical trials. This methodology allows prediction of future cognitive decline in subjects based on covariate profiles and cognitive functioning history

Collaborations & Partnerships

Dr Brian Tom, the Principal Investigator, additionally contributed to the following work:

- Investigator on Deep and Frequent Phenotyping Project
- Investigator on EPAD
- Investigator on Alzheimer's Society grant on Stratified Cohort based on Dementia Risk

Further Funding

Dr Brian Tom, the Principal Investigator is a Co-investigator on the following grants:

- MRC grant entitled "Deep and Frequent Phenotyping: combinatorial biomarkers for dementia experimental medicine"
- Alzheimer's Society grant entitled "Bioresource - Genes and Cognition. Establishing a stratified population cohort of 100000 people recallable for pre-clinical studies of neurodegeneration and dementia"

Next Destinations

- Anais Rouanet has gone to the University of Bordeaux to work with Cecile Proust-Lima and is continuing to work on dementia
- Mary Fortune is a Teaching Associate in Medical Statistics and Assessment in the Public Health Education Group
- Steven Hill is still at the MRC Biostatistics Unit

Engagement Activities

2019

- B. Tom. Non-parametric clustering for longitudinal cognitive measurements, baseline imaging and genetic data. EcoSta 2019, Taichung, Taiwan 25th June 2019 (Invited Speaker)

- A. Rouanet, B. Tom, S. Richardson. Nonparametric clustering approach for longitudinal cognitive measurements, baseline imaging and genetic data in precision medicine (oral presentation), Channel Network Conference 2019, Rothamsted, UK
- A.Rouanet, S. Richardson, B. Tom. Bayesian nonparametric clustering from longitudinal cognitive measurements, baseline imaging and genetic data for precision medicine (oral presentation), ISCB 2019, Leuven, Belgium

2018

- B. Tom. Mixture models for stratifcation in dementia research. DPUK Next Generation Seminar: Analytics, Royal Statistical Society, London, UK (Invited Speaker)
- A. Rouanet, R. Johnson, S. Richardson, B. Tom. Dirichlet process mixture model for longitudinal data and side information for precision medicine: Study of cognitive decline (oral presentation), IBC 2018, Barcelona, Spain
- A. Rouanet. Bayesian outcome-driven mixture modelling of a longitudinal marker and profile variables: Precision medicine in Alzheimer's Disease (oral presentation), RSS 2018, Cardiff, Wales
- M. Fortune and A. Mander. Designing Dementia Trials Embedded within a Cohort (poster) ARUK 2018
- S. M. Hill, S. Richardson and B. D. M. Tom. Risk stratification for cognitive decline using genetic data (poster) ARUK 2018
- A. Rouanet, R. Johnson, S. Richardson, B. Tom. Identification of subgroups with specific cognitive evolution patterns and brain imaging profiles for precision medicine (poster) ARUK 2018

Influence of policy, practice, patients & the public

None

Research Tools & Methods

The R software (**PReMiuMar**) which allows Bayesian clustering to be extended to longitudinal data is now available at https://github.com/anarouanet/PReMiuMar. This software implements both the multivariate normal and Gaussian Process extensions to handle longitudinal data in this framework. The accompanying paper to this R software package extension is near completion.

Research Databases & Models

• An investigation into the utility of genetic markers and polygenic risk scores for rapid cognitive decline for the purpose of risk stratification has been undertaken. IGAP genetic and cognitive data from CFAS I and corresponding data from ADNI was used to perform this investigation. • Latent class mixed modelling methodology has been compared to Bayesian profile regression methodology for AD research. Further extensions of these methodologies for handling multivariate longitudinal outcomes and event history outcomes, for incorporating prior knowledge and for improving the efficiency and scalability of the MCMC algorithm have been undertaken. • An investigation into how to design dementia trials embedded within a cohort through simulation using latent class mixed models and informed by work done within EPAD. Specific focus was on using information already collected on individuals within the cohort to inform selective recruitment and adaption of clinical trials. We explore how to sample from a single disease cohort to form a trial population to balance the desire to estimate a trial effect with a particular power and conditional on this specified power, to estimate an extrapolated treatment effect in the cohort population with the highest precision. • The Group has played an important role in DPUK meetings to improve the accessibility, functionality and relevance of the DPUK Data Portal for research. It has provided feedback and identified issues and hurdles, together with sharing its practical knowledge of usage of the Portal for this type of work.
Intellectual property & licencing
None
Medical products, interventions & clinical trials
None
Artistic & creative products
None
Software & technical products
R software (**PReMiuMar**) which allows Bayesian clustering to be extended to longitudinal data is found at https://github.com/anarouanet/PReMiuMar.
Spin outs
None

Awards & recognition
Steven Hill received a DPUK travel award for his presentation at the DPUK 2017 Annual Scientific Conference.
Other outputs & knowledge/future steps
None
Use of facilities & resources
This project was designed around use of the DPUK Data Portal.
Most successful outcome and what it means for future dementia research:
The development of methodology that integrates in a coherent way the multiple domains/dimensions of risk without having to summarise as a risk score. The model (i) allows the identification of subgroups that are linked to different risk profiles, thereby helping to identify potential individuals for recruitment into clinical trials; (ii) allows the prediction of disease course (iii) allows further personalising of management/treatment strategies; and (iv) allows improved understanding of the disease and potentially allows for the identification of biomarkers that could be targeted for therapeutic development
Lessons learned:
• Lessons were learnt with regard to access to data and the assumption that getting data from the DPUK portal would be forthcoming in a timely fashion. Even with contingencies in place to get access to cohort data externally rather than through the portal, this took an extremely long time to complete. The following steps were necessary: formal reapplication of the project to the various Data Access Committees, studying the various datasets and selecting the variables required (which was extremely time-consuming), clarification of the role of DPUK in the applications (in particular, with regard to obtaining ELSA genetics via METADAC) and adding appropriate experts to deal with possible incidental findings. Additionally, after approval the University needed to put the data transfer agreements in place (e.g. Whitehall II) after seeking appropriate clarification plus sort out with DPUK the payments for access to some of these cohorts (such as UK Biobank and METADAC ELSA Genetics). As an illustration of the time taken, the process of getting access to the linked METADAC ELSA Genetic and Cohort data took over a year to complete. • Interaction with other researchers in dementia research and gaining exposure of this work through meetings went well • Useful skills were developed using the remote secure portal and data environment

• Further lesson learnt included clarifying whether the genetic data of DPUK Cohorts (if they exist) would be made available through the DPUK portal or externally through a separate application and understanding the arrangements involved in getting access to UK Biobank data and subsequent use through the DPUK portal.
Other:
Nothing further to report
Date of Report:
28 April 2020