

Workshop on Flexible Models for Longitudinal
and Survival Data with Applications in
Biostatistics (27th-29th July, 2015)

Poster Session

Acute kidney injury amongst chronic kidney disease patients: a case-study in statistical modelling

Authors: Ozgur Asar, James Ritchie, Philip A. Kalra, Peter J. Diggle
Lancaster University, UK.

Abstract

Chronic kidney disease (CKD) and acute kidney injury (AKI) are two important kidney related health problems. The former is defined as enduring kidney damage or decreased kidney function, the latter as a sudden fall in the kidneys' excretion function. These two conditions are now recognised to be strongly associated. In this study, we investigate the influence of AKI on long-term kidney function amongst patients already diagnosed with CKD. We develop a longitudinal statistical model with Matern correlation structure and multivariate t-distributed stochastic components. We also specify three change-points that describe the typical trajectory of an AKI event. We use maximum likelihood estimation implemented with an expectation-maximisation algorithm to estimate the model parameters, and best linear unbiased prediction to predict the random effects. Our case-study uses data from an on-going cohort study of chronic kidney patients at Salford Royal NHS Foundation Trust, Salford, UK.

Comparison of different regression models of Joint Modelling for multivariate longitudinal data and survival process

Authors: Ipek Guler, Christel Faes, and Carmen Cadarso-Suárez
University of Santiago de Compostela, Spain.

Abstract

Many follow-up studies produce different types of outcomes including both longitudinal measurements and time-to-event outcomes. Commonly, it is of interest to study the association between them. The joint modelling approaches of longitudinal and survival data have gained an increasing attention recently. Various authors have proposed joint modeling for a single longitudinal marker and a time-to-event process [Wulfsohn and Tsiatis (1997); Tsiatis and Davidian (2004); Henderson et al (2000)]. However, many studies collect several longitudinal biomarkers and the interest may be in the relationship between these biomarkers and the survival outcome. Existent joint modelling approaches are difficult to implement when the number of random effects is large. Due to the computational complexity, it's hard to evaluate the joint likelihood of multivariate longitudinal and survival data. Fieuws and Verbeeke (2005) and Fieuws, Verbeke and Molenberghs (2007) have proposed an approach for the multivariate longitudinal processes where all the possible pairs of longitudinal data are separately modeled and then combined in a final step. We use this approach to model multivariate longitudinal biomarkers and then relate to the survival data with a two-stage approach. This paper presents a comparison of this two-stage approach with different regression models to study the multivariate longitudinal data and survival process.

Semiparametric extension to glm

Authors: Mark Hannay and Aeberhard, dalhousie
University of Geneva.

Abstract

Recently Alan Huang (2013), introduced an extension of glm, where the error distribution is treated as an infinite-dimensional parameter, which is to be estimated simultaneously with the mean model parameters. His approach relies on empirical likelihood, requiring an estimation of n weights on top of p parameters of the model mean, where n is the number of observations. As the number of observations increases the number of parameters to estimate increases linearly, making the estimation increasingly computationally intensive for bigger data sets. We propose a likelihood based method to reduce the number of parameters to estimate, making this extension computationally feasible for larger number of observations. There is a trade-off between the number of dimensions allowed and the accuracy of estimation, which we then discuss.

Joint modeling of autoantibodies and type 1 diabetes onset: substantive results and methodological challenges

Authors: Meike Köhler, Sonja Greven, Andreas Beyerlein, Ezio Bonifacio,
Anette Ziegler
Helmholtz Zentrum München

Abstract

Motivation. The onset of clinical symptoms of type 1 diabetes (T1D) is preceded by a preclinical phase in which insulin-autoantibodies (IAA) can be detected in the blood. Whereas the overall level of this marker has shown to be predictive of T1D onset we used joint modeling of the longitudinal marker levels to gain further insights into disease progression.

Methods. We applied shared random-effect models with direct, slope and cumulative link as implemented in the R-package JM-bayes as well as joint latent class models as implemented in the package lcmm on data of the prospective birth cohorts BABY-DIAB and BABYDIET. For the analysis we used 134 children (943 autoantibody measurements) of 2441 children who showed the marker of insulin autoantibodies persistently. Of these 76 progressed to clinical T1D.

Results. Both approaches indicated that higher IAA levels over time are associated positively with the hazard for T1D progression: In shared random-effect models higher IAA levels were associated with greater risk and a two group joint latent class model differentiated between trajectories in the subjects overall marker level. However fitting the models posed several challenges: Firstly, the heterogeneity of the marker profiles demanded a flexible modeling of the individual trajectories (e.g. modelling time with fixed and random spline effects) which was problematic with regards to stability and convergence. Secondly the discriminative ability of the shared random-effect models varied highly over time. Finally convergence of models with more than four groups and the most appropriate specification of the baseline hazard were unsolved issues in the latent class approach.

Discussion. Joint models allow for additional insights into the relationship between IAA trajectories and clinical onset of T1D. To allow for more flexible modeling of trajectories we aim at developing a P-spline based approach that may also include time-varying effects of the longitudinal marker on the hazard. A larger data set is available for the evaluation of more complex models.

Survival Analysis Models on MESS epilepsy data.

Authors: Boryana Cristina López Kolkovska
University of Warwick, UK.

Abstract

In the Multicentre study of early Epilepsy and Single Seizures (MESS) study, patients diagnosed with epilepsy were subjected to a randomized controlled trial policies of immediate versus deferred treatment. We are interested in providing a prognosis for the patients and clinicians from the study, and propose to apply Survival Analysis methods. A first approach proposed by J. Rogers is presented, where the times to a first seizure are modeled by a negative binomial mixture model, with a gamma distributed random effect. The model considers the existence of a proportion of patients who attain remission post-randomization, and terms such proportion of patients as a cure fraction of the population. For this model each patient is considered to have an individual seizure rate, which is assumed to change when randomized to an anti-epileptic drug. The underlying seizure rate, post-randomization rate change for each patient and the population's heterogeneity coefficient are estimated from Rogers model.

For this type of live recurrent events, Cox proportional hazards models are commonly used. We perform a residual analysis on Cox and Rogers models, in order to compare and study their goodness of fit. From such results and the inherent nature of patients diagnosed with epilepsy, we present a truncated negative binomial mixture model. An alternative survival analysis model for this problem arises by considering a competing risks model, where patients types of epilepsy, dropout and death are proposed as the competing risks in the population.

Joint Dispersion Model with a Flexible Link

Authors: Rui Martins

Centro de Investigação Interdisciplinar Egas Moniz (ciiEM).

Abstract

Our objective is to model and analyze longitudinal and survival data jointly in order to take into account the dependence between the two responses that are observed in one subject using a flexible link which is time-dependent and specified as a penalized spline (P-Spline) allowing for a considerable flexibility. In addition, longitudinal individual standard deviation is considered as a covariate in the hazard model.

Keywords: Longitudinal analysis; Survival analysis; Flexible link.

Time-dependent ROC analysis for censored event-time data

Authors: Adina Najwa Kamarudin, Trevor Cox, Ruwanthi
Kolamunnage-Dona
University of Liverpool, Liverpool, UK.

Abstract

Receiver operating characteristic (ROC) analysis is extensively used in the biomedical studies for evaluating the diagnostic accuracy of a continuous marker. The sensitivity and specificity of the corresponding diagnostic test are the basic accuracy measurements of a marker, used to plot the ROC curve for all possible cut-off points. It evaluates marker on the basis of the area under the ROC curve (AUC) as a concordance index, with a larger AUC indicating a more powerful marker. The disease status is considered a binary and time-fixed outcome in the classical ROC analysis. However, many disease outcomes change with time, and hence ROC analysis that varies as a function of time is more appropriate in health research. Often, event-time outcomes are censored which bring additional complexity in time-dependent ROC analysis. In this study, we will compare the standard and time-dependent ROC using a real dataset.

Keywords: ROC, AUC, event-time, time-dependent, censoring.

Correlated Reversible Multi-state Models with Random Effects

Authors: Aidan G. O’Keeffe, Brian D. M. Tom, and Vernon T. Farewell
University College London, UK.

Abstract

In many scenarios, observational data are collected on multiple continuous-time processes of interest for a number of subjects seen at discrete time points. Whilst each process can be modelled separately using a multi-state model, the processes are likely to be correlated within each subject. We explore the use of random effects to account for such correlation, an approach for which there is a limited literature. Specifically, we develop a modelling framework that allows the incorporation of random effects when fitting a series of general multi-state models, including reversible multi-state models. We describe the use of standard statistical software to fit such models and demonstrate this for the modelling of two renal processes in a cohort of systemic lupus erythematosus patients.

Joint modelling approach for analyzing the effect of background health status on elderly insureds' mortality risk

Authors: Ramon Alemany, Montserrat Guillen, Xavier Piulachs
Riskcenter, University of Barcelona.

Abstract

The developed world is experiencing a significant growth in its population aged 65 and over, which not only means people are living longer but that they tend to face a greater number of years with a range of health problems. Survival is quite clearly related to individuals age, but it is also dependent on health conditions and habits, as well as on subject-specific characteristics. In this regard, a jointly modelling of longitudinal and survival information enables us to estimate the degree of association between the demand for medical care within a health insurance company and time-to-death, serving at the same time as a useful tool for predicting risk of death based on historical medical background.

In our work we study both the medical resource usage and the longevity from a large dataset of insured individuals aged 65 years or older. In particular, yearly measurements of each subject's emergency room and ambulance service use are collected, and occurrence of mortality is also monitored. The study aims to capture the simultaneous dependence between their demand for healthcare and the corresponding time until policyholders death event. A shared-parameter joint model scheme is implemented by incorporating information on longitudinal demand for medical care through a weighted cumulative effects parameterization, as the demand impact decreases as further is a health reading with respect the current time point. Alternative weighting specifications are tested, showing in all cases that a joint approach indicates a positive and strong connection between health care demand and death point. Likewise it allows us to estimate individual survival curves which update dynamically as new information on the demand for services becomes known.

Comparison of methods for analyzing longitudinal data truncated by death

Authors: [Anaïs Rouanet](#), Dienabou Sylla, Hlne Jacqmin-Gadda
Université de Bordeaux – ISPED.

Abstract

In ageing studies, the collection of longitudinal markers is often truncated by dropout and death. Marginal models estimated by Generalized estimating equations (GEE) and mixed models estimated by maximum likelihood have been proposed to study longitudinal data and several articles (Robins et al., 1995, Williamson et al. 2009) focused on their properties in the presence of drop out. In particular, marginal models are not robust if data are not missing completely at random (MCAR) but this can be corrected by weighting methods, rendering them robust to data missing at random (MAR). On the contrary, mixed models are robust to data MCAR and MAR but the predicted outcome can be interpreted as the evolution of the longitudinal marker in a population with no attrition.

When data are truncated by dropout and/or death, the interpretation gets more complex, in particular regarding the fixed effects estimates. Kurland (2009) considers that the fixed effects in mixed models can be interpreted in the framework of immortal cohort only whereas we think that they also measure individual change in mortal cohort, since they quantify the effect of covariates on the marker, adjusted on the random effects.

On the other hand, the fixed effects of GEE represent the effect of the covariates on the marker in the population still alive and observed at the current time. Using weighting methods such as the inverse probability to be observed, given that the subject is alive, or the inverse probability to be observed, we can estimate either the effect of covariates on the marker in the alive population or in the population with no dropouts and no deaths, respectively.

Within different data generation patterns for dropout and death, we compared the consistency, efficiency and robustness of weighted GEE and mixed models through simulations. To check the above interpretation of predicted values of the outcome, we also compared the mean trajectories estimated with these methods to the conditional expectation of the outcome according to the time at which the subject is currently alive.

Scalable flexible survival analysis with Gaussian processes

Authors: [Alan D. Saul](#), Neil D. Lawrence
University of Sheffield.

Abstract

We propose a scalable flexible survival analysis approach within the Gaussian process framework. The method allows for non-linear responses to changes in the covariates, along with heterogeneity in the shape of the tails of the parametric distribution chosen. We rely on a variational approximation that can be trained on mini-batches of data, allowing it to scale well with the number of data available for training.

Analysing epidemiological data on bovine TB in a badger population using a Bayesian latent time-to-event model

Authors: Neil Walker
Oxford Biomedical Research Centre.

Abstract

Woodchester Park in Gloucestershire is the site of a long-term project looking at the epidemiology of bovine tuberculosis in a local badger population using mark-recapture methods. Here we present details of a framework for modelling epidemiological processes in this population using a Bayesian MCMC algorithm. Individual animals contribute a clinical “life-history” consisting of results from (i) ELISA antibody (ii) microbial culture tests taken on occasions where badgers are trapped. Using prior information on the sensitivity and specificity of the two diagnostic tests, a posterior distribution for a putative date of infection is derived for each animal. This distribution is then used as the basis of a latent time-to-event analysis (“latent” since the true date of infection is not known). The time to infection is modelled as a piecewise exponential process and regressed on a range of covariates, both variant and invariant in time. Model output in terms of both performance and results is discussed.

Previous Injury and Match Load as Risk Factors for Injury in Professional Rugby Union Players: Application of a Frailty Model for Recurrent Events

Authors: [Sean Williams](#), Grant Trewartha, Simon P. T. Kemp, John H. M. Brooks, Colin W. Fuller, Aileen E. Taylor, Matthew J. Cross, and Keith A. Stokes

University of Bath, Bath, UK

Abstract

Background: Given the relatively high risk of injuries in professional Rugby Union, there is a need to identify and understand risk factors for injury in this population, whilst also accounting for the correlated nature of this data.

Objective: To investigate the influence that previous injury and match loads have upon injury risk in professional Rugby Union players, through the application of a frailty model for recurrent events.

Design: A seven season (2006/7-2012/13) prospective cohort design was used to record time-loss injuries.

Setting: English Premiership Professional Rugby Union.

Participants: 1253 professional Rugby Union players.

Assessment of Risk Factors: A player's previous injury history (injury load; a product of the number of previous injuries and time-weighted average severity), the number of matches played in the preceding season (previous season match load), and the number of matches played in the previous 30 days (month match load) were assessed.

Main Outcome Measurements: The frailty model was applied to calculate adjusted hazard ratios (HRs) for each of the assessed risk factors.

Results: A two standard deviation (2-SD) increase in the injury load (2-SDs=170 AU) and previous season match load (2-SDs= 18.4 matches) variables was associated with HRs of 1.40 (95% CI: 1.25-1.58) and 1.22 (95% CI: 1.06-1.40), respectively. The relationship between injury load and injury risk followed an 'inverted-U' shaped curve, with an apparent peak in risk between 189-224 AU. Inferences relating to a 2 SD increase in month match loads were unclear (HR: 1.02, 95% CI: 0.89-1.16). Subset analyses revealed a reduced risk of non-contact injuries for players with month match loads above the median value (HR: 0.86, 95% CI: 0.76-0.98).

Conclusions: High previous injury loads and playing a high number of matches in the preceding season are overall risk factors for injury in this population, while high month match loads were associated with a reduced risk of non-contact injuries. These data may be used to identify players at an increased injury risk, so that appropriate interventions can be made to prevent/reduce injuries.