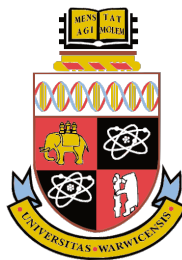


“Workshop on Flexible Models for Longitudinal and Survival Data with Applications in Biostatistics”

27th-29th July, 2015,
University of Warwick, Coventry, UK.



Contents

1	Schedule	3
2	Abstracts: Keynote Speakers	5
3	Abstracts: Contributed Talks	13
4	Poster titles	22

1 Schedule

The workshop will take place in the Zeeman Building, University of Warwick, Coventry, UK, 27th–29th July, 2015, (Building no. 38 on the campus map). All talks will take place in MS.03 room, Maths/Stats (Zeeman) Building.

Day 1

09:00 - 09:50	Registration (Tea and Coffee)
09:50 - 10:00	Welcome
10:00 - 10:55	Vern Farewell
11:00 - 11:55	Peter Müller
12:00 - 13:30	Lunch
13:30 - 14:00	Ziqi Chen
14:00 - 14:30	David Hughes
14:30 - 15:00	Francisco Javier Rubio
15:00 - 15:30	Tea and Coffee
15:30 - 16:20	Break
16:30 - 18:30	Poster session + wine reception

Day 2

09:00 - 09:55	Rebecca Betensky
10:00 - 10:30	Tea and Coffee
10:30 - 11:00	Walter Dempsey
11:00 - 11:30	Jose S. Romeo
11:30 - 12:00	Catalina A. Vallejos
12:00 - 13:30	Lunch
13:30 - 14:00	Paul Blanche
14:00 - 14:30	Markus C. Elze
14:30 - 15:00	Qiuju Li
15:00 - 15:30	Tea and Coffee
15:30 - 16:25	Arnošt Komárek
16:30 - 17:25	Bernard Rachet
19:00 -	Gala Dinner

Day 3

09:00 - 09:55	Michael J. Crowther
10:00 - 10:30	Tea and Coffee
10:30 - 11:25	Cecile Proust-Lima
12:00 - 13:30	Lunch
13:30 - 14:25	Dimitris Rizopoulos

2 Abstracts: Keynote Speakers

Analyzing the influence of a time-varying biomarker process on time to event

Speaker: Rebecca Betensky
Harvard School of Public Health

Abstract

The influence of biomarkers on the risk of diagnosis of Alzheimer's disease is of interest for understanding the pathological progression of the disease, as well as for drug development. Currently, PET scan imaging of amyloid and tau in the brain are of considerable interest, as previously measures of neuropathology were available only at autopsy. However, this imaging is expensive, and is typically obtained only at one or two time points during a study. This raises the challenge of how to analyze the role of this longitudinal biomarker that is measured at a single time-point, which may not be comparable across subjects relative to the time origin (e.g., onset of impairment or birth). We examine the implications of treating the time-varying biomarker measured at study entry as a baseline covariate when adjusting for delayed entry in a time to event analysis. We also consider sigmoidal models for the biomarker, with a variety of error models, to enable treatment of the biomarker as time varying. We conduct these investigations using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) as well as in simulation studies.

Flexible parametric joint modelling of longitudinal and survival data

Speaker: Michael J. Crowther

University of Leicester, Leicester, UK and Karolinska Institutet, Stockholm, Sweden

Abstract

Joint modelling of longitudinal and survival data is now widely used in a variety of clinical areas such as cancer, AIDS and cardiovascular disease, with the implementation of user friendly software making great progress in recent years. In this talk, I will describe recent developments of the Stata package `stjm`, which implements shared parameter joint models within a maximum likelihood framework. In particular, I will discuss the ability to model multiple longitudinal outcomes within the multivariate generalised linear mixed effects framework, incorporating delayed entry, and the calculation of conditional survival predictions. I will illustrate the package using a cardiovascular disease dataset, and will concentrate on the use of restricted cubic splines to model both the trends in the longitudinal outcome(s) over time, and the baseline hazard function, providing a highly flexible modelling framework. Finally, I will describe some current work incorporating sampling weights which opens up the possibility of using joint models more readily in large registry based clinical datasets.

Two Tools for the Analysis of Longitudinal Data: Motivations, Applications and Issues

Speaker: Vern Farewell
MRC Biostatistics Unit, Cambridge, UK

Abstract

Two tools for the analysis of longitudinal analysis will be discussed; multi-state models and two part models. The use of multi-state models for a variety of applications will be illustrated to demonstrate their usefulness in the specification of data structures and their flexibility. These applications will involve the challenges of panel data, adjustment for highly variable time-dependent covariates and correlated processes. The application of causal reasoning in the context of multi-state models will also be briefly discussed. With a primary focus on semi-continuous data, the use of two part models with random effects will also be examined. The role of correlated random effects and issues related to marginal and subject specific regression coefficients will be highlighted.

Regression modelling of misclassified correlated interval-censored data

Speaker: Arnošt Komárek
Charles University in Prague, Czech Republic

Abstract

Research presented in this talk is motivated by a need to assess the effect of different predictors on time to caries experience (CE) in the permanent dentition, research being motivated by a longitudinal oral health study, the Signal Tandmobiel study, conducted in Flanders (North of Belgium) in 1996-2001. For each child, time to event (time to CE) is recorded for several teeth which asks to deal with a survival regression model for correlated (clustered) data. Further, occurrence of the event is only checked at pre-specified (annual) dental examination. This would classically lead to interval-censored event times. Nevertheless, as soon as the occurrence of the event is diagnosed by a classification procedure with imperfect sensitivity and/or specificity we obtain so called misclassified interval-censored responses. This was also the case for the Signal Tandmobiel study where the occurrence of the CE was diagnosed by one of 16 dental examiners whose caries scoring was not free of classification errors. At the same time, a particular child was possibly examined by different examiners at different visits and more importantly, a particular examiner did not have information available on caries classification recorded at the previous visits. That is, it can be assumed that at each visit, event (caries) status was evaluated by a not necessarily perfect examiner whose evaluation was independent of all previous evaluations of the event status. Observed data on time to event (time to CE of one tooth) are then represented by a not necessarily monotone sequence of zeros and ones and corresponding visit times. Recorded zero means that at a particular visit, it was (possibly incorrectly) determined that the event has not occurred yet. Similarly, recorded one indicates that it was determined, again possibly incorrectly, that the event has already occurred. Analogous type of data is encountered whenever the event status is regularly checked at pre-specified occasions by an imperfect classification procedure (laboratory assessment etc.) In this talk, we show possibility of (a) regression modelling of such misclassified interval-censored data (which can additionally be correlated) and (b) estimation of characteristics of the classification process (its sensitivity and specificity).

Nonparametric Bayesian survival regression with variable dimension covariate vector

Speaker: Peter Mueller
University of Texas Austin

Abstract

Motivated by inference for a study of targeted therapies in cancer we develop a nonparametric Bayesian survival regression that allows for a variable number of covariates to be recorded for different subjects. That is, we do not require for all covariates to be recorded for all subjects. The proposed model is based on a random partition of patients, with the partition including a regression on covariates. The key feature of this construction is that a prior on patient-specific cluster membership can be specified on the basis of available covariates, without requiring the imputation of missing covariates. We introduce the model construction and outline strategies for posterior simulation. Finally we show how the proposed nonparametric survival regression is used in the design of the motivating trial for targeted therapies.

Examples of joint models for multivariate longitudinal and multistate processes in chronic diseases

Speaker: Cécile Proust-Lima
INSERM & University of Bordeaux

Abstract

Joint models for longitudinal and survival processes have become the keystone of the statistical analyses in chronic diseases where continuous processes along with times to progression are of interest. Most developments initially focused on a single Gaussian longitudinal marker and a right-censored time to event. However, chronic diseases usually involve much more complex data with multiple longitudinal markers and multiple causes and stages of progression. This is the case for example in prostate cancer progression after treatment, and in Alzheimers disease (AD). In this talk, we detail several extensions of the standard joint modelling framework to analyze the multivariate processes involved in these two chronic diseases. Multiple times of progression are taken into account in a competing event setting (for AD diagnosis and competing AD-free death) and more generally in a multistate setting (for AD onset along with death, or for the different types of recurrences and death in prostate cancer). Multiple longitudinal markers encountered in AD (repeated cognitive tests and repeated dependency indicators) are analyzed through latent process models. Joint models always rely on a conditional independence assumption which means that the longitudinal and survival processes are linked by a latent structure that captures the whole correlation between the two processes. In these works, the latent structure is either latent classes (shared discrete latent variable) or random effects (shared continuous latent variable). Our models are mostly parametric and are estimated in the Maximum Likelihood framework. In each application, the models are specifically shaped to fit at best the data: flexible distributions are considered, different specifications are compared and main assumptions are properly tested.

Missing data and net survival analysis

Speaker: Bernard Rachtel

The London School of Hygiene & Tropical Medicine

Abstract

Net survival from cancer, a key metric for cancer control policy, is the survival which would be observed if the patients could die only from their cancer. On the mortality scale, the excess hazard of death from cancer is the analogue of net survival. The overall hazard of death is assumed to be the sum of the excess hazard (due to cancer) and the expected hazard (due to other causes). When the cause of death is not reliably recorded (i.e. within the relative survival setting), the expected hazard is estimated from the general population with socio-demographic characteristics similar to the cancer patients. Unbiased estimation of net survival can be obtained using a non-parametric estimator accounting for informative censoring or using a multivariable, flexible excess hazard model, which also enables the effects of co-variables (e.g. tumour stage at diagnosis) to be estimated using the excess hazard ratio.

Incomplete data, a common concern in research, are an even more prevalent issue in routine data such as those collected by population-based cancer registries. The use of ad hoc methods for handling missing data (e.g. complete-case analysis, mean substitution, missing indicator) can severely affect the inferential validity of the analysis. More appropriate approaches have been developed, such as expectation-maximization algorithm, inverse probability weighting, full Bayesian analysis. Among them, multiple imputation (MI) based on Rubin's rules has demonstrated its broad applicability while being relatively simple. Caveats exist in specific situations such as net survival analysis.

The MI approach first requires the imputation model to be properly specified. This means that the imputation model contains the variables which determine the missing data and the substantive model has to be "nested" within the imputation model. In addition to common issues related to non-linearity of the relationships and interactions, further difficulties occur when the parameter estimated in the substantive model (here excess hazard) does not correspond to the final outcome of interest (net survival). These issues will be illustrated with missing tumour stage, using cancer registry data, and recommendations will be made.

Personalized Screening Intervals for Biomarkers using Joint Models for Longitudinal and Survival Data

Speaker: Dimitris Rizopoulos
Erasmus MC

Abstract

Screening and surveillance are routinely used in medicine for early detection of disease and close monitoring of progression. Biomarkers are one of the primary tools used for these tasks, but their successful translation to clinical practice is closely linked to their ability to accurately predict clinical endpoints during follow-up. Motivated by a study of patients who received a human tissue valve in the aortic position, in this work we are interested in optimizing and personalizing screening intervals for longitudinal biomarker measurements. Our aim in this paper is twofold: First, to appropriately select the model to use at time t , the time point the patient was still event-free, and second, based on this model to select the optimal time point $u \geq t$ to plan the next measurement. To achieve these two goals we develop measures based on information theory quantities that assess the information we gain for the conditional survival process given the history of the subject that includes both baseline information and his/her accumulated longitudinal measurements.

3 Abstracts: Contributed Talks

Dynamic predictions from joint models of longitudinal and time-to-event data: a note on R^2 -type curves

Speaker: Paul Blanche

University of Copenhagen, Department of Biostatistics.

Abstract

Boosted by the growing interest in personalized medicine, joint modeling of longitudinal marker and time-to-event data has recently started to be used for making dynamic predictions of patient prognosis. By dynamic predictions we mean predictions of a clinical event which are updated as soon as the time of making predictions changes. Such dynamic predictions take into account the whole information about the marker trajectory of the patient at the time of making predictions. Several authors have recently studied dynamic ROC curves and expected Brier scores to evaluate dynamic prediction capacities (Lawless 2010; Rizopoulos, 2012; Blanche et al. 2014).

In this work, we specifically discuss the use of R^2 -type curves to assess the accuracy of such dynamic predictions. We give several arguments to show that the R^2 -type curves provide reliable and useful predictive accuracy information which is easily interpretable. A non-parametric inverse probability weighting estimator is suggested to deal with censoring. Large sample results are provided. They enable the computation of confidence regions for pointwise and simultaneous inference.

Using data from the French DIVAT cohort (www.divat.fr), a detailed application on kidney transplantation is presented. Firstly, we jointly modeled serum creatinine evolution and risk of graft failure or death with a functioning graft from a learning sample of 2749 kidney recipients. Secondly, dynamic predictions of return to dialysis or death were estimated by using serum creatinine evolution and baseline characteristics from an independent sample of 1300 patients. Finally, we assessed the prognostic capacities of dynamic predictions by using dynamic ROC curves, expected Brier scores and R^2 -type curves.

A profile likelihood approach to longitudinal data

Speaker: Ziqi Chen

School of Mathematics and Statistics, Central South University, Changsha, China.

Abstract

Longitudinal data arise frequently in biomedical and health studies in which repeated measurements from the same subject are correlated. In order to attain efficient estimators, the existing papers need correct specification of the correlation structure. If the full likelihood function is known, we could achieve the maximum likelihood estimators of the regression parameters, which is efficient. However, it is difficult to know the true correlation matrix, and to specify the full likelihood function when responses are non-normal because of the correlated nature of longitudinal data. By regressing the error on its predecessors, we treat the prediction error density as an unknown nonparametric function and propose to estimate it by kernel smoothing. With the estimated prediction error density, we achieve the estimators of the regression parameters by maximizing the so-called profile likelihood. Our proposed method performs well without specification of correlation structure as well as the likelihood. We show that the proposed estimator is efficient in both theory and practice.

Survival models and health sequences

Speaker: Walter Dempsey
University of Chicago

Abstract

Medical investigations focusing on patient survival often generate not only a failure time for each patient but also a sequence of measurements on patient health at annual or semi-annual check-ups while the patient remains alive. Such a sequence of random length accompanied by a survival time is called a survival process. Ordinarily robust health is associated with longer survival, so the two parts of a survival process cannot be assumed independent.

This talk is concerned with a general technique—temporal realignment—for constructing statistical models for survival processes. A revival model is a regression model in the sense that it incorporates covariate and treatment effects into both the distribution of survival times and the joint distribution of health outcomes. It also allows the sequence of health outcomes to be used clinically for predicting the subsequent trajectory, including the residual survival time.

Incorporating reference ranges from healthy individuals in joint longitudinal and time-to-event modelling

Speaker: Markus C. Elze
London School of Hygiene & Tropical Medicine (UK)

Abstract

The incorporation of time-varying data in time-to-event models is a common objective in areas where longitudinal measurements are collected at arbitrary time points, such as clinical trials or the social sciences. Joint modelling of longitudinal measurements and time-to-event data is a natural solution to this problem, but the amount of available data may limit the use of joint models. Here, we show that transforming the longitudinal data using additional information from external sources may increase the amount of information gained from the data.

‘Bone marrow transplantation’ is a potentially curative treatment option for different hematologic disorders, such as severe leukaemia. However, it is still associated with high mortality rates due to complications after transplantation. Early identification of high-risk patients is crucial for successful intervention. Thus, predictive models are needed to assist clinical decision making. The development of longitudinal immune measurements is relevant for complication prediction and should be considered in modelling. As studies are often faced with limited patient numbers, assessing the immune recovery may be difficult. Here, we demonstrate how the use of reference data from healthy individuals may assist the model fitting.

Age- and sex-dependent reference ranges are created from a dataset of healthy children for several immune subpopulations using the LMS method (Cole et al., 1992). These are then employed to assess the immune recovery for paediatric patients who underwent bone marrow transplantation for severe leukaemias or similar diseases. The performances of joint models with and without the use of reference ranges are compared. The benefits and drawbacks of these models for the evaluation of clinical studies and for dynamic prediction in clinical practice are discussed.

Flexible Discriminant Analysis Using Multivariate Mixed Models

Speaker: David Hughes

Department of Biostatistics, University of Liverpool, England.

Abstract

Discriminant function analysis can be used to make predictions of the group to which an individual most likely belongs (e.g., disease/healthy, risk groups). Linear mixed models are well established for the analysis of longitudinal data. However, little is known about the extent to which the inclusion of longitudinal information improves classification. An approach to discrimination using longitudinal data via extension of the linear mixed models has been recently proposed (Komarek et al. 2010). Our aim is (i) to investigate how the accuracy of classification improves through the inclusion of longitudinal information and (ii) to investigate the effect of including patient covariate information on the accuracy of prediction. We extend the existing methods for discrimination by allowing longitudinal profiles to be modelled using multivariate generalized linear mixed models and allowing cross-sectional data to be used as discriminators, and not just as covariates of the longitudinal profiles. This framework for longitudinal discriminant analysis is flexible in three key ways: (i) the use of generalised linear mixed models allows for a variety of data types to be included in the same model as discriminators (developing the ideas of Fieuws et al. 2008 and Komarek and Komarekova 2013), (ii) the use of mixture distributions for the random effects relaxes typical distributional assumptions (Komarek et al. 2010) and (iii) the use of covariate information as potential discriminators provides a more patient specific approach to classification. We apply the proposed methodology to data from a large longitudinal study of over 12,000 patients diagnosed with diabetes and classify these patients into various groups based on their risk of developing sight threatening diabetic retinopathy. Our aim is to correctly identify patients who are at a high risk in order to provide treatment in time. Accurate prediction would allow patients at low risk to be screened less frequently, significantly reducing the cost to the NHS, whilst also maintaining high screening efficiency.

Accommodating informative dropout and death: a joint modelling approach of longitudinal and semi-competing risks data

Speaker: Qiuju Li
MRC Biostatistics Unit

Abstract

In longitudinal studies, both dropout and death can truncate the observations of outcomes of interest. In certain scenarios, it is of interest to obtain a longitudinal outcome profile of some population given they are being alive (mortal cohort) since extrapolation of the profile beyond death is not meaningful. Focusing on directly modelling the conditional longitudinal profile of being alive, partly conditional models (Kurland and Heagerty, 2005) with weighted estimating equations have been proposed. In this work, to accommodate both informative dropout and death, we develop a new likelihood-based approach to jointly model a longitudinal outcome and event times of dropout and death, where the dropout and death are treated as semi-competing risks that are associated with the longitudinal outcome process through random effects. A major advantage of our approach is that the conditional longitudinal profile of being alive can be obtained in a closed form. The estimation is implemented in both maximum likelihood and Bayesian frameworks. We illustrate the methods by CD4 count data with dropout and HIV-related death from the HIV epidemiology research study (HERS).

The Power Variance Function Copula Model in Bivariate Survival Analysis: An application to Twin Data

Speaker: Jose S. Romeo

Department of Mathematics, University of Santiago, Chile.

Abstract

In this work we present the Power Variance Function (PVF) copula family to model the dependence structure of multivariate lifetime data. The PVF copula family includes the Clayton, Positive Stable (Gumbel) and Inverse Gaussian copulas as special or limiting cases. Dependence properties of the copula models are explored and described. We adapt a new general simulation technique to simulate from the PVF copula. We suggest a Bayesian approach to parameter estimation of the PVF copula using MCMC for posterior computation. A simulation study is performed to investigate the small sample properties of the estimates. Finally, we illustrate the usefulness of the methodology using data from the Australian NH&MRC Twin registry. Parameters of the marginal distributions and the PVF copula are simultaneously estimated in a parametric as well as a semiparametric approach where the marginal distributions are modelled using Weibull and piecewise exponential distributions, respectively.

Linear mixed models with improper priors and flexible distributional assumptions for longitudinal and survival data

Speaker: Francisco Javier Rubio
University of Warwick, UK.

Abstract

We propose a Bayesian approach involving improper priors for hierarchical linear mixed models with flexible random effects and residual error distributions. The error distribution is modelled using scale mixtures of normals, which can capture tails heavier than those of the normal distribution. This generalisation is useful to produce models that are robust to the presence of outliers. We present general results for the propriety of the posterior that cover cases with point and censored observations, allowing for the use of these models in the contexts of longitudinal and survival analysis. We consider the use of copulas with flexible marginals for modelling the dependence between the random effects, but our results cover the use of any random effects distribution.

Incorporating unobserved heterogeneity in Weibull survival models: A Bayesian approach

Speaker: Catalina A. Vallejos
MRC Biostatistics Unit.

Abstract

Flexible classes of survival models are proposed that naturally deal with both outlying observations and unobserved heterogeneity. We present the family of Rate Mixtures of Weibull distributions, for which a random effect is introduced through the rate parameter. This family contains the well-known Lomax distribution and can accommodate flexible hazard functions. Covariates are introduced through an Accelerated Failure Time model and we explicitly take censoring into account. We construct a weakly informative prior that combines the structure of the Jeffreys prior with a proper prior on the parameters of the mixing distribution, which is induced through a prior on the coefficient of variation. This improper prior is shown to lead to a proper posterior distribution under mild conditions. The mixing structure is exploited in order to provide an outlier detection method. Our methods are illustrated using two real datasets, one concerning bone marrow transplants and another on cerebral palsy.

4 Poster titles

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.

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.

Semiparametric extension to glm

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

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

Survival Analysis Models on MESS epilepsy data.

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

Joint Dispersion Model with a Flexible Link

Authors: Rui Martins
Centro de Investigação Interdisciplinar Egas Moniz (ciiEM).

Time-dependent ROC analysis for censored event-time data

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

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.

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.

Comparison of methods for analyzing longitudinal data truncated by death

Authors: Anaïs Rouanet, Dienabou Sylla, Hlne Jacqmin-Gadda
Université de Bordeaux – ISPED.

Scalable flexible survival analysis with Gaussian processes

Authors: Alan D. Saul, Neil D. Lawrence
University of Sheffield.

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.

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