

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

## **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](http://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.