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

1 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 timepoint, 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 Rachet
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 primarily 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 ¿ 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.