Abstracts - Keynote speakers

- **Per Kragh Andersen, University of Copenhagen.**
  "Measures of lost years of life"

In survival analysis much emphasis is on the hazard function and hazard ratios are frequently used as summary measures when comparing the survival distributions among various groups. The main reason why expected values are used less frequently is of course the presence of right-censoring that often prevents inference for the tail of the survival time distribution. As an alternative parameter, the restricted mean life time may be studied. We will first discuss how (cause-specific) measures of life years lost can be defined via the restricted mean life time. Second, we will summarize a substantial number of suggestions that has been put forward to quantify life years lost for patients suffering from a given disease compared with the general population and we will give some recommendations. Data from Danish patients with bipolar disorder will be used as illustration.

- **Gerda Claeskens, Katholieke Universiteit Leuven.**
  "Variable selection methods for survival data"

Several variable selection methods will be considered for survival data, with particular attention to mixture cure models where part of the population is insusceptible to the event of interest.
Variable selection is relevant for each model part, the logistic regression model for the probability of insusceptible cases, and the Cox proportional hazard part to model the, often right-censored, survival times of the susceptible cases.
Next to the Akaike information criterion, a more directed search with the focused information criterion is explained, leading to selected models that produce more accurate estimators.

- **Emmanuel Lesaffre, Katholieke Universiteit Leuven.**
  "A stochastic EM algorithm for doubly interval censored data"

In clinical trials, it is frequently of interest to estimate the time between the onset of two events (e.g. duration of response in oncology). Here we consider the case where subjects are assessed at fixed visits and both the origin (denoted U) and the onset of the time of interest (V) happen in-between the visits. This type of data, called doubly interval censored (DI) is often analyzed with standard methods assuming that the onset and end of the time of interest are measured exactly. We investigate the performance of simplification of DI data and existing methods to analyse DI data.

We propose an estimation method using the Stochastic EM algorithm that overcomes the problems in the existing approaches. We illustrate our method on a clinical trial in Chronic Myeloid Leukemia (CML) in chronic phase.

- **Yi Li, University of Michigan.**
  "Modeling complex large-scale time-to-event data: An efficient quasi-Newton approach"

Nonproportional hazards models often arise in modern biomedical studies, as evidenced by a recent national kidney transplant study. During the follow up, the effects of baseline risk factors, such as patients' commorbidity conditions collected at transplantation, may vary over time, resulting in a weakening or strengthening of associations over time. Time-varying survival models have emerged as a powerful means of modeling the dynamic changes of covariate effects. Traditional methods of fitting time-varying effects survival model rely on an expansion of the original dataset in a repeated measurement format, which, even with a moderate sample size, leads to an extremely large working dataset. Consequently, the computational burden increases quickly as the sample size grows, and analyses of a large dataset such as our motivating example defy any existing statistical methods and software. We propose a novel application of quasi-Newton iteration method, via a refined line search procedure, to model the dynamic changes of covariates' effects in survival analysis. We show that the algorithm converges superlinearly and is computationally efficient for large-scale datasets. We apply the proposed methods to analyze the national kidney transplant data and study the impact of potential risk factors on post-transplant survival.

- **Thomas Scheike, University of Copenhagen.**
  "Modelling the dependence in time to event data"

I will review some recent work on how to assess the dependence for family data in the context of competing risks and survival data. The dependence can subsequently by decomposed into a genetic and environmental component using random effects models and genetic assumptions. I will present several models that can be used for this. One useful way of describing the consequences of these models on the observed data is to compute for example the concordance rate, i.e., the probability that two family members experience the event of interest, in the competing risks setting. This is a number that has traditionally been computed together with the casewise concordance, the probability of both having the event given that one is a case, in the genetic literature without careful consideration of censoring and time issues.

The motivation for this work has been the interest in characterizing the "heritabilty" of specific diseases such as for example prostate or breast cancer.

- **Hans C. van Houwelingen, Leiden University Medical Center.**
  "Reduced-rank modeling as an approach to reduce complexity in event history analysis"

Reduced-rank models are a well-established tool in applied data analysis. Generally speaking, it can be applied when a matrix of unknown parameters has be estimated. The assumption that the true matrix is of low rank can greatly reduce the number of parameters. One application is principal components analysis, another one is testing for interaction in a two-way analysis of variance with a single observation in each cell.

In event history two applications were explored in a NWO-project (Dutch NSF) "Survival Analysis for Complicated Data" at our department in the period 2002-2006.

1. Multistate models, where the matrix of parameters contains the regression coefficients of different covariates (rows) for different columns. These results are published with Marta Fiocco as first author.

2. Cox models with time-varying effect of covariates, where the matrix of parameters contains the regression coefficients of the covariates on a suitable basis time-functions. These results are published with Aris Perperoglou as first author.

The talk will concentrate on the second application (time-varying effects). It will review the history of the approach, the technical methodology, the software and the comparison with similar approaches (Perperoglou, SiM, 2007).

Finally, on a more general level, the relation well be discussed between this type of time-varying effect models and direct modeling of the predictive probabilities by "landmarking" as advocated in Van Houwelingen and Putter, Dynamic Prediction in Clinical Survival Analysis, Chapman and Hall, 2012.

- **Lan Wang, University of Minnesota.**
  "A unified approach of quantile regression analysis of survival data under biased sampling**"**

Biased sampling occurs frequently in economics, epidemiology and medical studies either by design or due to data collecting mechanism. Failing to take into account the sampling bias usually leads to incorrect inference. We propose a unified estimation procedure and a computationally fast resampling method to make statistical inference for quantile regression with survival data under general biased sampling schemes, including but not limited to the length-biased sampling, the case-cohort design and variants thereof. We establish the uniform consistency and weak convergence of the proposed estimator as a process of the quantile level. We also investigate more efficient estimation using the generalized method of moments and derive the asymptotic normality. We further propose a new resampling method for inference, which differs from alternative procedures in that it does not require to repeatedly solve estimating equations. It is proved that the resampling method consistently estimates the asymptotic covariance matrix. The unified framework proposed in this paper provides researchers and practitioners a convenient tool for analyzing data collected from various designs. Simulation studies and applications to real data sets are presented for illustration. (Joint work with Gongjun Xu, Tony Sit and Chiung-Yu Huang)

Abstracts - Invited speakers

- **Mailis Amico, Université catholique de Louvain.**
  "Estimation of a mixture cure model with a single-index structure"

In medicine (as in other fields), the classical survival analysis assumption that all patients will experience the event of interest (if followed long enough) is often not realistic (e.g. time to progression of a cancer, time to pregnancy…). Cure models extend classical survival analysis models to the case where a "cure" or a "immune" fraction is present. The mixture cure model is one of the two broad classes of cure models that have been proposed in the literature. Considering that the population of interest is a mixture of cured and uncured individuals, the model is composed of two elements, the incidence part referring to the probability of being uncured, and the latency part corresponding to the survival function of the uncured observations.

Most often, the probability of being uncured is modelled parametrically assuming a logistic regression model. In this research, we propose a more flexible approach. Our proposal is to consider a single index structure for the incidence part, which offers more flexibility than a parametric approach but avoids the curse of dimensionality phenomenon encountered in nonparametric modelling. We develop an estimation method based on the EM algorithm and we propose a kernel estimator for the unknown link function in the single-index model.

Based on simulations, we study the finite sample size performance of our proposed method. We also present an application of our methodology on a real medical dataset and contrast our result with those obtained assuming a logistic regression model for the incidence part.

- **Liesbeth de Wreede, Leiden University Medical Center.**
  "Prediction in multi-state models when data are missing"

One of the difficulties hampering the application of multi-state models in clinical studies is missing data, both for outcomes and for predictors. We will illustrate this problem and our proposals for solutions in an observational dataset describing a multi-center cohort of CLL (chronic lymphocytic leukemia) patients who have undergone a hematopoietic stem cell transplantation. This dataset, collected by the European Society for Blood and Marrow Transplantation, contains information about pre-transplant risk factors and post-transplant events. We focus on the probability of treatment success and the impact of covariates and intermediate events on this outcome. Missing outcome data were imputed based on clinical knowledge whereas missing baseline covariates were imputed by means of multiple imputation by chained equations (implemented in the mice package in R). A semi-parametric (Cox model based) Markov multi-state model was fitted in mstate in R. Covariates for the different transitions were selected on the basis of a combination of clinical and statistical arguments. Estimates of transition probabilities and their associated standard errors were then pooled by Rubin's rules to yield predicted outcome probabilities with pointwise confidence intervals for good and poor risk reference patients.

- **Candida Geerdens, Universiteit Hasselt.**
  "Nonparametric copula estimation for clustered right-censored event time data"

In many scientific studies the response of interest is the time until a predefined event (e.g., the time to tumor appearance or the time to death). Often, this event time is right-censored for some items in the study sample, i.e., only a lower time bound for the event is observed (e.g., by the end of the study still no tumor has appeared or the study item is still alive). Another difficulty that can arise is the clustering of data (e.g., the study might include only twins, each twin then acts as a group of size two). Since clustered items share common traits, their event times are correlated.

Copulas provide a useful tool to assess the dependence in grouped time-to-event data. In a data setting where it is less evident to predetermine a parametric copula, a nonparametric copula approach is preferable. We define a new nonparametric copula estimator for the joint survival function of grouped right-censored event time data. In here, we consider two right-censoring schemes: univariate censoring and copula censoring. For the new nonparametric copula estimator, we establish the consistency and we investigate the finite sample performance in various data settings via a simulation study. A comparison with the recent nonparametric copula estimator of Gribkova and Lopez (2015) is made. It is shown that depending on the percentage of censoring, the strength of the data association and the data dimension either the new estimator or the Gribkova-Lopez estimator is to be preferred.

- **Olivier Lopez, Université Pierre et Marie Curie.**
  "Censored regression trees with applications to nonlife insurance"

In this work, we provide consistency results for censored regression trees in view of applying them in nonlife insurance. The approach we develop is adapted to heterogeneous observations sets, since it produces a decomposition of the population into several classes, on which the regression function is estimated. We derive the consistency of the procedure we use to select the number of appropriate classes to correctly fit the data. The specificity of our technique is that it can be used for estimating the regression function, but also to estimate the value of the censored observations contained in the sample. We illustrate our method to forecast the final amount of medical malpractice claims that are not fully settled.

- **Valentin Patilea, ENSAI.**
  "Modified cox regression with right censored and current status data"

Let T be a lifetime or a failure time and Z a vector of covariates. To account the effect of covariates on the duration T we assume a proportional hazards model. This model has been largely studied in presence of incomplete data. We consider the case where the covariates Z are observed and where the duration T is either observed or it is left or right censored (current status). A latent variable model is introduced, that allows us to derive a Breslow type estimator of the cumulative baseline hazard rate function for a fixed value of the regression parameter, and then to write a profile likelihood function, which depends on the regression

parameter only. Finally, estimators of all the unknown parameters are derived. Based on general asymptotic results for profile likelihood estimators, and using empirical processes theory, several results of consistency and weak convergence for the estimators are obtained. The semiparametric efficiency of the regression parameter estimator is also derived. The extension to the case where data contain cured individuals is also considered. Finite distance properties of the estimators are investigated through a Monte Carlo study and an application to a real data set.

(This is joint work with M.L. Avendano, L. Bordes, M.C. Pardo)

- **Francois Portier, Université catholique de Louvain.**
  "On proportional hazards cure models"

In this paper we introduce a new semiparametric model that accounts for the presence of cure individuals in the data, i.e. subjects that will never experience the event of interest. We consider the nonparametric maximum likelihood estimator (NPMLE). First, by relying on a profile likelihood approach, we show that the NPMLE may be computed by a single maximization over a set whose dimension equals the dimension of the covariates plus one (as for the Cox estimation procedure). Following an empirical process approach, we derive the asymptotics of the NPMLE and show the semi-parametric efficiency of the procedure. Since the variance is difficult to estimate, we develop a weighted bootstrap procedure that allows for a consistent approximation of the asymptotic law of the estimators.

- **Sylvie Scolas, Université catholique de Louvain.**
  "Diagnostic checks in the presence of interval-censoring and cure"

In Alzheimer's disease, studying the time until Mild Cognitive Impairment (MCI) can be crucial since it may be a precursor of the disease. Contrary to what is generally assumed in survival analysis, not all of the patients will develop signs of MCI. This special feature is more and more encountered, and we refer to this as the presence of cure fraction. Also, as in many other medical studies, patients come at scheduled interviews. Thus, the event times are not exactly observed, but are interval-censored, in addition of the possibility to be right-censored as usual.

When considering diagnostic checks in survival analysis, the definition and use of residuals from linear regression must be adapted to take account of right-censoring. Proposals have been made, mainly in the context of the right-censoring Cox and AFT regression models. For example, the martingale residuals are commonly used to detect non-linearity in a covariate. While cure models are gaining into popularity, literature on model checking in this context is very sparse, and even sparser in the case of interval-censored data. Yet, an adaptation to mixture cure models is not straightforward and is worth further study.

In this talk, we briefly review the issues occurring when data are right-censored, in a parametric AFT or Cox regression model. Secondly, we discuss the difficulties encountered in the presence of interval-censoring, as well as in the cure mixture model, if there actually is a fraction of cured individuals in the data. We finally illustrate the results thanks to simulations and to the Alzheimer study.

Abstracts - Posters

- **Steven Abrams, Hasselt University.**
  "A general correlated gamma frailty model for bivariate time-to-event data"

Modelling the association among multivariate event times in survival analysis is often done using shared and correlated frailty models (see, e.g., Duchateau and Janssen, 2008, Wienke, 2010) [1,2]. Shared frailty models impose a perfect correlation among individuals within the same cluster, whereas correlated frailty models imply a less restrictive association structure. In spite of the flexibility of the correlated frailty model, correlated frailties are often defined through an additive decomposition in terms of independent shared and unshared components across the frailties. Although this `variable-in-common' approach proposed by Yashin et al. (1995) [3] is mathematically convenient and straightforward to implement, the implied non-negative correlation coefficient is bounded above due to the additive structure of the event- or cluster-specific frailties. More specifically, whenever the frailty variances differ substantially, the upper bound becomes very restrictive. We present a generalization of the additive correlated gamma frailty model to obtain a computationally tractable bivariate gamma distribution, useful to model correlated time-to-event data, thereby relaxing the potentially restrictive upper bound in the additive model. Our methodology is illustrated on right-censored Danish twin data on cause-specific mortality and bivariate current status data derived from cross-sectional serology on hepatitis A and B infections in Flanders, Belgium.

References :
[1] Duchateau, L. and Janssen, P. (2008). The Frailty Model. Berlin: Springer.
[2] Wienke, A. (2010). Frailty Models in Survival Analysis. Boca Raton: Chapman & Hall/CRC.
[3] Yashin, A.I., Vaupel, J.W., and Iachine, I.A. (1995). Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. Mathematical Population Studies, 5(2), 145-159.

- **Theodor A. Balan, Leiden University Medical Center.**
  "Event-based ascertainment of recurrent events data"

Recurrent events are increasingly common in clinical studies. Often, such data is collected retrospectively from hospital registries. In this case, the ascertainment of the patients can depend on the history of the recurrent event process; for example, only patients who have experienced at least one event during a pre-specified (calendar) time period are selected. We propose a flexible model and a likelihood-based procedure which takes the selection process into account when estimating the model parameters.

In the context of recurrent events, it is known that the selection conditions must be considered in developing an analysis (Cook and Lawless 2007) , however a general approach appears to be lacking. Furthermore, the event-based ascertainment criteria seem to be ignored in several studies of recurrent risks, and although it has been suggested that this might be the source of

some paradoxes in clinical research (Dahabreh and Kent 2011) , the consequences on the estimation procedure have not been extensively studied. The same holds in particular for frailty models, in which a random effect is used to explain unaccounted heterogeneity.

We adapt existing frailty models for recurrent events by accounting for the event-based ascertainment, while analysing the pitfalls associated with ignoring the selection process. We derive consistent estimators of the model parameters under different ascertainment schemes and study the small sample properties of these estimators through a simulation study. Finally, we illustrate the proposed methods on a data set comprising of recurrent pneumothoraces, for subjects who were ascertained if they had experienced at least one event during a certain time window.

- **Aurelie Bertrand, Université catholique de Louvain. "Correcting for the bias due to mismeasured covariates in a survival cure model"**

Cure models are survival models taking into account the existence of subjects who will never experience the event of interest. In addition to this first feature, dealing with covariates measured with some error (which causes biased estimators) happens quite often in practice.

Ma and Yin (2008) [2] suggested a method to correct the bias in one of the main cure models, the promotion time cure model. In Bertrand et al. (2015) [1], we propose an alternative approach, based on the SIMEX method, which has the advantage of being very intuitive. While both approaches have good asymptotic properties, they rely on the assumption that the measurement error must have a known and constant variance and its distribution must be Gaussian, assumptions which are not always met in practice.

After describing the Ma and Yin (2008) [2] approach as well as our proposed SIMEX-based approach we present the results of an extensive simulation study investigating the robustness of both approaches with respect to their assumptions.

Based on these simulations, we give some practical recommendations about whether to correct for the measurement error, about which method to be used depending on the objective of the study, and about the consequences to be expected when misspecifying the distribution of the measurement error, or its variance. We conclude by illustrating these issues in the analysis of real medical data.

References :
[1] Bertrand, A., Legrand, C., Carroll, R.J., De Meester, C., Van Keilegom, I. (2015) Inference in a Survival Cure Model with Mismeasured Covariates using a SIMEX Approach. Biometrika (under revision).
[2] Ma, Y., Yin, G. (2008). Cure Rate Model With Mismeasured Covariates Under Transformation. Journal of the American Statistical Association, 103, 743-756.

- **Justin Chown, Université catholique de Louvain. "Nonparametric estimation of the error distribution from a location-scale modeling of cure model data"**

A common problem encountered in survival studies is that some subjects never view an event of interest, and are said to always survive. Consequently, these subjects are always censored, and therefore create an identifiability problem. For example, it may be the case that participants in a heart attack study never experience a heart attack even though they are believed to be "at risk". A correct characterization of this data requires accounting for the proportion of those individuals who are censored because they will never experience a heart attack. Unlike general censoring models, the identifiability of the "cured proportion of individuals" implies an inherent structure in the cure model (mixture) that allows us to view it as a special case of a censored data model where censoring is never heavy enough to distort inference. We review some key concepts that allow us to define crucial assumptions that are necessary for data to be adequately described by a cure model (mixture). Then we study the distribution function of the standardized errors of a completely nonparametric location-scale modeling of this data. Our techniques are demonstrated on a data set from a leukemia study.

- **Mickaël De Backer, Université catholique de Louvain. "Copula Quantile Regression with Censored Data"**

Quantile regression is a common way to investigate the possible relationships between a d-dimensional covariate X and a response variable T. Since it was introduced by Koenker and Bassett (1978) as a robust (to outliers) and flexible (to error distribution) method, quantile regression has received notable interest in the literature of theoretical and applied statistics as a very attractive alternative to the classical mean regression method that captures only the central tendency of the data. In survival analysis, the quantile regression approach allows the analyst to estimate the functional dependence between variables for all portions of the conditional distribution of the (possibly) right-censored response variable. In that sense, quantile regression provides a more complete view of relationships between T and X and constitutes an alternative to popular regression techniques like the Cox proportional hazards model or the accelerated failure time model. In this work, under the usual assumption of conditional independence between the survival time and the censoring time, we consider a new class of estimators that would allow practitioners to analyse in a flexible way medium to high-dimensional censored data using quantile regression. Actually, our methodology is an extension of the recent work of Noh et al. (2013) [1] and Noh et al. (2015) [2] to allow for the presence of censoring. Accordingly, in a similar spirit as for the case without censoring, the methodology makes use of the advantages of copulas in dependence modelling as the main idea consists of expressing the characterization of the quantile regression in terms of a multivariate copula and marginal distributions. However, in order to bypass the effects of a possible misspecication of the underlying copula, we propose in this work an alternative semiparametric estimation scheme for the multivariate copula density, driven by the regression context. Asymptotic normality of our estimator is obtained under classical regularity conditions, and numerical examples are used to illustrate the validity and finite sample performance of our procedure.

References:
[1] Noh, H., El Ghouch, A. and Bouezmarni, T. (2013), Copula-Based Regression Estimation and Inference. Journal of the American Statistical Association, 108, 676-688.
[2] Noh, H., El Ghouch, A. and Van Keilegom, I. (2015), Semiparametric Conditional Quantile Estimation through Copula-Based Multivariate Models. Journal of Business and Economic Statistics, 33(2), 167-178.

- **Achim Dörre, Universität Rostock.**
  "Bayesian estimation of a proportional hazards model for double-censored durations"

Purpose: Double-censored data consist of uncensored, left-censored and rightcensored data and occur frequently in survival time contexts. Bayes estimators of a parametric proportional hazards model for random durations subject to double-censoring are investigated. In particular, we prove consistency and asymptotic normality of the Bayes estimators with respect to quadratic loss.

In addition, estimators of standard errors are derived. The proposed model is applied to rating class data from a large German bank. Initial rating class effect and time effects for rating transitions are analysed.

Design and Methodology: Bayes estimators are implicitly defined as the minimizing argument of expected loss with respect to the posterior distribution of the parameters. In the given setting, no conjugate prior distribution is available. Therefore, the computation is performed by use of MCMC algorithms for which logarithmic ratios of the posterior density are derived.

A simulation study is conducted in order to analyse the finite-sample performance in comparison to maximum likelihood estimation.

Results and Practical Implications: It turns out that parametric Bayes estimators and confidence intervals can be derived in a consistent manner and are asymptotically normal. Their performance in finite samples is satisfying and does not require asymptotic justifications. It is argued that the proposed Bayes estimators are a reasonable alternative to maximum likelihood estimation for small and moderate sample size. The analysis suggests that an upgrade of a rating increases the duration in that class by about ten days on average.

- **Juha Karvanen, University of Jyvaskyla.**
  "Visualizing complex study designs"

The first step in the analysis of complex time-to-event data is to understand the processes that generated the data. This means recognizing the study design and characterizing the missing data mechanism (including censoring and truncation mechanism). If the goal is to estimate causal effects, the causal structure should be specified as well. Communicating all this information is not a trivial task and the risk of misunderstandings exists. For instance, commonly used terms such as "case-control design" and "missing at random" are not specific enough to describe the study design and the data collection exactly.

Motivated by these difficulties I present an extension structural causal models that describes the missing data mechanism and the study design together with the causal structure [1]. These "causal models with design" define the study design formally in a mathematically precise way. The visualization as a directed acyclic graph describes the flow of the study by ordering the nodes of the causal diagram in two dimensions by their causal order and the time of the observation.

Two examples are presented. The first example illustrates the case-cohort design used in the MORGAM Project where the aim is to estimate the impact of classic and genetic risk factors on the risk of cardiovascular diseases. The second example illustrates the optimization of the data collection in follow-up studies where the longitudinal covariate measurements can be conducted only for a subset of the cohort [2].

References:
[1] J. Karvanen (2015). Study design in causal models. Scandinavian Journal of Statistics, 42(2), 361{377. Available also as arXiv:1211.2958v4.
[2] J. Reinikainen & J. Karvanen (2015). Bayesian subcohort selection for longitudinal covariate measurements in follow-up studies, Submitted.

- **Mia Klinten Grand, Leiden University Medical Center.**
  "Dynamic prediction of cumulative incidence functions by direct binomial regression"

In recent years there has been a development of methods for dynamic prediction of the cumulative incidence function (CIF) in a competing risk setting. These models enable the predictions to be updated as time progresses and more information becomes available, e.g. when a patient comes back for a follow-up visit after completing a year of treatment, the risk of death and adverse events may have changed since treatment initiation.

One approach to model the CIF is by direct binomial regression (Scheike et. al , 2008). We have extended this approach by combining it with landmarking to enable dynamic prediction of the CIF. The proposed models are very flexible as they allow the covariates to have complex time-varying effects and we introduce tests to investigate these possible time-varying structures. The method handles incompletely observed event times by inverse probability weighing of individuals who have experienced the event of interest and the models can be fitted by generalized estimating equations. One advantage of the method, compared to modeling the cause-specific hazards, is the direct interpretation of the effect of the covariates on the cumulative incidence. The method is illustrated with data from bone marrow transplant patients.

- **Carlo Lancia, Leiden University Medical Center.**
  "Counterfactual osteosarcoma survival by marginal structural proportional hazards"

To date, it is still unclear what role is played by different combinations of cytotoxic drugs, drug dosages, and time to complete chemotherapy in regimens for osteosarcoma on

survival outcomes. In analysing longitudinal data collected during randomised trials, particular attention is required due to the presence of a time-dependent confounder such as toxicity. This confounder is strictly patient's related as a consequence of the exposure to cytotoxic agents. Both drug doses and starting date of the next chemotherapy cycle are dynamically allocated to each patient depending on the toxicity levels through to the end of the last cycle.

Data used in this analysis come from EURAMOS-1, a large study in resectable osteosarcoma which includes two randomised controlled trials. The aim of the trial is to determine the effect of post-operative chemotherapy based on histological response on survival outcomes. Administered dose of (up to) 5 drugs were recorded for each patient and for each cycle, as well as the CTCAE (Common Terminology Criteria for Adverse Events) grade for more than 20 different toxicities.

The analysis performed focuses in particular on the role played by a specific cytotoxic agent (Methotrexate) on both histological response and survival outcomes. This is a specific clinical question addressed by the chairman of the protocol of EURAMOS-1. Answering to this question requires the use of ad-hoc statistical methodology such as Marginal Structural Models.

These models can deal with both complexity and longitudinal nature of the data, and the delicate but crucial interplay between toxicity and allocated treatment. Under the main assumption of no unmeasured confounding, these models give unbiased estimates for the parameters of the model of interest (in this context logistic regression for the histological response and Cox model for the survival outcome) by means of Inverse Probability of Treatment Weighting (IPTW). The effect of IPTW is to create a pseudo-population where (i) the allocated treatment is no longer confounded by toxicities, and (ii) the causal parameters of interest are the same as in the true population. In other words, unbiased estimates of the parameters of interest can be obtained by standard crude analysis on the pseudo-population generated by IPTW.

- **Samuel Maistre, Université catholique de Louvain.**
  "Spline backfitted kernel estimation of a nonparametric additive model in the presence of right-censored data."

Spline backfitted kernel (SBK) method is a two-step estimation method for nonparametric additive models that has been developed during the last decade. See [1, 2, 3, 4, 5, 6]. We adapt this method to the case of right censoring.

Suppose the random vector $(X,Y) \in [0,1]^D \times \mathbb{R}$ *follows the model* $Y = \mu + \sum_{d=1}^{D} m_d(X^d) + \varepsilon$ *where the response* Y *is subject to right censoring by the variable* C*. The new estimator we propose follows the two steps of SBK :*

1. *Assuming independence between* (X,Y) *and* C*, we use the Kaplan-Meier estimator of the c.d.f. of* C *to construct synthetic responses* $\hat{Y}_i$*.*
   Next, we resolve a least squares problem to obtain splines estimates $\hat{\mu}$ *and* $\hat{m}_d(.)$*,* $1 \leq d \leq D$*.*

2. *The final estimator for say* $m_1(.)$ *is obtained by using some kernel smoothing procedure with response* $\hat{Y} - \hat{\mu} - \sum_{d=2}^{D} \hat{m}_d(X^d)$ *(that mimics* $m_1(X^1) + \varepsilon$*) and covariate* X^1*. Other functions estimates are obtained in the same way.*

We present asymptotic results and a simulated example to illustrate the method.

References :
[1] J. Horowitz, J. Klemelä, E. Mammen. Optimal estimation in additive regression models. Bernoulli, 12(2):271–298, 2006.
[2] R. Liu, L. Yang. Spline-backfitted kernel smoothing of additive coefficient model. Econometric Theory, 26(1):29–59, 2010.
[3] R. Liu, L. Yang, W. K. Härdle. Oracally efficient two-step estimation of generalized additive model. J. Amer. Statist. Assoc., 108(502):619–631, 2013.
[4] S. Ma, L. Yang. Spline-backfitted kernel smoothing of partially linear additive model. J. Statist. Plann. Inference, 141(1):204–219, 2011.
[5] J. Wang, L. Yang. Efficient and fast spline-backfitted kernel smoothing of additive models. Ann. Inst. Statist. Math., 61(3):663–690, 2009.
[6] L. Wang, L. Yang. Spline-backfitted kernel smoothing of nonlinear additive autoregression model. Ann. Statist., 35(6):2474–2503, 2007.

- **Ivana Malá, University of Economics, Prague.**
  "The modelling of unemployment duration in the Czech Republic"

The Labour Force Sample Survey is performed quarterly by the Czech Statistical Office and provides thorough information about employment or unemployment in the Czech households. Data of all unemployed respondents from five consecutive quarters (IVQ 2013 - IVQ 2014) are used. Unemployment duration was derived as interval censored for those who found a job and right censored otherwise, 100 % censoring occurs and no exact values are included in the survey data. In the poster a probability distribution of unemployment spell is modelled with the use of finite mixtures of lognormal distributions. Known component membership (given by gender and education) is used to construct the mixture. Maximum likelihood method is used and all computations are performed in R. Strong positive impact of education on unemployment duration is shown (with the use of estimated quantile characteristics of unemployment duration and residual median). The difference between positions of men and women on the labour market (quantified by unemployment rate as well as by unemployment duration characteristics) is less apparent although the problem of unemployment is more serious for women than for men.

- **Bizhan Shabankhani, Mazandaran University.**
  "Factors affecting creatinin changes in hemodialysis patients using linear random effects model"

Survival analysis for patients with ESRD (ESRD) and factors influencing their survival, have become an important issue due to increase in number of these patients along with their high mortality rate (10-20 times compared to the general population), globally. Despite the development in renal replacement therapy, dialysis is still one of the main ways of care and

survival for the patients with ESRD. Around 70% of these patients are under haemodialysis treatment.

Despite recent advances, mortality in patients with ESRDs remains high all around the world. Some studies showed a one-year mortality rate of 20% among patients undergoing haemodialysis in the United States . However, as survival is depended on the level of care that patients received, survival rate varies in different countries.

Ministry of Health and Medical Education has reported the prevalence of renal failure as 2.5 per 1000 population in Iran. It was estimated that there is an annual growth of 12% in the prevalence of this disease in the year 2006. Low rate of renal replacement (24 grafts in 106 patients) make haemodialysis the main therapy for this patients in Iran. Haghighi et al. reported that in Iran, 53.7% of ESRD cases use haemodialysis as their treatment modality, peritoneal dialysis is very low (<1%) and home haemodialysis is not performed.

The purpose of this study was to analyse the survival rate of patients undergoing haemodialysis and factors influencing their survival using Cox proportional Hazard regression model, which is one of the main and the most common used methods for survival analysis.

- **Geert Silversmit, Belgian Cancer Registry.**
  "Cure of cancer in the Flanders region of Belgium"

Introduction :
A variable fraction of cancer patients will not die due to the cancer and can in terms of life expectancy be regarded as statistically cured. This so called cured proportion will experience the same death hazard as the general population. The remaining proportion of the cancer patients, the fatal cases, experiences an excess death hazard rate with respect to the general population. The cure of cancer within a population can therefore be quantified by the proportion cured and the mean survival time of the fatal cases.

Study Goals :
The cure of cancer for seven cancer sites (cervix, colon, corpus uteri, malignant melanoma of the skin, pancreas, stomach and oesophagus) in the Flemish Region for the incidence years 1999 to 2011 was estimated and gender and age differences were explored. This work represents the first results on cure of cancer in Belgium.

Methods :
Flemish residents older than 14 years diagnosed with at least one primary malignant tumour in the period 1999-2011 were considered. Vital status was obtained by linkage with the Belgian Crossroads Bank for Social Security with a follow-up up to the 1st of July 2013. Relative survival was calculated using the Ederer II method aggregated in time intervals of 0.5 year from 0 to 14 years of follow-up. The cured proportion and mean survival time of fatal cases were estimated from mixture cure models based on a non-linear fit to the aggregated relative survival curves. Exponential and Weibull distributions for the survival times of the fatal cases were evaluated.

Results :
The follow-up of 14 year was not sufficient to reach the cured plateau for cancer of the oesophagus, while statistical cure was observed for all other examined cancer sites.

The overall estimated cured proportion ranged from 6% for pancreatic cancer to 81% for melanoma of the skin. Cured proportions decreased with increasing age group; e.g. for cervix cancer more than 80% for the youngest age group (15-44 year) while only 35% for women older than 65 year. Higher cured proportions were observed for females compared to males for cancer of the colon, cancer of the stomach, cancer of the pancreas and melanoma of the skin.

Conclusions :
A follow-up period of 14 years was sufficient to estimate cure of cancer in the Flemish Region for six cancer sites (cervix, colon, corpus uteri, melanoma of the skin, pancreas and stomach) diagnosed from 1999 to 2011 using parametric mixture cure models applied to grouped relative survival curves. The expected proportion cancer survivors and the mean survival time of fatal cases are of direct interest for cancer patients.

- **Chien-Lin Su, National Chiao Tung University.**
  "Statistical analysis for two-level hierarchical clustered data"

Multi-level hierarchical clustered data are commonly seen in financial and biostatistics applications. In this talk, we introduce several modeling strategies for describing the dependent relationships for members within a cluster or between different clusters (in the same or different levels). In particular we will apply the hierarchical Kendall copula, first proposed by Brechmann (2014), to model two-level hierarchical clustered survival data. This approach provides a clever way of dimension reduction in modeling complicated multivariate data. Based on the model assumptions, we propose statistical inference methods, including parameter estimation and a goodness-of-fit test, suitable for handling censored data. Simulation and data analysis results are also presented.

- **Majda Talamakrouni, Université catholique de Louvain.**
  “Parametrically guided local quasi-likelihood with censored data”

It is widely pointed out in the literature that misspecification of a parametric model can lead to inconsistent estimators and wrong inference. However, even a misspecified model can provide little valuable information about the target function. This prior parametric knowledge has been noticed and used in a flexible way by a parametrically guided nonparametric approach in order to improve the fully nonparametric estimator. First introduced by Hjort and Glad (1995) in the context of density estimation, the guided approach has been later investigated in different frameworks due to its nice properties. This includes for example, nonparametric regression by Glad (1998), local quasi-likelihood by Fan et al. (2009) and very recently generalized additive models by Fan et al. (2014). Our contribution is concerned with parametrically guided local quasi-likelihood estimation that we adapt to randomly right censored data. The generalization to censored data involves synthetic data and the well known Beran (1981) estimator. To simplify our presentation, local linear fitting is investigated. The asymptotic properties of the guided estimator as well as its finite sample performance are studied and compared with the unguided quasi-likelihood estimator via simulation studies. The results confirm the bias reduction property and show that using an appropriate guide and the optimal bandwidth the proposed estimator outperforms the classical local quasi-likelihood estimator.