Workshop

Methods for time-to-event data from the life sciences with a special focus on clustered data

18./19. November 2021

der Arbeitsgruppen “Statistische Methoden in der Medizin“ (IBS-DR), „Statistische Methoden in der Epidemiologie“ (IBS- DR, DGEpi, DGSMP), „Statistische Methoden in der klinischen Forschung“ (GMDS), „Epidemiologische Methoden“ (DGEpi, GMDS, DGSMP)

Programm:

Donnerstag, 18. November

10:15 - 10:25	Begrüßung
	Session 1 *(Chair: Sarah Friedrich + Anne Lotz)*
10:25 - 11:05	Matthias Schmid (Bonn): Competing risks analysis for discrete time-to-event data
11:05 - 11:30	Sabrina Schmitt (Koblenz): Ereigniszeitanalyse mit konkurrierenden Risiken unter Berücksichtigung von Clusterstrukturen – Methodenvergleich anhand einer Simulationsstudie
11:30 - 11:55	Ann-Kathrin Ozga (Hamburg): Accelerated failure time models for semi-competing risk data with recurrent events
11:55 - 13:00	Mittagspause
	Session 2 *(Chair: Nicole Rübsamen + Kerstin Rubarth)*
13:00 - 13:25	Marc Ditzhaus (Dortmund): How to deal with nonproportional hazards in factorial survival designs?
13:25 - 13:50	Alexander Seipp (Oldenburg): Accelerated failure time models for crossing survival curves
13:50 - 14:15	Christoph Wies (Darmstadt): Testing VIMPs for Dependencies in Random Forest Analyses: Methods and an Application to Post Transplant Survival
14:15 - 14:30	Pause
	Session 3 *(Chair: Ralph Brinks + Irene Schmidtmann)*
14:30 - 15:10	Andreas Wienke (Halle-Wittenberg): Correlated random-effects models for clustered time-to-event data
15:10 - 15:30	Diskussion

Freitag, 19. November

9:00 - 9:55	AG-Sitzung und Wahlen *(Wahlleitung: Irene Schmidtmann)*
	Session 4 *(Chair: Philipp Mildenberger + Juliane Hardt)*
10:00 - 10:40	Niel Hens (Hasselt University & University of Antwerp, Belgium): Time-varying frailty models and the estimation of heterogeneities in transmission of infectious diseases

10:40 - 11:05 Maximilian Bardo (Göttingen): The Addams family of discrete frailty distributions for multivariate survival data

11:05 - 11:30 Markus Schepers (Mainz): How to model the spreading of infectious diseases using networks embedded into hyperbolic space?

11:30 - 11:45 Abschluss und Diskussion

Competing risks analysis for discrete time-to-event data

Matthias Schmid

Institut für Medizinische Biometrie, Informatik und Epidemiologie (IMBIE), Universität Bonn

Longitudinal studies often involve the statistical analysis of observation times that are measured on a discrete time scale $t = 1, 2, \ldots, q$. Typical examples are given by clinical and epidemiological studies with prespecified follow-up times where the values of t refer to fixed time intervals (e.g. 3-month or 6-month intervals). Discrete observation times are also encountered in studies with an intrinsically discrete time scale, for example, in clinical trials dealing with time to pregnancy (where the observation time is often defined by the number of menstrual cycles).

If the interest of a study is in one or more target events that may (or may not) occur at the end of the observation times, statistical investigations usually require the application of methods for failure time analysis with competing events. In contrast to established methods for competing risks analysis, which are based on the assumption of a continuous time scale, statistical techniques for the analysis of discrete event times have been less well explored. Consequently, as the latter methods are often more appropriate when the number of time points is small and/or when the discrete time scale cannot be approximated by a continuous one, discrete competing risks analysis has gained increasing interest in the research community.

The talk will present an overview of statistical methods for the analysis of possibly right-censored discrete failure times with competing events. We describe a set of modeling approaches for this type of data, including discrete versions of the cause-specific hazards model and the subdistribution hazard model. In addition to discussing the characteristics of these models, we present approaches to nonparametric estimation and model validation. All presented models have a straightforward interpretation and can be fitted using readily available software for multivariable regression.

Reference:

https://wires.onlinelibrary.wiley.com/doi/full/10.1002/wics.1529

Ereigniszeitanalyse mit konkurrierenden Risiken unter Berücksichtigung von Clusterstrukturen – Methodenvergleich anhand einer Simulationsstudie

Sabrina Schmitt[1], Ann-Kathrin Ozga[2]

[1] Hochschule Koblenz – RheinAhrCampus
[2] Institut für Medizinische Biometrie und Epidemiologie, Universitätsklinikum Hamburg-Eppendorf

In Studien der klinischen Forschung besteht häufig ein großes Interesse an dem Nachweis von Therapiewirksamkeiten. Zur Beurteilung werden meist Modelle der Ereigniszeitanalyse herangezogen, wobei die Zeit bis zum Eintritt eines primär interessierenden Ereignisses betrachtet wird. In der Praxis kann aber zusätzlich ein dazu konkurrierendes Ereignis zuvor auftreten, welches in der Analyse berücksichtigt werden sollte. Außerdem werden klinische Studien häufig an mehr als nur einer Klinik (Cluster) gleichzeitig durchgeführt. Es besteht die Annahme, dass diese hier entstehende Clusterstruktur zu einer potenziellen Abhängigkeit zwischen den Ereigniszeiten führt. Während die angesprochene Clusterstruktur bereits Anwendung in vielen anderen Gebieten, wie beispielsweise den gemischten linearen Modellen, findet, wird sie bei der Auswertung klinischer Studien mit konkurrierenden Risiken bisher meist noch vernachlässigt. Nun ist von Interesse, wie solche Clusterstrukturen in Ereigniszeitanalysen mit konkurrierenden Ereignissen berücksichtigt werden können. Verschiedene Methoden wurden dazu schon in der Literatur beschrieben, aber noch nicht systematisch verglichen. Ziel dieser Arbeit war daher anhand einer Monte-Carlo-Simulationsstudie diese Methoden zu vergleichen und daraus Empfehlungen für zukünftige Analyse von Ereigniszeiten unter Berücksichtigung von konkurrierenden Ereignissen und vorliegender Clusterstruktur abzuleiten. Die betrachteten Methoden basierten auf dem Cox proportional hazards model (z.B. Cox Modell mit Frailty [1]) oder Methoden, welche subdistribution hazards modellieren (Fine und Gray [2] und Erweiterungen davon: Katsahian et al. [3] und Zhou et al. [4])

Zusammenfassend ergaben sich nur marginale Unterschiede zwischen den betrachteten Modellen bezüglich Bias, mittlerem quadratischem Fehler und empirischer Power. Jedoch zeigte der Ansatz von Katsahian et al. in den meisten Szenarien die beste Performance basierend auf diesen Werten.

References:

[1] Hougaard P. (1995). Frailty models for survival data. Lifetime Data Anal. 1(3):255-73

[2] Fine, J.P. & Gray, R.J. (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risk, Journal of the American Statistical Association, 94:446, 496-509

[3] Katsahian, S., Boudreau, C. (2011). Estimating and testing for center effects in competing risks, Statistics in Medicine, 30(13):1608-1617.

[4] Zhou, B., Fine, J., Latouche, A., Laboin, M. (2011). Competing risks regression for clustered data, Oxford University Press, 13(3):371–383.

Accelerated failure time models for semi-competing risk data with recurrent events

Ann-Kathrin Ozga[1], Annika Hoyer[2], Oliver Kuß[3]

[1] Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf
[2] Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University
[3] German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich Heine University Düsseldorf, Institute for Biometrics and Epidemiology, Düsseldorf

Clinical trials often compare a treatment to a control with respect to two correlated time-to-event endpoints like time to hospitalization and time to death. Thereby, the first endpoint may occur more than once ("recurrent"), whereas the second endpoint is absorbing, that is, after observing the second endpoint an individual can no longer experience the first event. However, usually only the time until the first occurrence of an event for a patient is analyzed. Inclusion of all observed events in the analysis can increase the power and provides a more complete picture of the disease. Therefore, statistical methods for recurrent events are required that take into account that an absorbing event serves as a competing event for the recurrent event. In the literature, semi-parametric joint frailty models were proposed for this task [1, 2]. We propose a bivariate parametric accelerated failure time model that overcomes the standard weaknesses of hazard-based approaches and accomplishes this task. This method adequately models the semi-competing risks setting and recurrent events, i.e., uses all event information and properly accounts for censoring as well as the correlation between and the hierarchy of endpoints. The proposed approach is illustrated with an example from the Interdisciplinary Network for Heart Failure (INH) study by Angermann el al. [3]. This multi-centre randomized controlled trial investigated the efficacy of a nurse-coordinated disease management program (HNC) in heart failure compared to usual care for patients that were first hospitalized for systolic heart failure. A total of 1022 patients (513 in usual care, 509 in HNC group) with 663 deaths and 3016 rehospitalizations were observed.

References:

[1] Rogers, J. K., Yaroshinsky, A., Pocock, S. J., Stokar, D., & Pogoda, J. (2016). Analysis of recurrent events with an associated informative dropout time: Application of the joint frailty model. Statistics in Medicine, 35(13), pp. 2195-2205.

[2] Rondeau, V., Mathoulin-Pelissier, S., Jacqim-Gadda, H., Brouste, V., & Soubeyran, P. (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. Biostatistics, 8(4), pp. 708-721.

[3] Angermann CE, Störk S, Gelbrich G, Faller H, Jahns R, Frantz S, Loeffler M, Ertl G; Competence Network Heart Failure. (2012). Mode of action and effects of standardized collaborative disease management on mortality and morbidity in patients with systolic heart failure: the Interdisciplinary Network for Heart Failure (INH) study. Circ Heart Fail. 5(1), pp. 25-35.

How to deal with nonproportional hazards in factorial survival designs?

Marc Ditzhaus

TU Dortmund University

While the log-rank test and hazard ratios were the gold standard in time-to-event analysis for a long time, there is a recent trend towards alternative methods not relying on the proportional hazard assumption. The reason for this change are violations of the proportional hazard assumption frequently observed in real data. For example, Trinquart et al. (2016) analysed 54 phase III oncology clinical trials from five leading journals and in 13 (24assumption could be rejected significantly. In this talk, I present some solutions Ditzhaus et al. (2021a,b) to tackle the problem of crossing survival curves or, more generally, non-proportional hazards in general factorial designs. These not only allow the detection of main factor effects (e.g. treatment or gender) but also the inference of potential interaction effects as, e.g., stated by Lubsen and Pocock (1994): "it is desirable for reports of factorial trials to include estimates of the interaction between the treatments". The developed methodology is motivated by a recent study on asthma, for which the assumption of proportional hazards is not justifiable.

References:

M. Ditzhaus, D. Dobler, and M. Pauly. Inferring median survival differences in general factorial designs via permutation tests. Statistical Methods in Medical Research, 30(3):875–891, 2021a.

M. Ditzhaus, J. Genuneit, A. Janssen, and M. Pauly. CASANOVA: Permutation inference in factorial survival designs. Biometrics, early View, 2021b. doi: https://doi.org/10.1111/biom.13575.

J. Lubsen and S.J. Pocock. Factorial trials in cardiology: pros and cons. European Heart Journal, 15:585–588, 1994.

L. Trinquart, J. Jacot, S.C. Conner, and R. Porcher. Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. Journal of Clinical Oncology, 34(15):1813–1819, 2016.

Accelerated failure time models for crossing survival curves

A. Seipp, A. Timmer, F. Otto-Sobotka

Division of Epidemiology and Biometry, Carl von Ossietzky Universität Oldenburg; Oldenburg (Germany)

The most popular method for analysis of time-to-event data is the Cox proportional hazards model. However, the proportional hazards assumption does not always hold true. In a study of 152 oncological phase 3 trials, hazard rates of treatment effects were found to be non-proportional in 25% of cases (Rahman et al., 2019). Parametric accelerated failure time (AFT) models are an attractive alternative to the Cox model, which do not assume proportional hazards. Rather, covariates are assumed to accelerate (shift) the survival curve. However, both the proportional hazards and the AFT assumption can be incorrect at the same time. An example of this is the crossing of survival curves, indicating an influence of a covariate on the whole distribution. In this talk, we show that the AFT model can be extended to relax the assumption of simple location shifts. The AFT model can be viewed as a special case of a Generalized Additive Model for Location, Scale and Shape (GAMLSS, Rigby and Stasinopoulos, 2005). GAMLSS allows for modeling of the whole distribution by estimating multiple distributional parameters like mean, variance and skewness at the same time. Further flexibility can be achieved by modeling the parameters with semiparametric predictors (e.g. for nonlinear trends). While GAMLSS is mostly known in the analysis of uncensored data, it can be adapted to incorporate censored data as well, since estimation is likelihood-based. We present an application of GAMLSS to the overall survival of 590 colon cancer patients from a local cancer registry. Patients were treated at a specialized cancer center between September 2009 to March 2019. Bivariate analysis of chemotherapy treatment (yes/no) and survival times with the Kaplan-Meier estimator showed an advantage of chemotherapy for up to 3 years and converging survival curves after that. Similarly, we found a larger association of chemotherapy with the lower tail of the survival time distribution using expectile regression (Seipp et al., 2021). In the presented analysis, we use GAMLSS to account for this by modeling further parameters like the scale.

References:

Rahman, R., Fell, G., Ventz, S., Arfé, A., Vanderbeek, A. M., Trippa, L., and Alexander, B. M. (2019). Deviation from the proportional hazards assumption in randomized phase 3 clinical trials in oncology: prevalence, associated factors, and implications. Clinical Cancer Research, 25(21), 6339-6345.

Rigby, R. A. and Stasinopoulos, D. M. (2005): Generalized additive models for location, scale and shape. Applied Statistics, 54(3):507-554.

Seipp, A, Uslar, V, Weyhe, D, Timmer, A, Otto-Sobotka, F. (2021): Weighted expectile regression for right-censored data. Statistics in Medicine, 1–20.

Testing VIMPs for Dependencies in Random Forest Analyses: Methods and an Application to Post Transplant Survival

Christoph Wies

Hochschule Darmstadt

This work is motivated by an RF analysis of survival after kidney transplantation in USA, where a variable measuring kidney quality (KDPI) shows the largest permutation variable importance (VIMP). The VIMP is one way to measure the importance of a variable in a random forest (RF). It describes the expected increase in mean squared error by applying a random permutation on a variable. It is well known that dependencies between covariates affect their VIMPs. This could lead to a reduction in VIMPs of actual important variables or to an increase in VIMPs of non-relevant ones. Thus it can be hard to interpret VIMPs as measure for the association between outcome and variable. We construct a statistical test for the alternative hypothesis that the importance of a variable is partially or totally caused by shared information with associated variables. Therefore we compute an adjusted VIMP of a residual based model. Both are same only under the null hypothesis. Resampling samples are used to estimate the VIMPs distribution and afterwards the adjusted VIMP is compared to this distribution providing a p-value. This work is motivated by an RF analysis of survival after kidney transplantation in USA, where a variable measuring kidney quality (KDPI) shows the largest VIMP. KDPI is a key factor in the US allocation process thus it is strong association with recipient characteristics. We apply the proposed method to better understand the importance of KDPI to the post- transplant survival.

Correlated random-effects models for clustered time-to-event data

Andreas Wienke

Martin-Luther-Universität Halle-Wittenberg

Frailty models have become very popular during the last three decades and their applications are numerous. The introduction of this talk will shortly deal with univariate frailty models meaning random effects models applied to independent event times. Univariate frailty models are useful to adjust Cox regression analysis for unobserved heterogeneity (unobserved covariates). Different frailty distributions and their consequences are discussed. The main part of the talk is devoted to multivariate frailty models. Such multivariate frailty models account for correlations between event times within clusters (here, a cluster can consist of individuals from the same group, say a family, litter, clinic, community; or of multiple or recurrent events from the same individual). The most often applied model here is the shared frailty model. However, it does have some limitations. To avoid these limitations, correlated frailty models have been developed for the analysis of multivariate failure time data. The talk discusses advantages and limitations of different frailty models and is illustrated by real data applications in epidemiology and medicine. The first example deals with the application of frailty models to lung cancer data applying univariate as well as multivariate shared frailty models for these data. The second example is devoted to the analysis of Danish twin data. A correlated frailty model is suggested for analysis of the bivariate time-to-event data.

Time-varying frailty models and the estimation of heterogeneities in transmission of infectious diseases

Niel Hens

Hasselt University and University of Antwerp, Belgium

Frailty models are often used in survival analysis to model multivariate time-to-event data. In infectious disease epidemiology, frailty models have been used to model heterogeneity in contracting infections and to study the association in the occurrence of infections. Previously, Hens et al. studied the behavior of the bivariate correlated gamma frailty model for bivariate current status data whereas Farrington, Unkel and their collaborators introduced time-varying shared frailty models. In this talk we combine both approaches and consider an extension of the frailty methodology to account for age dependence in individual heterogeneity through the use of age-dependent shared and correlated gamma frailty models. The methodology is again illustrated using bivariate serological data. We further discuss an important feature in infectious disease epidemiology, which is to properly account for the underlying infection process of which the impact on baseline and effective reproductive rates has been quantified by Abrams and Hens.

The Addams family of discrete frailty distributions for multivariate survival data

Maximilian Bardo[1] , Niel Hens[2,3] and Steffen Unkel[1,4]

[1] Department of Medical Statistics, University Medical Center Goettingen, Germany
[2] I-BioStat, Data Science Institute, Hasselt University, Diepenbeek, Belgium
[3] Centre for Health Economics Research and Modelling Infectious Diseases, Vaccine & Infectious Disease Institute, University of Antwerp, Antwerp, Belgium
[4] Faculty of Life Sciences, University of Siegen, Germany

Frailty models provide a conceptually simple and appealing way of modelling heterogeneities resulting from factors which may be difficult or impossible to measure; examples are heterogeneity induced by genetics or through environmental exposure, governing the individual's survival. The frailty is usually assumed to have a continuous distribution. In some areas of application, however, the unobserved heterogeneity might be more of a discrete nature, such as the (unobserved) number of sexual partners for sexually transmitted diseases. In such scenarios, a discrete frailty model might be more appropriate for capturing the important differences in cluster-specific effects. In the present work, we model a family of discrete frailty distributions, which was outlined by Farrington et al. (2012), for multivariate current-status and right-censored, possibly left-truncated data. We suggest an interpretation of the discrete frailties as being ordered latent risk categories. The content-related interpretation of each distinct risk category and its estimated numerical value lead to points of comparison and analysis of the conditional model. From an analytical point of view, this becomes particularly interesting if one includes a model for the frailty distribution's parameters along, for example, categorical variables. Additionally, our estimation approach is built up such that the frailty distribution is chosen by the data among a set of distributions which might or might not imply a cure rate.

We further investigate admissible shapes of association that are covered by discrete shared frailty models. As an illustrative example, we utilize paired serological data on infections with the human papillomaviruses 16 & 18 (Mollema et al. (2010); Scherpenisse (2012)). We assume the distribution of individual heterogeneity to be distinct for males and females. We find that the distribution of risk categories for males is more scattered than for the female counterpart. Due to the frailty, females have a higher hazard in each of the risk-categories, however. This could be explained by more 'extreme' individuals in the male population due to behavioural reasons but a higher biological burden in that respect for females.

References:

[1] Farrington C. P., Unkel S. and Anaya-Izquierdo K. (2012): The relative frailty variance and shared frailty models, Journal of the Royal Statistical Society Series B, Vol. 74, pp. 673-696.

[2] Mollema L., de Melker, H.E., Hahne, S.J.M., van Weert, J.W.M., Berbers, G.A.M., and van der Klis, F.R.M. (2010): PIENTER 2-project: second research project on the protection against

infectious diseases offered by the national immunization programme in the Netherlands, National Institute for Public Health and the Environment (RIVM), Report 230421001.

[3] Scherpenisse, M., Mollers, M., Schepp, R., Boot, H., de Melker, H., Meijer, C., Berbers, G., and van der Klis, F. (2012): Seroprevalence of seven high-risk HPV types in The Netherlands, Vaccine, Vol. 30, pp. 6686-6693.

Acknowledgement
This research has received funding from the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation, grant number UN 400/2-1).

How to model the spreading of infectious diseases using networks embedded into hyperbolic space?

Markus Schepers

Universitätsmedizin der Johannes-Gutenberg-Universität Mainz

The global spread of covid-19 has triggered tremendous research efforts to gain a better understanding of infectious diseases, in particular their mathematical modelling. Indeed historically, mathematical reasoning helped, for instance, in the eradication of smallpox or the containment of malaria. Recently, random graphs and percolation have been used to model contact patterns in heterogeneous populations. This talk aims to discuss the potential of hyperbolic random graphs and their dynamics for the understanding of infectious diseases. As a part of this, we will give an overview of previous modelling approaches and highlight some of the most significant insights. The main part of the discussion will revolve around the impact of the contact network on epidemic spread and the assessment of the effects of intervention methods.