technische universität
dortmund

fakultät
statistik

I B S

CENTER FOR TRUSTWORTHY
DATA SCIENCE AND SECURITY
UA RUHR
RESEARCH ALLIANCE

Book of Abstracts
Workshop on Causal Inference in Time-to-Event Analysis

September 17 and 18, 2024, at TU Dortmund University

Organized by

Dennis Dobler
(Department of Statistics, TU Dortmund University & Research Center Trustworthy Data Science and Security, University Alliance Ruhr)

and

Jan Feifel
(Merck Healthcare KGaA),

the speakers of the
Working Group Statistics of Stochastic Processes of the International Biometric Society, German Region.

Cooperation partners from University Alliance Ruhr:
Prof. Dr. Jens Kleesiek
(Chair of the Medical Machine Learning Section of the Institute for AI in Medicine (IKIM) at University Hospital Essen)

and

Dr. Michael Kamp
(Research Group leader Trustworthy Machine Learning at IKIM Essen, located at Ruhr-University Bochum).

Further supported by **Ulrike Guba, Erik Weber, and Simon Mack**, all from TU Dortmund University.

Kindly funded by the International Biometric Society, German Region,
and the Research Center Trustworthy Data Science and Security, University Alliance Ruhr.

Block 1: 13:00–14:30 on Tuesday, September 17, 2024. Chair: Dennis Dobler

Ruth Keogh (London School of Hygiene and Tropical Medicine)
Censoring-and-weighting for estimating effects of longitudinal treatment strategies: What are our options for analysis?

The censoring-and-weighting approach is popular for enabling estimation of the effects of longitudinal treatment strategies on time-to-event outcomes. It can be used, for example, when individuals are observed as following one of the treatment strategies of interest from time 0, but where they may later be observed to switch treatment status. The censoring-and-weighting approach involves censoring a person's follow-up when they deviate from their treatment status at time 0, resulting in a modified data set in which each individual's treatment status is constant over time. The artificial censoring is likely to be informative and this should be addressed in the analysis. An advantage of the censor-weight approach is that it means we can then make use of analysis methods for point treatments. Analysis approaches that have been used involve combining inverse probability of censoring weights (IPCW) with inverse probability of treatment weights (IPTW), or combining IPCW with a conditional Cox regression model for the outcome. This talk will discuss how we could combine the censoring-and-weighting design with doubly robust analysis methods. Different options will be discussed, covering scenarios where the artificial censoring depends on baseline covariates only, or where it depends on time-updated covariates.

Tijn Jacobs (Amsterdam University Medical Center)
A causal investigation into short- and long-term effects of treatment for osteosarcoma

We present the challenges and complexities in execution and interpretation of novel causal inference methods to survival data with a cured fraction encountered in our re-analysis of the EURAMOS-I trial. This trial compares a standard and experimental treatment for osteosarcoma, a rare but malignant bone tumor in young individuals. Our reanalysis focused on the causal effect of an experimental treatment on both cure probability and survival prolongation, with an emphasis on causal estimands and their interpretation. We employ a mixture cure model that comprises a latency and an incidence component. The latency component describes how quickly patients experience a recurrence of osteosarcoma over time through an accelerated failure time (AFT) model combined with a Bayesian causal forest. The incidence component describes the probability of cured patients by a probit regression model using a Bayesian causal forest. The mixture cure model thus facilitates a nuanced, causal understanding of the treatment effects on both the short- and long-term outcomes of osteosarcoma patients. Our reanalysis revealed that the experimental treatment has no effect on either the incidence of cure or the latency of osteosarcoma recurrence, i.e., it provides no short- or long-term benefits for osteosarcoma patients. Despite the challenges and complexities in handling and interpreting causal estimands, our approach offers a more nuanced and precise understanding of the treatment effects. Our study shows, by example, the value of incorporating modern causal inference techniques in survival analysis, particularly for diseases with a good prognosis, to enhance the accuracy of clinical decision-making.

Joint work with Marta Fiocco, Eni Musta, Stéphanie van der Pas, and Wessel van Wieringen.

Block 2: 15:00–16:30 on Tuesday, September 17, 2024. Chair: Simon Mack

Eni Musta (University of Amsterdam)
Bounds on a causal hazard ratio: an approach to understand waning of treatment effect

Understanding the evolution of the causal effect of a treatment over time, including the potential for treatment effect waning, is important for informed decisions on treatment repetition or interruption. However, this is a challenging problem. It is well known by know that, even within randomized controlled trials, a naïve comparison of hazard functions between two groups can lead to incorrect causal conclusions due to inherent selection bias. While comparing survival curves is often recommended as a safer measure of causal effect, it only represents a cumulative effect over time and does not address treatment effect waning.

We propose using the causal hazard ratio, introduced by Martinussen et al (2020), as a relevant estimand. This ratio contrasts hazards for the same group of individuals who would have survived up to a certain time regardless of the received treatment. However, identifying such an effect requires untestable modelling assumptions. Instead, we derive bounds for this causal hazard ratio in a general setting that accounts for both observed and latent covariates. These bounds can be tightened with the assumption of monotonicity, meaning the treatment does not cause any harm. We discuss when the bounds are informative and how they contribute to understanding treatment effect waning. As an application, we consider vaccine efficacy studies from malaria and HIV vaccine trials.

Joint work with Joris Mooij and Jesús Garcia Garcia.

Sebastian Starke (Helmholtz-Zentrum Dresden-Rossendorf)
(Not yet) causal approaches for deep learning-based outcome models in personalized radiotherapy

A personalized approach to radiotherapy bears promising potential for improving treatment responses of cancer patients across various tumor entities. Medical imaging-based biomarkers, in particular, have indicated favorable prospects for the creation of outcome models in a process called radiomics. In recent work [1,2], we combined right- censored time-to-event data with deep learning approaches using the Cox proportional hazards model for the prediction of tumor recurrence risks after radiotherapy. Using such outcome models, patients identified to be at high risk of tumor recurrence might benefit from treatment adaptation using higher dose prescriptions. Likewise, patients predicted to be at lower risk might receive treatment with reduced dose, which would go along with a potential reduction of side effects after therapy. Currently, our neural network models are black boxes, which are hard to interpret. Moreover, they are optimized based solely on correlations in the available data, making their adoption in clinical workflows challenging. Therefore, we aim to explore the applicability of causality-focused machine learning models with an emphasis on (but not restricted to) imaging data for a further personalization of radiotherapy. We envision possible use cases of counterfactual modelling for treatment adaptation, side effect evaluation, biomarker identification [3] and model explainability. As we have very limited prior experience in this field, we are looking forward to engage in fruitful discussions on the requirements and the realizability of our vision with causality experts in the upcoming workshop. References:

[1] Starke et al., Scientific Reports, 2020. (https://doi.org/10.1038/s41598-020-70542-9)
[2] Starke et al., Cancers, 2023. (https://doi.org/10.3390/cancers15194897)
[3] Feuerriegel et al., Nature Medicine, 2024. (https://doi.org/10.1038/s41591-024-02902-1)

Joint work with Peter Steinbach and Steffen Löck

Daniel Klippert (TU Dortmund University)
An Analysis of Causal Inference Methods for Competing Risks

The prediction of heterogeneous treatment effects (HTEs) is crucial for treatment decision-making in personalized medicine. HTEs can be estimated using randomized controlled trials (RCTs). Often, however, researchers cannot conduct RCTs, and instead rely on observational data. In this talk, I address the challenge of estimating HTEs from observational survival data with competing risks, where the occurrence of an event precludes the occurrence of any other event. Judging the effectiveness of a treatment in preventing the event of interest requires accounting for all competing events, especially when patients have comorbidities. Focusing on right-censored survival times and binary treatment, I consider HTEs in the form of conditional average treatment effects (CATEs), defined as the difference in the absolute risk of the event of interest between treatment groups at a specific time-point, conditioned on covariates. Since the CATE is a causal quantity, traditional supervised learning approaches cannot be used for its estimation. Instead, various meta-learners have been developed, which re-purpose arbitrary machine learning algorithms for CATE estimation. The goal is to provide practical guidance on model selection for CATE estimation in competing risks scenarios. Therefore, I compare six meta-learners for CATE estimation, utilizing Cox regression and random survival forests for absolute risk modeling, and elastic net regression and random forests for direct CATE modeling. I evaluate their performance throughout multiple simulation settings, differing in the hazard function complexity, event type distribution, strength of treatment heterogeneity, and treatment assignment mechanisms. Additionally, I demonstrate how a fit-the-fit approach can be used for the identification of subgroups with distinct average treatment effects based on a meta-learning model.

Joint work with Sarah Friedrich and Markus Pauly.

Block 3: 17:00–18:30 on Tuesday, September 17, 2024. Chair: Ruth Keogh

Vanessa Didelez (Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen)

Causal mediation and separable treatments in time-to-event analyses

In this presentation I will discuss 'separable effects' as an alternative approach to causal mediation analyses with view to time-dependent settings. The basic idea was introduced by Robins & Richardson (2011) and refers to a way of elaborating our causal model in order to better motivate a mediational research question. Specifically, separable effects are concerned with situations where the treatment (or exposure) can be decomposed into two or more components that could (more or less hypothetically) be intervened upon separately and thus separately 'activate' different causal pathways. It is appealing to formulate mediational research questions in this way as it sidesteps 'cross-world' notions and assumptions involved in the analysis of natural (in)direct effects based on nested counterfactuals (Andrews & Didelez, 2021).

Crucially, separable effects turn out to be especially useful in time-dependent setting, e.g. when the outcome is survival and mediation is via a process - in this case it can be shown that natural (in)direct effects are neither well-defined nor identifiable while separable effects are (Didelez, 2019). Further, separable effects lead to a new estimand for competing events settings: Well-known concepts are based on cause-specific incidence or on treating the competing event as censoring; these can be seen as total causal effect or as effect under an additional intervention eliminating the competing event. Instead, the separable direct effect on the event of interest is the effect of the treatment component that only activates causal paths avoiding the competing event (Stensrud et al., 2020).

Examples from the field of cancer research and drug development will be given. The presentation will focus on basic principles and concepts rather than technical details.

References

- Aalen OO, Stensrud MJ, Didelez V, Daniel R, Roysland K, Strohmaier S. Time-dependent mediators in survival analysis: Modeling direct and indirect effects with the additive hazards model. Biometrical Journal. 2020;62(3):532-549.
- Didelez V. Defining causal mediation with a longitudinal mediator and a survival outcome. Lifetime Data Analysis. 2019;25(4):593-610.
- Roysland K, Ryalen PC, Nygård M, Didelez V. Graphical criteria for the identification of marginal causal effects in continuous-time survival and event-history analyses. Journal of the Royal Statistical Society - Series B (Statistical Methodology). 2024 (e-pub).
- Stensrud MJ, Young JG, Didelez V, Robins JM, Hernán MA. Separable effects for causal inference in the presence of competing events. Journal of the American Statistical Association. 2022;117(537):175-183.
- Stensrud MJ, Hernán MA, Tchetgen Tchetgen E, Robins JM, Didelez V, Young JG. A generalized theory of separable effects in competing event settings. Lifetime Data Analysis. 2021;27(4):588-631

Jasmin Rühl (Augsburg University)

Resampling-based inference for the average treatment effect in time-to- event data

The g-formula can be used to estimate treatment effects while accounting for confounding bias in observational studies. For time-to-event endpoints, statisticians need to take additional difficulties into account. It is for example not advisable to answer causal questions by hazard ratios, which is why we consider the risk difference instead. This way, a competing risks framework is accommodated on top. The distribution of the associated stochastic process is rather complicated, and hence, confidence intervals are commonly constructed by means of the classical non-parametric bootstrap. In certain situations, e.g., when the data lack independence, the classical bootstrap suffers from limitations, though. Furthermore, its execution can be rather time-consuming. This work investigates alternative resampling methods proceeding from the underlying stochastic process of the average treatment effect. Apart from Efron's classical bootstrap, we consider an approach that is based on the influence function [1] and, since counting processes are inherent to time-to-event analysis, a bootstrap version based on the martingale representation of the process. It is shown that the bootstrap methods approximate the distribution of the stochastic process at hand. We further compare the precision of the different techniques in a simulation study considering confidence intervals and bands for the average treatment effect. Our simulations imply that the wild bootstrap generally achieves accurate coverage levels if the sample size is small and sufficient data on the event of interest have been accrued.

References

[1] Ozenne, B. M. H. and Scheike, T. H. and Stærk, L. and Gerds, T. A. (2020). On the estimation of average treatment effects with right-censored time to event outcome and competing risks. Biometrical Journal, 62: 751-763

Block 4: 8:30–10:00 on Wednesday, September 18, 2024.
Chair: Thomas Alexander Gerds

Pål Christie Ryalen (University of Oslo)
The role of martingales for identifying causal effects in survival analysis

Many researchers use potential outcome variables for causal reasoning, and there is an extensive epidemiology-oriented literature on dynamic treatment regimes and identification of causal effects in discrete time. A corresponding development in marked point process (MPP) models, which deals with events that occur in continuous time, has not yet occurred. Such a development is of interest in fields such as survival and event history analysis, where 1) there is a large literature of statistical methods formulated in continuous time, e.g., counting process models, and 2) interest is frequently in the effects of time-varying treatment strategies, which may often be most appropriately viewed as time-continuous processes.

In this presentation, I aim to fill this gap by providing rigorous results on dynamic regimes, potential outcomes, and the identification of effects for data described by point processes. I describe analogues of the existing discrete-time conditions 'exchangeability,' 'consistency,' and 'positivity' which ensure that outcomes of interest are identified in the presence of unobserved components. I establish continuous-time identification formulas and discuss connections with related work. I show how the results can be used to reason about the identification of commonly studied parameters in survival analysis and highlight some similarities and differences with existing methods.

If time permits, I will discuss a general strategy for estimating a range of marginal parameters in survival and event history analysis that solve differential equations. The method gives researchers easy access to parameters that are easier to interpret causally than hazard (ratios).

Jannis Guski (Fraunhofer IAIS)
Does a SARS-CoV-2 infection increase the risk of dementia? An application of causal time-to-event analysis on real-world patient data

Despite the relative recency of the global pandemic, there is evidence that a SARS-CoV-2 infection may act as a catalyst in the gradual manifestation of neurodegenerative diseases. For example, comparisons of biomarkers before and after an infection revealed signs of changes in brain structure, neuroinflammation and disruptions of the blood-brain barrier.

As part of the EU-funded COMMUTE project, we deploy and adapt time-to-event models on real-world patient data to investigate whether a SARS-CoV-2 infection increases the risk or accelerates the process of developing Alzheimer's or Parkinson's disease, and which factors from a patient's history may play a role in this assumed co-pathology.

Any findings are potentially affected by confounding bias and informative censoring. These problems may be amplified by the problem that control patients have to be sampled from before the pandemic because presumably, almost everyone has had a documented or undocumented SARS-CoV-2 infection at some point after 2020.

In a first step, we estimate average exposure effects of COVID-19 on event risks with statistical models that address the potential sources of bias via (augmented) IPTW, IPCW, doubly robust methods or genetic instrumental variables.

Furthermore, machine learning models are trained that will be able to make risk predictions for new patients. These models may derive encodings of structured data from electronic health records, e.g., via transformer architectures.

At the workshop, we would like to present preliminary results from our medical application of causal inference in time-to-event analysis, and discuss with the expert audience how our methodology may be further refined.

Block 5: 10:15–11:45 on Wednesday, September 18, 2024. Chair: Jan Feifel

Richard Post (Technical University Eindhoven)

Formalization of the complex causal interpretation of hazard contrasts

In recent years, the hazard ratio's causal interpretation (or lack thereof) has gained significant attention. In the presence of unobserved heterogeneity, even for data from a randomized controlled trial, the observed hazard ratio (OHR) suffers from built-in selection bias as, over time, the individuals at risk in the exposed and unexposed are no longer exchangeable. In this talk, I will examine a general structural causal model (SCM) to formalize how the observed hazard rate evolves. When the causal effect of an exposure on the hazard is multiplicative, the expected OHR can be shown to equal the ratio of expectations of the latent variables (frailty and modifier) conditioned on survival with and without exposure. Thus, in the presence of frailty and/or unobserved effect modifiers, the expected OHR deviates from the population effect of interest. Moreover, if the causal effect is additive, the expected observed hazard difference (OHD) evolves by selecting favorable levels of effect modifiers in the exposed group. Consequently, in the presence of unobserved effect heterogeneity, the OHD deviates from the population causal effect of interest. For both scenarios, a homogeneous time-varying causal effect on the hazard is indistinguishable from a time-invariant heterogeneous causal effect. Finally, I will explain why contrasts of survival functions offer simpler causal interpretations and are thus more suitable for describing causal effects.

Joint work with Edwin R. van den Heuvel and Hein Putter.

Mari Brathovde (Oslo University Hospital)

Formalization of the causal interpretation of accelerated failure time models in the presence of unmeasured heterogeneity

In the presence of unmeasured heterogeneity, the hazard ratio for exposure is known to have a complicated causal interpretation. It is often suggested that this problem can be overcome by comparing the survival curves or by modeling the treatment's effect on the survival time ratio scale, specifically by employing accelerated failure time (AFT) models. Using an AFT model has the advantage of allowing straightforward incorporation of confounder adjustment into the model. In this work, we formalize the causal interpretation of the acceleration factor estimand in AFT models. To do so, we study systems describing the causal effect of a binary exposure parameterized using structural causal models, and data observed under independent censoring. We prove that the acceleration factor yields an appropriate causal effect measure in the presence of frailty and treatment effect heterogeneity. For illustration, we simulate a system where both AFT and proportional hazard models apply and demonstrate how the acceleration factor better reflects the causal effect than the hazard ratio. Moreover, we extend the formalization of the causal interpretation for causal systems with time-dependent acceleration factors. For this scenario, we highlight an important finding: the inability to differentiate between a time-varying but homogeneous effect and the existence of unmeasured effect heterogeneity. Despite the positive findings on the causal interpretation of acceleration factors, we reveal challenges that may arise when solving the estimating equations in the presence of effect heterogeneity. Thus, practitioners should exercise caution when employing parametric estimators of the AFT estimands.

Joint work with Hein Putter, Morten Valberg, and Richard Post.

Sandra Schmeller (Ulm University)

A "what if"-interpretation of the Kaplan-Meier estimator

It should be well known that in a competing risk setting the use of the Kaplan Meier estimator counting only one type of event is biased for estimation of the cumulative incidence probability (Schuhmacher et al., Schmeller et al.). Several analytical proofs of this exist which confirm examples from real data. Therefore, it is clear that the false-Kaplan-Meier is not an estimator for the cumulative incidence but it is questioned if this estimator still has a meaningful interpretation. The aim of this talk is to show a proof that the Kaplan-Meier estimator has a causal interpretation for the survival probability in a time to combined endpoint analysis that would have been observed if the random censoring had been avoided. However, the talk will show that this cannot be adapted to a competing risk setting and the Kaplan-Meier estimator counting only one type of event and censoring the other cannot be interpreted as the estimator for the cumulative incidence of the event of interest if the competing events had been avoided (the intervention distribution with intervention "no competing event"). The reason is that the occurrence of the competing events are not independent. The simple proof is based on a simple causal graph and a straightforward application of the g-computation rule. It is conceptually related to but technically simpler than the recent argument by Young et al. however, without requiring that one cause of death precedes the other as in Young et al. .

Schuhmacher et al. : Schumacher M, Ohneberg K, Beyersmann J (2016) Competing risk bias was common in a prominent medical journal. J Clin Epidemiol 80:135–136

Schmeller et al. : Schmeller S, Fürst D, Beyersmann J (2023). Konkurrierende Risiken Modelle. In Jan Gertheis, Matthias Schmid, and Martin Spindler, editors, Moderne Verfahren der Angewandten Statistik. Springer Spektrum, Berlin, Heidelberg, 2023.

Young et al.: Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA (2020) A causal framework for classical statistical estimands in failure-time settings with competing events. Stat Med 39(8):1199–1236

Joint work with Daniel Fürst and Jan Beyersmann.

Block 6: 13:00–14:30 on Wednesday, September 18, 2024.
Chair: Pål Christie Ryalen

Ilaria Prosepe (Leiden University Medical Center)
Causal multi-state models: estimating the effect of treatment delay with an illness-death model

Multistate models allow for the study of scenarios where individuals can experience different events over time. While effective for descriptive and predictive purposes, multistate models are not typically used for causal inference. We propose an estimator that combines a multistate model with g-computation to estimate the causal effect of treatment strategies on the marginal probability of recovery. Our focus are delay strategies, such as initiating treatment after waiting for 3 months. The impact of specific delay strategies is often unknown, but can provide doctors and patients with valuable information. It helps them decide on whether and for how long to "wait-and-see" if recovery occurs naturally before initiating treatment. We formulate this problem as an illness-death model, where illness and death represent respectively treatment and recovery. We study delay strategies by setting the transition intensity for one of the transitions. We: (i) formulate the causal assumptions needed for identification and the modelling assumptions needed to estimate the quantities of interest; (ii) propose an estimator that combines multi-state model with g-computation. In a simulation study, we present scenarios where the proposed method can make more efficient use of data compared to the clone-censor-reweight approach, where patients are artificially censored at the moment they deviate from the treatment strategy of interest and reweighted to correct for the subsequent dependent censoring. We then showcase the proposed methodology using real data from a study of 1896 couples with unexplained subfertility considering to start intrauterine insemination treatment. This analysis estimates the effect of delaying treatment on their probability of getting pregnant.

Joint work with Saskia le Cessie, Hein Putter, and Nan van Geloven.

Thomas Alexander Gerds (University of Copenhagen)
Emulating target trials in longitudinal register data with time-to-event outcomes

In this talk I will discuss the longitudinal analysis of electronic health records by emulating a target trial. Pharmacoepidemiology studies the use and effects of drugs in a large population. Here target trial emulation enables replication of results from a randomized trial and beyond. For example, we can identify all diabetes patients who started one of several alternative medical treatments in the Danish registers. We could be interested in the cardiovascular disease risk if hypothetically all patients had been assigned to using one of the drugs continuously for 5 years. To compare longitudinal treatment regimens we use a series of emulated trials which define estimands as contrasts of the cardiovascular disease risks. Longitudinal targeted minimum loss estimation projects the observed data onto a discrete time grid, estimates sequences of nuisance parameter models for treatment propensity and censoring probabilities, and then estimates the target parameters by updating a sequential regression formula backward in time. The estimation procedure accounts for treatment changes, censoring and non-cardiovascular mortality as a competing risk.