technische universität
dortmund

61. Biometrisches Kolloquium

Biometrics and Communication: From Statistical Theory to Perception in the Public

Abstract Volume

March 15–18, 2015

Contents

Tutorials

MiSeq, HiSeq, RNA-Seq, ChiP-Seq, ...-Seq: Was heißt das, und was fange ich damit an?

Zeller, Tanja [1] Mazur, Johanna [2] Dering, Carmen [3]

[1] Universitäres Herzzentrum Hamburg
[2] Institut für Medizinische Biometrie, Epidemiologie und Informatik, Universitätsmedizin Mainz
[3] Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck

Organisatoren:
Harald Binder, Institut für Medizinische Biometrie, Epidemiologie und Informatik, Universitätsmedizin Mainz Inke König, Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck

Inhalt:
Zum Schlagwort NGS (Next Generation Sequencing) gehören Messplattformen zur Hochdurchsatzsequenzierung, Genexpressionsmessung und Bestimmung weiterer molekularer Charakteristika, aber auch dazugehörige neue Techniken der Bioinformatik und Biostatistik. Auf technologischer Seite haben sich einige Plattformen (z.B. MiSeq und HiSeq der Firma Illumina) etabliert. Demgegenüber steht jedoch eine kaum überschaubare Zahl von Werkzeugen zur Datenverarbeitung. Bisher gibt es keinen Standard-Weg zur Analyse, sondern je nach Projektanforderung müssen Komponenten zu einem Analyseworkflow kombiniert werden, der sowohl bioinformatische als auch biostatistische Teile beinhaltet. Als Basis dafür behandelt dieses Tutorium

1. Biologische Grundlagen der Sequenzierungsmethoden und eingesetzten Plattformen zur Messung
2. Workflow-Komponenten vom Alignment bis zum statistischen Testen, illustriert an RNA-Seq und ChIP-Seq-Anwendungen
3. Ansätze zur Analyse von seltenen Varianten

Das Tutorium legt eine breite Wissensbasis, kann aber nicht den konkreten Software-Einsatz einüben. Es ist prinzipiell für einen Neueinstieg in die Thematik geeignet. Einige Grundlagen (z.B. aus dem Tutorium im Rahmen des Biometrischen Kolloquiums 2014) und erste Erfahrungen erlauben aber eine vertiefte Diskussion

Reproduzierbare Forschung

Hofner, Benjamin [1] Edler, Lutz [2]

[1] Institut für Medizininformatik, Biometrie und Epidemiologie der FAU Erlangen-Nürnberg
[2] Biometrical Journal

Dozenten:
Benjamin Hofner ist PostDoc am Institut für Medizininformatik, Biometrie und Epidemiologie der FAU Erlangen-Nürnberg. Er ist derzeit Reproducible Research Editor des Biometrical Journal und Autor/Co-Autor diverser R-Pakete (u.a. mboost, gamboostLSS, stabs, opm und papeR). Er beschäftigt sich auch in seiner täglichen Arbeit mit reproduzierbarer Forschung. Weitere Forschungsschwerpunkte sind statistische Modellbildung und Modellwahl, Modelle für biologische High-Throughput Experimente und die Entwicklung und Erweiterung moderner, modelbasierter Boostingansätze für biomedizinische Daten.
Lutz Edler leitete bis 2010 die Abteilung Biostatistik des Deutschen Krebsforschungszentrums (DKFZ) und ist derzeit einer der beiden Herausgeber des Biometrical Journal. Forschungsbereiche sind mathematische und statistische Modellbildung und Datenanalyse, Design von experimentellen und klinischen Studien und Methoden der Risikoanalyse zur Krebsentstehung und ihrer Anwendung zur Sicherheit von Nahrungs- und Futtermitteln in der EU.

Reproduzierbare Forschung ...
... bezeichnet die Idee, dass das letztendliche Resultat der Wissenschaft die Publikation zusammen mit der kompletten Computerumgebung (d.h. Code, die Daten, Computerprogramme etc.) ist, welche benötigt wird um die Ergebnisse zu reproduzieren. Dies soll die Nachvollziehbarkeit und Überprüfbarkeit der Ergebnisse garantieren und zu neuen wissenschaftlichen Erkenntnissen führen.

Inhalt:
Ziel des Tutoriums ist die Vermittlung der Einsicht in die Notwendigkeit reproduzierbarer Forschung in der biometrischen Forschung und ihren Anwendungen, sowie die Befähigung zur Nutzung reproduzierbarer Forschung als einen übergreifenden Ansatz moderner Datenanalyse, um somit die Reproduzierbarkeit der Ergebnisse zu einem ein essentieller Bestandteil der täglichen Arbeit zu machen. Wir beginnen mit einer Einführung in das Konzept der reproduzierbaren Forschung und geben Einblicke in die Problematik reproduzierbarer Forschung. Dazu wird die Praxis der Biometrischen Zeitschrift (Biometrical Journal) bei der Umsetzung von reproduzierbarer Forschung herangezogen und es werden mögliche Hürden und Fallstricke aus der Sicht von Zeitschriften der angewandten Statistik und der Biometrie besprochen. Alltägliche Herausforderungen reproduzierbarer Forschung werden aufgezeigt und Lösungsansätze vermittelt. *Literate Programming*, d.h. das "verweben" von Text und Code ist eine hilfreiche Methode. Hierzu wird Sweave vorgestellt, welches die Verknüpfung von R-Code und LaTeX erlaubt. Hiermit können einfach und schnell Berichte oder Publikationen generiert werden. Ändern sich die Daten, so ändert sich auch das fertige Dokument. Ein kurzer Einblick in knitr, eine Weiterentwicklung von Sweave wird gegeben. *Versionskontrolle* ist ein weiteres wichtiges Werkzeug zur Unterstützung reproduzierbarer Forschung und im Projektmanagement welche es erlaubt Änderungen an Dateien zu protokollieren und zu speichern (Backup). Somit werden Änderungen nachvollzieh-

bar und können jederzeit Rückgängig gemacht werden. Auch das Arbeiten in Teams, sogar mit Unterschiedlichen Standorten, wird dabei erleichtert. Hierzu werden Subversion (SVN) und Git/GitHub vorgestellt. Diese Systeme finden häufig in der Softwareentwicklung Verwendung, erleichtern jedoch auch die Erstellung und Verwaltung von Dissertationen, Publikationen und sonstigen Dokumenten.

Ablauf:

12:00 - 13:00	Notwendigkeit und Ziele reproduzierbarer Forschung und ihre praktische Umsetzung in Publikationen
13:00 - 13:30	Reproduzierbare Forschung im Arbeitsalltag (z.B. bei der Erstellung von Berichten, Dissertation, Veröffentlichungen)
13:30 - 14:00	Kaffeepause
14:00 - 15:30	Einführung in Sweave und knitr mit Hilfe von RStudio
15:30 - 16:00	Kaffeepause
16:00 - 17:30	Einführung in "ProjektmanagementSoftware: Subversion (SVN) und Git/GitHub für eigene Projekte und in Kollaborationen

Erforderliche Vorkenntnisse:
Die Teilnehmer sollten grundlegende Computerkenntnisse haben. (Grund-) Kenntnisse in LaTeX und der Programmiersprache R sind von Vorteil.
Teile des Kurses erfolgen interaktiv. Hierzu ist es von Vorteil wenn jeder Teilnehmer einen eignen Laptop zur Verfügung hat auf dem RStudio (1), R (2) und LaTeX (3) installiert sind.

Softwarequellen:
(1) http://www.rstudio.com/products/rstudio/download/
(2) http://cran.rstudio.com
(3) z.B. MiKTeX: http://miktex.org

Main conference talks

Control charts for biosignals based on robust two-sample tests

Abbas, Sermad Fried, Roland

TU Dortmund, Deutschland
sermad.abbas@tu-dortmund.de fried@statistik.tu-dortmund.de

The detection of abrupt jumps in the signal of a time series is an important task in biosignal analysis. Examples are the plasmon-assisted microscopy of nano-size objects (PAMONO) for virus detection in a sample fluid or the online monitoring of individual vital signs in intensive care. In the PAMONO context, a jump in the time series indicates the presence of a virus. In intensive care, sudden jumps point at the occurrence of a clinically relevant event.

Control charts are a reasonable tool to monitor such processes. They work in two phases. First, the chart is established based on gathered data or reference values. Then, the process is monitored. However, in the above applications no target values or historical data exist. Furthermore, vital signs of the critically ill are typically not stationary, so that a chart with fixed control limits does not seem to be appropriate.

We use two-sample tests in a moving window to detect the jumps. The window is divided into a reference and a test window. The location difference between them is estimated and standardized to construct robust two-sample tests. Tests based on the sample median, the one-sample and the two-sample Hodges-Lehmann estimator are compared in a simulation study regarding their Average Run Length (ARL) with some non-robust competitors and Shewhart-type control charts. A relationship between the significance level and the ARL is specified. In addition, the detection speed under different distributions is assessed. The simulations indicate that the construction principle can lead to two-sample tests with a nearly distribution-free ARL. Strengths and weaknesses of the methods are discussed.

Adjusted excess length-of-stay in hospital due to hospital-acquired infections

Allignol, Arthur [1] Schumacher, Martin [2] Harbarth, Stephan [3]
Beyersmann, Jan [1]

[1] Institute of Statistics, Ulm University, Germany
arthur.allignol@uni-ulm.de
[2] Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Germany
[3] Infection Control Program, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

The occurrence of an hospital-acquired infection (HAI) is a major complication leading to both increased morbidity and mortality. It also leads to a prolonged length of stay (LoS) in hospital, which is one of the main driver of extra costs induced by HAIs. Excess LoS associated with an HAI is often used in cost-benefit studies that weighs the costs of infection control measures against the costs raised by HAIs.

Estimation of the excess LoS is complicated by the fact that HAI is time-dependent. Cox models that include HAI as time-dependent variable could be fitted, but do not allow for a direct quantification of the number of extra days spent in hospital. Using multistate models and landmarking, Schulgen and Schumacher (1, 2) proposed to quantify the excess LoS by comparing the mean LoS given current HAI status.

The impact of covariates on the excess LoS can either be investigated through stratified analyses or through predicted excess LoS obtained via Cox transition hazards models. As an alternative, we propose to fit a direct regression model to the excess LoS using the flexible pseudo values regression technique (3).

Motivated by a recent study on HAI, we investigate the use of pseudo-values regression for identifying risk factors that influence the excess LoS. The proposed model is also compared to the predicted excess LoS obtained through Cox models in a simulation study.

References:

(1) Schulgen, G. and Schumacher, M. (1996). Estimation of prolongation of hospital stay attributable to nosocomial infections. Lifetime Data Analysis, 2, 219–240.

(2) Allignol, A., Schumacher, M. and Beyersmann J. (2011). Estimating summary functionals in multistate models with an application to hospital infection data. Computational Statistics, 26(2):181–197.

(3) Andersen, P. K. and Perme, M. P. (2010). Pseudo-observations in survival analysis. Statistical methods in medical research, 19(1):71–99.

Wissenschaftlichkeit in der Medizin und Patientenpräfenzen: ein unüberwindbarer Gegensatz?

Antes, Gerd

Universitätsklinik Freiburg, Deutschland
antes@cochrane.de

Gemeinsame, partizipative Entscheidungsfindung zwischen Arzt und Patient steht seit zwei Jahrzehnten als grundlegende politische Forderung im Raum. Die Realisierung bedeutet jedoch für beide Seiten Herausforderungen, bei deren Bewältigung sich alle Beteiligten schwer tun. Auf der wissenschaftlichen Seite hat eine beschleunigte Zunahme an durchgeführten randomisierten kontrollierten Studien zu einer Wissensexplosion geführt, die vom Einzelnen nicht mehr verarbeitet werden kann. Das Institute of Medicine (IOM, USA) hat daraus 2011 in einer fundamentalen Publikation “Finding what works in Health Care: Standards for Systematic Reviews"den Rahmen vorgegeben, einzelne Studien nicht mehr isoliert zu betrachten, sondern sie jeweils als schrittweisen Beitrag zur Wissensakkumulation zu betrachten. Dem entspricht die Forderung einer führenden medizinischen Zeitschrift The Lancet, jeden Bericht einer klinischen Studie mit dem vorhandenen Wissen zu beginnen und das Studienergebnis am Schluss ins vorhandene Wissen einzubetten. Systematic Reviews haben sich als Standardtechnologie für diesen globalen Wissensprozess etabliert. Von den 5 Schritten Frageformulierung – systematische Suche – Qualitätsbewertung der Funde – Synthese der hochwertigen Studien – Interpretation/Anwendung fordern vor allem der dritte und vierte Schritt Methoden von Biometrie und Biostatistik. Leitprinzip ist die Forderung nach Minimierung des Biasrisikos.

Auf der anderen Seite stehen Patienten, Gesunde (bei Vorsorge und Früherkennung) und deren Angehörige mit ihren Werten den wissenschaftlich begründeten Bewertungen von Diagnostik und Therapie oft verständnis- und hilflos gegenüber. Zu dem Gefühl der Überforderung kommt ein oft nur zu berechtigtes Misstrauen gegenüber dem Wissenschaftsbetrieb. Um zu einer tatsächlich informierten, gemeinsamen Entscheidungsfindung zu kommen, sind erheblich mehr Anstrengungen notwendig. Gerade auch auf der wissenschaft-lichen Seite sind die durch umfangreiche Begleitforschung inzwischen gut bekannten Defizite im gesamten Wissenschaftsprozess durch eine Qualitätsoffensive anzugehen, wie sie zu Beginn 2014 von The Lancet in einem Sonderband vorgeschlagen wurde. Für die Patientenseite ist es Zeit für neue Formen des Verständnisses klinischer Forschung und für mehr dafür notwendige Unterstützung. Dass Wissenschaft in der Mitte der Gesellschaft ankommt, ist gerade für die patientenorientierte Forschung von zentraler Bedeutung.

A Bayesian Nonparametric Regression Model with Normalized Weights: A Study of Hippocampal Atrophy in Alzheimer's Disease

Antoniano, Isadora

Università Bocconi, Italy
isadora.antoniano@unibocconi.it

Hippocampal volume is one of the best established biomarkers for Alzheimer's disease. However, for appropriate use in clinical trials research, the evolution of hippocampal volume needs to be well understood. Recent theoretical models propose a sigmoidal pattern for its evolution. To support this theory, the use of Bayesian nonparametric regression mixture models seems particularly suitable due to the flexibility that models of this type can achieve and the unsatisfactory fit of semiparametric methods.

We to develop an interpretable Bayesian nonparametric regression model which allows inference with combinations of both continuous and discrete covariates, as required for a full analysis of the data set. Simple arguments regarding the interpretation of Bayesian nonparametric regression mixtures lead naturally to regression weights based on normalized sums. Difficulty in working with the intractable normalizing constant is overcome thanks to recent advances in MCMC methods and the development of a novel auxiliary variable scheme. We apply the new model and MCMC method to study the dynamics of hippocampal volume, and our results provide statistical evidence in support of the theoretical hypothesis.

Blinded Sample Size Reestimation for Time Dependent Negative Binomial Observations

Asendorf, Thomas [1] Schmidli, Heinz [2] Friede, Tim [1]

[1] Universitätsmedizin Göttingen, Deutschland
thomas.asendorf@med.uni-goettingen.de tim.friede@med.uni-goettingen.de
[2] Novartis AG, Basel, Switzerland
heinz.schmidli@novartis.com

Sample size estimation procedures are a crucial element in planning clinical trials. However, sample size estimates strongly depend on nuisance parameters, which have to be guestimated from previous trials, as illustrated in the context of trials in relapsing multiple sclerosis (RMS) in (1). Blinded sample size reestimation procedures allow for an adjustment of the calculated sample size within a running trial, by using gathered data to estimate relevant nuisance parameters without unblinding the trial (2). We consider a model for statistical inference of time dependent count data and provide sample size estimation and reestimation techniques within this model. The model presented allows for time dependent discrete observations with marginal Poisson or negative binomial distribution (3). Examples of this type of data include longitudinally collected MRI lesion counts in RMS trials. Procedures will be presented for sample size estimation and blinded sample size reestimation in clinical trials with such data. A simulation study is conducted to assess the properties of the proposed procedures.

References:

1. Nicholas et al., 2011. “Trends in annualized relapse rates in relapsing-remitting multiple sclerosis and consequences for clinical trial design". Multiple Sclerosis Journal, Vol. 17, pp. 1211-1217. SAGE.

2. Friede Tim and Schmidli Heinz, 2010. “Blinded sample size reestimation with count data: Methods and applications in multiple sclerosis". Statistics in Medicine, Vol. 29, pp. 1145-1156. John Wiley and Sons.

3. McKenzie Ed, 1986. “Autoregressive Moving-Average Processes with Negative-Binomial and Geometric Marginal Distributions". Advances in Applied Probability, Vol. 18, No. 3, pp. 679-705. Applied Probability Trust.

Efficient use of omics data for time to event prediction models: selective sampling and its alternatives

Becker, Natalia Benner, Axel

DKFZ, Heidelberg, Deutschland
natalia.becker@dkfz.de benner@dkfz.de

To improve prediction of patients' clinical outcome a new approach is to use individual omics data to develop a prognostic signature. However, due to resource or budget limitations omics data can often only be measured for a subset of patient samples. It is therefore important to choose the most efficient sampling design with respect to model selection in high dimensions.

A simple and frequently used approach in practice is selective sampling. For survival endpoints, for example, samples are selected from "poor prognosis" (early deaths) and "good prognosis" (long term survivors) subsets. Data analysis is then performed by two-sample comparisons ignoring individual survival times, e.g. using logistic regression models.

Reasonable alternatives are nested case-control and case-cohort sampling, well known in epidemiological applications [1,2]. However, these sampling techniques are suited only for studies investigating rare events, which is not typical for clinical research. The need to perform simultaneous variable selection via penalization presents an additional difficulty.

Here we consider 'generalized case-cohort sampling', which is a generalization of nested case-control and case-cohort sampling that allows the sampling of non-rare events [3]. We provide an implementation of generalized case-cohort sampling for high dimensional data and compare this approach with selective and simple random sampling. Prediction models using the full cohort are considered as gold standard. Comparisons of sampling designs will be performed by a simulation study. True positive rates and false selection rates as proposed by G'Sell et al. [4] will be used as model selection criteria.

Finally, we present the application of the different sampling designs on a real data example.

References:

1. Prentice R L. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 1986; 73: 1–11.

2. Liddell F D K, McDonald J C, Thomas D C. Methods of cohort analysis: appraisal by application to asbestos mining. Journal of the Royal Statistical Society. Series A 1977; 140: 469–491.

3. Cai J, Zeng D. Power calculation for case-cohort studies with nonrare events. Biometrics 2007; 63: 1288-1295.

4. G'Sell M G, Hastie T, Tibshirani R. False variable selection in regression, arXiv preprint 2013, http://arxiv.org/abs/1302.2303

Bewertung von randomisierten kontrollierten Studien unter Berücksichtigung von Treatment Switching: biometrische Anforderungen aus der Sicht des IQWiGs

Beckmann, Lars Grouven, Ulrich Guddat, Charlotte
Köhler, Michael Vervölgyi, Volker Bender, Ralf

IQWiG, Deutschland
lars.beckmann@iqwig.de
ulrich.grouven@iqwig.de
charlotte.guddat@iqwig.de
michael.koehler@iqwig.de
volker.vervoelgyi@iqwig.de
ralf.bender@iqwig.de

In onkologischen Studien ist im Studienverlauf häufig ein vorzeitiger Wechsel der Behandlung von der Kontrollbehandlung auf die neue Therapie, möglich (Treatment Switching). Dies geschieht zumeist nach (radiologischer) Progression der Erkrankung. Ein Treatment Switching nach Progression kann für Endpunkte, die nach der Progression auftreten (z. B. das Gesamtüberleben) zu einer Verzerrung des Effektschätzers führen. Auch vor dem Hintergrund, dass diese Verzerrung sowohl zugunsten als auch zuungunsten der neuen Therapie möglich ist, kann ein Treatment Switching dazu führen, dass die Ergebnisse nicht mehr sinnvoll interpretierbar sind.

Im Vortrag werden die Aspekte diskutiert, die aus biometrischer Sicht für eine adäquate Beurteilung der Ergebnisse im Rahmen einer Nutzenbewertung notwendig sind.

Die Ergebnisse der ITT-Analyse sollen immer dargestellt werden. Weitere Ergebnisse aus Analysen, die Treatment Switching berücksichtigen können dargestellt werden. Die in der Literatur vorgeschlagenen Methoden beruhen auf Annahmen bzw. Voraussetzungen, die teilweise nicht überprüfbar sind und deren Plausibilität zu begründen ist. Daher sind für eine Bewertung dieser Analysen zusätzliche Informationen notwendig, die über die im CONSORT Statement geforderten Angaben hinausgehen. Dazu gehören unabhängig von den gewählten Methoden detaillierte Angaben zum Mechanismus, der zum vorzeitigen Behandlungswechsel führt sowie eine genaue Beschreibung der Patientenpopulation, die vorzeitig wechselt. Des Weiteren werden methodenspezifische Anforderungen gestellt. Die Ergebnisse sollten hinsichtlich einer möglichen Verzerrung und der Präzision der Effektschätzung diskutiert werden. Weiterhin soll diskutiert werden, inwiefern diese Analysen geeignet sind, einen von der Nullhypothese verschiedenen Effekt nachzuweisen, bzw. sogar Aussagen zum Ausmaß dieses Effektes erlauben. Die notwendigen Informationen müssen ggf. zusätzlich für interessierende Subgruppen ausreichend beschrieben werden.

Benefit risk assessment in the drug approval process

Benda, Norbert

BfArM, Deutschland
norbert.benda@bfarm.de

Marketing authorization of a new medicine requires that its benefits outweigh it undesired effects or risks. Balancing benefits against risk is a difficult task, where the main issues relate to reasonable quantitative measures that incorporate different kinds of outcomes and the uncertainty related to the information given by the relevant clinical studies. In order to avoid inconsistencies related to subjective case-by-case decisions a methodological framework is needed to ensure a transparent and undisputable assessment process. The talk will discuss different proposals together with the main pitfalls and challenges and a roadmap that could be followed to come up with a sensible and traceable decision process.

Florence Nightingale, William Farr and competing risks

Beyersmann, Jan Schrade, Christine

Uni Ulm, Deutschland
jan.beyersmann@uni-ulm.de christine.schrade@uni-ulm.de

Aiming for hospital sanitary reforms, Florence Nightingale (1820-1910), a pioneer of nursing and modern hospital hygiene, joined forces with William Farr (1807-1883), whose 1885 book Vital Statistics is an early important contribution to statistics for the life sciences. Today, Farr's book is well remembered for his clear recognition of time-dependent (immortal time) bias. He mockingly argued: "make [young men] Generals, Bishops, and Judges - for the sake of their health!" As in Vital Statistics, Nightingale and Farr used incidences densities to quantify hospital mortality, that is, the number of hospital deaths divided by the hospital population time at risk. Nightingale's and Farr's methodological choice was criticized by their contemporaries, one issue being that incidence densities may exceed 100%. Interestingly, these questions are still with us today. In hospital epidemiology, quantifying incidence of in-hospital events typically uses incidence densities, although hospital mortality is nowadays almost exclusively expressed as the number of hospital deaths divided by the size of the hospital population, the so-called incidence proportion. In randomized clinical trials, the risk of adverse events is often quantified via incidence densities. However, their use is also criticized in this context, one issue being possibly informative censoring by the primary trial outcome. This talk outlines that the modern survival analytical concepts of competing risks and cause-specific hazards overcome such criticism and, e.g., reconcile the concepts of incidence densities and incidence proportion [1]. We will show that Nightingale and Farr were essentially aware of these connections.

References:

1. Beyersmann, J, Gastmeier, P, Schumacher, M: Incidence in ICU populations: how to measure and report it? Intensive Care Medicine 2014; 40: 871-876

Comparing multi-state approaches for reducing the bias of relative risk estimates from cohort data with missing information due to death

Binder, Nadine Herrnböck, Anne-Sophie Schumacher, Martin

Institute for Medical Biometry and Statistics, Medical Center - University of Freiburg
nadine@imbi.uni-freiburg.de as.herrnboeck@googlemail.com ms@imbi.uni-freiburg.de

In clinical and epidemiological studies information on the outcome of interest (e.g. disease status) is usually collected at a limited number of follow-up visits. The disease status can often only be retrieved retrospectively in individuals who are alive at follow-up, but will be missing for those who died before. Right-censoring the death cases at the last visit (ad-hoc analysis) yields biased hazard ratio estimates of a potential risk factor, and the bias can be in either direction [1].

We focus on (semi-)parametric approaches that use the same likelihood contributions derived from an illness-death multi-state model for reducing this bias by including the death cases into the analysis: first, a penalized likelihood approach by Leffondré et al. [2] and second, an imputation based approach by Yu et al. [3]. We compare the two approaches in simulation studies and evaluate them on completely recorded real data, where missing information due to death is artificially induced. All considered approaches can reduce the bias resulting from an ad-hoc analysis to a different extent and under different circumstances depending on how well the underlying assumptions are fulfilled.

References:

1. Binder N, Schumacher M. Missing information caused by death leads to bias in relative risk estimates. J Clin Epidemiol. 2014;67(10):1111–20.

2. Leffondré K, Touraine C, Helmer C, Joly P. Interval-censored time-to-event and competing risk with death: is the illness-death model more accurate than the Cox model? Int J Epidemiol. 2013;42(4):1177–86.

3. Yu B, Saczynski JS, Launer L. Multiple imputation for estimating the risk of developing dementia and its impact on survival. Biom J. 2010;52(5):616–27.

Detecting SNP interactions associated with disease using model-based recursive partitioning

Birke, Hanna Schwender, Holger

Heinrich-Heine-Universität, Düsseldorf, Deutschland
hanna.birke@hhu.de holger.schwender@hhu.de

In genome-wide association studies, it is presumed that not only individual single nucleotide polymorphisms (SNPs), but also combinations of SNPs are associated with complex disease.

To deal with these high-dimensional data, ensemble methods such as Random Forest as well as logic regression have been used to analyze SNP interactions related to disease risk. A flexible alternative to these procedures is model-based recursive partitioning (MOB) [1]. This algorithmic method combines non-parametric classification trees with parametric models, and includes an automated variable selection and interaction detection. Linear and generalized linear models as well as the Weibull model have been integrated into the MOB algorithm. In this presentation, we illustrate how MOB can be used to detect SNP interaction in case-control data.

These MOB approaches assume independence of the study participants. However, in case-parent trio studies considering genotype data of children affected by a disease and their parents, the individuals are related. Only very few methods have been proposed for the detection of disease-associated SNP interactions in such case-parent trio data. Here, we extend the MOB algorithm to case-parent trio data by incorporating a conditional logistic regression model, which allows to account for the dependency structure in the data. This adaptation can be applied to a wide range of genetic models, including a dominant, recessive, or additive model.

The performance of the suggested technique is evaluated in a simulation study and applied to case-parent trios from the International Cleft Consortium.

References:

[1] Zeileis, A., Hothorn, T., and Hornik, K. (2008). Model-Based Recursive Partitioning. Journal of Computational and Graphical Statistics, 17(2), 492–514.

Detecting time-dependency in live cell-imaging

Bissantz, Nicolai

Ruhr-Universität Bochum, Deutschland
nicolai.bissantz@rub.de

We introduce a method for the monitoring and visualization of statistically significant structural changes over time between frames in fluorescence microscopy of living cells.

The method can similarly be adopted for use in other imaging systems and helps to detect significant structural changes between image frames which are not attributable to random noise in the image. It delivers valuable data to judge about time-resolved small scale structural changes e.g. in the nanometer range.

Here we consider the question of statistical recovery of time-resolved images from fluorescence microscopy of living cells.

The question if structural changes in time-resolved images are of statistical significance and therefore of scientific interest or just due to random noise e.g. intracellular diffusion processes is a well ignored point in practical applications such as live cell fluorescence microscopy.

The proposed method is applicable to both images and image changes and it is based on data reconstruction with a regularization method as well as new theoretical results on uniform confidence bands for the function of interest in a two-dimensional heteroscedastic nonparametric convolution-type inverse regression model.

In particular a new uniform limit theorem for the case of Poisson-distributed observations is presented and a strong approximation result for the two- dimensional array of non-identically distributed Poisson-residuals by an array of independent and identically distributed Gaussian random variables is derived.

A data-driven selection method for the regularization parameter based on statistical multiscale methods is also discussed.

The theoretical results are used to analyse real data of fluorescently labelled intracellular transport compartments in living cells.

Confidence Bands for Nelson-Aalen Estimates in a Multistate Model: The Wild Bootstrap Approach with Applications in Health Services Research

Bluhmki, Tobias [1] Buechele, Gisela [2] Beyersmann, Jan [1]

[1] Institute of Statistics, Ulm University, Germany
tobias.bluhmki@uni-ulm.de jan.beyersmann@uni-ulm.de
[2] Institute of Epidemiology and Medical Biometry, Ulm University, Germany
gisela.buechele@uni-ulm.de

Based on a large Bavarian health insurance data set, we investigate the impact of a preceding non-femoral (osteoporotic) index fracture on the incidence of and mortality after a femoral fracture within an elderly population. In this medical context, risks are often expressed in terms of hazards as a function of "time-on-study". The commonly used incidence and mortality rates are not suitable for estimation in this data, since the required constant hazard assumption is violated. We present an alternative approach based on a complex fracture-death multistate model allowing for non-constant hazards. We include "long-term care" as a time-dependent variable. "Age" is taken into account as the underlying time-scale leading to left-truncated data. Therefore, classical Kaplan-Meier estimation is not appropriate. In our setting, risks are non-parametrically estimated by means of Nelson-Aalen estimates and time-simultaneous statistical inference can be based on a flexible wild bootstrap approach. The idea is to keep the data fixed and to approximate the underlying limit processes by substitution of the unknown martingales with standard normal random variables. Repeated simulations enable the construction of confidence bands. Proofs are based on martingale limit theorems [1] and recent results on linear resampling statistics in martingale difference arrays [2]. We show that a prior index fracture increases the risk of a femoral fracture, but does not affect mortality after a femoral fracture. The interpretation is that index fractures increase absolute mortality risk via increasing the femoral fracture hazard without altering the mortality hazard.

References:

[1] Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N.: Statistical models based on counting processes. Springer New York (1993).

[2] Pauly, M. (2011): Weighted resampling of martingale difference arrays with applications. Electronical Journal of Statistics. 5, 41–52.

Real-time detection of trends in time series of carbon dioxide concentration in exhalation air

Borowski, Matthias [1] Kourelas, Laura [2] Bohn, Andreas [3]

[1] Institute of Biostatistics and Clinical Research, University of Muenster, Germany
matthias.borowski@ukmuenster.de
[2] Department of Anesthesiology, Intensive Care and Pain Medicine, University Hospital Muenster, Germany
[3] City of Muenster Fire Department, Muenster, and Department of Anesthesiology, Intensive Care and Pain Medicine, University Hospital Muenster, Germany

Even experienced physicians cannot predict the success of reanimations. However, there are clear indications that a return of spontaneous circulation (ROSC) is preceded by a rising concentration of carbon dioxide in the patient's exhalation air [1]. We propose a procedure which is able to detect current trends in time series online (i.e. in real-time) and therefore offers the opportunity to assess the success of a reanimation. Thus, frequent interruptions of the reanimation (which are necessary to check if a ROSC has occurred) can be avoided. Our method for online trend detection is based on the SCARM (Slope Comparing Adaptive Repeated Median), a procedure for robust online signal extraction [2]. A robust Repeated Median regression line is fitted within a moving window sample, where the window width is chosen automatically w.r.t. the current data situation at each time point. Then the slope estimate of the regression line can be standardized according to the current variability of the time series which is determined using a trend-invariant scale estimator [3]. Hence, extreme values of the standardized slope estimate indicate a current trend.

In a clinical example of carbon dioxide time series from 77 successful and 92 unsuccessful reanimations, the potential of the trend detection procedure is illustrated and findings from literature are confirmed.

References:

[1] Pokorná, M., Nečas, E., Kratochvíl, J., Skřipský, R., Andrlík, M., Franěk, O. (2010): A sudden increase in partial pressure end-tidal carbon dioxide (PETCO2) at the moment of return of spontaneous circulation. The Journal of Emergency Medicine 38, 614–621.

[2] Borowski, M., Fried, R. (2014): Online signal extraction by robust regression in moving windows with data-adaptive width selection. Statistics and Computing 24, 597–613.

[3] Gelper, S., Schettlinger, K., Croux, C., Gather, U. (2009): Robust online scale estimation in time series: a model-free approach. Journal of Statistical Planning and Inference 139(2), 335–349.

Publication bias in methodological computational research

Boulesteix, Anne-Laure [1] Stierle, Veronika [1] Hapfelmeier, Alexander [2]

[1] Ludwig-Maximilians-Universität München, Deutschland
boulesteix@ibe.med.uni-muenchen.de veronika.stierle@web.de
[3] Technische Universität München, Deutschland
alexander.hapfelmeier@tum.de

The problem of publication bias has long been widely discussed in research fields such as medicine. There is a consensus that publication bias is a reality and that solutions should be found to reduce it. In methodological computational research publication bias may also be at work and the publication of negative research findings is certainly also a relevant issue, but has attracted only very little attention to date. Discussions of these issues mostly occur in blogs of individual scientists or letters to the editors, but have been much less investigated than in the context in biomedicine. To briefly sketch the importance of the topic, imagine that ten teams of computational-methodological researchers around the world work on the same specific research question and have a similar promising idea that, in fact, is not valid. Eight of the ten teams obtain "disappointing" results, i.e. recognize the futility of the approach. The ninth team sees a false positive, i.e. observes significant superiority of the new promising method over existing approaches, although it is in fact not better. The tenth team optimizes the method's characteristics and thus also observes significant superiority by "fishing for significance". The two latter teams report the superiority of the promising idea in papers, while the eight other studies with negative results remain unpublished: a typical case of publication bias. This scenario is certainly caricatural, but similar occurrences are likely to happen in practice. This is problematic because literature may give a distorted picture of current scientific knowledge, leading to suboptimal standards or wasted time when scientists pursue ideas that have already been found useless by others. In this talk I discuss this issue in light of first experiences from a pilot study on a possible publication bias with respect to the ability of random forests to detect interacting predictors.

Entwicklung und Implementierung eines "Prüfarzt-Tracks" im Studiengang Humanmedizin

Braisch, Ulrike Mayer, Benjamin Meule, Marianne
Rothenbacher, Dietrich Muche, Rainer

Universität Ulm, Institut für Epidemiologie und Medizinische Biometrie, Deutschland
ulrike.braisch@uni-ulm.de benjamin.mayer@uni-ulm.de marianne.meule@uni-ulm.de
dietrich.rothenbacher@uni-ulm.de rainer.muche@uni-ulm.de

Hintergrund An der Universität Ulm können Studierende der Humanmedizin im Rahmen des sogenannten Track-Programms ihre individuellen Schwerpunkte bereits im Studienverlauf setzen. Gemäß dem Konzept des "Forderns und Förderns" erfahren die Studierenden im Rahmen der Tracks eine enge Anbindung an die jeweiligen Fachdisziplinen mit intensiver Betreuung in Lehre, (vor-)klinischer Praxis und Forschung. Im Gegenzug erbringen die Studierenden eine Mehrleistung gegenüber den Regelstudierenden. Bislang wurden Studientracks in den Bereichen Neurologie, Herz-Lunge-Gefäß, Traumaversorgung und Traumaforschung, Experimentelle Medizin und Lehren Lernen implementiert.

Methoden Derzeit erfolgt die Entwicklung und Implementierung eines methodisch ausgerichteten "Prüfarzt-Tracks", der zu Beginn des Sommersemesters 2015 erstmals von maximal fünf Ulmer Studierenden pro Studienjahr der Humanmedizin belegt werden kann. Den Studierenden sollen dabei sowohl die statistischen Methoden des wissenschaftlichen Arbeitens im Rahmen der klinischen Forschung, als auch die ethischen Aspekte vermittelt werden. Hierzu muss ein entsprechendes Curriculum erstellt werden, welches mit dem regulären Curriculum des Humanmedizinstudiums zeitlich kompatibel ist. Dazu werden u.a. von uns angebotene Lehrexportfächer an die Studiengänge Mathematische Biometrie, Molekularmedizin und Medizininformatik genutzt. Zudem müssen die Voraussetzungen für eine individuelle Betreuung der Trackstudierenden geschaffen werden.

Durchführung Der Studientrack ist für eine Dauer von vier Semestern konzipiert und beinhaltet verschiedene Pflicht- und Wahlveranstaltungen, auf deren Basis nach Abschluss des Tracks das Prüfarzt-Zertifikat verliehen werden kann [1,2]. Zusätzlich werden die Trackstudierenden eine durch uns unterstützte, methodisch orientierte Dissertation mit einem klinischen Kooperationspartner anfertigen. Parallel zu den regulären Veranstaltungen werden die Teilnehmer des Tracks ein Doktorandenseminar besuchen, in dessen Rahmen die Studierenden unter anderem in den Bereichen Literaturrecherche und wissenschaftliches Schreiben und Präsentieren eingehend und individuell geschult werden sollen. Derzeit ausstehend sind noch Überlegungen zur Evaluation des entwickelten Tracks sowie zur Konzeption entsprechender Werbemaßnahmen, um die Studierenden auf den Prüfarzt-Track aufmerksam zu machen.

Literatur:

[1] Bundesärztekammer: Curriculare Fortbildung - Grundlagenkurs für Prüfer/Stellvertreter und Mitglieder einer Prüfgruppe bei klinischen Prüfungen nach dem Arzneimittelgesetz (AMG) und für Prüfer nach dem Medizinproduktegesetz (MPG). Deutsches Ärzteblatt 110 (2013), Heft 23-24, A1212–A1219.

[2] Neussen N, Hilgers RD: Neue Wege in der biometrischen Ausbildung im Rahmen des Medizinstudiums – individuelle Qualifikationsprofile. GMS Med

Inform Biom Epidemiol (2008), 4:Doc01.

On the error of incidence estimation from prevalence data

Brinks, Ralph Hoyer, Annika Landwehr, Sandra

Deutsches Diabetes-Zentrum, Institut für Biometrie und Epidemiologie, Deutschland
ralph.brinks@ddz.uni-duesseldorf.de annika.hoyer@ddz.uni-duesseldorf.de
sandra.landwehr@ddz.uni-duesseldorf.de

In epidemiology, incidence rates are typically estimated by lengthy, and often costly, follow-up studies. Cross-sectional studies usually require less effort. Recently, we have shown how to estimate the age-specific incidence rate of an irreversible disease by two cross-sectional studies in case the mortality rates of the diseased and the healthy population are known [1]. The cross-sections are needed to estimate the age-specific prevalences at two points in time. We use a simulation study [2] to examine the sources of errors affecting the incidence estimation. The rates are chosen to mimic the situation of type 2 diabetes in German males.

Three sources of errors are identified: (i) a systematic error which is given by the study design (or the available data), (ii) the age sampling of the prevalences in the cross-sections, and (iii) by the error attributable to sampling the population.

In the simulation, the error source (i) leads to a relative error between -2.4% and 0.7% (median -0.5%) in the age range 40 to 95 years. The error (ii) has a similar magnitude. These errors are negligible compared to the error source (iii), in which the coefficient of variation reaches up to 10% (in the simulated settings). Thus, the sampling error of the population seems to be the limiting factor to obtain accurate estimates of the incidence. As a conclusion, special care about the sampling should be taken in practical applications.

References:

[1] Brinks R, Landwehr S: Relation between the age-specific prevalence and incidence of a chronic disease and a new way of estimating incidence from prevalence data, Mathematical Medicine and Biology (accepted)

[2] Brinks R et al: Lexis Diagram and Illness-Death Model: Simulating Populations in Chronic Disease Epidemiology, PLoS One 2014, DOI 10.1371/journal.pone.0106043

Utilizing surrogate information in adaptive enrichment designs with time-to-event endpoints

Brückner, Matthias Brannath, Werner

Universität Bremen, Deutschland
mwb@math.uni-bremen.de brannath@math.uni-bremen.de

In two-stage adaptive enrichment designs with subgroup selection the aim of the interim analysis at the end of the first stage is to identify promising subgroups which have a sufficiently large treatment effect. The decision whether or not to continue with recruitment in a given subgroup is based on the estimated treatment effect. When the estimate is above a given threshold the subgroup is selected and otherwise dropped. The trade-off between sensitivity (true positive rate) and specificity (true negative rate) of a decision rule can be quantified independently of the threshold by the area under the ROC curve (AUC).

At the time of the interim analysis, long-term endpoints such as overall survival are usually not yet observed for many of the recruited patients resulting in a large percentage of censored observations. Often surrogate endpoints, such as progression-free survival or tumor response, are available instead. These surrogate endpoints are themselves not observed immediately after randomisation and may be missing for some patients.

We will discuss how estimators of the treatment effect can be constructed using surrogate and primary endpoint information. We investigate how much can be gained from using these estimators in the interim decision. The finite-sample performance, especially for small sample sizes, is compared in Monte-Carlo simulations using the AUC as performance measure.

Stichprobenplanung für allgemeine Rangtests

Brunner, Edgar

Universitätsmedizin Göttingen, Deutschland
ebrunne1@gwdg.de

Es wird eine allgemeine Methode vorgestellt, mit der man für Rangtests die benötigte Fallzahl für beliebige Datentypen berechnen kann. Dazu wird die Rangstatistik durch eine asymptotisch äquivalente Statistik von unabhängigen (nicht beobachtbaren) Zufallsvariablen dargestellt, die asymptotisch normalverteilt ist. Darauf werden die bekannten Methoden zur Bestimmung der Fallzahl angewendet. Die aus den vorhandenen Vorinformationen bekannte Verteilung F_1 der Kontrollbehandlung wird durch einen künstlichen Datensatz so nachgeahmt, dass dessen empirische Verteilung genau F_1 entspricht. Der sich aus praktischen Überlegungen ergebende relevante Unterschied erzeugt dann die Verteilung F_2 der experimentellen Behandlung, die ebenfalls durch einen künstlichen Datensatz nachgeahmt wird. Die für die Fallzahlplanung benötigten Größen werden dann aus den Mittelwerten und empirischen Varianzen der Differenzen von den Global- und Intern-Rängen (Placements, [1], [3]) bestimmt. Für verbundene Stichproben wird noch eine Vorinformation über die Korrelation der Ränge bzw. über die Varianz der Rangdifferenzen benötigt.

Das Verfahren gilt allgemein in beliebigen Designs und ist für stetige Daten, Zähldaten sowie geordnet kategoriale Daten anwendbar. Als Spezialfall ergibt sich für den Wilcoxon Test die Noether-Formel [2] für stetige Daten und für geordnet kategoriale Daten das Resultat von Tang [4]. Das Verfahren wird anhand von Beispielen erläutert.

Literatur:

[1] Brunner, E. and Munzel, U. (2000). The nonparametric Behrens-Fisher-Problem: Asymptotic theory and a small sample approximation. Biometrical Journal 42,17–25.

[2] Noether, G.E. (1987). Sample Size Determination for Some Common Nonparametric Tests. Journal of the American Statistical Association 82, 645–647.

[3] Orban, J. and Wolfe, D.A. (1982). A Class of Distribution-Free Two-Sample Tests Based on Placements. Journal of the American Statistical Association 77, 666–672.

[4] Tang, Y. (2011). Size and power estimation for the Wilcoxon-Mann-Whitney test for ordered categorical data. Statistics in Medicine 30, 3461–3470.

Analyse der Machbarkeit der Surrogatvalidierung nach IQWiG-Methodik: Ergebnisse von Simulationsstudien

Buncke, Johanna [1] Goertz, Ralf [2] Jeratsch, Ulli [2]
Leverkus, Friedhelm [1]

[1] Pfizer Deutschland GmbH johanna.buncke@pfizer.com friedhelm.leverkus@pfizer.com [2] AMS Advanced Medical Services GmbH Ralf.Goertz@ams-europe.com Ulli.Jeratsch@ams-europe.com

In onkologischen Studien wird oftmals statt des patientenrelevanten Endpunkts Gesamtüberleben (overall survival, OS) der Endpunkt progressionsfreies Überleben (progression-free survival, PFS) erfasst. Für eine Anerkennung von PFS als patientenrelevant im Verfahren zur Nutzenbewertung, gilt es dieses als Surrogatendpunkt für OS in der betrachteten Indikation zu validieren. Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) [1] hat Methoden zur Validierung von Surrogatendpunkten dargestellt und Empfehlungen zur Verwendung von korrelationsbasierten Verfahren ausgesprochen. Neben der Einschätzung der Aussagesicherheit muss für den Nachweis der Validität für das Surrogat auf Studienebene ein gleichgerichteter Zusammenhang zwischen den Effekten des Surrogats und des patientenrelevanten Endpunkts vorliegen. Dieser wird durch die Korrelation mit dem Korrelationskoeffizienten ρ bzw. dem Bestimmtheitsmaß R^2 gemessen. Die Validität des Surrogats kann in drei Kategorien hinsichtlich der Korrelation mit dem entsprechenden Konfidenzintervall eingestuft werden: hoch, mittel und niedrig. Im Falle einer mittleren Korrelation kann das Konzept des Surrogate Threshold Effects (STE) [2] zur Validierung angewandt werden.

In Simulationsstudien wurde nun untersucht, welche Bedingungen für eine erfolgreiche Surrogatvalidierung mit korrelations-basierten Verfahren erfüllt sein müssen. Variierende Parameter sind die Effekte der Endpunkte, die Korrelation zwischen den Effekten, die Patientenanzahl sowie die Anzahl der Studien. Es wird analysiert, in welchen Szenarien der Nachweis einer hohen Korrelation gelingt und falls nicht, welche Voraussetzungen für die erfolgreiche Durchführung des STE-Konzept vorliegen müssen. Die Herausforderungen der vom IQWiG präferierten Methodik zur Surrogatvalidierung in der Praxis werden analysiert.

References:

1: IQWiG, 2011: Aussagekraft von Surrogatendpunkten in der Onkologie (Rapid Report).

2: Burzykowski T, Molenberghs G, Buyse M, 2005: The evaluation of surrogate endpoints. Springer.

The use of hidden Markov-models to analyze QTL-mapping experiments based on whole-genome next-generation-sequencing data

Burzykowski, Tomasz

Hasselt University, Belgium
tomasz.burzykowski@uhasselt.be

The analysis of polygenetic characteristics for mapping quantitative trait loci (QTL) remains an important challenge. QTL analysis requires two or more strains of organisms that differ substantially in the (poly-)genetic trait of interest, resulting in a heterozygous offspring. The offspring with the trait of interest is selected and subsequently screened for molecular markers such as single nucleotide polymorphisms (SNPs) with next-generation sequencing (NGS). Gene mapping relies on the co-segregation between genes and/or markers. Genes and/or markers that are linked to a QTL influencing the trait will segregate more frequently with this locus. For each identified marker, observed mismatch frequencies between the reads of the offspring and the parental reference strains can be modeled by a multinomial distribution with the probabilities depending on the state of an underlying, unobserved Markov process. The states indicate whether the SNP is located in a (vicinity of a) QTL or not. Consequently, genomic loci associated with the QTL can be discovered by analyzing hidden states along the genome.

The methodology based on the aforementioned hidden Markov-model (HMM) is quite flexible. For instance, a non-homogenous HMM, with a transition matrix that depends on the distance between neighboring SNPs, can be constructed. In the presentation, this and other advantages of the proposed approach will be discussed and illustrated on data from a study of ethanol tolerance in yeast.

Methods to monitor mental fatigue in operators using EEG signals

Charbonnier, Sylvie

ENSE3, France
sylvie.charbonnier@gipsa-lab.grenoble-inp.fr

During the realization of monotonous and repetitive tasks, mental fatigue, or reduced alertness, arises with growing time-on-task (TOT). This is a gradual and cumulative process that leads to shallow or even impaired information processing and can therefore result in a significant decrease in performance. Due to its major safety applications, mental fatigue estimation is a rapidly growing research topic in the engineering field.

EEG signals, recorded from several electrodes positioned on the scalp, provide information on cerebral activity and can be used to monitor mental fatigue on operators that monitor complex systems during long periods of time, such as air traffic controllers or nuclear plants operators, who have to concentrate during long periods on information displayed on a screen. The new technology that is now emerging to record EEG in an easy and practical way, such as EEG headsets or caps with dry electrodes, makes it possible to envision an EEG system that would monitor operators' mental state for long periods.

During this talk, we will present different methods designed to monitor mental fatigue using EEG signals from 32 electrodes. Different strategies will be presented. Some are inspired from active brain computer interface (BCI) methods, using spatial filters to enhance the difference between signals recorded during different fatigue states and classifiers to assign a mental state to an epoch of signals. Others provide indicators of the evolution of the operator's fatigue state from a reference state learnt at the beginning of the recording. They use statistical distances between covariance matrices or feature vectors. The results obtained by the different methods will be evaluated and discussed on a data base. This data base is formed of EEG signals recorded on several subjects during an experiment where each subject spent a long time on a demanding cognitive task.

Boosting in Cox regression: a comparison between likelihood-based and model-based approaches with focus on the R packages CoxBoost and mboost

De Bin, Riccardo

University of Munich, Germany
debin@ibe.med.uni-muenchen.de

Despite the limitations imposed by the proportional hazards assumption, the Cox model is probably the most popular statistical tool in analyzing survival data, thanks to its flexibility and ease of interpretation. For this reason, novel statistical/machine learning techniques are usually adapted to fit it. This is the case with boosting, originally developed in the machine learning community to face classification problems, and later extended to several statistical fields, including survival analysis. In a parametric framework, the basic idea of boosting is to provide estimates of the parameters by updating their values iteratively: at each step, a weak estimator is fit on a modified version of the data, with the goal of minimizing a loss function. In the Cox model framework, the loss function is the negative partial log-likelihood. A critical tuning parameter, namely the number of boosting steps, serves as the stopping criterion and influences important properties such as resistance to overfitting and variable selection. The latter is relevant in the case of component-wise boosting, in which the explanatory variables are treated separately, allowing the handling of high-dimensional data. The popularity of boosting has been further driven by the availability of user-friendly software such as the R packages *mboost* and *CoxBoost*, both of which allow the implementation of component-wise boosting in conjunction with the Cox model. Despite the common underlying boosting principles, these two packages use different techniques: the former is an adaption of model-based boosting, while the latter adapts likelihood-based boosting. Here we compare and contrast these two boosting techniques, as implemented in the R packages. We examine solutions currently only implemented in one package and explore the possibility of extending them to the other, in particular, those related to the treatment of mandatory variables.

Bayesian Augmented Control Methods for Efficiently Incorporating Historical Information in Clinical Trials

DiCasoli, Carl Kunz, Michael Haverstock, Dan

Bayer Healthcare Pharmaceuticals, Berlin, Germany, and Whippany, New Jersey USA
carl.dicasoli@bayer.com michael.kunz@bayer.com daniel.haverstock@bayer.com

In new clinical trials for anti-infective therapies, there is often historical clinical data available and a potential challenge is to derive a valid method for testing non-inferiority taking the historical information into account. Recently Viele et al. (5) have presented approaches that incorporate this historical clinical data into an analysis procedure. We expand upon their idea of dynamic borrowing in the framework of various additional Bayesian hierarchical modeling strategies and derive the resulting Type I error, power, and DIC. These additional strategies may include estimating parameters for each trial separately versus pooling, weighting the prior distribution corresponding to each historical study based on the sample size, and incorporating historical borrowing on the control arm by a separate random effect parameter. Furthermore, one critical assumption in non-inferiority trials is constancy; that is, the effect of the control in the historical trial population is similar to the effect in the current active control trial population. The constancy assumption can be at risk due to the potential heterogeneity between trial populations primarily related to subject characteristics, and secondarily to other sources of heterogeneity resulting from differences in patient management (e.g., usage of concomitant medications). As a further refinement, we present how in the setting of a non-inferiority trial, a covariate adjustment approach can be implemented to recalibrate the non-inferiority margin based on the effect of active control minus placebo from the current study data.

References:

1. CBER and CDER FDA Memorandum. Guidance for Industry: Non-Inferiority Clinical Trials, March 2010.

2. CBER and CDER FDA Memorandum. Guidance for Industry: Antibacterial Therapies for Patients with Unmet Medical Need for the Treatment of Serious Bacterial Diseases, July 2013.

3. Jones, Ohlssen, Neuenschwander, Racine, and Branson. Bayesian models for subgroup analysis in clinical trials. Clinical Trials 2011; 8:129.

4. Nie, L and Soon G. A covariate-adjustment regression model approach to noninferiority margin definition. Statistics in Medicine 2010; 29: 1107–1113.

5. Viele, Berry, Neuenschwander, Amzal, Chen, Enas, Hobbs, Ibrahim, Kinnersley, Lindborg, Micallef, Roychoudhury, Thompson. Use of historical control data for assessing treatment effect in clinical trials. Pharmaceutical Statistics 2013.

SPINA und SEPIA: Algorithmen für die Differentialdiagnostik und personalisierte Therapieplanung in der Endokrinologie

Dietrich, Johannes W.

BG Universitätsklinikum Bergmannsheil, Deutschland
johannes.dietrich@ruhr-uni-bochum.de

Die Plasma-Konzentrationen fast aller Hormone werden von teils komplexen Regelkreisen konstant gehalten. Da Sollwerte und Strukturparmeter der Regelkreise meist in erheblichem Ausmaße genetisch bestimmt sind, ist die intraindividuelle Variabilität der Hormonkonzentration meist geringer als die interindividuelle Variabilität. Im Falle von Folgereglern oder bei pulsatiler Sekretion kann die intraindividuelle Variabiltät ausnahmsweise auch größer sein. In jedem Falle ergibt sich daraus die diagnostische Konsequenz, dass klassische Referenzbereiche, die als 95%-Toleranzintervalle definiert sind, in der Endokrinologie nur mit Einschränkung anwendbar sind.

Dies könnte ein wesentlicher Grund für die reduzierte Lebensqualität von Patientinnen und Patienten mit Schilddrüsenerkrankungen sein. In 5 bis 10% der Fälle einer substituierten Hypothyreose, in denen eine Therapie mit Levothyroxin erfolgt und ein normaler TSH-Spiegel eine hinreichende Substitution suggeriert, klagen die Betroffenen dennoch über charakteristische und wiederkehrende Beschwerden wie Müdigkeit, Depression und Palpitationen.

Aus theoretischen Erwägungen erscheint es daher sinnvoll, als Therapieziel nicht mehr den breiten Referenzbereich des TSH-Spiegels sondern den Set-Point, d. h. den individuellen Gleichgewichtspunkt des Regelkreises heranzuziehen. Bislang war das jedoch nicht umsetzbar, da im Falle einer Hypothyreose, also eines aufgetrennten Regelkreises, der ursprüngliche Set-Point unbekannt ist.

Zwei neue Berechnungsverfahren, die auf einer systembiologischen Theorie der Schilddrüsenhomöostase aufbauen, erlauben es nun, konstante Strukturparameter und den Set-Point auch am aufgetrennten Regelkreis zu rekonstruieren. In Studien konnte gezeigt werden, dass damit Parameter wie die Sekretionsleistung der Schilddrüse mit einer hohen Reliabilität ermittelt werden können. Meist kann auch der wahre Set-Point des Regelkreises rekonstruiert werden, wobei die Streuung weniger als 50% der des univariaten TSH- oder FT4-Spiegels ausmacht.

Zusammenfassend ist es mit fortschrittlichen systembiologischen Methoden möglich, die individuelle Sensitivität der Diagnostik wesentlich zu verbessern und damit den Weg zu einer personalisierten Endokrinologie zu ebnen. Im Rahmen des Vortrags sollen diese Methoden im Detail und die Ergebnisse erster Studien vorgestellt werden.

A data-dependent multiplier bootstrap applied to transition probability matrices of inhomogeneous Markov processes

Dobler, Dennis Pauly, Markus

University of Ulm, Deutschland
dennis.dobler@uni-ulm.de markus.pauly@uni-ulm.de

The analysis of transition probability matrices of non-homogeneous Markov processes is of great importance (especially in medical applications) and it constantly gives rise to new statistical developments. While observations may be incomplete, e.g. due to random left-truncation and right-censoring, estimation of these matrices is conducted by employing the Aalen-Johansen estimator which is based on counting processes. However, results of weak convergence towards a Gaussian process cannot be utilized straightforwardly since the complicated limiting covariance structure depends on unknown quantities.

In order to construct asymptotically valid inference procedures, we insert a set of bootstrap multipliers (from a large class of possible distributions) into a martingale representation of this estimator. A new aspect to this approach is given by the possibility to choose these multipliers dependent on the data, covering, for instance, the Wild bootstrap as well as the Weird bootstrap. In doing so, we gain conditional weak convergence towards a Gaussian process with correct covariance functions resulting in consistent tests and confidence bands.

For small samples the performance in the simple competing risks set-up is assessed via simulation studies illustrating the type I error control and analyzing the power of the developed tests and confidence bands for several bootstrap multipliers.

Proper handling of over- and underrunning in phase II designs for oncology trials

Englert, Stefan [1] Kieser, Meinhard [2]

[1] Heidelberg, Deutschland
stefan.englert@gmx.net
[2] Universität Heidelberg, Deutschland
meinhard.kieser@imbi.uni-heidelberg.de

Due to ethical considerations, phase II trials in oncology are typically performed with planned interim analyses. The sample sizes and decision boundaries are determined in the planning stage such that the defined significance level and power are met. To assure control of the type I rate, these rules have to be followed strictly later on. In practice, however, attaining the pre-specified sample size in each stage can be problematic and over- and underrunning are frequently encountered.

The currently available approaches to deal with this problem either do not guarantee that the significance level is kept or are based on assumptions that are rarely met in practice. We propose a general framework for assuring type I error control in phase II oncology studies even when the attained sample sizes in the interim or final analysis deviate from the pre-specified ones. The method remains valid for data-dependent changes.

We show that the nominal type I error rate must be reduced in case of overrunning to ensure control of the significance level while this does not apply to underrunning. Further, we will investigate the power of the proposed method and show that practically relevant regions exist where the gained flexibility comes at no or almost no cost (<1% reduction in power).

Application of the proposed procedure and its characteristics are illustrated with a real phase II oncology study.

Propensity score methods to reduce sample selection bias in a large retrospective study on adjuvant therapy in lymph-node positive vulvar cancer (AGO CaRE-1)

Eulenburg, Christine [1] Suling, Anna [1] Neuser, Petra [2]
Reuss, Alexander [2] Wölber, Linn [3] Mahner, Sven [3]

[1] Medizinische Biometrie und Epidemiologie, Universitätsklinikum Hamburg-Eppendorf
C.Eulenburg@uke.de A.Drabik@uke.de
[2] KKS Philipps-Universität Marburg
Petra.Neuser@KKS.uni-marburg.de
Alexander.Reuss@KKS.uni-marburg.de
[3]Gynäkologie und Gynäkologische Onkologie, Universitätsklinikum Hamburg-Eppendorf
linn_woelber@gmx.de Sven.Mahner@gmx.de

Propensity score (PS) methods are frequently used to estimate causal treatment effects in observational and other non-randomized studies. In cancer research, many questions focus on time-to-event outcomes. We demonstrate the use of different propensity score methods, such as covariate adjusting, PS matching, PS stratification and inverse-probability-of-treatment-weighting (IPTW), in a time-to-event study of current clinical relevance. Analyses are applied to data from AGO CaRE-1, a large observational study, investigating the effect of adjuvant radiotherapy on disease recurrence and survival in lymph-node positive vulvar cancer. Additionally, the treatment effect on the cumulative incidences of disease-related mortality and death from other causes is studied in a competing risk setting. Several sensitivity analyses are performed.

The results obtained by the different PS methods agree that adjuvant radiotherapy is associated with improved prognosis. In the present case of a very large retrospective data set on a rare disease it appears that the IPTW method is the most appropriate technique, as it enables interpretable estimations of the average treatment effect under the assumption of no unmeasured confounders.

Nonparametric Mixture Modelling of Dynamic Bayesian Networks Derives the Structure of Protein-Networks in Adhesion Sites

Fermin, Yessica [1] Ickstadt, Katja [1] Sheriff, Malik R. [2]
Imtiaz, Sarah [2] Zamir, Eli [2]

[1] TU Dortmund University, Faculty of Statistics, Germany
fermin@statistik.uni-dortmund.de ickstadt@statistik.tu-dortmund.de
[2] Max Planck Institute of Molecular Physiology, Department of Systemic Cell Biology, Germany
sheriff@mpi-dortmund.mpg.de sarah.imtiaz@mpi-dortmund.mpg.de
eli.zamir@mpi-dortmund.mpg.de

Cell-matrix adhesions play essential roles in important biological processes including cell morphogenesis, migration, proliferation, survival and differentiation [2][3]. The attachment of cells to the extracellular matrix is mediated by dynamic sites along the plasma membrane, such as focal adhesions, at which receptors of the integrin family anchor the actin cytoskeleton to components of the extracellular matrix via a large number of different proteins [6]. Focal adhesions can contain over 100 different proteins, including integrins, adapter proteins, and intracellular signaling proteins [5]. Due to the large number of components and diversity of cell–matrix adhesion sites, a fundamental question is how these sites are assembled and function.

In systems biology graphical models and networks have been widely applied as a useful tool to model complex biochemical systems. In this work we propose a nonparametric mixture of dynamic Bayesian networks [1] to study dynamic dependencies among interacting proteins in the presence of heterogeneity among focal adhesions. Nonparametric mixture modelling of dynamic Bayesian networks is developed by a combination of dynamic Bayesian networks and of nonparametric Bayesian networks [4]. This approach provides further grouping of focal adhesions according to their protein–network structures. We apply and illustrate our approach using multicolor live cell imaging datasets, in which the levels of four different proteins are monitored in individual focal adhesions.

References:

1. Ghahramani, Z. (1997) Learning Dynamic Bayesian Networks. Lecture Notes In Computer Science 1387, 168–197.

2. Gumbiner,B. M. (1996) Cell adhesion: the molecular basis of tissue architecture and morphogenesis, Cell 84(3), 345–357.

3. Hynes, R. O. and Lander, A. D. (1992) Contact and adhesive specificities in the associations, migrations, and targeting of cells and axons, Cell 68(2), 303–322.

4. Ickstadt, K., Bornkamp, B., Grzegorczyk, M., Wieczorek, J., Sheriff, M.R., Grecco, H.E. and Zamir, E. (2011) Nonparametric Bayesian Networks. In: Bernardo, Bayarri, Berger, Dawid, Heckerman, Smith and West (eds.): Bayesian Statistics 9, Oxford University Press, 283–316.

5. Zaidel-Bar, R., Itzkovitz, S., Ma'ayan, A., Iyengar, R. and Geiger, B. (2007) Functional atlas of the integrin adhesome, Nat Cell Biol 9(8), 858–67.

6. Zamir, E. and Geiger, B. (2001) Molecular complexity and dynamics of cell-matrix adhesions, Journal of Cell Science 114 (20), 3583–3590.

Heterogeneity in multi-regional clinical trials and subgroup analyses

Framke, Theodor Koch, Armin

Medizinische Hochschule Hannover, Deutschland
framke.theodor@mh-hannover.de koch.armin@mh-hannover.de

Heterogeneity is not only an important aspect of every meta-analysis; it is also of great relevance in multi-regional clinical trials and subgroup analyses. Consistency of regions or subgroups is a desired aspect of a trial, however, in many studies it is dealt with this issue only at the analysis stage. Problems arise in the justification and interpretation of clinical trials with inconsistent results. It remains a challenge to find well-founded strategies to handle these issues. Furthermore, heterogeneity may lead to an elevated number of false-positive decisions and thus jeopardizes sound conclusions.

This presentation highlights crucial situations in which conclusions on efficacy are not reliable. Based on several actual trial results, the impact of heterogeneity on the empirical type I error will be presented in a simulation study. We investigated how much the nominal significance level has to be lowered in order to keep the type I error probability. Both fixed and random effects models are considered.

References:

[1] Koch A, Framke T (2014): Reliably basing conclusions on subgroups of randomized clinical trials. J Biopharm Stat 24(1), 42–57.

[2] Pocock S et al. (2013): International differences in treatment effect: do they really exist and why? European Heart Journal 34(24), 1846–1852.

Estimation of pregnancy outcome probabilities in the presence of heavy left-truncation

Friedrich, Sarah [1] Beyersmann, Jan [1] Winterfeld, Ursula [2]
Allignol, Arthur [1]

[1] Institute of Statistics, Universität Ulm, Deutschland
sarah.friedrich@uni-ulm.de, jan.beyersmann@uni-ulm.de, arthur.allignol@uni-ulm.de
[2] STIS and Division of Clinical Pharmacology and Toxicology, University Hospital, Lausanne, Switzerland
ursula.winterfeld@chuv.ch

Estimation of pregnancy outcome probabilities in a competing risks setting - a pregnancy may end in a spontaneous or induced abortion or a life birth - is complicated by the fact that the data are left-truncated, i.e. women enter observational studies at different time points several weeks after conception. This leads to small risk sets, especially at the beginning of a study, causing unreliable estimates. In our pregnancy data analysis, standard estimation methods showed a protective effect of statin use during pregnancy on the risk of induced abortion, which is medically implausible.

Since events like spontaneous or induced abortion often happen early during pregnancy, these problems are a major concern in pregnancy data applications.

We present a modification of the well-known Nelson-Aalen and Aalen-Johansen estimator, which eliminates too small risk sets, generalizing concepts used by Lai and Ying (1) for the standard survival case to the competing risks setting.

Making use of martingale arguments, uniform strong consistency and weak convergence against a Gaussian martingale are obtained for the modified estimator and a new variance estimator is derived. The modified estimator turns out to eliminate small risk sets without altering the asymptotics of the usual estimators.

Extensive simulation studies and a real data analysis illustrate the merits of the new approach. Reanalyzing the pregnancy data with the modified estimator leads to plausible results concerning the effect of statin use on the risk of abortion.

References:

1 T. L. Lai, Z. Ying: Estimating a distribution function with truncated and censored data, The Annals of Statistics 1991, 19(1): 417-442.

2 U. Winterfeld, A. Allignol, A. Panchaud, L. Rothuizen, P. Merlob, B. Cuppers-Maarschalkerweerd, T. Vial, S. Stephens, M. Clementi, M. Santis, et al. Pregnancy outcome following maternal exposure to statins: a multicentre prospective study. BJOG: An International Journal of Obstetrics & Gynaecology 2012, doi: 10.1111/1471-0528.12066.

Multiblock redundancy analysis: Recommendations for the number of dimensions in veterinary epidemiological observational studies with small sample sizes

Frömke, Cornelia [1] Kreienbrock, Lothar [1] Campe, Amely [1]
Bougeard, Stéphanie [4]

[1] Stiftung Tierärztliche Hochschule Hannover, Deutschland
cornelia.froemke@tiho-hannover.de lothar.kreienbrock@tiho-hannover.de
amely.campe@tiho-hannover.de
[2] French Agency for Food, Environmental and Occupational Health Safety, France
stephanie.bougeard@anses.fr

In veterinary epidemiology it is common to investigate associations between factors on multiple outcomes under real farm conditions using observational studies. Recently, the number of investigated factors and outcomes has grown, which increased the complexity of associations. However, standard procedures such as generalized linear models do not take adequate account of such complex structures. An appropriate analysis for multiple explanatory variables and outcomes is the multiblock redundancy analysis (mbRA) [1].

In the mbRA, the explanatory variables are allocated in thematic blocks and all outcomes in one outcome block. The idea is to create a latent variable for each block out of a linear combination of variables in the specific block, such that a criterion reflecting the association of each latent variable from the explanatory blocks to the latent variable from the outcome block is maximized. The mbRA enables to perform linear regression explaining each outcome with all explanatory variables. It combines tools from factor analysis and tools from regression analysis to e.g. visualization tools helping to detect relationships. Further, it offers epidemiological results, such as the importance of each explanatory variable on the individual outcome and the importance of each explanatory block on the outcome block.

The latent variables are computed with eigenanalysis. The number of dimensions chosen in the eigenanalysis impacts the type I error and the power. So far, no recommendations exist on the choice of the number of dimensions for the mbRA. Simulation results for an appropriate number of dimensions will be presented for data with small sample sizes, varying numbers of explanatory and outcome variables and correlation structures among the variables. Further, results of power simulations will be presented comparing mbRA, mbPLS and linear regression.

References:

[1] Bougeard, S., Qannari, E.M., Rose N. (2011): Multiblock Redundancy Analysis: interpretation tools and application in epidemiology. J.Chemometrics; 25: 467–475.

A Contribution to Variance Estimation of Resampling Procedures in Classification and Regression

Fuchs, Mathias Krautenbacher, Norbert

[1] Ludwig-Maximilians-Universität München, Deutschland
fuchs@ibe.med.uni-muenchen.de krautenbacher@helmholtz-muenchen.de

We are motivated by the problem of estimating the variance of K-fold cross-validation. This is important because a good variance estimator would allow to set up, at least approximately, a confidence interval for the true error of the classification or regression procedure performed in each fold. Thus, the problem is of far-reaching practical importance.

A famous paper by Bengio and Grandvalet [1] states that there is no unbiased estimator for K-fold cross-validation. Here, we investigate the problem in general and decompose the variance of any resampling procedure into a short linear combination of regular parameters in the sense of Hoeffding, by expanding the variance into covariances and counting the number of occurrences of each possible value.

We take the simple toy example of a linear regression where only the intercept is estimated, derive a formula for the variance of cross-validation, and show that there is an unbiased estimator of that variance, in contrast to [1].

We show in the toy example as well as in a numerical experiment that balanced incomplete block designs can have smaller variance than cross-validation, and show how to derive an asymptotic confidence interval using a Central Limit Theorem.

References:

[1] Yoshua Bengio and Yves Grandvalet. No unbiased estimator of the variance of K-fold cross-validation. Mach. Learn. Res., 5:1089–1105, 2003/04. ISSN 1532-4435.

Quantitative Methoden zur Risiko-Nutzenbewertung

Gebel, Martin

Bayer Healthcare, Deutschland
martin.gebel@bayer.com

Die Evaluierung der Balance zwischen Nutzen und Risiken eines Medikaments ist fundamental für ein Medikament. Pharmafirmen müssen entscheiden, ob sie ein Produkt (weiter-)entwickeln und zur Zulassung beantragen, Regulatoren, ob es zuzulassen ist und schließlich Ärzte und Patienten, ob es in dem individuellen Fall anzuwenden ist.

Die Risiko-Nutzenbewertung verlief in der Vergangenheit häufig primär qualitativ. Allerdings hat in den letzten Jahren bei den Zulassungsbehörden ein Umdenken stattgefunden und quantitative Methoden werden vermehrt gefordert. Die neuen regulatorischen Vorgaben werden daher hier kurz vorgestellt.

Anschließend wird die relativ einfache, grundlegende Methodik für quantitative Methoden für die Risiko-Nutzenbewertung basierend auf Inzidenzraten oder Schätzern aus der Ereigniszeitanalyse vorgestellt.

Zur Veranschaulichung gibt es Anwendungsbeispiele von aktuellen Zulassungen neuer Medikamente mit Anwendung im kardiovaskulären Bereich. Diese basieren in erster Linie auf Material von FDA Advisory Committees der letzten Jahre.

Incorporating geostrophic wind information for improved space-time short-term wind speed forecasting and power system dispatch

Genton, Marc G.

King Abdullah University of Science and Technology, Saudi Arabia
marc.genton@kaust.edu.sa

Accurate short-term wind speed forecasting is needed for the rapid development and efficient operation of wind energy resources. This is, however, a very challenging problem. We propose to incorporate the geostrophic wind as a new predictor in an advanced space-time statistical forecasting model, the trigonometric direction diurnal (TDD) model. The geostrophic wind captures the physical relationship between wind and pressure through the observed approximate balance between the pressure gradient force and the Coriolis acceleration due to the Earth's rotation. Based on our numerical experiments with data from West Texas, our new method produces more accurate forecasts than does the TDD model using air pressure and temperature for 1- to 6-hour-ahead forecasts based on three different evaluation criteria. For example, our new method obtains between 13.9% and 22.4% overall mean absolute error improvement relative to persistence in 2-hour-ahead forecasts, and between 5.3% and 8.2% improvement relative to the best previous space-time methods in this setting. By reducing uncertainties in near-term wind power forecasts, the overall cost benefits on system dispatch can be quantified. The talk is based on joint work with Kenneth Bowman and Xinxin Zhu.

Risks of predictive modelling in survival analysis

Gerds, Thomas

University of Copenhagen, Denmark
tag@biostat.ku.dk

In health and clinical research statistical models are used to identify individuals who have a high risk of developing an adverse outcome over a specific time period. A risk prediction model thus provides an estimate of the risks for a single subject based on predictor variables. The statistical challenges lie in the choice of of modelling strategy which can be used to train a risk prediction model based on a data set, and in the estimation of prediction prediction accuracy parameters. In this talk I will survey the applied and theoretical side of the estimation problem in situations with right censored competing risk outcome and then discuss several pitfalls in the attempt to find an optimal model.

Using Prediction Performance as a Measure for optimal Mapping of Methylation and RNA-Seq Data

Gerhold-Ay, Aslihan Mazur, Johanna Binder, Harald

University Medical Center Mainz, Deutschland
gerholda@uni-mainz.de mazur@uni-mainz.de binderh@uni-mainz.de

High Dimensional data like RNA-Seq and methylation data are becoming more and more important for medical research. They enable us to develop gene signatures for prediction of clinical endpoints like death, via the integration of the information presented in RNA-Seq data on gene expression and methylation data on CpG sites. However, the challenge, which CpG sites should be considered as being related to one specific gene, has not been solved yet.

Our aim is to tackle the question how prediction performance can be used as a measure to find the optimal mapping of CpG sites to their related genes.

For finding this optimal mapping of the gene information of RNA-Seq data and the CpG sites of methylation data, we define a length of nucleotides around all genes, a so-called window. In a two-step approach, we first use a likelihood-based componentwise boosting approach with 50 resampling steps to estimate a gene signature based on the RNA-Seq data only. Genes with an inclusion frequency (IF) of > 0.1 are used for the further analysis. In the following step, we take the methylation data of the CpG sites that are falling in this window to estimate a new signature. For varying window sizes, we decide to be the optimal mapping the one with the best prediction performance for the latter signature

To show the effect of the window sizes on the prediciton performance of the clinical endpoint, we use data of kidney tumor patients.

Methods for the combination of RNA-Seq and methylation data can be powerful tools for the classification of cancer patients. To underpin these tools, we propose the prediction performance measure as a criterion to find the optimal mapping window for RNA-Seq and methylation data and show its usefulness.

Ein universelles Bayes-Design für einarmige Phase II-Studien mit binärem zeitlich erfasstem Endpunkt

Gerß, Joachim

Westfälische Wilhelms-Universität Münster, Deutschland
joachim.gerss@ukmuenster.de

In der Phase II der klinischen Forschung werden häufig einarmige klinische Studien mit binärem Endpunkt und einer Zwischenauswertung durchgeführt. In der Zwischenauswertung wird die Studie abgebrochen, falls sich die Studientherapie als aussichtslos erweist. Das klassische Studiendesign geht auf Simon (1989) zurück. Dabei wird vorausgesetzt, dass der Endpunkt unmittelbar nach der Behandlung eines Patienten erfasst wird. Dies ist nicht immer der Fall. In einer aktuellen klinischen Studie bei Patienten mit einem Retinoblastom ist der binäre Endpunkt die Erhaltung des erkrankten Auges in einem Zeitraum von 6 Monaten nach dem Beginn der Therapie. In der Zwischenauswertung wird es Patienten geben, bei denen innerhalb von 6 Monaten ein Verlust des Auges eingetreten ist, Patienten, bei denen das Auge 6 Monate lang erhalten werden konnte, und zensierte Fälle, d.h. Patienten mit erhaltenem Auge, die allerdings den Nachbeobachtungszeitraum noch nicht abgeschlossen haben und ein Verlust des Auges noch eintreten kann. In der Literatur existieren einige Vorschläge geeigneter Studiendesigns für derartige zeitlich erfasste binäre Endpunkte oder Endpunkte in Form von Ereigniszeiten. Ein Nachteil dieser Designs ist, dass sie gewisse Annahmen oder Vorkenntnisse erfordern, die nicht immer vorliegen. Dazu zählt die Nutzung historischer Kontrollen, aus denen die Überlebensfunktion geschätzt wird, sowie die Annahme einer bestimmten Verteilung der Überlebenszeiten, z.B. einer Exponential- oder Weibull-Verteilung. Im Vortrag wird ein neu entwickeltes Bayes-Verfahren vorgestellt, das in dieser Hinsicht vorteilhaft ist. Es erfordert nur ein Minimum an Annahmen und Vorkenntnissen. Dies gilt auch in Hinblick auf die A-priori-Verteilung, die nicht-informativ ist. Das Verfahren beruht auf einem Beta-Binomial-Modell. Eine wesentliche Erweiterung besteht in dem Umgang mit zensierten Fällen. Die entsprechende Wahrscheinlichkeit für einen zukünftigen Augenverlust vor Abschluss des 6-monatigen Beobachtungszeitraums wird so gut wie möglich anhand sämtlicher bisher vorliegender Daten geschätzt. Das entwickelte Verfahren stellt eine natürliche Erweiterung des klassischen Ansatzes nach Simon für zeitlich erfasste Endpunkte dar. D.h. falls keine zensierten Fälle vorliegen, so ist es identisch zu einem Simon-Design. Nach der theoretischen Vorstellung des Verfahrens wird dessen praktische Umsetzung in der oben erwähnten Studie bei Patienten mit einem Retinoblastom gezeigt. Mit Hilfe simulierter Daten werden die Gütekriterien des Verfahrens untersucht und es wird gezeigt, welche Unterschiede zu bestehenden Verfahren bestehen.

Causal inference in practice: Methods for the construction and comparison of standardized risks to measure quality of care across centers

Goetghebeur, Els

Ghent University, Belgien
Els.Goetghebeur@UGent.be

In two consecutive 80 minute presentations, we will discuss estimation of the causal effect of discrete (treatment) choice on a binary or failure time outcome from observational data. In the absence of unmeasured confounders, fitting traditional as well as more novel causal effect models has become almost straightforward. It is often less apparent how the distinct approaches target different effects for varying types of exposure in specific (sub) populations. Each method has in addition technical (dis)advantages in terms of its reliance on (untestable) assumptions, the degree of data extrapolation involved and the amount of information drawn from a given data set. Motivated by our work on the evaluation of quality of care over treatment centers we discuss these topics in turn [7, 8], clarify how they help guide the method choice in practice and show results for a case study on acute stroke care in Sweden [1]. We will present a new R-package (available soon) that implements our method of choice and returns output with supporting evidence for specific treatment choices. 1. No unmeasured confounding and effect estimators building on outcome regression We consider studies that register key confounders of the association between observed exposure and outcome. When estimating the effect of treatment center, this would typically include the patient's age and disease stage at entry. A patient-specific adjusted risk or log odds ratio under one treatment choice versus another can then be derived from outcome regression. To estimate the total effect of treatment center choice one should not adjust for hospital characteristics such as its size or the presence of specific equipment. Through direct averaging, inverse weighting or double robust methodology, exposure-specific conditional risks can be marginalized over default (sub)populations [3,4]. Traditionally, indirect estimation contrasts the outcome in a given center with what may be expected were its patients to chose their treatment center at random. On the other hand, direct standardization estimates how the entire study population would fare under the potential care of each individual center. With either approach an auxiliary propensity score model, regressing exposure (treatment choice) on baseline confounders, may enter the estimating equations. This has varying consequences for the bias-precision tradeoff leading to a natural choice of method in many practical situations. We will show how depending on the analysis goal, one or the other method appears more appropriate. The choice of method will matter more when center by baseline covariate interactions exist. With limited information per center, fitting a fixed effect for each of many centers can become a challenge. In such cases, normal random effects models have been adapted as a solution for convergence problems at a cost of power to diagnose deviant behavior in especially the small (low information) centers. We propose the Firth correction [2] with its corresponding random effects interpretation as an alternative that combines several advantages. The convergence problems increase dramatically when interactions between centers and covariates enter the model. Fortunately, ignoring interactions has little impact when patient mixes are similar across centers and treatment effects are summarized through di-

rect or indirect standardization. We explain why for (in)directly standardized risks the (largest) smallest centers are most affected by bias. Quality evaluation focuses on diagnosing 'possible problems' or 'room for improvement' to help improve care. Bench marking standardized risks is however non trivial [5]. Equivalence regions and confidence levels first require consensus. Binary decisions can then build on this with corresponding sensitivity and specificity for the diagnosis of clinically relevant problems. As an alternative, we display center-specific results in funnel plots [6] which draw the line for flagging deviant results in a traditional way that protects the null, or in a way that balances type I and type II errors. We will demonstrate how much of this can be achieved with the help of a new dedicated R-package. 2. Further practical questions and challenges Even with the great registers available in Sweden, covariate values are sometimes missing. We will show how multiple imputation (MICE) can recover consistent estimators when missingness occurs at random. We will also explain the likely bias of the complete case analysis in our set-up. Having established the causal effect of center on a binary outcome based on cross-sectional data, we will explore trends over time. We first adapt methods to cover time to (cause-specific) event outcome, such as time to death or dependent living, rather than a simple binary outcome. We will then discuss how the descriptive funnel plot and original analysis on binary outcomes can be used to track the center's standardized risk evolution over calendar time. We compare results between two distinct time points and extend methods for more general causal effect evaluation over time. Having focused on the total effect of center on outcome, the next question concerns a decomposition of the total effect into controlled direct and indirect effects. Through which path is a deviating risk being reached and how are the controlled effects meaningfully estimated? In summary, we aim to equip researchers with an understanding of the theoretical and practical properties of different methods for causal effects estimation on risk. He or she will be able to make a motivated choice among these methods and know where to find the software to implement this. This will be illustrated on a case study, evaluating the quality of (acute stroke) care over hospitals in Sweden.

References:

[1] Asplund, K., Hulter Asberg, K., Appelros, P., Bjarne, D., Eriksson, M. et al. (2011). The Riks-Stroke story: building a sustainable national register for quality assessment of stroke care. International Journal of Stroke 6, 6–99.

[2] Firth D. (1993). Bias reduction of maximum likelihood estimates. Biometrika, 80, 27–38.

[3] Goetghebeur E., Van Rossem R. et al. (2011). Quality insurance of rectal cancer–phase 3: statistical methods to benchmark centers on a set of quality indicators, Good Clinical Practice (GCP). Belgian Health Care Knowledge Centre. KCE Report 161C, D/2011/10.273/40, 1–142.

[4] Hernan, M. and Robins, J. (2006). Estimating causal effects from observational data. Journal of Epidemiology and Community Health, 60, 578–586.

[5] Normand, S-L. and Shahian, D. (2007). Statistical and Clinical Aspects of Hospital Outcomes Profiling. Statistical Science, 22, 206–226.

[6] Spiegelhalter, D. (2005). Funnel plots for comparing institutional performance. Statistics in medicine 24, 1185–1202.

[7] Van Rompaye, B., Eriksson, M. and Goetghebeur E. (2014). Evaluating hospital performance based on excess cause-specific incidence. Statistics in Medicine, In press.

[8] Varewyck, M., Goetghebeur, E., Eriksson M. and Vansteelandt S. (2014). On shrinkage and model extrapolation in the evaluation of clinical center performance. Biostatistics,15: 651–64.

Detecting interaction effects in imputed datasets

Gola, Damian König, Inke R.

Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Lübeck, Deutschland
gola@imbs.uni-luebeck.de inke.koenig@imbs.uni-luebeck.de

Common complex traits may be determined by multiple genetic and environmental factors and their interactions. Many methods have been proposed to identify these interaction effects, among them several machine learning and data mining methods. These are attractive for identifying interactions since they do not rely on specific (genetic) model assumptions. One of these methods is Model-Based Multifactor Dimensionality Reduction (MB-MDR) [1], a semi-parametric machine learning method that allows to adjust for confounding variables and lower level effects. Current implementations enable MB-MDR to analyze data from genome-wide genotyping arrays [2]. In these days, it has become common practice to combine the genotype information with further databases to also impute genotypes at loci that are not on the array. The result of this imputation is a probability for each of the possible genotypes.

However, MB-MDR requires hard genotypes, removing any uncertainty in imputed genotypes, e.g. by using “best-guess” imputed genotypes.

We propose an extension of MB-MDR to analyze imputed data directly using weighted statistics (iMB-MDR).

Using simulations, we consider a range of sample sizes, minor allele frequencies, and imputation accuracies to compare the performance of iMB-MDR with MB-MDR using “best-guess" imputed genotypes.

References:

[1] Malu Luz Calle et al. MB-MDR: Model-Based Multifactor Dimensionality Reduction for detecting interactions in high-dimensional genomic data. Tech. rep. Department of Systems Biology, Universitat de Vic, 2007, pp. 1–14. URL: http://www.recercat.net/handle/ 2072/5001.

[2] Franşois Van Lishout et al. “An efficient algorithm to perform multiple testing in epistasis screening.” In: BMC Bioinformatics 14.1 (Jan. 2013), p. 138. ISSN: 1471–2105. DOI: 10.1186/1471-2105-14-138.

Adaptive power priors for using co-data in clinical trials

Gravestock, Isaac Roos, Małgorzata Held, Leonhard

Isaac Gravestock, Małgorzata Roos and Leonhard Held on behalf of COMBACTE consortium
Epidemiology, Biostatistics and Prevention Insitute, University of Zurich, Switzerland
isaac.gravestock@uzh.ch mroos@ifspm.uzh.ch leonhard.held@ifspm.uzh.ch

Clinical trials are traditionally analysed on the basis of no prior information. However, drug-resistant infections pose problems that make traditional clinical trials difficult to perform. The use of additional information is an important topic in regulatory discussions [2] about the development of new medications for drug-resistant bacterial infections. Making use of previous trial data can help overcome some barriers of antibiotics trials by increasing the precision of estimates with fewer patients needed.

The power prior is an elegant technique for the construction of a prior based on the likelihood of the historical data, which is raised to a power between 0 and 1. However, the original publications do not correctly normalise the prior which leads to misleading results [3] and unfortunately these methods are still being used. The correct methods[1] have not been thoroughly examined in the literature.

We present some new results for this under-appreciated technique. We look at the construction of priors for survival outcomes in clinical trials of the same agent in different populations. We show fully Bayesian and Empirical Bayes estimation of the power parameter and how these estimates behave under prior-data conflict. We re-examine the connection to hierarchical models and the prior assumptions this implies, as well as extensions to multiple historical datasets.

COMBACTE is supported by IMI/EU and EFPIA.

References:

[1] Yuyan Duan, Keying Ye, and Eric P Smith. Evaluating water quality using power priors to incorporate historical information. Environmetrics, 17(1):95–106, 2006.

[2] Center for Drug Evaluation and Research (FDA). Guidance for Industry: Antibacterial Therapies for Patients With Unmet Medical Need for the Treatment of Serious Bacterial Diseases. 2013.

[3] Beat Neuenschwander, Michael Branson, and David J Spiegelhalter. A note on the power prior. Statistics in medicine, 28(28):3562–3566, 2009.

An investigation of the type I error rate when testing for subgroup differences in the context of random-effects meta-analyses

Guddat, Charlotte

IQWiG, Deutschland
charlotte.guddat@iqwig.de

There are different approaches to test for differences between two or more subsets of studies in the context of a meta-analysis. The Cochrane Handbook refers to two methods. One is a standard test for heterogeneity across subgroup results rather than across individual results. The second is to use meta-regression analyses. Assuming the random-effects model, we have conducted a simulation study to compare these two approaches with respect to their type I error rates when only 10 or less studies are in the pool. Besides the number of studies and amongst further parameters, we have varied the extent of heterogeneity between the studies, the number of subgroups and the distribution of the studies to the subgroups.

The heterogeneity test gives extremely high error rates, when the heterogeneity between the studies is large and the distribution of the studies to the subgroups is unequal. In contrast, the error rates of the F-test in the context of a meta-regression are acceptable irrespective of the chosen parameters.

A Dunnett-Type Test for Response-Adaptive Multi-Armed Two-Stage Designs

Gutjahr, Georg

Uni Bremen, Deutschland
georg.gutjahr@math.uni-bremen.de

Consider the following two-stage design for comparing multiple treatments against a single control: initially, control and treatments are allocated by response-adaptive randomization during the first stage; after completion of the first stage, some treatments are selected to proceed to the second stage; finally, control and selected treatments are allocated by block randomization during the second stage.

To protect the familywise error rate in such designs, one possible approach is to view the trial as a data-dependent modification of a simpler design, for which we know the distributions of its test statistics and to account for the data-dependent modification, by the conditional invariance principle. Gutjahr et al. (2011) used this approach with a preplanned pooled test statistic.

In this talk, we show that it is also possible to work with a preplanned maximum (Dunnett-type) test statistic. We examine the operating characteristics of the resulting multiple test procedure and compare it with the pooled test.

Welchen Einfluss hat die korrekter Modellierung des Versuchsdesigns bei Versuchsserien/Projektkooperationen?

Hartung, Karin

Universität Hohenheim, Deutschland
karin.hartung@uni-hohenheim.de

Immer wieder werden Versuchsserien und Projektkooperationen durchgeführt, bei denen ein bestimmtes Versuchsdesign vom Projekt vorgegeben ist. Die Daten werden von den Projektpartnern erhoben und übermittelt. Entsprechend des vorgegebenen Designs werden dann die Daten ausgewertet. Bei genauer Recherche zeigt sich dann, dass die Exaktversuche jedoch mit völlig anderen Versuchsdesigns durchgeführt wurden. Anhand eines Projektdatensatzes, der laut Projektvorgabe auf randomisierten vollständigen Blockanlagen mit 3 bis 4 Wiederholungen je Jahr beruhen sollte und drei Jahre, ca. 40 Standorten und ca. 70 Sorten sowie verschiedene Umweltdaten enthält, sollen die sich ergebenden Probleme betrachtet werden:

Von wie vielen Orten sind die Versuchsdesigns sicher rekonstruierbar?

Welche Versuchsdesigns finden sich tatsächlich?

Lassen sich diese in ein statistisch korrektes Modell für eine Varianzanalyse überführen?

Welche Unterschiede ergeben sich, wenn die Daten, statt je Standort statistisch korrekt, generell mit der Projektvorgabe “Blockanlageäusgewertet werden?

Was lässt sich daraus für Versuchsserien/Projektkooperationen ableiten?

Confounder-selection strategies in (environmental) epidemiology: Classical variable selection approaches vs. a DAG-approach

Hennig, Frauke [1,2] Ickstadt, Katja [2] Hoffmann, Barbara [3]

[1] IUF-Institut für umweltmedizinische Forschung, Deutschland
frauke.hennig@iuf-duesseldorf.de
[2] Faculty of Statistics, TU Dortmund University, Dortmund, Germany
ickstadt@statistik.uni-dortmund.de
[3] Heinrich Heine University of Düsseldorf, Medical Faculty, Düsseldorf, Germany
Barbara.Hoffmann@IUF-Duesseldorf.de

Background: Identification of confounding variables is essential for unbiased exposure effect estimation. Directed acyclic graphs (DAGs) were introduced as a non-parametric method for confounder identification based on expert knowledge1. Bayesian networks (BN) could provide an applicable link between expert knowledge and observed data. We aim to learn DAG-structures from data and perform a DAG-based confounder-selection in comparison to classical variable selection.

Methods: We simulated multivariate normally distributed data based on a given (expert) DAG (DAG-E), including 1 exposure (X), 1 outcome (Y), 1 Confounder (C1) and 4 other variables, including predictors of Y (V1, V2), predictors of X (V3), and other (V4). Beta coefficients were randomly chosen from a uniform distribution on [0, 0.5] with random sign. We learned the DAG-structure (DAG-L) using score-based and hybrid structure learning algorithms, implemented in the R-package bnlearn2 and a Bayesian Gaussian Equivalent-score-based greedy search algorithm with MCMC simulation. Performance was evaluated by sensitivity and specificity. We compared the confounder-selection proposed by Greenland1 to classical stepwise forward variable selection strategies, including change-in-point-estimate (CIPE5%, CIPE10%), significance testing (ST5%, ST15%), Bayesian and Akaike Information Criteria (BIC, AIC).

Results: First results show that learning algorithms perform very differently and rather poor. Best results were reached by the hybrid structure-learning algorithm and the BGe-MCMC simulation with a sensitivity of 77.8-100% and a specificity of 93.9-100%, yet the correct DAG-structure was only detected in <30% of simulation runs. As a result, the confounder C1 was uniquely identified in only 50%. The CIPE5% and CIPE10% detected C1 in 87.3% and 99.9%, while ST5%, ST15%, BIC and AIC detected C1 only in combination with predictors of Y (C1,V1,V2) (90.6%, 73.9%,73.2% and 99.6%).

Conclusion: Learning algorithms perform poorly leading to poor confounder identification. CIPE10% yielded the best performance in detecting the confounder, while ST, BIC and AIC approaches detected the confounder only in combination with predictors of Y.

References:

1 Greenland S, Pearl J, Robins JM (1999): Causal Diagrams in Epidemiologic Research, Epidemiology, Vol.10, pp. 37–48.

2 Scutari M (2010): Learning Bayesian Networks with the bnlearn R Package. Journal of Statistical Software, 35(3), 1–22.
http://www.jstatsoft.org/v35/i03/.

Detection and modeling of seasonal variation in blood pressure in patients with diabetes mellitus

Hermann, Julia M. [1] Rosenbauer, Joachim [2] Dost, Axel [3]
Steigleder-Schweiger, Claudia [4] Kiess, Wieland [5] Schöfl, Christof [6]
Schwandt, Anke [1] Holl, Reinhard W. [1]

[1] University of Ulm, Germany
julia.hermann@uni-ulm.de anke.schwandt@uni-ulm.de reinhard.holl@uni-ulm.de [2] German Diabetes Centre, Leibniz Centre at Heinrich-Heine University Düsseldorf, Germany
Joachim.Rosenbauer@ddz.uni-duesseldorf.de
[3] University Hospital Jena, Germany
Axel.Dost@med.uni-jena.de
[4] University Hospital of Salzburg, Paracelsus Medical University, Austria
c.steigleder-schweiger@salk.at
[5] University of Leipzig, Germany
wieland.kiess@medizin.uni-leipzig.de
[6] Friedrich-Alexander-University Erlangen-Nuremberg, Germany
christof.schoefl@uk-erlangen.de

Various epidemiologic studies observed seasonally varying blood pressure (BP) in different cohorts, most of them using descriptive methods or simple winter-summer comparisons. This study aimed to use time series methods and regression models in order to examine seasonal patterns in BP in patients with diabetes.

A total of 732,179 BP values from 62,589 patients with type 1 diabetes (T1D) and 254,639 BP values from 99,546 patients with type 2 diabetes (T2D) were extracted from the German/Austrian observational multicenter DPV database. Fourier periodograms were used to identify dominant cyclic behavior in monthly mean BP from 2003 to 2012. Autocorrelation of the BP time series was examined by means of seasonal autoregressive integrated moving average models (SARIMA). In addition, cross-correlation functions were applied to assess the association between BP and potentially related factors such as outdoor temperature. Harmonic regression models comprising sine and cosine terms to account for periodicity were used to estimate amplitude and phase shift of BP variation. First-order interaction terms between the trigonometric functions and age, gender or duration of diabetes were added to perform subgroup comparisons.

BP was significantly higher in winter than in summer in both types of diabetes, inversely related to outdoor temperature. Fourier and SARIMA analysis revealed seasonal cycles of twelve months. According to harmonic regression models, the estimated systolic BP difference throughout the year was 2.28/2.48 mmHg in T1D/T2D patients (both $p<0.001$). A gender difference was observed in T1D patients only, while age differences occurred in both types of diabetes. Phase shift differed between T1D and T2D ($p<0.001$).

All methods used seemed suitable to gain new insights into BP seasonality. Subgroup comparisons indicated that reasons underlying this phenomenon are likely to be complex and vary by type of diabetes, age and gender. Particular attention must be paid to time series data requirements (e.g. stationarity).

A comparison of clustering approaches with application to dual colour data

Herrmann, Sabrina [1] Ickstadt, Katja [1] Verveer, Peter [2]

[1] TU Dortmund, Deutschland
herrmann@statistik.tu-dortmund.de ickstadt@statistik.tu-dortmund.de
[2] Max-Planck Institut für molekulare Physiologie, Dortmund
peter.verveer@mpi-dortmund.mpg.de

The cell communicates with its environment via proteins, which are located at the plasma membrane that separates the interior of a cell from its surroundings. The spatial distribution of these proteins in the plasma membrane under different physiological conditions is of interest, since this may influence their signal transmission properties. In our work, we compare different methods such as hierarchical clustering, extensible Markov models and the Gammics method for analysing such a spatial distribution.

We apply these methods to a single colour simulation as well as real data and also give an insight whether these methods work on dual colour data.

Analyse und Synthese von sensorgesteuerten Herzschrittmacheralgorithmen

Hexamer, Martin

Ruhr-Universität Bochum, Deutschland
martin.hexamer@rub.de

Bestimmte Formen einer beeinträchtigten Herz-Kreislaufregulation haben ihre Ursachen in Defekten des Reizbildungs-/Reizleitungssystem des Herzens. Die Folgen sind Herzrhythmusstörungen, die in der Regel mit implantierbaren Herzschrittmachern therapiert werden. Dieser Beitrag erläutert einige grundlegende physiologische Zusammenhänge der intakten intrinsischen Herz-Kreislaufregulation und beleuchtet die Folgen bestimmter pathophysiologischer Veränderungen, die letztlich die Unterbrechung eines internen Regelkreises darstellen. Das Krankheitsbild, das dieser Beitrag im weiteren vertieft, ist die chronotrope Inkompetenz, ein Zustand, in dem das Herz-Kreislauf-System nicht mehr in der Lage ist, die Herzschlagfrequenz an die Belastung anzupassen. Infolge sind die Patienten mehr oder weniger stark leistungslimitiert und in ihrer Lebensqualität beeinträchtigt. Im Beitrag wird eine Systematik vorgestellt mit der man die therapierelevanten Eigenschaften des Patienten messtechnisch erfassen und, basierend auf diesen Informationen, Herzschrittmachersysteme patientenindividuell parametrieren kann.

Conditional survival as framework to identify factors related to long-term survival

Hieke, Stefanie [1,2] Kleber, Martina [3,4] König, Christine [4]
Engelhardt, Monika [4] Schumacher, Martin [1]

[1] Institute of Medical Biometry and Statistics, Medical Center – University of Freiburg, Freiburg, Germany
hieke@imbi.uni-freiburg.de ms@imbi.uni-freiburg.de
[2] Freiburg Center for Data Analysis and Modeling, University Freiburg, Freiburg, Germany
[3] Clinic for Internal Medicine, University Hospital Basel, Basel, Switzerland
martina.kleber@usb.ch
[4] Department of Medicine I, Hematology, Oncology and Stem Cell Transplantation, Medical Center - University of Freiburg, Germany
christine.koenig@uniklinik-freiburg.de monika.engelhardt@uniklinik-freiburg.de

Disease monitoring based on genetics or other molecular markers obtained by noninvasive or minimally invasive methods will potentially allow the early detection of treatment response or disease progression in cancer patients. Investigations in order to identify prognostic factors, e.g. patient's baseline characteristics or molecular markers, contributing to long-term survival potentially provide important information for e.g. patients with multiple myeloma. In this disease substantial progress with significantly improved prognosis and long-term survival, even cure, has been obtained. Conditional survival offers an adequate framework for considering how prognosis changes over time since it constitutes the simplest form of a dynamic prediction and can therefore be applied as starting point for identifying factors related to long-term survival. Conditional survival probabilities can be calculated using conditional versions of Cox regression models by taking additional information on the patient's baseline characteristics and molecular markers into account. Predicted (conditional) survival probabilities of such prognostic models can be judged with regard to prediction performance. In an application to survival data from multiple myeloma, we estimate (conditional) prediction error curves to assess prediction accuracy of prognostic models in survival analysis using a (conditional) loss function approach (Schoop et al., 2008). Conditional survival allows us to identify factors related to long-term survival of cancer patients since it gives a first insight into how prognosis develops over time.

References:

1. Rotraut Schoop, Erika Graf, and Martin Schumacher. Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. Biometrics, 2008; 64:603–610.

Translational Statistics

Hinde, John Newell, John

NUI Galway, Ireland
john.hinde@nuigalway.ie john.newell@nuigalway.ie

Translational medicine, often described as “bench to bedside", promotes the convergence of basic and clinical research disciplines. It aims to improve the flow from laboratory research through clinical testing and evaluation to standard therapeutic practice. This transfer of knowledge informs both clinicians and patients of the benefits and risks of therapies.

In an analogous fashion, we propose the concept of Translational Statistics to facilitate the integration of biostatistics within clinical research and enhance communication of research findings in an accurate and accessible manner to diverse audiences (e.g. policy makers, patients and the media). Much reporting of statistical analyses often focuses on methodological approaches for the scientific aspects of the studies; translational statistics aims to make the scientific results useful in practice.

In this talk we will consider some general principles for translational statistics that include reproducibility, relevance, and communication. We will also consider how modern web-based computing allows the simple development of interactive dynamic tools for communicating and exploring research findings. Various examples will be used to illustrate these ideas.

Improved cross-study prediction through batch effect adjustment

Hornung, Roman [1] Causeur, David [2] Boulesteix, Anne-Laure [1]

[1] Department of Medical Informatics, Biometry and Epidemiology, University of Munich, Marchioninistr. 15, 81377 Munich, Germany
hornung@ibe.med.uni-muenchen.deboulesteix@ibe.med.uni-muenchen.de
[2] Applied Mathematics Department, Agrocampus Ouest, 65 rue de St. Brieuc, 35042 Rennes, France
causeur@agrocampus-ouest.fr

Prediction of, say, disease status or response to treatment using high- dimensional molecular measurements is among the most prominent applications of modern high-throughput technologies. Many classification methods which can cope with large numbers of predictors have been suggested in a wide body of literature. However, the corresponding prediction rules are only infrequently in use in clinical practice. A major stumbling block hampering broader application is the lack of comparability between the data from patients whose outcomes we wish to predict - the "test data" - and the data used to derive the prediction rules - the "training data". The divergences, commonly known as batch effects, are of varying magnitude and are unrelated to the biological signal of interest. Several methods for batch effect removal have been suggested in the literature with the aim of eliminating these systematic differences.

A fact that is under acknowledged in the scientific literature, however, is that prediction performance can often be improved when a sufficient amount of test data is available all from the same source. The improvement is achieved by making the test data more similar to the training data by using slightly modified versions of existent batch effect removal methods. We recently developed a batch effect removal method which can be effectively used for this purpose. It combines location-and-scale batch effect adjustment with data cleaning by adjustment for latent factors.

In this talk we critically discuss this procedure and related batch effect removal methods and present an extensive real-data study in which we compare the procedures with respect to their performances in cross-study prediction validation (CSV). By CSV, we mean that prediction rules are derived using the data from a first study and applied to test data from another study, the most common practice when applying prediction rules.

Statistical methods for the meta-analysis of ROC curves – A new approach

Hoyer, Annika Kuß, Oliver

Deutsches Diabetes-Zentrum (DDZ), Leibniz-Zentrum für Diabetes-Forschung an der Heinrich-Heine-Universität Düsseldorf, Deutschland
annika.hoyer@ddz.uni-duesseldorf.de oliver.kuss@ddz.uni-duesseldorf.de

Meta-analyses and systematic reviews are the basic tools in evidence-based medicine and there is still an interest in developing new meta-analytic methods, especially in the case of diagnostic accuracy studies. Whereas statistical methods for meta-analysis of a diagnostic test reporting one pair of sensitivity and specificity are meanwhile commonly known and well-established, approaches producing meta-analytic summaries of complete ROC curves (that is use of information from several thresholds with probably differing values across studies) have only recently been proposed [1, 2]. Most often in these situations researchers pick one (in the majority of cases the optimal) pair of sensitivities and specificities for each study and use standard methods. This imposes untestable distributional assumptions, the danger of overoptimistic findings and a considerable loss of information. We propose an approach for the meta-analysis of complete ROC curves that uses the information for all thresholds by simply expanding the bivariate random effects model to a meta-regression model. We illustrate the model by an updated systematic review on the accuracy of HbA1c for population-based screening for type 2 diabetes. The updated data set includes 35 studies with 122 pairs of sensitivity and specificity from 26 different HbA1c thresholds and a standard analysis would discard more than 70% of the available observations.

References:

[1] Martinez-Camblor P. Fully non-parametric receiver operating characteristic curve estimation for random-effects meta-analysis. Stat Methods Med Res. 2014; Epub ahead of print.

[2] Riley RD, Takwoingi Y, Trikalinos T, Guha A, Biswas A, Ensor J, Morris RK, Deeks JJ. Meta-Analysis of Test Accuracy Studies with Multiple and Missing Thresholds: A Multivariate-Normal Model. J Biomet Biostat. 2014; 5:3

Relative improvement of lung function in elderly German women after reduction of air pollution: Results from the SALIA cohort study

Hüls, Anke [1] Krämer, Ursula [1] Vierkötter, Andrea [1]
Stolz, Sabine [1] Hennig, Frauke [1] Hoffmann, Barbara [1,2]
Ickstadt, Katja [3] Schikowski, Tamara [1,4]

[1] IUF - Leibniz-Institut für umweltmedizinische Forschung GmbH, Deutschland
Anke.Huels@IUF-Duesseldorf.de, Ursula.Kraemer@IUF-Duesseldorf.de,
Andrea.Vierkoetter@IUF-Duesseldorf.de, Sabine.Stolz@IUF-Duesseldorf.de,
Frauke.Hennig@IUF-Duesseldorf.de, Barbara.Hoffmann@IUF-Duesseldorf.de,
Tamara.Schikowski@IUF-Duesseldorf.de
[2] Medizinische Fakultät der Heinrich-Heine Universität Düsseldorf
Barbara.Hoffmann@IUF-Duesseldorf.de
[3] Fakultät Statistik der Technischen Universität Dortmund
ickstadt@statistik.uni-dortmund.de
[4] Swiss Tropical and Public Health Institute, Basel
Tamara.Schikowski@IUF-Duesseldorf.de

Background and aims: Air pollution is a major environmental risk factor influencing lung function decline. Until now, a common method for analysing this association is to determine the absolute decline in lung function. This approach however does not consider that the absolute decline in lung function accelerates in adults over age which makes it incomparable between different age groups. Using reference values allows to determine whether reduction in air pollution relatively improves lung function over time independent of age. The aim of this master thesis was to test the applicability of the cross-sectional based GLI (Global Lung Initiative) reference values for lung function [1] in a cohort study and, if applicable, apply them to an association analysis.

Methods: We used data from the SALIA cohort of elderly German women (baseline: 1985-1994 (age 55 years), first follow-up: 2008/2009, second follow-up: 2012/2013; n=319). The GLI reference values were based on generalized additive models for location, scale and shape (GAMLSS) which allow very general distribution families for the outcome [2]. Furthermore, mean, skewness and kurtosis can be modelled independently by different predictors including smoothing terms. Using the GLI reference values for the mean, skewness and kurtosis we transformed the lung function values (FEV1, FVC) to GLI z-scores which should be standard normally distributed for every time point of examination if the GLI reference values are valid for our population. After evaluation of the cross-sectional and longitudinal validity of the GLI reference values a linear mixed models with a random participant intercept was used to analyse the association between a reduction of air pollution and change in z-scores between baseline and the two follow-up examinations.

Results: The GLI reference values for FEV1 and FVC were valid cross-sectionally and longitudinally for our data. Between baseline and first follow-up examination air pollution decreased strongly (e.g. PM10 from 47.4 ± 7.9 $\mu g/m^3$ to 26.6 ± 2.0). This reduction of air pollution was associated with a relative improvement of lung function over age, considering an on-going increase in z-scores up to 9.6% [95% CI: 6.6; 12.5] between baseline and second follow-up per 10 $\mu g/m^3$ decrease in PM10.

Conclusions: GLI reference values are valid cross-sectionally and longitudinally for lung function values of elderly German women. Reduction of air pollution was associated with an on-going relative improvement of lung function in these women between 55 and 83 years of age.

References:

[1] Quanjer, Philip H et al. 2012. "Multi-Ethnic Reference Values for Spirometry for the 3-95-Yr Age Range: The Global Lung Function 2012 Equations." The European Respiratory Journal 40(6): 1324-43.

[2] Rigby, R. A., and D. M. Stasinopoulos. 2005. "Generalized Additive Models for Location, Scale and Shape (with Discussion)." Journal of the Royal Statistical Society: Series C (Applied Statistics) 54(3): 507-54.

Statistical Analysis of the In Vivo Pig-a Gene Mutation Assay

Igl, Bernd-Wolfgang [1] Sutter, Andreas [1] Dertinger, Stephen [2]
Vonk, Richardus [1]

[1] Bayer Pharma AG, Deutschland
bernd-wolfgang.igl@bayer.com andreas.sutter@bayer.com richardus.vonk@bayer.com
[2] Litron Laboratories, USA
sdertinger@litronlabs.com

The in vivo Pig-a gene mutation assay is a new method for evaluating the genotoxic potential of chemicals. In the rat-based assay, a lack of CD59 on the surface of circulating erythrocytes is used as a reporter of Pig-a mutation. The frequency of these mutant phenotype cells is measured via fluorescent anti-CD59 antibodies in conjunction with flow cytometric analysis. This assay is currently under validation; it has, nevertheless, already achieved regulatory relevance as it is suggested as an in vivo follow-up test for Ames mutagens in the recent ICH M7 step 4 document. However, there is still no recommended standard for the statistical analysis of this assay. On these grounds, we propose a statistical strategy involving a down-turn protected Poisson regression in order to determine a dose-dependency of rare mutant phenotype erythrocytes. We demonstrate suitability of this strategy using example data sets obtained with the current blood labeling and analysis protocol.

Medical Device Alarms – A Clinical Perspective

Imhoff, Michael [1] Fried, Roland [2]

[1] Ruhr-Universität Bochum, Deutschland; 2: TU Dortmund, Deutschland
mike@imhoff.de
[2] TU Dortmund, Deutschland
fried@statistik.tu-dortmund.de

The alarms of medical devices are a matter of concern in critical and perioperative care. The high rate of false alarms is not only a nuisance for patients and caregivers, but can also compromise patient safety and effectiveness of care. The development of alarm systems has lagged behind the technological advances of medical devices over the last 25 years. From a clinical perspective, major improvements of alarm algorithms are urgently needed. This requires not only methodological rigor and a good understanding of the clinical problems to be solved, but also adequate matching of statistical and computer science methods to clinical applications. It is important to see the broad picture of requirements for monitoring and therapy devices in general patient populations. This can only be achieved in close cooperation between clinicians, statisticians, and engineers.

While already "simple" approaches such as alarm delays and context sensitivity may result in tangible improvements, only few solutions have been implemented in commercially available devices. Reasons for this include regulatory barriers, concerns about potential litigation, and other development and marketing priorities.

Existing standards, norms and guidelines cannot solve the problems of medical device alarms. Therefore, even with the advances in statistical and computational research including robust univariate and multivariate algorithms, there is still a long way toward improved and intelligent alarms.

Extrapolation of internal pilot estimates for sample size re-assessment with recurrent event data in the presence of non-constant hazards

Ingel, Katharina [1] Schneider, Simon [2] Binder, Harald [1]
Jahn-Eimermacher, Antje [1]

[1] University Medical Center Mainz, Deutschland
ingel@uni-mainz.de binderh@uni-mainz.de jahna@uni-mainz.de
[2] University Medical Center Göttingen, Deutschland
schneider.med.goettingen@gmail.com

In clinical trials with recurrent event data, as for example relapses in multiple sclerosis or acute otitis media, the required sample size depends on the cumulative hazard at time T, with $[0, T]$ being the patients' follow-up period. As in the planning phase of a trial there might be uncertainty about the recurrent event process over $[0, T]$, internal pilot studies have been proposed to re-estimate the sample size accordingly (Schneider et. al., Statistics in Medicine 2013; Ingel and Jahn-Eimermacher, Biometrical Journal 2014). However, for trials with a long follow-up period relative to a short recruitment period, at the time of the interim analysis patient data might only be available over a shorter period $[0, t]$, questioning the use of these approaches.

For these situations, we will investigate an extrapolation of the interim estimate for the cumulative hazard to the full follow-up period. Based on simulated data, different parametric recurrent event models will be applied to blinded interim data and interim estimates will be extrapolated to the end of the planned follow-up T for the purpose of sample size re-estimation. The timing of the interim analysis will be evaluated as one determinant for the accuracy of that extrapolation. Results will be compared with respect to the re-estimated sample size and the corresponding power of the trial. For simulation, recurrent event data under various hazard functions defined on a calendar time scale are required and we propose a general simulation algorithm for this purpose.

Results are used to identify situations, where an extrapolation of interim results can be useful to improve the accuracy of sample size planning in contrast to situations with a potential risk for misleading conclusions. Clinical trial examples on episodes of otitis media and relapses in multiple sclerosis are used to validate these findings.

Models for network meta-analysis with random inconsistency effects

Jackson, Daniel

Cambridge Institute of Public Health, United Kingdom
daniel.jackson@mrc-bsu.cam.ac.uk

In network meta-analysis, multiple treatments are included in the analysis, using data from trials that compare at least two of these treatments. This enables us to compare all treatments of interest, including any that may not have been compared directly. Perhaps the greatest threat to the validity of a network meta-analysis is the possibility of what has been called "inconsistency" or "incoherence". In this talk I will explain why I prefer to use random inconsistency effects to model the extent of the inconsistency in the network. I will describe my new model for network meta-analysis and I will explain how inference can be performed using this model in both the classical and Bayesian frameworks. I will also describe methods that can be used to quantify the impact of any inconsistency and I will illustrate my methods using some challenging medical datasets.

Simulating recurrent event data in the presence of a competing terminal event with an application to cardiovascular disease

Jahn-Eimermacher, Antje Ingel, Katharina Preussler, Stella
Binder, Harald

University Medical Center Mainz, Deutschland
jahna@uni-mainz.de ingel@uni-mainz.de spreussl@students.uni-mainz.de
binderh@uni-mainz.de

Simulation techniques are an important tool for planning clinical trials with complex time-to-event structure. Motivated by clinical trials evaluating hospitalization rates in fatal diseases, we specifically consider simulation of a recurrent event process in the presence of a competing terminal event. In a gap time approach the hazard for experiencing events is defined on the time since last event, which simplifies the simulation. However, in many clinical settings the hazard depends on the time since study start, e.g. the time since beginning treatment of some progressive disease. This calls for a calendar time approach.

Accordingly, we propose a method for simulating recurrent and terminal event data that follow a time-inhomogeneous multistate Markov model with transition hazards defined on calendar time. We identify the distributions in the nested competing risks models conditional on the time of the previous event by deriving the conditional transition hazard functions. Using the conditional cumulative hazards we can recursively simulate nested competing risks data by inversion sampling for survival data and combine these data to a full dataset. Closed form solutions are provided for common total time distributions as Weibull or Gompertz. We extend our methods to incorporate fixed and random covariates into proportional transition hazards models.

Our methods are illustrated by simulating clinical trial data on heart failure disease subject to recurrent hospitalizations and cardiovascular death. Simulations are applied to compare a multistate likelihood-based analysis using robust standard errors with a joint frailty modeling approach. Results are used to recommend statistical analysis methods for clinical trials in heart failure disease that go beyond the usually applied time to first combined event analysis.

Functional analysis of high-content high-throughput imaging data

Jiang, Xiaoqi [1] Wink, Steven [2] Kopp-Schneider, Annette [1]

[1] German Cancer Research Center, Deutschland
xiaoqi.jiang@dkfz-heidelberg.de kopp@dkfz.de
[2] Faculty of Science, Leiden Academic Centre for Drug Research, Toxicology
s.wink.3@lacdr.leidenuniv.nl

High-content automated imaging platforms allow the multiplexing of several targets simultaneously to address various biological processes. Fluorescence-based fusion constructs permit to quantitatively analyze physiological processes and hence to generate multi-parametric single-cell data sets over extended periods of time. As a proof-of-principle study a target of Nrf2 (NFE2L2), Srxn1, was fused to fluorescent EGFP in HepG2 cells to be able to monitor the effect of drug-induced liver injury related compound treatment using high content live cell imaging. Images were acquired of single cells over time capturing the Nrf2-response to the compound set. Several response variables were recorded, e.g. cell speed, cytosol integrated intensity and nuclei Hoechst staining intensity. Typically, standard simple measures such as mean value of all observed cells at every observation time point were calculated to summarize the whole temporal process, however, resulting in loss of time dynamics of single cells. In addition, three independent experiments are performed but observation time points are not necessarily identical, leading to difficulties when integrating summary measures from different experiments. Functional data analysis (FDA) is a collection of statistical techniques specifically developed to analyze continuous curve data. In FDA, the temporal process of a response variable for each single cell can be described using a smooth curve, which forms the basis unit for subsequent analyses. This allows analyses to be performed on continuous functions, rather than on the original discrete data points. Functional regression models were applied to determine common temporal characteristics of a set of single cell curves as well as to evaluate temporal differences between treated and control samples. Functional random effects were employed in regression models to explain variation between experiments. Functional principal component analysis (FPCA) was used to study the temporal variation in fitted smooth curves for different compounds. Two FPCA scores explained more than 99% variance between different compounds, thus these scores could be treated as "data" to evaluate compounds in a multivariate way. In conclusion, the FDA approach was superior to traditional analyses of imaging data in terms of providing important temporal information.

Filling the gaps – back-calculation of lung cancer incidence

Jürgens, Verena [1,2,7] Ess, Silvia [3] Phuleria, Harish C. [1,2,8]
Früh, Martin [4] Schwenkglenks, Matthias [5] Frick, Harald [6]
Cerny, Thomas [4] Vounatsou, Penelope [1,2]

[1] Swiss Tropical and Public Health Institute, Basel, Switzerland
verena.juergens@uni-oldenburg.de phuleria@iitb.ac.in penelope.vounatsou@unibas.ch
[2] University of Basel, Switzerland E-Mail
[3] Cancer Registry of St. Gallen & Appenzell, St. Gallen, Switzerland silvia.ess@kssg.ch
[4] Kantonsspital St. Gallen, St. Gallen, Switzerland martin.frueh@kssg.ch
thomas.cerny@kssg.ch
[5] Institute of Pharmaceutical Medicine (ECPM), University of Basel, Switzerland
m.schwenkglenks@unibas.ch
[6] Cancer Registry Grisons & Glarus, Kantonsspital Graubünden, Chur, Switzerland
harald.frick@kssg.ch
[7] Current affiliation: Carl von Ossietzky Universität Oldenburg, Deutschland
[8] Current affiliation: Indian Institute of Technology, Mumbai, India

Background: Data on lung cancer incidence provide important information on the burden of the disease. Often, incidence data are estimated by cancer registries. However, these registries may only cover parts of the country (50% in Switzerland). On the other hand, mortality data, obtained from death certificates, might be more comprehensive. While there are several suggestions for back-calculation methods, these need to be extended for analyses at a lower geographical level or sparse disease, where the number of cases is low and a rather skewed distribution is expected.

Methods: Bayesian back-calculation models were developed to estimate incidence from survival distributions and lung cancer deaths. The latter was extracted from the national mortality database which is maintained by the Federal Statistical Office (FSO). Mortality and incidence data from the Cancer Registry of St. Gallen-Appenzell (SGA) were used to estimate lung cancer survival probabilities. The proportion of miss-reported cause of death in the FSO data was calculated from the SGA cancer registry data and considered in the analyses. A gamma autoregressive process was adapted for the incidence parameter and survival was assumed to follow a mixed Weibull distribution. Conditional autoregressive models were employed to provide gender-specific smooth maps of age standardized incidence ratios.

Results: Validation comparing observed and estimated incidence for cantons with cancer registries indicated good model performance. Smooth maps of lung cancer incidence for females showed higher estimates in the urbanized regions, while for males a rather homogeneous distribution was observed.

Conclusion: The proposed models improve earlier methodology and are important not only in mapping the spatial distribution of the disease but also in assessing temporal trends of lung cancer incidence in regions without registries.

Detecting Significant Changes in Protein Abundance

Kammers, Kai Ruczinski, Ingo

Johns Hopkins University, USA
kai.kammers@jhu.edu iruczin1@jhu.edu

The problem of identifying differentially expressed proteins in mass- spectrometry based experiments is ubiquitous, and most commonly these comparisons are carried out using t-tests for each peptide or protein. Sample sizes are often small however, which results in great uncertainty for the estimates of the standard errors in the test statistics. The consequence is that proteins exhibiting a large fold change are often declared non-significant because of a large sample variance, while at the same time small observed fold changes might be declared statistically significant, because of a small sample variance. Additionally, adjustments for multiple testing reduce the list of significant peptides or proteins.

In this talk we review and demonstrate how much better results can be achieved by using "moderated" t-statistics, arguably the analytical standard for gene expression experiments. This empirical Bayes approach shrinks the estimated sample variances for each peptide or protein towards a pooled estimate, resulting in far more stable inference particularly when the number of samples is small. Using real data from labeled proteomics experiments (iTRAQ and TMT technology) and simulated data we show how to analyze data from multiple experiments simultaneously, and discuss the effects of missing data on the inference. We also present easy to use open source software for normalization of mass spectrometry data and inference based on moderated test statistics.

Securing the future: genotype imputation performance between different Affymetrix SNP arrays

Kanoungi, George Nürnberg, Peter Nothnagel, Michael

Cologne Center for Genomics, University of Cologne, Cologne, Germany
gkanoung@uni-koeln.de peter.nuernberg@uni-koeln.de michael.nothnagel@uni-koeln.de

Hundreds of thousands of individuals have been genotyped in the past using genotyping arrays and present a valuable data resource also for future biomedical research. Novel chip designs and their altered sets of single-nucleotide polymorphisms (SNPs) pose the question of how well established data resources, such as large samples of healthy controls, can be combined with newer samples genotyped on those novel arrays using genotype imputation. We investigated this question based on the data of 28 HapMap CEU samples that had been genotyped on each of three Affymetrix SNP arrays, namely GeneChip 6.0, Axiom-CEU and Axiom UK-BioBank. Pursuing two parallel approaches, we performed a cross-comparison of the imputation accuracy among the three different arrays and a concordance analysis of the imputed SNPs between the arrays. Imputation was based on the 1000 Genomes reference and conducted using the Impute2 software. We observed that between 6% and 32% of all SNPs on the arrays could not reliably be imputed, depending on the chosen array pair and on the required quality of the imputation. In particular, UK-BioBank performed worst when used for imputing the other two arrays' markers. On the other hand, concordance in predicted genotypes for imputable markers was high for any pair of arrays, ranging from 84% to over 92% for different pairs and quality criteria. We conclude that, given the substantial losses of covered SNPs on the arrays, the re-genotyping of existing samples sets, in particular those of healthy population controls, would be a worthwhile endeavor in order to secure their continued use in the future.

Design and analysis issues in non-traditional epidemiological studies

Kass, Philip H.

UC Davis, United States of America
phkass@ucdavis.edu

Refinements in epidemiological study design methods developed in the last half-century have largely fallen into the domain of the two major observational study designs: cohort and case-control studies. The justification for these two designs arises from their ability to approximate – albeit non-experimentally – randomized studies. But the control through randomization that study investigators can experimentally achieve to approach comparability of groups, and hence (in the absence of other biases) validity, cannot in general be commensurately achieved in non-experimental research. More recently, innovations in study design have been employed to circumvent some of the limitations of classical epidemiological studies. These innovations have included sampling strategies to make case-control designs more efficient or flexible (e.g., case-cohort and nested case-control studies). In addition, a suite of case-only designs that draw from the principles of the potential outcomes model of causal inference have gained popularity. The premise underlying such designs is that individuals potentially live in a simultaneous duality of exposure and non-exposure, with a corresponding duality of observable and hypothetical outcomes. That individuals living under one exposure condition should, in theory, be compared with themselves under other counterfactual (i.e., counter to fact) conditions leads in turn to the premise of "case-only" studies (e.g., self-controlled case-series, case-specular studies). The basis for these designs, as well as examples of their use in the epidemiological literature, will be presented to illustrate their utility.

From treatment selection studies to treatment selection rules: A comparison of four approaches

Kechel, Maren Vach, Werner

Department für Medizinische Biometrie und Medizinische Informatik, Deutschland
kechel@imbi.uni-freiburg.de wv@imbi.uni-freiburg.de

Today, biomarkers often promise to assist in choosing between two different therapeutic alternatives. Clinical trials randomizing all patients to the two alternatives and measuring the biomarker in all patients allow to check such a promise by establishing a (qualitative) interaction. Moreover, they may allow to determine a cut point or a more complex decision rule to decide on the treatment for each patient. Several statistical approaches have been suggested to estimate a marker dependent treatment effect and to assess whether we can state such a dependence. Many of these approaches also allow to assess the precision of the marker dependent treatment effect, e.g. by use of pointwise or simultaneous confidence intervals. However, in most cases no explicit suggestions are made to translate these findings into a treatment selection rule.

In our talk we consider four different approaches to derive at treatment selection rules based on

- the estimated marker dependent treatment effect alone
- pointwise confidence bands
- simultaneous confidence bands
- confidence intervals for the marker value corresponding to a null effect.

We study the behaviour of these different approaches for both linear and quadratic models for the marker dependent treatment effect. To approach this, we consider a framework focusing on the long term impact on all patient populations when following the suggested approach to construct treatment selection rules.

Designing dose finding studies with an active control for exponential families

Kettelhake, Katrin [1] Dette, Holger [1] Bretz, Frank [2]

[1] Ruhr-Universität Bochum, Deutschland
katrin.kettelhake@rub.de olger.dette@rub.de
[2] Novartis Pharma AG, Schweiz
frank.bretz@novartis.com

Often dose finding studies are focused on the comparison of a new developed drug with a marketed standard treatment, refered to as active control. The situation can be modeled as a mixture of two regression models, one describing the new drug, the other one the active control. In a recent paper Dette et al. [1] introduced optimal design problems for dose finding studies with an active control, where they concentrated on regression models with normal distributed errors (with known variance) and the problem of determining optimal designs for estimating the target dose, that is the smallest dose achieving the same treatment effect as the active control.

This talk will discuss the problem of designing active-controlled dose finding studies from a broader perspective considering a general class of optimality criteria and models arising from an exponential family, which are frequently used to analyze count data.

The talk will start with a short introduction to optimal design theory for active-controlled dose finding studies. Next we will show under which circumstances results from dose finding studies including a placebo group can be used to develop optimal designs for dose finding studies with an active control, where we will focus on commonly used design criteria. Optimal designs for several situations will be indicated and the disparities arising from different distributional assumptions will be investigated in detail. The talk will continue considering the problem of estimating the target dose which will lead to a c-optimal design problem. We will finish with the illustration of several examples and a comparison of some designs via their efficiencies.

References:

[1] Dette, H., Kiss, C., Benda, N. and Bretz, F. (2014). Optimal designs for dose finding studies with an active control. *Journal of the Royal Statistical Society, Ser. B*, 76(1):265–295.

Comparing different approaches for network meta-analysis – a simulation study

Kiefer, Corinna Sturtz, Sibylle Sieben, Wiebke Bender, Ralf

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Deutschland
Corinna.Kiefer@iqwig.de Sibylle.Sturtz@iqwig.de Wiebke.Sieben@iqwig.de
Ralf.Bender@iqwig.de

Network meta-analysis is an extension of pairwise meta-analysis. It's becoming more and more popular in systematic reviews and HTA. Whereas for pairwise meta-analysis the properties of different approaches are well examined, little is known for the approaches in network meta-analysis.

Thus, we conducted a simulations study, which evaluates complex networks with up to 5 interventions and different patterns. We analyzed the impact of different network-sizes, different amounts of inconsistency and heterogeneity on MSE and coverage of established and new approaches. This exceeds previous simulations studies, which have been conducted by others [1-3].

We found that, with a high degree of inconsistency in the network, none of the evaluated effect estimators produced reliable results. For a network with no or just moderate inconsistency the Bayesian MTC estimator [4] and the estimator introduced by Rücker et al. [5] both showed acceptable properties, whereas the latter one showed slightly better results. We also found a dependency on the amount of heterogeneity in the network.

Our results highlight the need to check for inconsistency in the network reliably. But measures for inconsistency may be misleading in many situations. We therefore conclude that instead it is especially important to check for similarity (regarding populations, intervention, etc.) and heterogeneity to reduce inconsistency in the network beforehand and to use effect estimators, which can handle a moderate degree of inconsistency.

References:

1. Song, F. et al. (2012): Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. BMC Medical Research Methodology, 12, 138.

2. Jonas, D. E. et al. (2013): Findings of Bayesian Mixed Treatment Comparison Meta-Analyses: Comparison and Exploration Using Real- World Trial Data and Simulation. AHRQ Publication No. 13-EHC039-EF. Rockville, MD: Agency for Healthcare and Quality.

3. Veroniki, A. A. et al. (2014): Characteristics of a loop of evidence that affect detection and estimation of inconsistency: a simulation study. BMC Medical Research Methodology, 14, 106.

4. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Statistics in Medicine 2004; 23(20): 3105–3124.

5. Rücker, G. (2012): Network meta-analysis, electrical networks and graph theory. Research Synthesis Methods, 3, 312–324.

Determining optimal phase II and phase III programs based on expected utility

Kirchner, Marietta [1] Kieser, Meinhard [1] Götte, Heiko [2]
Schüler, Armin [2]

[1] Institut für Medizinische Biometrie und Informatik, Heidelberg
kirchner@imbi.uni-heidelberg.de meinhard.kieser@imbi.uni-heidelberg.de
[2] Merck KGaA, Darmstadt
Heiko.Goette@merckgroup.com Armin.Schueler@merckgroup.com

Traditionally, sample size calculations for pivotal phase III trials are based on efficacy data of preceding studies and are tailored to achieve a desired power. However, large uncertainty about the treatment effect can result in a high risk of trial failure [1]. To address this problem, the concept of assurance (probability of success) has been proposed [2]. It has been shown that the probability of success can be increased by increasing the sample size of the phase II study from which the effect estimate was derived, e.g., [2]. This leads to the idea of program-wise planning meaning that phase II is planned by already taking phase III into account.

We propose an approach for program-wise planning that aim at determining optimal phase II sample sizes and go/no-go decisions based on a utility function that takes into account (fixed and variable) costs of the program and potential gains after successful launch. It is shown that optimal designs can be found for various combinations of values for the cost and benefit parameters occurring in drug development applications. Furthermore, we investigate the influence of these parameter values on the optimal design and give recommendations concerning the choice of phase II sample size in combination with go/no-go decisions. Finally, expected utility for traditional sequential conduct of phase II and III is compared to a seamless approach with a combined phase II/III trial.

In summary, a quantitative approach is presented which allows to determine optimal phase II and III programs in terms of phase II sample sizes and go/no-go decisions and which is based on maximizing the expected utility. Application is illustrated by considering various scenarios encountered in drug development.

References:

[1] Gan HK, You B, Pond GR, Chen EX (2005). Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer. J Natl Cancer Inst 104:590–598.

[2] O'Hagan A, Stevens JW, Campbell MJ (2005). Assurance in clinical trial design. Pharm Stat, 4:187–200.

Gestaltungsmöglichkeiten und Herausforderungen in der Lehre eines berufsbegleitenden Studiengangs am Beispiel des Masterstudiengangs Medical Biometry/Biostatistics

Kirchner, Marietta Rauch, Geraldine Kieser, Meinhard

IMBI, Deutschland
kirchner@imbi.uni-heidelberg.de
rauch@imbi.uni-heidelberg.de
meinhard.kieser@imbi.uni-heidelberg.de

Ein berufsbegleitendes Studium bietet die Möglichkeit, neben dem Beruf eine akademische Qualifikation zu erwerben und sich somit neue Karrierewege zu eröffnen. Das Konzept ist daher für Berufstätige sehr attraktiv, stellt aber an die Studiengangorganisatoren besondere Herausforderungen. Die Besonderheit der berufsbegleitenden Struktur und die sich ergebenden Gestaltungsmöglichkeiten in der Lehre sollen in diesem Vortrag am Beispiel des Masterstudiengangs Medical Biometry/ Biostatistics vorgestellt werden.

Am Institut für Medizinische Biometrie und Informatik der Universität Heidelberg wird seit 2006 der berufsbegleitende Masterstudiengang Medical Biometry/ Biostatistics angeboten. Ziel des Studiengangs ist es, Biometriker auszubilden, die in dem interdisziplinären Kontext von Medizin und Statistik erfolgreich arbeiten. Die berufsbegleitende Struktur des Studiums bietet die Möglichkeit, die berufliche Tätigkeit in das Studium zu integrieren. So kann das Gelernte einerseits im Beruf direkt angewendet werden, andererseits können bestimme berufliche Tätigkeiten für das Studium angerechnet werden. Das Studium ist so organisiert, dass es mit einer Berufstätigkeit in Teil- oder Vollzeit vereinbar ist. Somit finden die Lehrveranstaltungen in der Regel in Blockkursen außerhalb des normalen Semesterbetriebs statt. Ein Kurskoordinator stimmt die Inhalte innerhalb eines Blockkurses mit den jeweiligen Dozenten ab, die Fachexperten auf dem entsprechenden Gebiet sind. So können aufeinander aufbauende Kurseinheiten gewährleistet und Redundanzen vermieden werden. Zudem werden die Inhalte kursübergreifend durch ein zentrales Organisationsteam geprüft. Ein zentraler Baustein zur stetigen Weiterentwicklung der Kurse ist unser Evaluationssystem, über das die Studierenden Rückmeldung zu den einzelnen Kursen auf organisatorischer und inhaltlicher Ebene geben können. Im Vortrag wird dargestellt, welche Chancen sich für die Lernenden durch ein solches System bieten, und mit welchen organisatorischen Herausforderungen dies verbunden ist.

Efficiency and Robustness of Bayesian Proof of Concept / Dose Finding Studies with Weakly Informative Priors

Klein, Stefan Mager, Harry

Bayer Healthcare, Deutschland
stefan.klein@bayer.com harry.mager@bayer.com

In early drug development estimation of dose-response relationships plays a crucial part. Although sample sizes are usually rather low, Bayesian data augmentation using historical studies may be hampered due to the lack of predecessor studies for the respective compound. Therefore, well established procedures like the meta-analytic-predicitve approach (1) cannot or can only partially be applied.

In many cases however, there exists at least informal prior knowledge on the parameters of such a dose response relationship in form of some upper and lower bounds i.e. intervals which should cover the expected parameter values with high but unknown probability.

Based on such truncated flat priors the efficiency and robustness of a Bayesian modeling approach is examined for sigmoidal dose response relationship in proof of concept (POC) / dose finding trials.

Essentially we will explore if an informative prior distribution derived from such information may have substantial impact on the probability to correctly determine a minimal effective dose (MED) as defined in (2).

The robustness of this approach is reviewed under several scenarios including such scenarios where central model assumptions like the shape of the dose response curve and the validity of the assumed prior distribution are violated.

The Bayesian approach adopted here will be compared to its respective classical statistics counterpart as well as to a Bayesian procedure using a vague prior (3) and illustrated by an example from pharmaceutical development.

References:

(1) Neuenschwander B, Capkun-Nigglia G, Branson M and Spiegelhalter DJ. Summarizing historical information on controls in clinical trials. Clinical Trials 2010; 7: 5–18.

(2) Bretz F, Hsu J, Pinheiro J and Liu Y. Dose Finding-A Challenge in Statistics. Biometrical Journal 2008; 50:4, 480–504.

(3) Thomas N. Hypothesis Testing and Bayesian Estimation using a sigmoid Emax model applied to Sparse Dose-Response Designs. Journal of Biopharmaceutical Statistics 2006; 16:5, 657–677.

Simultaneous Inference in Structured Additive Conditional Copula Regression Models: A Unifying Bayesian Approach

Kneib, Thomas Klein, Nadja

Georg-August-Universität Göttingen, Deutschland
tkneib@uni-goettingen.de nklein@uni-goettingen.de

While most regression models focus on explaining aspects of one single response variable alone, interest in modern statistical applications has recently shifted towards simultaneously studying of multiple response variables as well as their dependence structure. A particularly useful tool for pursuing such an analysis are copula-based regression models since they enable the separation of the marginal response distributions and the dependence structure summarised in a specific copula model. However, so far copula-based regression models have mostly been relying on two-step approaches where the marginal distributions are determined first whereas the copula structure is studied in a second step after plugging in the estimated marginal distributions. Moreover, the parameters of the copula are mostly treated as a constant not related to covariates and most regression specifications for the marginals are restricted to purely linear predictors. We therefore propose simultaneous Bayesian inference for both the marginal distributions and the copula using computationally efficient Markov chain Monte Carlo simulation techniques. In addition, we replace the commonly used linear predictor by a generic structured additive predictor comprising for example nonlinear effects of continuous covariates, spatial effects or random effects and also allow to make the copula parameters covariate-dependent. To facilitate Bayesian inference, we construct proposal densities for a Metropolis Hastings algorithm relying on quadratic approximations to the full conditionals of regression coefficients, thereby avoiding manual tuning and thus providing an attractive adaptive algorithm. The flexibility of Bayesian conditional copula regression models is illustrated in two applications on childhood undernutrition and macroecology.

Vergleich von co-primären und kombinierten binären Endpunkten in klinischen Studien

Kohlhas, Laura [1] Zinserling, Jörg [2] Benda, Norbert [2]

[1] Universität Heidelberg, Deutschland
laura.kohlhas@web.de
[2] Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), Bonn, Germany
joerg.zinserling@bfarm.de, norbert.benda@bfarm.de

In vielen Therapiebereichen werden zwei Variablen benötigt, um den Behandlungseffekt adäquat zu beschreiben. Hierbei sprechen wir von co-primären Endpunkten, falls für den Studienerfolg ein positiver Behandlungseffekt in beiden Endpunkten gezeigt werden muss. Eine andere Möglichkeit ist es, die Variablen zu einem kombinierten Endpunkt zusammenzufassen.

Zwischen der Verwendung von co-primären Endpunkten und einem kombinierten Endpunkt zu entscheiden, ist oft nicht einfach. Die Endpunkte können eine unterschiedliche klinische Bedeutung haben, ein Effekt könnte erst dann als abgesichert gelten, wenn beide Endpunkte erfolgreich sind, und es können sich Unterschiede in der Power oder der benötigten Fallzahl ergeben. Dass die Entscheidung nicht immer eindeutig ist, zeigt sich auch in nicht übereinstimmenden Vorgaben zur Verwendung von co-primären und kombinierten Endpunkten der europäischen und der US-amerikanischen Arneimittelzulassungsbehörde (EMA bzw. FDA) für verschiedene Indikationen. Insbesondere werden in Studien zur Behandlung des Reizdarmsyndroms die binären Endpunkte des Ansprechens bezüglich des Schmerzes einerseits und der abnormen Stuhlentleerung andererseits unterschiedlich verwendet.

Abhängig von dem Typ der Variablen gibt es verschiedene Methoden diese zu kombinieren. Hier werden binäre Variablen betrachtet, für die eine Response im kombinierten Endpunkt als Response in beiden Komponenten definiert ist. In diesem Beitrag werden die Gütefunktionen von kombinierten und co-primären Endpunkten basierend auf dem Z-Test für Anteile miteinander verglichen. Insbesondere wird die Auswirkung von verschiedenen Korrelationen zwischen den beiden Variablen untersucht.

Die Ergebnisse zeigen, dass die Wahrscheinlichkeit für eine Signifikanz bei Verwendung des kombinierten Endpunktes höher ist, wenn die Korrelationen in beiden Gruppen übereinstimmen oder die Korrelation in der Interventionsgruppe größer ist. Für den Fall einer höheren Korrelation in der Kontrollgruppe, hängt der Vergleich zusätzlich von der Differenz der Effekte zwischen beiden co-primären Endpunkten ab. Sind diese Effekte ähnlich, wird mit den co-primären Endpunkten eine höhere Power erreicht; unterscheiden sie sich deutlich voneinander, liefert der kombinierte Endpunkt eine höhere Power.

Literatur

Corsetti M, Tack J.: FDA and EMA end points: which outcome end points should we use in clinical trials in patients with irritable bowel syndrome? Neurogastroenterol Motil 2013; 25: 453-457.

Unimodal spline regression for detecting peaks in spectral data

Köllmann, Claudia Ickstadt, Katja Fried, Roland

Faculty of Statistics, TU Dortmund, Germany
koellmann@statistik.tu-dortmund.de, ickstadt@statistik.tu-dortmund.de,
fried@statistik.tu-dortmund.de

Unimodal regression as a type of nonparametric shape constrained regression is a suitable choice in regression problems where the prior information about the underlying relationship between predictor and response is vague, but where it is almost certain that the response first increases and then (possibly) decreases with higher values of the predictor. A semi-parametric spline regression approach to unimodal regression was derived in [1] and its usefulness in dose-response analysis was verified.

The method is based on the fact that using the B-spline basis, a spline can be restricted to be unimodal by choosing a unimodal sequence of B-spline coefficients with fixed mode. The mode is then selected from all possibilities based on the least squares criterion. The use of spline functions guarantees the continuity of the fit and smoothness can be achieved by using penalized spline regression.

In this talk we demonstrate with real data examples that unimodal regression is also useful in situations where the relationship between two variables is not unimodal, but multimodal. Explicitly, we use additive models, where each component is a unimodal regression, to detect peaks in spectral data from ion mobility spectrometry (IMS) or nuclear magnetic resonance spectroscopy (NMR). Since such data may contain several peaks and the unimodal regressions are fitted repeatedly with a backfitting algorithm, we propose to replace the unimodality constraint on the B-spline coefficients by a unimodality penalty to decrease the computational burden.

References

[1] Köllmann C, Bornkamp B, Ickstadt K (2014). Unimodal regression using Bernstein-Schoenberg-splines and penalties. Biometrics.
doi:10.1111/biom.12193.

Facing Uncertainty in Heterogeneity Parameters in Bayesian Network Meta-Analysis

König, Jochem [1] Krahn, Ulrike [1,2] Binder, Harald [1]

[1] Institut für Medizinische Biometrie, Epidemiologie und Informatik, Universitätsmedizin Mainz. Germany koenigjo@uni-mainz.de, krahnu@uni-mainz.de, binderh@uni-mainz.de
[2] Institut für Medizinische Informatik, Biometrie und Epidemiologie, Universitätsklinikum Essen, Germany

In network meta-analysis, direct evidence of different studies is pooled, each comparing only few treatments. The results are network based effect estimates for all pairs of treatments, taking indirect evidence into account. We have provided tools to see how parts of the network influence specific treatment comparisons [1]. These tools are satisfactory in the absence of heterogeneity within designs (i.e. studies comparing the same set of treatments) or when the heterogeneity variance is known. Unfortunately, often in network meta-analysis, evidence on heterogeneity and consistency is limited. Bayesian network meta-analysis, which is current practice to cope with uncertainty in heterogeneity parameters, averages over a range of parameter settings that is compatible with data in the light of prior knowledge. Then the results may be unduly depending on the choice of the prior distribution.

We therefore suggest new graphical tools to disclose the dependency of interval estimates of treatment effects on the heterogeneity parameters or, in particular, on the choice of the prior distribution. Some subsets of studies sharing the same design may show more heterogeneity than others and may, by consequence, inflate confidence intervals for other treatment comparisons. Our graphical tools aim at showing two aspects of influence of single studies or designs: the influence on heterogeneity parameter estimates and on effect estimates. More specifically, we plot interval estimates of treatment effects against the square root of the overall heterogeneity variance parameter and overlay its restricted likelihood. Various point estimates resulting from different priors or exclusion of groups of studies are marked as vertical reference lines.

Thus we illustrate sensitivity of effect estimates to uncertainty in modelling heterogeneity. Methods are illustrated with some published network meta-analyses.

References:

[1] Krahn U, Binder H, König J (2015). Visualizing inconsistency in network meta-analysis by independent path decomposition, BMC Medical Research Methodology, to appear.

Modellierung von Themenkarrieren in Printmedien

Koppers, Lars Boczek, Karin von Nordheim, Gerret Müller, Henrik
Rahnenführer, Jörg

TU Dortmund, Germany
koppers@statistik.tu-dortmund.de, karin.boczek@tu-dortmund.de,
gerret.vonnordheim@tu-dortmund.de, henrik.mueller@tu-dortmund.de,
rahnenfuehrer@statistik.tu-dortmund.de

Das Ausmaß an Berichterstattung zu einem bestimmten Thema in den (Print-)Medien über die Zeit hinweg wird Themenkarriere genannt und unterliegt oft einem charakteristischen Verlauf. Typischerweise lässt zunächst ein einzelnes Ereignis (etwa ein herausragendes Ereignis, eine Pressekonferenz oder eine wissenschaftliche Veröffentlichung) die Aufmerksamkeit für dieses Thema in den Medien stark ansteigen. Diese starke Aufmerksamkeit lässt aber nach einem kurzen Zeitraum wieder nach und fällt auf ein Grundrauschen ab, das im Vergleich zum Rauschen vor dem Ereignis etwas erhöht ist. Um den Verlauf von Themenkarrieren verstehen zu können, sollen diese in Textsammlungen von Zeitungen und Zeitschriften mit statistischen Methoden modelliert werden.

Grundlage unserer Analysen sind alle Artikel der Wochenzeitschrift "DER SPIEGEL" der Jahrgänge 1947–2013. Es werden zwei verschiedene Ansätze verfolgt. Als erste Annäherung an Themen werden einzelne repräsentative Wörter betrachtet. Eine typische zeitlich begrenzte Aufmerksamkeitskurve kann mit einer shifted Gompertz-Verteilung beschrieben werden. Auf Grund der Größe des betrachteten Zeitraums besteht eine typische Wortverteilung dann aus einer Mischung von gleichverteiltem Grundrauschen und einer oder mehrerer shifted Gompertz-verteilter Aufmerksamkeitsverteilungen. Im Vortrag zeigen wir eine Strategie zum Modellieren von Wortverteilungen auf unterschiedlich zeitlich aufgelösten Daten.

Im zweiten Ansatz wird nicht der Verlauf für einzelne Wörter oder Wortfamilien modelliert, sondern für Themen, mit Hilfe von LDA (latent dirichlet allocation). Die LDA ist ein topic model, mit dem, ausgehend von einer Artikelsammlung, eine Anzahl von Themen geschätzt wird, denen die Wörter eines Artikels und damit auch die Artikel selbst mit bestimmten Wahrscheinlichkeiten zugeordnet werden können. Die geschätzte Themenverteilung über die Zeit repräsentiert dann eine Themenkarriere und kann direkt als Grundlage für die Modellierung der zeitlichen Verteilung genutzt werden.

Evaluating two-level data: Comparison of linear mixed model analysis to ANOVA based on averaged replicate measurements

Kopp-Schneider, Anette Link, Mirko

DKFZ, Germany
kopp@dkfz.de, mirko.link@dkfz.de

Biomedical studies often use replicate measurements of biological samples, leading to a two-level situation of technical replicates of biological replicates. If a treatment effect is investigated and hence groups of biological replicates are to be compared, the appropriate analysis method in this situation is the application of a linear mixed model. Since linear mixed model analysis requires either use of statistical programming or of commercial statistical analysis software, a substitution used in biomedical research is to evaluate the mean of technical replicates, hence reducing the analysis from a two-level to a single level approach ("averaging approach").

In the case of a balanced design, i.e. equal number of technical replicates for every biological replicate, the linear mixed model and the averaging approach lead to identical results. Diverging results are obtained for unbalanced situations. Different linear mixed model approaches are available, such as the approximate ANOVA-based test by Cummings and Gaylor (1), the exact test by Khuri (2) and the Likelihood ratio test. Using simulated data with various constellations of technical and biological variability, the performance of linear mixed model approaches is compared to the averaging approach.

References:

(1) Cummings, W.B. and Gaylor, D.W. (1974). Variance Component Testing in Unbalanced Nested Designs. J. Amer. Statist. Assoc. 69: 765–771

(2) Khuri, A.I. Linear Model Methodology. CRC Press 2010.

Heterogenität in Subgruppen: Statistische Eigenschaften von Regeln zur Signalgenerierung

Kottas, Martina Gonnermann, Andrea Koch, Armin

Medizinische Hochschule Hannover, Germany
Kottas.Martina@mh-hannover.de, Gonnermann.Andrea@mh-hannover.de, Koch.Armin@mh-hannover.de

In randomisierten, kontrollierten Phase-III-Studien soll die Wirksamkeit und Sicherheit eines Prüfpräparats in einer breiten Studienpopulation nachgewiesen werden. Auch wenn das Prüfpräparat sich in der gesamten Studienpopulation gegenüber der Kontrolle als wirksam erweist, kann nicht daraus geschlossen werden, dass es in allen Subpopulationen gleichermaßen wirksam ist. So ist es durchaus denkbar, dass in einigen Subgruppen das Prüfpräparat wirksamer oder sogar aufgrund von Nebenwirkungen schädlicher sein könnte als im Vergleich zu anderen Subgruppen. Daher sollte die Konsistenz des Gesamtbehandlungseffektes in verschiedenen Subgruppen überprüft werden. Um festzustellen, ob der Behandlungseffekt in den Subgruppen variiert, können beispielweise Heterogenitäts- oder Interaktionstests verwendet werden. Diese werden jedoch für eine zu geringe Power kritisiert [1]–[3].

Im Vortrag werden vier Regeln vorgestellt, mit denen Signale für Heterogenität entdeckt werden können. Eine der Regeln liegt auf Cochran's Q, einen Heterogenitätstest bekannt aus Meta-Analysen [4], zugrunde. Zwei weitere Regeln basieren auf dem Behandlungseffekt und ob dieser in einer Subgruppe entweder halb oder doppelt bzw. ein Viertel oder viermal so groß wie der Gesamtbehandlungseffekt ist. Die letzte Regel betrachtet, ob der Punktschätzer einer Subgruppe nicht im Konfidenzintervall des Gesamtbehandlungseffektes liegt. Die Regeln wurden in einer Simulationsstudie in einem Modell mit festen Effekten und einem Modell mit zufälligen Effekten untersucht. Hierbei wurden die Anzahl und auch die Größe der Subgruppen in verschiedenen Simulationsszenarien variiert und die Regeln unter der Annahme von Homogenität und Heterogenität miteinander verglichen. Basierend auf den Simulationsergebnissen sollen die Eigenschaften dieser Regeln anhand des empirischen Fehlers 1. Art und der Power diskutiert werden.

References:

[1] Committee for Medicinal Products for Human Use, "Guideline on adjustment for baseline covariates - DRAFT," 2013. [Online]. Available: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/06/WC500144946.pdf.

[2] D. Jackson, "The power of the standard test for the presence of heterogeneity in meta-analysis," Stat. Med., vol. 25, pp. 2688–2699, Aug. 2006.

[3] R. J. Hardy and S. G. Thompson, "Detecting and describing heterogeneity in metaanalysis," Stat. Med., vol. 17, pp. 841–856, 1998.

[4] W. G. Cochran, "The combination of estimates from different experiments," Biometrics, vol. 10, pp. 101–129, 1954.

Communication between agricultural scientists and statisticians: a broken bridge?

Kozak, Marcin

University of Information Technology and Management in Rzeszów, Poland
nyggus@gmail.com

Modern agricultural research requires a lot of statistics. Unfortunately, most up-to-date statistical methods need advanced knowledge, often out of reach for agricultural researchers. Thus, efficient communication between them and statisticians is important for agricultural knowledge to develop. Many agricultural researchers claim that communication with statisticians is difficult; but also many statisticians claim that communication with agricultural researchers is not easy. This being true, such poor communication can be a barrier to efficient agricultural research. The aim of this research is to study this phenomenon. Do agricultural researchers and statisticians see problems in their communication? What kinds of problems these are? I will try to answer these questions based on a study among scientists representing both these groups.

On functional data analysis applied to interpretation of next-generation sequencing experimental results

Krajewski, Pawel [1] Madrigal, Pedro [2]

[1] Polish Academy of Sciences, Poland
pkra@igr.poznan.pl
[2] Wellcome Trust Sanger Institute, UK
pm12@sanger.ac.uk

An increasing number of research questions in biology is today answered by high-throughput sequencing of DNA or RNA. The experiments utilizing those techniques grow in size and complexity, and the next-generation sequencing (NGS) data volume expands. We show how the statistical processing of NGS data can be done by application of the functional data analysis (FDA) methods. The possibility of such application follows from the fact that in many situations the raw NGS data, the short reads produced by the sequencing device, are mapped to a reference genome, thus providing profiles of genome coverage containing signals. We treat those processed data as realizations of functions over the domain of genomic sequence. The FDA method that turned out to be especially useful is the functional principal component analysis (FPCA). It starts by approximating the observed profiles by linear combinations of some basis functions (e.g., B-splines), and then proceeds to estimating the eigenfunctions representing most of the data variability. In a number of experimental situations, the eigenfunctions and the FPCA scores can be conveniently interpreted to provide biological interpretation. The examples of data analysis that will be presented concern application of NGS techniques in problems related to transcription factor binding (ChIP-Seq assay), chromatin structure assessment (4C-Seq assay), and analysis of histone modifications (RRB-Seq assay).

Optimal Decision Rules for Subgroup Selection for a Targeted Therapy in Oncology

Krisam, Johannes Kieser, Meinhard

Universität Heidelberg, Germany
krisam@imbi.uni-heidelberg.de, meinhard.kieser@imbi.uni-heidelberg.de

Throughout the recent years, there has been a rapidly increasing interest regarding the evaluation of so-called targeted therapies. These therapies are assumed to show a greater benefit in a pre-specified subgroup of patients (commonly identified by a predictive biomarker) as compared to a total patient population of interest. This complex situation has led to the necessity to develop statistical tools which allow an efficient evaluation of such treatments. Amongst others, adaptive enrichment designs have been proposed as a solution (see, e.g., [1-2]). These designs allow the selection of the most promising subgroup based on an efficacy analysis at interim. As has recently been shown, the performance of the applied interim decision rule in such a design plays a crucial role in ensuring a successful trial [3].

We investigate the situation that the primary outcome of the trial is binary or a time-to-event variable. Statistical methods are developed that allow an evaluation of the performance of decision rules in terms of correct selection probability at interim. Additionally, optimal decision rules are derived which incorporate the uncertainty about several design parameters, such as the treatment effects and the sensitivity and specificity of the employed bioassay. These optimal decision rules are evaluated regarding their performance in an adaptive enrichment design in terms of correct selection probability, type I error rate and power and are compared to ad-hoc decision rules proposed in the literature.

References:

[1] Wang S J., O'Neill R T, Hung H.M J. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. Pharmaceutical Statistics 2007; 6:227-244.

[2] Jenkins M, Stone A, Jennison C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. Pharmaceutical Statistics 2011; 10:347–356.

[3] Krisam J, Kieser M. Decision rules for subgroup selection based on a predictive biomarker. Journal of Biopharmaceutical Statistics 2014; 24:188-202.

Simulation of Correlated Count Data from Sequencing or Mass Spectrometry Experiments

Kruppa, Jochen Beißbarth, Tim Jung, Klaus

Universitätsmedizin Göttingen, Germany
Jochen.Kruppa@med.uni-goettingen.de, Tim.Beissbarth@med.uni-goettingen.de, Klaus.Jung@med.uni-goettingen.de

Count data occur frequently in next-generation sequencing experiments, for example when the number of reads mapped to a particular region of a reference genome is counted [1]. Count data are also typical for proteomics experiments making use of affinity purification combined with mass spectrometry, where the number of peptide spectra that belong to a certain protein is counted [2]. Simulation of correlated count data is important for validating new statistical methods for the analysis of the mentioned experiments. Current methods for simulating count data from sequencing experiments only consider single runs with uncorrelated features, while the correlation structure is explicitly of interest for several statistical methods.

For smaller sets of features (e.g. the genes of a pathway) we propose to draw correlated data from the multivariate normal distribution and to round these continuous data to obtain discrete counts. Because rounding can affect the correlation structure – especially for features with low average counts – we evaluate the use of shrinkage covariance estimators that have already been applied in the context of simulating high-dimensional expression data from DNA microarray experiments.

In a simulation study we found that there are less deviations between the covariance matrices of rounded and unrounded data when using the shrinkage estimators. We demonstrate that the methods are useful to generate artificial count data for certain predefined covariance structures (e.g. autocorrelated, unstructured) but also for covariance structures estimated from data examples.

References:

[1] Liao, Y., Smyth, G.K., Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics, 30, 923-930.

[2] Choi, H., Fermin, D., Nesvizhskii, A.I. (2008) Significance Analysis of Spectral Count Data in Label-free Shotgun Proteomics, Moll Cell Proteomics, 7, 2373-2385.

On responder analyses in the framework of within subject comparisons - considerations and two case studies

Kunz, Michael

Bayer Pharma AG, Deutschland
michael.kunz@bayer.com

A responder analysis is a common tool that dichotomizes at least ordinal scaled data. In this paper, we generalize the definition of responders to within subject comparisons. Two large case studies conducted for regulatory purposes are presented that originate from the field of diagnostic imaging trials. Each subject was investigated using two different imaging modalities which enabled a subsequent within subject comparison. In this particular case the generalized responder analysis resulted in a better understanding of the data and even indicated a better performance of the investigational product vs. the comparator. We get further insights how the generalized responder analysis and the analysis based on the arithmetic mean compare in terms of power and identify the (rather rare) constellations where the generalized responder analysis outperforms the analysis based on the arithmetic mean.

References:

M. Kunz 'On responder analyses in the framework of within subject comparisons – considerations and two case studies', Statistics in Medicine, 2014, 33, 2939-2952.

Bias Reduction for Point Estimation of Treatment Effects in Two-Stage Enrichment Designs

Kunzmann, Kevin Kieser, Meinhard

Universität Heidelberg, Germany
kunzmann@imbi.uni-heidelberg.de, meinhard.kieser@imbi.uni-heidelberg.de

Over the last years, interest in so-called personalized medicine has risen steeply. In contrast to classical drug development, therapies are not tailored to the complete patient population but to specific biomarker-defined subgroups. Even in the case of biologically reasonable, pre-defined biomarkers, designing suitable trials remains a challenging task. As an appealing solution to this problem, adaptive enrichment designs have been proposed [1], where the most promising target population can be selected for further study during an interim analysis. As is well known [2], selection based on the variable of interest introduces bias to the classical maximum likelihood estimator of the treatment effect. Lack of a proper assessment of the problem's extent potentially puts investigators in an unsatisfactory situation.

We investigate the bias of the maximum likelihood estimator over a range of scenarios. The relationship to parameter estimation in the situation of treatment selection is established which allows to extend recent bias-reduction methods in this field [3 – 5] to the setting of enrichment designs. The performance of the competing approaches is compared in a simulation study, and recommendations for practical application are given.

References:

[1] Wang, S. J., O'Neill, R. T., Hung, H. M. J. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. Pharmaceutical Statistics 2007; 6:227–244.

[2] Bauer, P., Koenig, F., Brannath, W., Posch, M. Selection and bias – two hostile brothers. Statistics in Medicine 2010; 29:1–13.

[3] Luo, X., Mingyu, L., Weichung, J. S., Ouyang, P. Estimation of treatment effect following a clinical trial with adaptive design. Journal of Biopharmaceutical Statistics 2012; 22:700–718.

[4] Carreras, M., Brannath, W. Shrinkage estimation in two-stage adaptive designs with midtrial treatment selection, Statistics in Medicine 2013; 32:1677–1690.

[5] Pickard, M. D., Chang, M., A flexible method using a parametric bootstrap for reducing bias in adaptive designs with treatment selection. Statistics in Biopharmaceutical Research 2014; 6:163–174.

Automatic model selection for high dimensional survival analysis

Lang, Michel Bischl, Bernd Rahnenführer, Jörg

TU Dortmund University, Germany
lang@statistik.tu-dortmund.de, bischl@statistik.tu-dortmund.de,
rahnenfuehrer@statistik.tu-dortmund.de

Many different models for the analysis of high-dimensional survival data have been developed over the past years. While some of the models and implementations come with an internal parameter tuning automatism, others require the user to accurately adjust defaults, which often feels like a guessing game. Exhaustively trying out all model and parameter combinations will quickly become tedious or unfeasible in computationally intensive settings, even if parallelization is employed. Therefore, we propose to use modern algorithm configuration techniques like model based optimization (MBO) to efficiently move through the model hypothesis space and to simultaneously configure algorithm classes and their respective hyperparameters.

We apply MBO on eight high dimensional lung cancer data sets extracted from the gene expression omnibus database (GEO). The R package `mlr` is used here as an interface layer for a unified access to survival models. All models are fused with a feature selection filter whose hyperparameters get tuned as well.

MBO fits a regression model in the algorithm's hyperparameter space where the prediction performance on test data is the regressand. Interesting configurations are iterative suggested by a classical exploration-exploitation trade-off which prefers regions with either good performance estimates or high uncertainty.

In order to avoid tuning generally bad performing models, all considered survival models are tuned simultaneously so that the optimizer can avoid these regions in the hyperspace and use its resources in more promising regions. The resulting algorithm configuration is applied on an independent test set and compared with its auto-tuned counterparts.

MCDM im deutschen HTA Umfeld

Leverkus, Friedhelm Volz, Fabian

Pfizer Deutschland GmbH, Germany
friedhelm.leverkus@pfizer.com, fabian.volz@pfizer.com

In den letzten Jahren sind Patientenpräferenzen bei der Nutzen-Risiko-Bewertung von Arzneimitteln im Rahmen der Zulassung verstärkt in den Fokus der europäischen Arzneimittelbehörde (European Medicines Agency, EMA) getreten. Auch im Bereich des Health Technology Assessments (HTA) wird verstärkt über die Rolle der Patientenpräferenzen im Rahmen der Bewertung eines Zusatznutzens von Arzneimitteln diskutiert. Zur quantitativen Erhebung von Patientenpräferenzen eignen sich verschiedene Verfahren aus dem Multiple Criteria Decision Making (MCDM), wie bspw. Analytic Hierarchy Process (AHP) oder Discrete Choice Experimente (DCE). Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) hat zwei Pilotstudien durchgeführt, um diese beiden Verfahren für die Erstellung eines Maßes des Gesamtnutzens im Rahmen der Kosten-Nutzen-Bewertung zu prüfen. Die Ergebnisse der beiden Pilotstudien werden vorgestellt und es soll geprüft werden, ob diese beiden Verfahren bei der frühen Nutzenbewertung nach AMNOG eingesetzt werden können, insbesondere um den Zusatznutzen und ggf. Schaden in verschiedenen patientenrelevanten Endpunkten eines neuen Arzneimittels zu einem Gesamtmaß zu aggregieren.

Monitoring of count time series following generalized linear models

Liboschik, Tobias [1] Fried, Roland [1] Fokianos, Konstantinos [2]

[1] Technische Universität Dortmund, Germany
liboschik@statistik.tu-dortmund.de, fried@statistik.tu-dortmund.de
[2] University of Cyprus
fokianos@ucy.ac.cy

Detecting changes in time series of counts is of interest in various biostatistical applications, particularly in public health. Surveillance of infectious diseases aims at detection of outbreaks of epidemics with only short time delays in order to take adequate action promptly. Further goals are the validation of emergency and prevention measures as well as the prediction and identification of relevant trends for planning [1]. An appealing and flexible class to model such data are count time series following generalized linear models [4], for instance the so-called INGARCH models. As compared to other modelling approaches [2], this class accounts for the serial dependence typically observed.

Our study compares possible statistics for CUSUM or EWMA charts, for example we use transformed observations, different types of residuals and some likelihood-based measures. Considering situations where outliers might occur in the in-control data, we investigate how robust model fitting in Phase I can ensure a more reliable performance in Phase II.

References:

[1] Dresmann, J., Benzler, J. (2005). Surveillance übertragbarer Krankheiten auf der Grundlage des Infektionsschutzgesetzes in Deutschland durch den öffentlichen Gesundheitsdienst. Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz 48, 979–989.

[2] Höhle, M. (2007). surveillance: An R package for the monitoring of infectious diseases. Computational Statistics 22, 571-582.

[3] Liboschik, T., Kerschke, P., Fokianos, K., Fried, R. (2014). Modelling interventions in INGARCH processes. International Journal of Computer Mathematics, published online.
http://dx.doi.org/10.1080/00207160.2014.949250.

[4] Liboschik, T., Probst, P., Fokianos, K., Fried, R. (2014). tscount: An R package for analysis of count time series following generalized linear models. http://tscount.r-forge.r-project.org.

[5] Weiß, C.H. and Testik, M.C. (2012). Detection of Abrupt Changes in Count Data Time Series: Cumulative Sum Derivations for INARCH(1) Models. Journal of Quality Technology 44, 249-264.

Kritische Anmerkungen zur Anwendung von Propensity Score Methoden für die Analyse des Auftretens schwerwiegender unerwünschter Ereignisse

Listing, Joachim Richter, Adrian Klotsche, Jens

Deutsches Rheuma-Forschungszentrum, Germany
Listing@drfz.de, richter@drfz.de, Jens.Klotsche@drfz.de

Im Falle von Beobachtungsstudien sind simple Vergleiche von Behandlungswirkungen nicht zulässig. Behandlungsentscheidungen für oder gegen eine bestimmte Therapie werden nicht zufällig gefällt, sondern den spezifischen Erfordernissen der Patienten angepasst. Patienten, die mit Medikament A behandelt wurden, können a priori ein ungünstigeres Risikoprofil aufweisen als jene, die mit Medikament B behandelt wurden. Zur Kontrolle dieser verzerrenden Effekte werden in zunehmenden Maße Propensity Score (PS) Methoden eingesetzt. Teilweise wird ihre Anwendung von Gutachtern sogar gefordert. Dies ist problematisch, da die Verwendung von PS Methoden insbesondere bei der Analyse schwerwiegender unerwünschter Ereignisse (SUE) zu weiteren Verzerrungen und Fehlinterpretationen führen kann.

Der PS wird üblicherweise mittels logistischer Regression als bedingte Wahrscheinlichkeit einer Therapiezuordnung berechnet. Unter der Annahme großer Fallzahlen lassen sich balancierende Eigenschaften der PS Methoden nachweisen; die Asymptodik greift im Falle von Mittelwerts Vergleichen eher als im Falle der Analyse von SUEs.

Mit wachsender Fallzahl und Trennschärfe der Kovariablen steigt aber auch das Risiko, dass die empirischen PS Verteilungen sich nur teilweise überlappen oder es zumindest für Patienten unter A mit hohem PS nur sehr wenige Patienten unter Therapie B mit ähnlich hohem PS gibt. Da Analysen von SUEs auf Basis der PS Strata dann nicht mehr durchführbar sind, werden häufig gewichtete Analysen (inverse probablity weighting methods) auf Basis des PS gerechnet ohne das immanente Risiko von Verzerrungen ausreichend zu berücksichtigen.

Besonders kritisch sind PS basierte Längsschnittanalysen mit längerer Nachbeobachtungszeit zu beurteilen. Verzerrungen durch Dropouts und Therapiewechsel werden hier häufig ebenso wenig beachtet wie die Tatsache, dass Randverteilungs Odds Ratios häufig nicht geeignet sind, das Risiko einzelner Patienten zu schätzen.

Literatur:

Rosenbaum PR, Rubin DB (1983): The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41-45.

D'Agostino RB (1998): Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Statist. Med 17, 2265-2281.

Lunt M et. al.: Different methods of balancing covariates leading to different effect estimates. Am J Epidemiol 2009, 909-17.

On confidence statements associated to P-values

Mattner, Lutz Dinev, Todor

Universität Trier, Deutschland
mattner@uni-trier.de, dinev@uni-trier.de

We start with the simple but apparently not too well-known observation that from every P-value for a testing problem $((P_\vartheta : \vartheta \in \Theta), \Theta_0)$ we can directly compute a realized confidence region "for ϑ," thus making precise in a way the idea that small P-values provide stronger evidence against the hypothesis. By considering, however, a few examples, we show that such confidence regions are typically far from optimal.

In teaching statistics, the above may contribute to the understanding of P-values and, more importantly, suggest constructing tailor-made confidence regions anyway rather than being content with testing.

Schätzung von Halbwertszeiten in nichtlinearen Daten mit Fractional Polynomials

Mayer, Benjamin

Universität Ulm, Deutschland
benjamin.mayer@uni-ulm.de

Hintergrund
Regressionsmodelle werden häufig eingesetzt, um den Zusammenhang zwischen einer interessanten Zielgröße und einer oder mehrerer potentieller Einflussgrößen zu modellieren. Die Bestimmung von Halbwertszeiten stellt ein konkretes Anwendungsbeispiel dar. Das spezifische Problem hierbei ist allerdings, dass die zeitabhängige Dosiskonzentration eines untersuchten Präparats oftmals nichtlinear ist. Gleichzeitig stellt jedoch gerade die Linearitätsannahme eine entscheidende Limitierung für den Einsatz gewöhnlicher Regressionsverfahren dar, weshalb die Notwendigkeit für flexiblere Modellierungsansätze besteht.

Methoden
Die Anwendung eines univariaten Fractional Polynomials (FP) wird als Modellierungsansatz vorgestellt, der anschließend unter Verwendung von Ridders' Algorithmus als Schätzmodell zur Bestimmung der Halbwertszeit verwendet werden kann. Die Methode von Ridders ist ein Standardansatz der Numerik zur näherungsweisen Berechnung von Nullstellen und vermeidet im betrachteten Schätzmodell die Bestimmung einer inversen FP-Funktion, welche zur Schätzung der Halbwertszeit nötig wäre. Der vorgeschlagene Ansatz wird auf der Basis realer Datensätze verglichen mit einfacheren Methoden zum Umgang mit nichtlinearen Daten (ln bzw. Wurzeltransformation der Originalwerte). Für die Analysen wird Stata (Version 12) bzw. das rootSolve Paket in R verwendet.

Ergebnisse
Die Anwendung des kombinierten Ansatzes über FP und Ridders' Algorithmus führt bei allen Beispieldaten zur besten Modellgüte und zu den plausibelsten Schätzungen der jeweiligen Halbwertszeit. Teilweise unterscheiden sich die berechneten Halbwertszeiten deutlich zwischen den untersuchten Ansätzen. Der vorgeschlagene kombinierte Ansatz ermöglicht eine hohe Flexibilität hinsichtlich der zugrunde liegenden Datenstruktur.

Diskussion
FPs sind eine angemessene Methode um nichtlineare zeitabhängige Dosiskonzentrationen zu modellieren. Da im Allgemeinen die Bestimmung einer inversen FP-Funktion schwierig ist, stellt die Methode von Ridders eine hilfreiche Alternative dar, um ein Schätzmodell für Halbwertszeiten zu erhalten. Allerdings setzt die Anwendung des Ridders-Verfahren eine geeignete Wahl von Startwerten voraus. Im Rahmen einer Weiterentwicklung des vorgeschlagenen Modells kann ein Bootstrap-Ansatz dazu verwendet werden, entsprechende Konfidenzintervalle für die geschätzte Halbwertszeit zu konstruieren.

Combining gene expression measurements from different platforms using simultaneous boosting to identify prognostic markers

Mazur, Johanna Zwiener, Isabella Binder, Harald

University Medical Center Mainz, Deutschland
mazur@uni-mainz.de, isabella.zwiener@gmx.de, binderh@uni-mainz.de

Development of gene expression risk prediction signatures in a survival setting is typically severely constrained by the number of samples. An approach which analyzes several data sets simultaneously is therefore highly desirable.

However, different platforms, like RNA-Seq and microarrays, are often used to obtain gene expression information. Therefore, a pooled analysis of individual patient data to identify prognostic markers is typically not possible anymore.

Our approach for this problem is an analysis with simultaneous boosting for regularized estimation of Cox regression models. First, a componentwise likelihood-based boosting algorithm is performed for each study simultaneously, where in each step the variable selection is combined, such that the variable with the largest score statistic across studies is updated.

For assessment of simulated data, where a pooled analysis is possible, the prediction performance is compared to the prediction performance of the pooled analysis. Additionally, for simulated data we test the performance with respect to identify important genes for our simultaneous boosting approach, the pooled analysis and a setting where only gene lists are available.

Finally, we use RNA-Seq and gene expression microarray data from "The Cancer Genome Atlas" for kidney clear cell carcinoma patients to validate our strategy.

Our newly proposed simultaneous boosting approach performs close to the pooled analysis where the latter is feasible. In addition it is possible to easily combine gene expression studies from different molecular platforms to find prognostic markers for medical research.

Time matters – analyzing prenatal adverse drug reactions under non-continuous exposure

Meister, Reinhard [1] Allignol, Arthur [2] Schaefer, Christof [3]

[1] Beuth Hochschule für Technik Berlin, Deutschland;
meister@bht-berlin.de
[2] Universität Ulm, Deutschland
arthur.allignol@uni-ulm.de
[3] Charité Berlin, Deutschland
christof.schaefer@charite.de

Early prenatal development during organogenesis is a highly dynamic process that follows an extremely stable schedule. Therefore, susceptibility to toxic agents such as some maternal medication varies over time with identical periods of high vulnerability for all exposed pregnancies. Exposure time dependent specificity of developmental defects (malformations) has been described in detail for thalidomide (Contergan). When analyzing the effects of potentially embryotoxic drugs, the exposure interval has to be taken into account. Using event history methods for statistical analysis requires careful modelling of these time-dependencies, avoiding any conditioning on future events, e.g. start or stop of a medication. We will present examples using Cox-regression in order to assess the influence of the exposure interval on pregnancy outcome. Points to consider for checking validity of results will be addressed.

Design and analysis of non-replicated genetic experiments

Mejza, Stanisław

Poznań University of Life Sciences, Poland
smejza@up.poznan.pl

This presentation deals with selection problems in the early stages of a breeding program. During the improvement process, it is not possible to use an experimental design that satisfies the requirement of replicating the treatments, because of the large number of genotypes involved, the limited quantity of seed and the low availability of resources. Hence unreplicated designs are used. To control the real or potential heterogeneity of experimental units, control (check) plots are included in the trial.

There are many methods of using the information resulting from check plots. Each of them is appropriate for some specific structure of soil fertility. The problem here is that we do not know what kind of soil structure occurs in a given experiment. Hence we cannot say which of the existing methods is appropriate for a given experimental situation. The method of inference presented here is always appropriate, because a trend in soil variability is identified and estimated.

To begin with we identify the response surface characterizing the experimental environments. We assume that the observed yield results directly from two components: one due to soil fertility, the other due to genotype effects. This means that the difference between the observed and forecast yields can be treated as an estimate of genotype effects. The response surface will then be used to adjust the observations for genotypes. Finally, the adjusted data are used for inference concerning the next stages of the breeding program. Moreover, the arrangements and density of check plots play an important role in the selection of a response function; this problem is discussed also. The theoretical considerations are illustrated with an example involving spring barley.

Reference:

Mejza S., Mejza I. (2013): Check plots in field breeding experiments. Biometrical Letters, 137-149, DOI:10.2478/bile-2013-0024.

Der p-Wert – Eine elementare Herausforderung für die Lehre

Merkel, Hubert

HAWK Hildesheim/Holzminden/Göttingen, Deutschland
hubert.merkel@hawk-hhg.de

Der Umgang mit dem p-Wert,wenn nicht der p-Wert selber, wird in jüngster Zeit zum Teil fundamental kritisiert. Diese Kritik hat mittlerweile auch Eingang in die populärwissenschaftliche Diskussion gefunden (vgl. z. B. Spektrum der Wissenschaft, September 2014). Nach Darstellung und Analyse der wesentlichen Kritikpunkte werden mögliche Konsequenzen für die Lehre aufgezeigt und der Versuch unternommen, didaktische Lösungen für diese "neuen" (?) Probleme zu formulieren.

Variable selection in large scale genome wide association studies

Mészáros, Gábor [1] Sölkner, Johann [1] Waldmann, Patrik [2]

[1] University of Natural Resources and Life Sciences, Vienna, Austria
gabor.meszaros@boku.ac.at, johann.soelkner@boku.ac.at
[2] Linköping University, Linköping, Sweden
patrik.waldmann@liu.se

Genome wide association studies (GWAS) became a common strategy to identify regions of interest in human and non-human genomes, by estimating the significance and effect size of the single nucleotide polymorphisms (SNP). This data is collected using many thousands of SNP markers on a much lower number of individuals, leading to the so called p » n problem. In addition we expect only a portion of SNPs to be truly associated with the phenotype of interest, therefore the implementation of variable selection methods is necessary.

Penalized multiple regression approaches have been developed to overcome the challenges caused by the high dimensional data in GWAS. In this study we compare the statistical performance of lasso and elastic net using data sets of varying complexity. The goal of the first simulated data set was to evaluate the influence of different levels of correlation (linkage disequilibrium) on the performance of different variable selection methods. A second example was evaluating the same properties using a simulated, but biologically more complex QTLMAS data set. Finally we have used the regularization methods on a genome wide association study of milk fat content on the entire German-Austrian Fleckvieh genotype pool of 5570 individuals and 34,373 SNPs after quality control.

Cross validation was used to obtain optimal values of the regularization parameter. In the first data set the best results, were obtained with elastic net using the penalty weight of 0.1. This weight however gave some false positive results in the QTLMAS data. In the cattle data set the elastic net with penalty factor 0.1 identified major regions on four chromosomes, along with many other regions with a potential influence on influencing milk fat content. Finally, we present some results from ongoing research on implementation of Bayesian versions of the lasso and elastic-net.

Combining social contact data with spatio-temporal models for infectious diseases

Meyer, Sebastian Held, Leonhard

Universität Zürich, Schweiz
Sebastian.Meyer@ifspm.uzh.ch, Leonhard.Held@uzh.ch

The availability of geocoded health data and the inherent temporal structure of communicable diseases have led to an increased interest in statistical models for spatio-temporal epidemic data. [1] The spread of directly transmitted pathogens such as influenza and noroviruses is largely driven by human travel [2] and social contact patterns [3]. To improve predictions of the future spread, we combine an age-structured contact matrix with a spatio-temporal endemic-epidemic model for disease counts. Using age-stratified public health surveillance time series on noroviral gastroenteritis in the 12 districts of Berlin, 2011-2014, we investigate if accounting for age-specific transmission rates improves model-based predictions.

References:

[1] Meyer S, Held L, Höhle M (2014): Spatio-temporal analysis of epidemic phenomena using the R package surveillance. arXiv:1411.0416.

[2] Meyer S, Held L (2014): Power-law models for infectious disease spread. Annals of Applied Statistics, 8 (3), pp. 1612-1639.

[3] Mossong J, Hens N, Jit M, Beutels P, Auranen K, et al. (2008): Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Medicine, 5, e74.

Design and Analysis of Adaptive Confirmatory Trials using MCPMod

Mielke, Tobias

ICON PLC, Deutschland
tobias.mielke@iconplc.com

The MCPMod approach for dose-finding in the presence of model uncertainty was in 2014 the first statistical methodology to receive a positive qualification opinion of the EMA. MCPMod combines in Phase 2 clinical trials optimized trend tests with a dose-response modelling step to support dose selection for Phase 3. Alternatively, MCPMod may be applied in adaptive seamless Phase 2/3 studies with treatment arm selection. Seamless Phase 2/3 studies combine data of both stages of drug development, allowing in particular situations a quicker and more efficient drug development program. MCPMod may be utilized in these studies to decrease bias in dose selection and to increase the power of the confirmatory test. The design and analysis of seamless Phase 2/3 trials, using MCPMod and pairwise testing and dose selection procedures will be compared to the standard approach of separate Phase 2 and 3 trials based on a case study. Benefits and limitations of the proposed approach will be discussed.

Rückgekoppelte medizintechnische Systeme

Misgeld, Berno Leonhardt, Steffen

RWTH Aachen University, Deutschland
misgeld@hia.rwth-aachen.de, leonhardt@hia.rwth-aachen.de

Rückgekoppelte oder geschlossene Kreise werden heutzutage bereits in modernen medizintechnischen Systemen auf Geräteebene verwendet, um so ein automatisches Einstellen interner, technischer Größen zu ermöglichen. Bei Geräten, in denen interne Größen mit dem physiologischen System des Menschen gekoppelt, oder der Mensch sogar einen Teil der Strecke darstellt, gestaltet sich der Entwurf geeigneter Rückkopplungsalgorithmen ungleich schwieriger. In einem ersten Schritt wird daher in diesem Beitrag eine strukturelle Gruppierung von Rückkopplungen in medizintechnischen Systemen vorgenommen, die anhand ausgewählter praktischer Beispiele illustriert werden. Hierbei werden zudem Herausforderungen identifiziert, die an den Rückkopplungsalgorithmus in Form von Nichtlinearitäten, Totzeiten und parametrischen Unsicherheiten gestellt werden. In einem zweiten Schritt werden Grenzen erreichbarer Regelgüte für medizintechnische Systeme hergeleitet und anhand des Beispiels der künstlichen Blutzuckerregelung erläutert. Neben robuster Stabilität, ist bei der künstlichen Blutzuckerregelung die Garantie von robuster Güte von sicherheitskritischer Relevanz, da bei Auftreten von starker Hypo- bzw. Hyperglykämie die Gefahr von Koma bis zum Tod des Patienten droht. In einem neuen Entwurfsansatz zur Störgrößenunterdrückung werden über die Verwendung von a-priori Prozessinformation und der online Adaptation des Rückkopplungsalgorithmus die Grenzen erreichbarer Regelgüte erweitert. Im neuen Ansatz wird hierzu als optimaler Zustandsbeobachter ein erweiterter Kalman-Filter verwendet. Durch die Erweiterung des Zustandsvektors durch einen bandbreitenbegrenzten Gaußschen Prozess wird das Verfolgen ausgewählter zeitvarianter Parameter zur Adaption des Rückkopplungsalgorithmus möglich. Die Validierung erfolgt in einer in-silico Studie auf Basis von in-vivo parametrierten Modellen des Göttinger-Minischweins.

A new method for testing interaction in different kinds of block designs

Moder, Karl

Universität für Bodenkultur, H851, Österreich
karl.moder@boku.ac.ate

A new method for testing interaction in different kinds of block designs based on a linear model

Randomized complete block designs (RCBD) introduced by [1] are probably the most widely used experimental designs. Despite many advantages they suffer from one serious drawback. It is not possible to test interaction effects in ANOVA as there is only one observation for each combination of a block and factor level. Although there are some attempts to overcome this problem none of these methods are used in practice, especially as most of the underlying models are non-linear. A review on such tests is given by [2] and [3].

Here a new method is introduced which permits a test of interactions in block designs. The model for RCBDs is linear and identical to that of a two factorial design. The method as such is not restricted to simple block designs, but can also be applied to other designs like Split-Plot-design, Strip-Plot-design, ... and probably to incomplete block designs. ANOVA based on this method is very simple. Any common statistical program packages like SAS, SPSS, R ... can be used.

References:

[1] W. G. Cochran, and D. R. Cox, Experimental Designs, John Wiley & Sons, Inc, New York, London, Sidney, 1957, 2 edn.

[2] G. Karabatos, "Additivity Test," in Encyclopedia of Statistics in Behavioral Science, edited by B. S. Everitt, and D. C. Howell, Wiley, 2005, pp. 25–29.

[3] A. Alin, and S. Kurt, Stat. in Medicine 15, 63–85 (2006).

Selecting the block model

Möhring, Jens Piepho, Hans-Peter

Universität Hohenheim, Deutschland
moehring@uni-hohenheim.de, piepho@uni-hohenheim.de

Blocking is commonly used in field trials to reduce experimental error and to compare treatments more precisely. As block designs try to reduce heterogeneity of plots within blocks, the optimal block size depends on the variability of the field. The field variability of an experiment is often unknown during planning the experiment. Therefore, optimal block size for creating the optimal design is unknown. Repeatedly nested block designs, which can be created at http://www.expdesigns.co.uk/, are designs with a multiple hierarchical block structure. The design is optimized iteratively for each stage given all larger stages. A stage is characterized by adding a block structure nested within previous blocks. While the idea behind these multiple nested block effects is that the optimal block model can be chosen after performing the experiment, it is not clear how to select the best block model. We simulated field trials of different size and different true error model including spatial and non-spatial error structures. Additionally, we used nested block designs with varying number of stages and block sizes. For each design we selected the optimal design according to AIC or likelihood-ratio-tests for random effects and F-values for fixed effects. We compared different selections methods by calculating the mean square error of treatment differences between estimated and true treatment effect differences and the Pearson correlation between estimated and true treatment effects.

Efficient tests for the similarity of dose response curves

Möllenhoff, Kathrin Dette, Holger

Ruhr Universität Bochum, Deutschland
kathrin.moellenhoff@rub.de, holger.dette@rub.de

In clinical trials the observation of different populations and their reaction on medicinal drugs is of great interest. In this regard regression models (especially non-linear models) are very important. Concerning these models providing dose response information, in many cases the question occurs whether two dose response curves can be assumed as equal. This problem also appears in the situation of detecting noninferiority and/or equivalence of different treatments, shown by Liu et al.[1].

The main goal was to achieve a higher accuracy and to reduce the computational effort of confidence bands and in addition the improvement of the power of hypotheses tests measuring the similarity of dose response curves with various metrics. The currently available methods are based on the union-intersection test of Berger [2] which is not optimal and yields to procedures with extremely low power. Therefore an alternative methodology for testing the similarity of two dose response curves is developd, using asymptotic theory and a parametric bootstrap approach. It is demonstrated that the new test yields to an substantial increase in power compared to the currently available state of the art (Gsteiger et al.[3]).

Acknowledgement: The research is embedded in the IDeAl EU-FP7 project, Grant-Agreement No. 602 552

References:

[1] Liu, W., Bretz, F., Hayter, A. J. and Wynn, Henry (2009): Assessing nonsuperiority, noninferiority, or equivalence when comparing two regression models over a restricted covariate region., Biometrics, 65 (4). pp. 1279-1287. ISSN 1541-0420 725.

[2] Berger, R. L. (1982): Multiparameter hypothesis testing and acceptance sampling., Technometrics 24, 295–300.

[3] Gsteiger, S. , Bretz, F. and Liu, W.(2011): Simultaneous Confidence Bands for Nonlinear Regression Models with Application to Population Pharmacokinetic Analyses, Journal of Biopharmaceutical Statistics, 21: 4, 708–725.

Random forests for functional covariates

Möller, Annette [1] Tutz, Gerhard [2] Gertheiss, Jan [1,3]

[1] Department of Animal Sciences, University of Göttingen, Germany
annette.moeller@agr.uni-goettingen.de, jan.gertheiss@agr.uni-goettingen.de
[2] Department of Statistics, University of Munich, Germany
gerhard.tutz@stat.uni-muenchen.de
[3] Center for Statistics, University of Göttingen, Germany

Applied researchers often employ intuitive procedures to deal with functional data. A popular approach is to discretize a signal $x(t)$, $t \in J \subset \mathbb{R}$, by partitioning its domain into (disjunct) intervals and employ the mean values computed across each interval as covariates in a regression or classification model.

As the choice of the interval pattern in this approach is arbitrary to some extent, we investigate methods with a fully randomized choice for the intervals.

Based on the idea of computing mean values over intervals, we propose a special form of random forests to analyze functional data. The covariates used for the single (regression or classification) trees are the mean values over intervals partitioning the functional curves. The intervals are generated at random using exponentially distributed waiting times. The rate parameter λ of the exponential distribution determines whether the functions' domain tends to be split into just a few (small λ) or many (large λ) intervals.

The predictive performance of our functional random forests is compared to the performance of parametric functional models and a non-functional random forest using single measurements $x(t_j)$ at measurement points t_j as predictors.

We apply our method to data from Raman spectroscopy on samples of boar meat, where the objective is to utilize the spectra to predict the concentration of the hormones skatole and androstenone.

Further, we investigate data from cell chips sensors exposed to dilutions with four different concentrations of paracetamol (AAP), where the aim is to employ the signals of the sensors to predict the paracetamol concentration.

Consulting Class – Ein Praktikum für Biometrie-Studierende

Muche, Rainer [1] Dreyhaupt, Jens [1] Stadtmüller, Ulrich [2]
Lanzinger, Hartmut [2]

[1] Institut für Epidemiologie und Medizinische Biometrie, Universität Ulm, Deutschland
rainer.muche@uni-ulm.de, jens.dreyhaupt@uni-ulm.de
[2] Institut für Zahlentheorie und Wahrscheinlichkeitstheorie, Universität Ulm, Deutschland
ulrich.stadtmueller@uni-ulm.de, hartmut.lanzinger@uni-ulm.de

Die Lehrveranstaltung "Consulting Class" ist ein anwendungsbezogenes Praktikum für Studierende im Studiengang Mathematische Biometrie an der Universität Ulm (5. Semester Bachelor). Voraussetzungen sind Grundkenntnisse der Wahrscheinlichkeitsrechnung und Statistik, wie sie in den Grundvorlesungen der Stochastik vermittelt werden sowie in Epidemiologie und klinischen Studien. Hauptsächlich besteht das Praktikum aus einer (möglichst) selbstständigen Lösung und Auswertung eines praktischen medizinischen Sachproblems. Die Studierenden sollen nach Abschluss der Veranstaltung in der Lage sein, das statistische Problem zu identifizieren, mit Statistiksoftware die Analyse umzusetzen sowie die Ergebnisse zu kommunizieren (Präsentation und Bericht). Die Auswertung wird in studentischen Arbeitsgruppen (etwa 4-6 Studierende) bearbeitet und den anderen Studierenden vorgestellt.

Ein weiterer Teil von "Consulting Class" ist die Teilnahme an einer statistischen Erstberatung im Institut für Epidemiologie und Medizinische Biometrie, um die Kommunikation in einer Beratungssituation im Umgang mit Fachwissenschaftlern kennenzulernen. Die Studierenden schreiben ein Protokoll, in dem die biometrischen Aspekte der Beratung darzustellen sind. Des Weiteren sollen die Studierenden den Umgang mit Fachliteratur anhand einer Buchbesprechung über einschlägige Literatur für angewandte Statistik erlernen.

Im Beitrag werden die Überlegungen bei der Planung und Durchführung des Praktikums sowie spezielle Aspekte in Bezug auf

- Auswertung eines realen Datensatzes
- Statistische Erstberatung
- Buchbesprechung

vorgestellt und Erfahrungen diskutiert.

Literatur:

R. Muche, J. Dreyhaupt, U. Stadtmüller, H. Lanzinger: Consulting Class: Ein Praktikum für Biometrie-Studierende. In: G. Rauch, R. Muche, R. Vontheim (Hrsg.): Zeig mir Biostatistik, Springer Verlag, Berlin (2014), S. 67-79.

Non-parametric Bayesian prediction of the time to reach a given number of events, with consideration of prior knowledge on closeness to the exponential distribution

Nehmiz, Gerhard Birgit, Gaschler-Markefski

Boehringer Ingelheim Pharma GmbH & Co. KG, Deutschland
gerhard.nehmiz@boehringer-ingelheim.com,
birgit.gaschler-markefski@boehringer-ingelheim.com

Event times, possibly independently censored, are investigated in a Bayesian non-parametric manner ([1], [2]). We proceed to the related question: When will a given number of events be reached, based on the interim information that is available so far? Close examination shows that the approach of [3] considers only the K–M estimate of the survival curve and neglects any prior information about its shape. In other words, the weight of the prior information remains = 0, and the approach of [3] is not truly Bayesian. Taking up the method of [4], we investigate a broad range of weighting factors between prior information and data, and show the consequences for the predicted time to reach a given number of events, and for its uncertainty. Estimation is done by MCMC ([5]).

We investigate theoretical examples as well as a randomized Phase III trial treating patients in non-small cell lung cancer ([6]).

References:

[1] Hellmich M: Bayes'sche Untersuchung von zensierten Daten [Bayesian investigation of censored data]. Presentation at the Workshop of the Bayes WG: "Hierarchical models and MCMC methods", Homburg/Saar 2001-03-19. http://www.imbei.uni-mainz.de/bayes/TutoriumHOM.htm

[2] Laud PW, Damien P, Smith AFM: Bayesian Nonparametric and Covariate Analysis of Failure Time Data. In: Dey D, Müller P, Sinha D (eds.): "Practical Nonparametric and Semiparametric Bayesian Statistics." New York / Heidelberg: Springer Verlag 1998, 213-225

[3] Ying G-s, Heitjan DF, Chen T-T: Nonparametric prediction of event times in randomized clinical trials. Clinical Trials 2004; 1: 352-361

[4] Susarla V, Van Ryzin J: Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observations. Journal of the American Statistical Association 1976; 71: 897-902

[5] Lunn D, Jackson C, Thomas A, Best N, Spiegelhalter D: The BUGS Book. Boca Raton/FL: Chapman&Hall / CRC 2012

[6] Reck M, Kaiser R, Mellemgaard A, Douillard J-Y, Orlov S, Krzakowski M, von Pawel J, Gottfried M, Bondarenko I, Liao M, Gann C-N, Barrueco J, Gaschler-Markefski B, Novello S: Docetaxel plus nintedanib versus docetaxel plus placebo in patients with previously treated non-small-cell lung cancer (LUME-Lung 1): a phase 3, double-blind, randomised controlled trial. Lancet Oncology 2014; 15: 143-155

Bayes-Methoden für sequentielle Designs bei kleinen Fallzahlen: Eine Simulationsstudie

Neumann, Konrad Grittner, Ulrike Piper, Sophie Dirnagl, Ulrich

Charité, Universitätsmedizin Berlin, Deutschland
konrad.neumann@charite.de, ulrike.grittner@charite.de,
sophie.piper@charite.de, ulrich.dirnagl@charite.de

Tierversuche mit Nagetieren spielen in der vor- und klinischen medizinischen Forschung eine zentrale Rolle. Da die Anzahl der eingesetzten Tiere insgesamt sehr groß ist, wird insbesondere von Seiten des Tierschutzes die Forderung nach sparsameren Studiendesigns, die mit weniger Versuchstiere auskommen, erhoben.

Zur Erreichung dieses Ziels werden frequentistisch sequentielle Studiendesigns, wie z.B. das nach O'Brien & Flemings oder Pocock, vorgeschlagen. Die Möglichkeit frühzeitig die Nullhypothese ablehnen zu können, kann zu einer Reduktion der benötigten Versuchstiere führen. Gerade bei starken Effektstärken liegt die erwartete Einsparung von Versuchstieren oft über 20%.

Im Gegensatz zu den sequentiell frequentistischen Ansätzen sind Bayes-Methoden in diesem Bereich noch weitgehend unbekannt.

Hier soll eine Simulationsstudie vorgestellt werden, in der verschiedene Bayes Entscheidungsregeln für sequentielle Studiendesigns verglichen und neben ihrem frequentistischen Verhalten auf Eignung für die Praxis untersucht werden. Insbesondere wird der Einfluss des Priors diskutiert. Drei verschiedene Herangehensweisen, den Prior zu wählen, werden gegenübergestellt:

1. Nicht-informativer Prior für alle Studienphasen.
2. Nicht informativer Prior nur in der ersten Phase. Danach Adaption des Priors gesteuert durch die Daten der vorangegangenen Stufen.
3. Informativer Prior in allen Studienphasen.

Da die Stichproben bei Tierversuchen in der medizinischen Forschung in der Regel klein sind, und auch die Anzahl der Studienphasen aus logistischen Gründen nicht größer als drei sein kann, liegt allen Simulationen eine typische Gesamtfallzahl von $N = 36$ gleichmäßig verteilt auf drei Stufen und zwei Studiengruppen zugrunde.

Mit Hilfe der Simulationsergebnisse werden für typische von Fachwissenschaftlern vermuteten Verteilungen der Effektstärke die Erwartung von signifikanten Ergebnissen und der Anteil der Ergebnisse mit Effektstärke > 0 unter den signifikanten Ergebnissen errechnet.

Die Simulationen wurden mit OpenBugs zusammen mit dem R-Paket `R2WinBUGS` durchgeführt.

Exposure density sampling in clinical cohorts

Ohneberg, Kristin [1] Wolkewitz, Martin [1] Beyersmann, Jan [2]
Schumacher, Martin [1]

[1] Institute for Medical Biometry and Statistics, Medical Center – University of Freiburg, Germany
ohneberg@imbi.uni-freiburg.de, wolke@imbi.uni-freiburg.de, ms@imbi.uni-freiburg.de
[2] Institute of Statistics, Ulm University, Germany
jan.beyersmann@uni-ulm.de

Cohort sampling designs like a nested case-control or case-cohort design are an attractive alternative to a full cohort analysis, especially in the situation where the event of interest is rare and the collection of covariate information is expensive. These cohort sampling designs require much less resources, while they are sufficient to provide results comparable to the analysis of the full cohort. For nested case-control studies incidence density sampling is applied, where controls for each case are randomly selected from the individuals at risk just prior to the occurrence of a case event. Incidence density sampling hence yields a dynamic matching with respect to an observed outcome.

If interest is in the impact of a rare time-dependent exposure on the time until some specific endpoint, a dynamic matching with respect to exposure occurring over time is required. For this purpose exposure density sampling has been suggested as an efficient sampling method [1]. The resulting subcohort may save resources if exposure and outcome data are available for the full cohort but additional covariate information is required that is rather costly or time-consuming to obtain. For some simplistic scenarios, exposure density sampling has shown to yield unbiased results. Yet the analysis investigated so far considered constant hazards and resetting the entry time, the latter being reasonable for exposed individuals, but probably less adequate for the unexposed individuals. Our aim is to further examine the exposure density sampling and possible methods of analysis [2] in real and simulated data to obtain a deeper insight.

References:

[1] Wolkewitz, M., Beyersmann, J., Gastmeier, P., and Schumacher, M. (2009). Efficient risk set sampling when a time-dependent exposure is present: matching for time to exposure versus exposure density sampling. Methods Inf Med, 48:438–443.

[2] Savignoni, A., Giard, C., Tubert-Bitter, P., and De Rycke, Y. (2014). Matching methods to create paired survival data based on an exposure occurring over time: a simulation study with application to breast cancer. BMC Medical Research Methodology, 14:83.

A score-test for testing zero-inflation in Poisson regression models under the identity-link

Oliveira, Maria [1,2] Einbeck, Jochen [1]

[1] Durham University, United Kingdom;
jochen.einbeck@durham.ac.uk
[2] University of Santiago de Compostela, Spain
maria.oliveira@usc.es

Van den Broek [Biometrics 51, 1995] proposed a score test for testing zero-inflation in Poisson regression models. The test assumes that the Poisson means are related to predictors through a log-link. In some applications, for instance, when modelling radiation-induced chromosome aberrations as a function of dose in biological dosimetry, the identity link is preferred to the log-link. When testing for zero-inflation in such models, it would be adequate to do this based on the model being employed, that is, using the identity link.

In this work we attempt to adjust van den Broek's work in order to test for zero-inflation in Poisson models with identity link. This comes with some difficulties due to the special characteristics of this model. Specifically, the Poisson mean function needs to be non-negative, which requires the use of constrained ML techniques. This, in turn, renders the components of the score function which relate to the regression parameters to be non-zero, leading to a more complex test statistic than under the log-link model. We derive this test statistic, and apply it onto radiation data of the type described above. The asymptotic chi-squared distribution of this test under the null hypothesis (no zero-inflation) appears to be remain valid in certain circumstances, but not generally.

Simultaneous comparisons between treatments and measurement occasions in longitudinal trials

Pallmann, Philip

Leibniz Universität Hannover, Deutschland
pallmann@biostat.uni-hannover.de

We propose multiple tests and simultaneous confidence intervals (SCIs) for linear combinations of Gaussian means in longitudinal trials. If scientific interest focuses on a set of clinically relevant measurement occasions, then one might wish to compare multiple treatments simultaneously at several occasions, or multiple occasions simultaneously within several treatment groups.

We compare two possible modeling strategies: a joint model covering the entire span of measurement occasions (typically a linear mixed-effects model), or a combination of occasion-specific marginal models i.e., one linear model fit per occasion [3].

Upon parameter and covariance estimation using either modeling approach, we employ a variant of multiple contrast tests [2] that acknowledges correlation over time. We approximate the joint distribution of test statistics as multivariate t for which we determine suitable degrees of freedom that ensure approximate control of the family-wise type I error rate even with small sample sizes.

Our method can be extended to binary and Poisson outcomes, at least in an asymptotic fashion; the joint model is now fitted by generalized estimating equations, and the marginal models approach combines occasion-specific generalized linear models.

SCI bounds and adjusted p-values are conveniently obtained using the R package `multcomp` [2]. We illustrate the method's application with data from a clinical trial on bradykinin receptor antagonism [1].

References:

[1] J. M. Balaguer, C. Yu, J. G. Byrne, S. K. Ball, M. R. Petracek, N. J. Brown, M. Pretorius (2013) Contribution of endogeneous bradykinin to fibrinolysis, inflammation, and blood product transfusion following cardiac surgery: a randomized clinical trial. Clinical Pharmacology and Therapeutics, 93(4), 326-334.

[2] T. Hothorn, F. Bretz, P. Westfall (2008) Simultaneous inference in general parametric models. Biometrical Journal, 50(3), 346-363.

[3] C. B. Pipper, C. Ritz, H. Bisgaard (2012) A versatile method for confirmatory evaluation of the effects of a covariate in multiple models. Journal of the Royal Statistical Society, Series C: Applied Statistics, 61(2), 315-326.

Approximation Procedures for High-Dimensional Repeated Measures Designs

Pauly, Markus [1] Ellenberger, David [2] Brunner, Edgar [2]

[1] Universität Ulm, Deutschland
markus.pauly@uni-ulm.de
[2] Universitätsmedizin Göttingen, Detuschland
david.ellenberger@med.uni-goettingen.de, ebrunne1@gwdg.de

High-dimensional repeated measures designs occur, e.g., in time profiles, where the number of measurements d is larger than the number of independent subjects n. In this framework the time profile is allowed to have a factorial structure and null hypotheses of specific contrasts (as, e.g., equal time profiles) are of interests. For such settings we propose a new test for repeated measures designs and derive its limit distribution in the Bai and Saranadasa (1996) model for the situation where $\min(n, d) \to \infty$. The test is based on a quadratic form together with novel unbiased and dimension-stable estimators of different traces of the underlying unrestricted covariance structure. The asymptotic distributions of the statistic are analyzed under different conditions and it turns out that they may be normal, standardized χ^2 and even non-pivotal in some situations. This motivates the use of an additional small sample approximation technique. We discuss its theoretical properties and show in simulations that the resulting test achieves a good type-I-error control. Moreover, the application is illustrated with a practical data set from a sleep-laboratory trial (Jordan et al., 2004).

References:

1. Bai, Z. and Saranadasa, H. (1996), Effect of High Dimension:by an Example of a Two Sample Problem. *Statistica Sinica* **6**, 311–329.

2. Chen, S. X. and Qin, Y.-L. (2010). A Two-Sample Test for High-Dimensional Data with Applications to Gene-Set Testing. *The Annals of Statistics* **38**, 808–835.

3. Jordan, W., Tumani, H., Cohrs, S., Eggert, S., Rodenbeck, A., Brunner, E., Rüther, E., and Hajak, G. (2004). Prostaglandin D Synthase (β-trace) in Healthy Human Sleep. *Sleep* **27**, 867–874.

4. Pauly, M., Ellenberger, D., and Brunner, E. (2013). Analysis of High-Dimensional One Group Repeated Measures Designs, Submitted Paper.

5. Srivastava, M.S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis* **100**, 518–532.

Multiplicative interaction in network meta-analysis

Piepho, Hans-Peter [1] Madden, Laurence V. [2] Williams, Emlyn R. [3]

[1] Universität Hohenheim, Deutschland
hans-peter.piepho@uni-hohenheim.de
[2] Ohio State University, Wooster, USA
madden.1@osu.edu
[3] Australian National University, Canberra, Australia
emlyn@alphastat.net

Meta-analysis of a set of clinical trials is usually conducted using a linear predictor with additive effects representing treatments and trials. Additivity is a strong assumption. In this paper, we consider models for two or more treatments that involve multiplicative terms for interaction between treatment and trial. Multiplicative models provide information on the sensitivity of each treatment effect relative to the trial effect. In developing these models, we make use of a two-way analysis-of-variance approach to meta-analysis, and consider fixed or random trial effects. It is shown using two examples that models with multiplicative terms may fit better than purely additive models and provide insight into the nature of the trial effect. We also show how to model inconsistency using multiplicative terms.

Reference:

Piepho, H.P., Madden, L.V., Williams, E.R. (2014): Multiplicative interaction in network meta-analysis. Statistics in Medicine (appeared online)

Analysis of a complex trait with missing data on the component traits

Piepho, Hans-Peter [1] Müller, Bettina [2] Jansen, Constantin [2]

[1] Universität Hohenheim, Deutschland
hans-peter.piepho@uni-hohenheim.de
[2] Strube Research, Germany
Be.Mueller@strube-research.net, C.Jansen@strube-research.net

Many complex agronomic traits are computed as the product of component traits. For the complex trait to be assessed in a field plot, each of the component traits needs to be measured in the same plot. When data on one or several component traits are missing, the complex trait cannot be computed. If the analysis is to be performed on data for the complex trait, plots with missing data on at least one of the component traits are discarded, even though data may be available on some of the component traits. This paper considers a multivariate mixed model approach that allows making use of all available data. The key idea is to employ a logarithmic transformation of the data in order to convert a product into a sum of the component traits. The approach is illustrated using a series of sunflower breeding trials. It is demonstrated that the multivariate approach allows making use of all available information in the case of missing data, including plots that may have data only on one of the component traits.

Reference:

Piepho, H.P., Müller, B.U., Jansen, C. (2014): Analysis of a complex trait with missing data on the component traits. Communications in Biometry and Crop Science 9, 26-40.

Blinded sample size reestimation in adaptive enrichment designs with multiple nested subgroups

Placzek, Marius Schneider, Simon Friede, Tim

Universitätsmedizin Göttingen, Deutschland
marius.placzek@med.uni-goettingen.de, schneider.med.goettingen@gmail.com, tim.friede@med.uni-goettingen.de

Growing interest in personalized medicine and targeted therapies is leading to an increase in the importance of subgroup analyses. Since adaptive designs carry the promise of making drug development more efficient, enrichment designs with adaptive selection of the population (i.e. predefined subpopulation or full population) at an interim analysis have gained increased attention. In confirmatory trials under regulatory conditions it is important that the statistical analysis controls the familywise type I error rate. This can be achieved by application of adequate testing strategies, e.g. the combination test approach (Brannath et al, 2009 [1]; Jenkins et al, 2011 [2]) or the conditional error function approach (Friede et al, 2012 [3]). Due to uncertainties about the nuisance parameters which are needed for sample size calculations, a sample size review can be performed in order to make the study more robust against misspecifications (Internal Pilot Study Design, Wittes & Brittain, 1990 [4]). Here blinded methods that properly control the type I error rate are preferred by regulatory authorities. We will present a method that combines both the blinded sample size reestimation and adaptive enrichment methods. This will include considerations of when to perform the blinded review and investigations on the optimal timepoint for the interim analysis. Simulations on power and type I error rates for the blinded sample size reestimation and the combined method will be shown.

References:

[1] Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, Racine-Poon A (2009). Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. Statistics in Medicine 28:1445–1463.

[2] Jenkins M, Stone A, Jennison C (2011). An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. Pharmaceutical Statistics 10: 347–356.

[3] Friede T, Parsons N, Stallard N (2012). A conditional error function approach for subgroup selection in adaptive clinical trials. Statistics in Medicine 31: 4309–4320.

[4] Wittes J, Brittain E (1990). The role of internal pilot studies in increasing the efficiency of clinical trials. Statistics in Medicine 9: 65–72.

The average hazard ratio – a good effect measure for time-to-event endpoints when the proportional hazard assumption is violated?

Rauch, Geraldine [1] Brannath, Werner [2] Brückner, Mathias [2]
Kieser, Meinhard [1]

[1] Universität Heidelberg, Deutschland
rauch@imbi.uni-heidelberg.de, meinhard.kieser@imbi.uni-heidelberg.de
[2] Universität Bremen, Deutschland
brannath@uni-bremen.de, mwb@math.uni-bremen.de

In many clinical trial applications, the endpoint of interest corresponds to a time-to-event endpoint. In this case, group differences are usually expressed by the hazard ratio which is the standard effect measure in a time-to-event setting. The commonly applied test is the logrank-test, which is optimal under the proportional hazard assumption. However, there are many situations in which the proportional hazard assumption is violated. Especially in applications were a full population and several subgroups or a composite time-to-first-event endpoint and several components are considered, the proportional hazard assumption usually does not simultaneously hold true for all test problems under investigation.

For the case of non-proportional hazards, Kalbfleisch and Prentice (1981) proposed the so-called average hazard ratio as an alternative effect measure which can also be used to construct a corresponding test for group differences. Furthermore the average hazard ratio has a meaningful interpretation even in the case of non-proportional hazards. However, it is hardly ever used in practice, whereas the standard hazard ratio is commonly reported in clinical trials analyzing time-to-event data.

The aim of this talk is to give a systematic comparison of the two approaches for different time-to-event settings corresponding to proportional and non-proportional hazards and illustrate the pros and cons of both methods in application. In particular, we will discuss how to apply the average hazard ratio in the presence of competing risks. Comparisons will be made by Monte-Carlo simulations and by a real clinical trial example.

Reference:

[1] Kalbfleisch D., Prentice R. L. Estimation of the average hazard ratio. Biometrika 1981, 68:105-112.

Bayesian modelling of marker effects in backcross experiments by means of their covariance matrix

Reichelt, Manuela Mayer, Manfred Teuscher, Friedrich
Reinsch, Norbert

Leibniz-Institut für Nutztierbiologie, Deutschland
reichelt@fbn-dummerstorf.de, mmayer@fbn-dummerstorf.de,
teuscher@fbn-dummerstorf.de, reinsch@fbn-dummerstorf.de

QTL-mapping in experiments where two inbred lines are crossed were considered. For this type of experiments we have derived a covariance matrix for the effects of genetic markers. It is also valid for equivalent designs with double haploids. Further, it is possible to directly compute the inverse of this covariance matrix from prior knowledge on recombination frequencies in the genome, as it is represented by distances in a genetic map. Statistical models making use of this inverse covariance matrix were compared to a previously published method, were all marker effects are treated as independent. Alternative models were fitted to simulated data in several scenarios, thereby taking a Bayesian smoothing perspective. Dependent on the simulated scenario inclusion of the precision matrix resulted in improvements in terms of mean squared errors for estimated genetic values and estimated genetic contributions from individual chromosomes to the total genetic variance.

Samplingstrategien in Bayeschen Methoden zur Schätzung von Markereffekten

Reichelt, Manuela Mayer, Manfred Teuscher, Friedrich
Reinsch, Norbert

Leibniz-Institut für Nutztierbiologie, Deutschland
reichelt@fbn-dummerstorf.de, mmayer@fbn-dummerstorf.de,
teuscher@fbn-dummerstorf.de, reinsch@fbn-dummerstorf.de

Quantitative Merkmale und deren Lokalisierung im Genom (QTL) sind in der Tierzucht von zentraler Bedeutung. Verfahren zu QTL-Kartierung (Gianola [1]) nutzen häufig den Gibbs-Sampler um QTL-Regionen zu identifizieren. Bei Einzelkomponenten-Samplern werden die genetischen Effekte der Marker nacheinander einzeln aus einer Verteilung gezogen. In Simulationsexperimenten zeigte sich, dass eine sehr hohe Anzahl an Iterationen nötig ist, um zur Konvergenz zu gelangen.

Um speziell diesen Nachteil auszugleichen, verwendeten wir im Gegensatz zu Einzelkomponenten-Samplern eine blockweise Vorgehensweise (Rue [2]). Hierbei werden die zu schätzenden genetischen Markereffekte chromosomenweise (also im Block) mit Hilfe von Inversenbildung und Choleskyzerlegung gezogen. Rechenaufwand und Speicherplatzbedarf sind zwar höher, es werden jedoch wesentlich weniger Iterationen benötigt.

Wir simulierten Datensätze für das Versuchsdesign der Rückkreuzung und wandelten verschiedene Einzelkomponenten-Modelle in ein Block-Modell um. Als erstes wurde die Methode von Xu [3] (Marker werden unabhängig voneinander behandelt) als Einzelkomponenten- und Block-Sampler gegenüber gestellt. Darüber hinaus wurde eine von uns neu entwickelte Methode, die eine Korrelationsmatrix der Markereffekte einbezieht, als Einzelkomponenten- und Block-Sampler verglichen. Insbesondere Markereffekte auf Chromosomen ohne QTL fallen beim Blocksampling deutlich variabler aus als beim Sampeln einzelner Komponenten, selbst bei einer sehr hohen Anzahl von Iterationen. Die langsame Konvergenz der Einzelkomponenten-Sampler wird also durch Block-Sampling umgangen und die Iterationslänge erheblich gekürzt.

References:

[1] Gianola, D., 2013: Priors in Whole-Genome Regression: The Bayesian Alphabet Returns, Genetics Society of America, 194: 573-596

[2] Rue, H., Held, L. ,2005: Gaussian Markov Random Fields: Theory and Applications, Chapman & Hall/CRC

[3] Xu, S., 2003: Estimating Polygenic Effects Using Markers of the Entire Genome, Genetics Society of America, 163: 789-801

Sequentielle Tests bei Monitoringverfahren zum Nachweis erhöhter Inzidenz – eine Simulationsstudie

Reinders, Tammo Konstantin [1] Timmer, Antje [2] Kieschke, Joachim [3]
Jürgens, Verena [2]

[1] Carl-von-Ossietzky-Universität Oldenburg, Deutschland
tammo.konstantin.reinders@uni-oldenburg.de
[2] Department für Versorgungsforschung,Carl von Ossietzky Universität Oldenburg, Deutschland
antje.timmer@uni-oldenburg.de,verena.juergens@uni-oldenburg.de
[3] Epidemiologisches Krebsregister Niedersachsen, Registerstelle, Oldenburg
kieschke@offis-care.de

Einleitung
Im Rahmen eines Pilotprojektes wurde ein prospektives Monitoringverfahren für das Land Niedersachsen eingeführt, um diagnosespezifische Fallzahlen zu überwachen und gegebenenfalls erhöhte Inzidenzen zu melden. In dieser Arbeit wurden die statistischen Werkzeuge, insbesondere der Sequential Probability Ratio Test (SPRT), in einer Simulationsstudie einer kritischen Prüfung unterzogen. Von besonderem Interesse waren die Anzahl falsch positiver und falsch negativer Meldungen, die Beobachtungsdauer und die Heterogenität der Gemeindegröße.

Material und Methoden
Auf Grundlage der realen Bevölkerungsstruktur Niedersachsens wurden die durchschnittlichen Fallzahlen dreier Diagnosen (Mesotheliom, Nierenzellkarzinom und akute myeloische Leukämie) im Zeitraum 2008–2012 zufällig auf die Gemeinden verteilt. Im nächsten Schritt wurde ein zweistufiges Monitoringverfahren[1], bestehend aus Standardized Incidence Ratio (SIR) und SPRT, angewandt. Dabei wurden verschiedene Szenarien, die sich durch die Parameterwahl unterscheiden, u.a. fester vs. flexibler (d.h. in der Suchphase durch den SIR geschätzter) Testparameter, erstellt und verglichen.

Ergebnisse
Unterschiede in der Performance waren vor allem in Bezug auf die Wahl des Testparameters festzustellen. Eine feste Wahl führte zu längerer Beobachtungsdauer bei Gemeinden mit geringer Population. Die flexible Variante führte bei Diagnosen mit geringer Fallzahl zu kürzerer, bei Diagnosen mit hoher Fallzahl zu längerer Beobachtungsdauer.

Diskussion
Die Methoden unterscheiden sich hinsichtlich der Beobachtungsdauer und der Zahl der falsch positiven Meldungen. Ergebnisse zu Simulationen mit künstlichen Clustern (und damit zu falsch negativen Meldungen) stehen noch aus. Die Heterogenität der Gemeindegrößen stellt ein Problem dar, welches durch eine flexible Wahl des Testparameters gemindert werden kann. Ein denkbarer Lösungsansatz hierfür ist die Unterteilung großer Gemeinden und Städte.

References:

[1] J Kieschke und M Hoopmann. Aktives Monitoring kleinräumiger Krebshäufungen. Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz, 57 (1) : 33–40, 2014.

Dose–response studies: simulation study for the comparison of four methods under design considerations

Rekowski, Jan [1] Bornkamp, Björn [2] Ickstadt, Katja [3]
Köllmann, Claudia [3] Scherag, André [4]

[1] Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University of Duisburg-Essen, Germany jan.rekowski@uk-essen.de
[2] Statistical Methodology, Novartis Pharma AG, Basel, Switzerland bjoern.bornkamp@novartis.com
[3] Faculty of Statistics, TU Dortmund, Germany ickstadt@statistik.uni-dortmund.de, koellmann@statistik.tu-dortmund.de
[4] Center for Sepsis Control and Care (CSCC), University Hospital Jena, Germany andre.scherag@med.uni-jena.de

Key aims of phase II trials are the establishment of the efficacy of the treatment under investigation, recommendation of one or more doses for a subsequent phase III trial, and gaining information on the general shape of the dose–response curve. Traditionally, analysis of these dose–response curves has been performed using either multiple comparison procedures, or parametric modeling. More recently, numerous more elaborate methods have become available and guidance is needed to select the most appropriate approach given a specific study design. We compare and discuss Dunnett's test [1] in an analysis of variance setting, MCP-Mod [2], Bayesian nonparametric monotone regression [3], and Penalized unimodal spline regression [4]. We set up a simulation study to evaluate the performance of these approaches under four design scenarios that vary by sample size, number of doses, and allocation strategy. To account for uncertainties regarding the underlying dose–response model, we assume five different dose–response profiles. We assess performance of the methods using the criteria used in [5] that include empirical power for the efficacy establishment and the target dose estimation as well as metrics to evaluate the accuracy of the target dose estimator and the fitted dose–response curve. Furthermore, we illustrate the investigated methods with an example from clinical practice. Our major conclusion is that all investigated modern approaches outperform Dunnett's test, but that method performance varies considerably by design scenario and underlying dose–response model.

References:

[1] Dunnett (1955). A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association, 50(272), 1096-1121.

[2] Bretz, Pinheiro, and Branson (2005). Combining multiple comparisons and modeling techniques in dose–response studies. Biometrics, 61(3), 738-748.

[3] Bornkamp and Ickstadt (2009). Bayesian nonparametric estimation of continuous monotone functions with applications to dose–response analysis. Biometrics 65(1), 198-205.

[4] Köllmann, Bornkamp, and Ickstadt (2014), Unimodal regression using Bernstein–Schoenberg splines and penalties. Biometrics. DOI:10.1111/biom.12193

[5] Bornkamp et al. (2007). Innovative approaches for designing and analyzing

adaptive dose-ranging trials. Journal of Biopharmaceutical Statistics, 17(6), 965-995.

Punktewolken und Streuungsellipsen als graphische Hilfsmittel bei der biometrischen Beratung

Rempe, Udo

Zoologisches Institut, Uni Kiel, Deutschland
rempe-udo@t-online.de

Anspruchsvollere multivariate Verfahren können in der Regel im einführenden Biometrieunterricht für Biologen oder Mediziner nicht oder nur unzureichend behandelt werden. Sobald Bachelor- oder Masterarbeiten oder Dissertationen anzufertigen sind, verlangt die Beantwortung der interessierenden biologischen Fragen häufig den Einsatz multivariater Verfahren. Dann ist es erforderlich, dass die biometrisch zu Betreuenden auch verstehen, was bei den Berechnungen geschieht, damit sie die Ergebnisse interpretieren können und Grenzen der Verfahren erkennen. Sehr hilfreich sind dabei Punktediagramme. Beispielsweise kann man "Punktewolken innerhalb" erzeugen, indem man die Punktewolken verschiedener Gruppen farblich kennzeichnet und so parallel verschiebt, dass alle Mittelpunkte der Punktewolken zur Deckung kommen. Solche Punktewolken kann man drehen stauchen und dehnen. Als Zwischenschritt lässt sich die Achsendrehung um 45° einschieben. Zusätzlich zu einer ursprünglich vorhandenen x- und y-Achenskala erhält man dann ein neues rechtwinkeliges Achsensystem $(y - x)$ und $(x + y)$. Bei jeder der vier Achsen kann man im Abstand von einer Standardabweichung links und rechts vom gemeinsamen Mittelpunkt Senkrechte errichten. So bekommt man ein Achteck als Zwischenstufe für die nicht parametrische Herleitung von Streuungsellipsen und Covariantzmatrizen. Am Beispiel von Zahnkronen, deren Längen x, Breiten y und Höhen z man durch Quader darstellen kann, kann man besonders gut die Verbindung zwischen abstrakten Punkten und realen Objekten demonstrieren. Der zu jedem Zahn gehörige Quader könnte durch ein Modell dargestellt werden, bei dem die Kanten durch Drähte veranschaulicht werden. Die Drahtmodelle aller Zähne können zum Vergleich so zusammengestellt werden, dass ihr linker unterer und vorderer Eckpunkt sich deckt und alle Seiten parallel angeordnet sind. Dann werden die Unterschiede voll durch die rechten, oberen und hinteren Eckpunkte dargestellt, die eine Punktewolke sind. Sofern man beispielsweise die Punktewolke des ersten menschlichen Mahlzahns mit jener des ersten Mahlzahns von Rhesusaffen vergleicht, erfüllt die Punktewolke des größeren Menschen einen viel größeren Raum als jene des Rhesusaffen. Transformation der Punktewolken durch Logarithmieren $X = \ln(x)$, $Y = \ln(y)$ und $Z = \ln(z)$ führt dazu, dass sich die Formen der Punktewolken einander angleichen. $X + Y$ entspricht dem Logarithmus der Mahlfläche; $Y - X$ ist der Logarithmus der relativen Zahnbreite usw. Viele Matrizenoperationen lassen sich über Punktediagramme veranschaulichen.

Statistical analysis of toxicogenomics data

Rempel, Eugen [1] Hengstler, Jan [2] Rahnenführer, Jörg [1]

[1] TU Dortmund, Deutschland
rempel@statistik.tu-dortmund.de, rahnenfuehrer@statistik.tu-dortmund.de
[2] Leibniz-Institut für Arbeitsforschung an der TU Dortmund, Deutschland
hengstler@ifado.de

Analysis of the in-vitro reaction of embryonic human stem cells (eHSC) to toxicological stress allows us to reduce the laborious in-vivo animal tests. To recapitulate early neuronal development two embryonic stem cell systems were investigated. Whereas the first studies addressed the question of general applicability of transciptome analyses (published in [1]), the follow-up explorations were devoted to the issues of discrimination of compounds, transcriptome changes due to concentration increase, and their reversibility after different exposure periods. For these purposes a pipeline for a computational downstream analysis of high-dimensional gene expression data was established: from descriptive analysis to batch correction to detection of differentially expressed genes (DEG) and to enrichment analysis.

Another important issue is the discrimination of various types of samples based on genomic data. To predict the type of compound with which eHSC were treated an SVM classifier was built. The good generalization properties were confirmed within a cross-validation testing procedure. The results of this project are about to be submitted.

The reaction of eHSC to increasing concentrations of toxicants represents a further goal within toxicogenomics. Important questions are to define at which concentration the compound becomes detectable and which biological signatures emerge at specific dose levels. The results of this study were published in [2].

References:

[1] Krug, Anne K., et al. Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach. Archives of toxicology 87(1): 123-143, 2013.

[2] Waldmann, Tanja, et al. Design principles of concentration-dependent transcriptome deviations in drug-exposed differentiating stem cells. Chemical research in toxicology 27(3): 408-420, 2014.

Heterogeneity in random-effects meta-analysis

Röver, Christian [1] Neuenschwander, Beat [2] Wandel, Simon [2]
Friede, Tim [1]

[1] University Medical Center Göttingen, Germany
christian.roever@med.uni-goettingen.de, tim.friede@med.uni-goettingen.de
[2] Novartis Pharma AG, Basel, Switzerland
beat.neuenschwander@novartis.com, simon.wandel@novartis.com

The between-study heterogeneity plays a central role in random-effects meta-analysis. Especially when the analysis is based on few studies, which is a common problem not only for rare diseases, external a-priori information on heterogeneity may be helpful. In case of little information, the use of plausible weakly informative priors is recommended. Computational simplifications (using the `bmeta` R package) helped to speed up computations for Bayesian standard random-effects meta-analysis to explore the frequentist properties of Bayesian estimators for different priors. We investigated a range of scenarios (heterogeneities, numbers of studies), to compare bias, MSE and coverage of the Bayesian and classical estimators.

Modelling biomarker distributions in meta-analysis of diagnostic test accuracy studies

Rücker, Gerta Schumacher, Martin

Universitätsklinikum Freiburg, Deutschland
ruecker@imbi.uni-freiburg.de, ms@imbi.uni-freiburg.de

Background: In meta-analysis of diagnostic test accuracy (DTA) studies, it is often assumed that each study reports only one pair of specificity (Sp) and sensitivity (Se). The bivariate model [1] considers the joint distribution of Sp and Se across studies. However, in primary studies (Sp, Se) pairs are often reported for two or more cut-offs, and the cut-off values are reported as well.

Objective: To use this additional information for modelling the distributions of the underlying biomarker for diseased and non-diseased individuals.

Methods: We assume that for each DTA study in a meta-analysis, a number of cut-offs is reported, each with corresponding estimates of Sp and Se. These provide direct information about the empirical cumulative distribution function (ecdf) for both groups of individuals. A hierarchical (parametric or non-parametric) model for the distribution can be fitted, including study as a fixed or random factor. The model leads to average ecdfs for both groups of individuals. As the difference between these is the Youden index as a function of the cut-off, an optimal cut-off can be found by maximising this function. A summary ROC curve is estimated based on the distributions.

Result: The approach is demonstrated on a meta-analysis of procalcitonin as a marker for sepsis [2].

Conclusion: If there are at least two cut-offs with (Sp, Se) given per study, we can determine an optimal cut-off and estimate a summary ROC curve based on all available information from the primary studies.

References:

[1] Reitsma JB, Glas AF, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J. Clin. Epidemiol. 58(10):982-990.

[2] Wacker C, Prkno A, Brunkhorst FM, Schlattmann P (2013). Procalcitonin as a diagnostic marker for sepsis: a systematic review and meta-analysis. Lancet Infect. Dis. 13(5):426-435.

Bayesian Outbreak Detection in the Presence of Reporting Delays

Salmon, Maëlle [1] Schumacher, Dirk [1] Stark, Klaus [1]
Höhle, Michael [2]

[1] Robert Koch Institute, Berlin, Deutschland;
salmonm@rki.de, schumacherd@rki.de, starkk@rki.de
[2] Stockholm University, Stockholm, Sweden
hoehle@math.su.se

Nowadays, health institutions such as hospitals and public health authorities collect and store large amounts of information about infectious disease occurrences. The collected individual case reports can be aggregated into time series of counts which are then analyzed by statistical methods in order to detect aberrations, since unexpectedly high counts can indicate an emerging outbreak. If detected early enough, such an outbreak may be controlled. However, inherent reporting delays of surveillance systems make the considered time series incomplete.

In the presented work, we synthesize the outbreak detection algorithms of Noufaily et al. (2013) and Manitz and Höhle (2013) while additionally addressing the right-truncation of the disease counts caused by reporting delays as in, e.g., Höhle and An der Heiden (2014). We do so by considering the resulting time series as an incomplete two-way contingency table which we model using negative binomial regression. Our approach is defined in a Bayesian setting allowing a direct inclusion of both estimation and prediction uncertainties in the derivation of whether an observed case count is to be considered an aberration. Altogether, our method aims at allowing timely aberration detection and hence underlines the need of statistical modelling to address complications of reporting systems.

The developed methods are illustrated using both simulated data and actual time series of Salmonella Newport cases in Germany 2002–2013 from the German national surveillance system SurvNet@RKI.

References:

Noufaily A et al. (2013). An improved algorithm for outbreak detection in multiple surveillance systems. Statistics in Medicine, 32(7), pp. 1206–1222.

Manitz J, Höhle M (2013): Bayesian model algorithm for monitoring reported cases of campylobacteriosis in Germany. Biom. J. 55 (4): 509–526.

Höhle M, an der Heiden M (2014), Bayesian Nowcasting during the STEC O104:H4 Outbreak in Germany, 2011 (2014), Biometrics. E-pub ahead of print.

Interaction of treatment with a continuous variable: simulation study of significance level and power for several methods of analysis

Sauerbrei, Wil [1] Royston, Patrick [2]

[1] Universitätsklinikum Freiburg, Deutschland
wfs@imbi.uni-freiburg.de
[2] MRC Clinical Trials Unit, London, United Kingdom
j.royston@ucl.ac.uk

Interactions between treatments and covariates in RCTs are a key topic. Standard methods for modeling treatment–covariate interactions with continuous covariates are categorization (considering subgroups) or linear functions.

Spline based methods and multivariable fractional polynomial interactions (MFPI) have been proposed as an alternative which uses full information of the data. Four variants of MFPI, allowing varying flexibility in functional form, were suggested.

In order to work toward guidance strategies we have conducted a large simulation study to investigate significance level and power of the MFPI approaches, versions based on categorization and on cubic regression splines. We believe that the results provide sufficient evidence to recommend MFPI as a suitable approach to investigate interactions of treatment with a continuous variable. If subject-matter knowledge gives good arguments for a non-monotone treatment effect function, we propose to use a second-degree fractional polynomial (FP2) approach, but otherwise a first-degree fractional polynomial (FP1) function with added flexibility (FLEX3) has a power advantage and therefore is the method of choice. The FP1 class includes the linear function and the selected functions are simple, understandable and transferable.

References:

Royston P., Sauerbrei W. (2013): Interaction of treatment with a continuous variable: simulation study of significance level for several methods of analysis. Statistics in Medicine, 32(22):3788-3803.

Royston P., Sauerbrei W. (2014): Interaction of treatment with a continuous variable: simulation study of power for several methods of analysis. Statistics in Medicine, DOI: 10.1002/sim.6308

Assessing noninferiority in meta-analyses of clinical trials with binary outcomes

Saure, Daniel Jensen, Katrin Kieser, Meinhard

Meinhard Institut für Medizinische Biometrie und Informatik, Heidelberg, Deutschland
saure@imbi.uni-heidelberg.de, jensen@imbi.uni-heidelberg.de,
kieser@imbi.uni-heidelberg.de

Meta-analysis is a quantitative tool for aggregating data from several trials resulting in an overall effect estimate and its related confidence interval. Usually, the aim is to demonstrate superiority. Nevertheless, in some clinical situations noninferiority is to be assessed within meta-analyses, for example when pooling safety data.

In the present contribution we focus on the choice of the test statistic when extending noninferiority from a single to several trials in a meta-analysis. For a single noninferiority trial, well-known test statistics were proposed by Blackwelder [1] and Farrington and Manning [2]. The respective test statistics differ in estimating the null variance of the effect measure chosen as the risk difference, risk ratio, or odds ratio. While the Blackwelder approach [1] incorporates the unrestricted maximum likelihood estimator for the event rates in both groups, Farrington and Manning [2] proposed to use the restricted maximum likelihood estimator instead. The latter estimator is recommended for most practical scenarios [3].

In a simulation study, we investigated both the Blackwelder and the Farrington-Manning noninferiority test adapted to fixed effect and random effects meta-analysis with respect to type I error rate and power. Various scenarios concerning theoretical event rate in the control group, noninferiority margin, number of trials, number of patients per trial and allocation ratio have been considered. The simulation results are compared to the established conclusions for single trials.

References:

[1] Blackwelder WC (1982). Proving the null hypothesis in clinical trials. Controlled Clinical Trials, 3:345-353.

[2] Farrington CP, Manning G (1990). Test Statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero difference or non-unity relative risk. Statistics in Medicine, 9: 1447-1454.

[3] Roebruck P, Kühn A (1995). Comparison of tests and sample size formulae for proving therapeutic equivalence based on the difference of binomial probabilities. Statistics in Medicine, 14: 1583-1594.

Structured Benefit-risk assessment: A review of key publications and initiatives on frameworks and methodologies

Schacht, Alexander

Lilly Deutschland, Germany
schacht_alexander@lilly.com

On behalf of the EFSPI BR SIG with contributions from Shahrul Mt-Isa, Mario Ouwens, Veronique Robert, Martin Gebel, and Ian Hirsch.

Introduction

The conduct of structured benefit-risk assessment (BRA) of pharmaceutical products is a key area of interest for regulatory agencies and the pharmaceutical industry. However, the acceptance of a standardized approach and implementation are slow. Statisticians play major roles in these organizations, and have a great opportunity to be involved and drive the shaping of future BRA.

Method

We performed a literature search of recent reviews and initiatives assessing BRA methodologies, and grouped them to assist those new to BRA in learning, understanding, and choosing methodologies. We summarized the key points and discussed the impact of this emerging field on various stakeholders, particularly statisticians in the pharmaceutical industry.

Results

We provide introductory, essential, special interest, and further information and initiatives materials that direct readers to the most relevant materials, which were published between 2000 and 2013. Based on recommendations in these materials we supply a toolkit of advocated BRA methodologies.

Discussion

Despite initiatives promoting these methodologies, there are still barriers, one of which being the lack of a consensus on the most appropriate methodologies. Further work is needed to convince various stakeholders. But this opens up opportunities, for statisticians in the pharmaceutical industry especially, to champion appropriate BRA methodology use throughout the pharmaceutical product lifecycle.

Conclusions

This article may serve as a starting point for discussions and to reach a mutual consensus for methodology selection in a particular situation. Regulators and pharmaceutical industry should continue to collaborate to develop and take forward BRA methodologies, ensuring proper communication and mutual understanding.

Integrative analysis of histone ChIP-seq, RNA-seq and copy number data using Bayesian models

Schäfer, Martin Schwender, Holger

Heinrich-Heine-Universität Düsseldorf, Deutschland
m.schaefer@uni-duesseldorf.de, holger.schwender@udo.edu

Besides DNA copy number, epigenetic variations such as methylation and acethylation of histones have been investigated as one source of impact on gene transcription. A key challenge are the small sample sizes common in data arising from next-generation sequencing studies. In such situations, while several approaches have been proposed for analysis of single inputs and some for analysis of two inputs, there are still only very few statistical methods to analyse more than two genomic variations in an integrative, model-based way.

In this contribution, we propose an approach to find genes that display consistent alterations in gene transcription, histone modification, and copy number in a collective of individuals presenting a condition of interest (e.g., cancer patients) when compared to a reference group. We define a coefficient Z inspired by previous work [1,3] that allows to quantify the degree to which transcripts present consistent alterations in the three variates.

To perform inference on the gene level, information is borrowed from functionally related genes. This process is guided by a published genomewide functional network [2]. An hierarchical model relates the values of the Z coefficient and the network information, where the latter is represented by a spatially structured term given an intrinsic Gaussian conditionally autoregressive (CAR) prior.

Both the ranking implied by values of the Z coefficient and the inference based on the model lead to a plausible prioritizing of cancer genes when analyzing gene transcription, histone modification and copy number measurements from a data set consisting of a prostate cancer cell line and normal primary prostate cells.

References:

[1] Klein, H.-U., et al. (2014): Integrative analysis of histone ChIP-seq and transcription data using Bayesian mixture models. Bioinformatics 30(8): 1154-1162.

[2] Lee, I., et al. (2011): Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Research 21: 1109–1121.

[3] Schäfer, M., et al. (2009): Integrated analysis of copy number alterations and gene expression: a bivariate assessment of equally directed abnormalities. Bioinformatics 25(24): 3228-3235.

Schichtungseffekte bei Neymann-Allokation in den Stichproben nach § 42 Risikostrukturausgleichsverordnung (RSAV)

Schäfer, Thomas

Westfälische Hochschule, Deutschland
thomas.schaefer@w-hs.de

Nach § 42 RSAV sind die Datenmeldungen für den morbiditätsbezogenen Risikostrukturausgleich mindestens alle zwei Jahre auf ihre Richtigkeit zu prüfen. Die Prüfungen umfassen einerseits die allgemeine und spezielle Anspruchsberechtigung im Ausgleichsjahr und andererseits die Korrektheit der Zuweisung zu den Risikogruppen des versichertenbezogenen Klassifikationssystems im Vorjahr. Die gemeldeten Daten werden anhand von Stichproben geprüft, für jeden Stichprobenversicherten wird ein Korrekturbetrag (KB) ermittelt und die Summe der KBe anschließend auf die Grundgesamtheit (GG) der Versicherten der geprüften Kasse hochgerechnet, welche diesen Betrag dem Gesundheitsfonds zurückerstatten muss.

Die Planung der Stichproben erfolgt im Spannungsfeld zwischen den Krankenkassen, den Prüfdiensten (PD) und dem Bundesversicherungsamt. Der Stichprobenumfang muss bei vorgegebener Genauigkeit für die Schätzung vom KB so klein wie möglich sein, da die Kapazitäten der Prüfdienste sonst nicht ausreichen. Der Anteil der fehlerbehafteten Datensätze ist sehr klein (in der Pilotuntersuchung lag er unter 1%), so dass der KB als Merkmal der Versicherten eine extrem schiefe Verteilung aufweist und für einfache Zufallsstichproben Stichprobenumfänge erforderlich werden, die das Verfahren sprengen würden. § 42 RSAV ist daher nur dann umsetzbar, wenn es gelingt, den Stichprobenumfang durch optimale Schichtung radikal zu reduzieren.

Zur Beschaffung der erforderlichen Planungsunterlagen wurde im Rahmen einer Piloterhebung aus den Daten von zehn sich nach Größe, Kassenart und Lage unterscheidenden Kassen je eine einfache Zufallsstichprobe gezogen. Dabei hat sich die versichertenbezogene Zuweisung aus dem Gesundheitsfonds als eine ausgezeichnete Schichtungsvariable erwiesen.

Aus den gepoolten Daten der Piloterhebung wurden geschätzt bzw. berechnet:

a) Die Gesamtvarianz und die Schichtvarianzen vom KB;

b) Der für eine vorgegebene Genauigkeit benötigte Stichprobenumfang zur Schätzung vom KB b1) aus einer einfachen Zufallsstichprobe (n) und b2) aus einer nach der Zuweisungssumme (mit Neymann-Allokation) geschichteten Stichprobe (n').

Der Schichtungseffekt der geschichteten gegenüber der einfachen Zufallsstichprobe wurde berechnet als $n'/n - 1$ und lag in der Pilotuntersuchung bei -97%.

Referenz:

Schäfer (2013): Stichproben nach § 42 RSAV,
http://www.bundesversicherungsamt.de/fileadmin/redaktion/
Risikostrukturausgleich/Gutachten_Stichprobe_42_RSAV.pdf

Simultaneous statistical inference for epigenetic data

Schildknecht, Konstantin [1] Olek, Sven [2] Dickhaus, Thorsten [1]

[1] Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Deutschland
schildkn@wias-berlin.de, thorsten.dickhaus@wias-berlin.de
[2] Ivana Türbachova Laboratory for Epigenetics, Epiontis GmbH, Berlin, Germany
sven.olek@epiontis.com

Epigenetic research leads to complex data structures. Since parametric model assumptions for the distribution of epigenetic data are hard to verify, we introduce in the present work a nonparametric statistical framework for two-group comparisons. Furthermore, epigenetic analyses are often performed at various genetic loci simultaneously. Hence, in order to be able to draw valid conclusions for specific loci, an appropriate multiple testing correction is necessary. Finally, with technologies available for the simultaneous assessment of many interrelated biological parameters (such as gene arrays), statistical approaches also need to deal with a possibly unknown dependency structure in the data. Our statistical approach to the nonparametric comparison of two samples with independent multivariate observables is based on recently developed multivariate multiple permutation tests, see [1]. We adapt their theory in order to cope with families of hypotheses regarding relative effects, in the sense of [2].

Our results indicate that the multivariate multiple permutation test keeps the pre-assigned type I error level for the global null hypothesis. In combination with the closure principle, the family-wise error rate for the simultaneous test of the corresponding locus/parameter-specific null hypotheses can be controlled. In applications we demonstrate that group differences in epigenetic data can be detected reliably with our methodology.

References:

[1] Chung EY, Romano JP (2013) Multivariate and multiple permutation tests. Technical report No. 2013-05, Stanford University.

[2] Brunner E, Munzel U (2013) Nichtparametrische Datenanalyse. Unverbundene Stichproben. Heidelberg: Springer Spektrum, 2nd updated and corrected edition.

A Linked Optimization Criterion for the Assessment of Selection and Chronological Bias in Randomized Clinical Trials

Schindler, David Hilgers, Ralf-Dieter

RWTH Aaachen, Deutschland
dv.schindler@googlemail.com, rhilgers@ukaachen.de

Selection and chronological bias pose pressing problems in clinical trials due to their distorting effect on the estimation of the treatment effect. The scope of the presentation is to recommend for a two-armed clinical trial with parallel group design a randomization procedure which is neither highly susceptible for selection bias nor to chronological bias. As linked optimisation criterion an adjustable desirability function is introduced.

The type of selection bias investigated in the presentation arises due to unsuccessful masking of previous treatment assignments in a randomized clinical trial. Chronological bias is based on changes in patients' characteristics over time. Several criteria for assessing the extent of selection bias or the chronological bias in the case of a continuous response variable are used to describe the behaviour of individual randomization sequences of a given randomization procedure. The established criteria use different scales, hence adjustable desirability functions are used to map all the criteria to $[0, 1]$. Finally, the transformed values are summarized using the (weighted) geometric mean to assess the overall performance of an individual randomization sequence. The desirability functions thus serve as linked optimization criterion. Finally, all sequences generated by one randomization procedure are weighted with their probability of appearance and the results are illustrated in a histogram. This approach is used to compare randomization procedures according to their behaviour to selection bias and chronological bias. Special emphasis is placed on the performance of randomization procedures for trials with a small number of (available) patients.

Acknowledgement: The research is embedded in the IDeAl EU-FP7 project, Grant-Agreement No. 602 552

Sample size planning for recurrent event analyses in heart failure trials – A simulation study

Schlömer, Patrick Fritsch, Arno

Bayer Pharma AG, Deutschland
patrick.schloemer@bayer.com, arno.fritsch@bayer.com

In the past, the standard (composite) primary endpoint in heart failure (HF) trials was the time to either HF hospitalization or cardiovascular death, whichever occurs first. With improved treatments on the market HF has now changed from a short-term and quickly fatal condition to a chronic disease, characterized by recurrent HF hospitalizations and high mortality. Therefore, there is interest to move from the standard 'time-to-first' event to recurrent events as the primary efficacy endpoint. Through this, one hopes to better characterize and quantify the full disease burden of HF because recurrent hospitalizations have a substantial impact on both the patients' well-being and the health systems. Moreover, one expects to achieve practical gains by means of sample size savings due to incorporating more statistical information.

As the literature on sample size planning for recurrent event analyses is limited and does e.g. not capture the specific feature of combining recurrent HF hospitalizations and the terminal event cardiovascular death to one endpoint, we conducted an extensive simulation study to quantify potential power gains and sample sizes savings by using recurrent events instead of standard 'time-to-first' event methods. By means of a clinical trial example we illustrate how to identify reasonable planning assumptions and investigate the influence of different factors such as treatment discontinuation after hospitalization. Finally, we examine the advantages and disadvantages of using recurrent event methods, also with respect to practical implications.

Statistical modeling of high dimensional longitudinal methylation profiles in leukemia patients under DNA demethylating therapy.

Schlosser, Pascal [1,2] Blagitko-Dorfs, Nadja [2] Lübbert, Michael [2]
Schumacher, Martin [1]

[1] Center for Medical Biometry and Medical Informatics, Institute for Medical Biometry and Statistics, Medical Center University of Freiburg, Germany
schlosser@imbi.uni-freiburg.de,ms@imbi.uni-freiburg.de
[2] Department of Hematology, University of Freiburg Medical Center, Freiburg, Germany
schlosser@imbi.uni-freiburg.de, adja.blagitko-dorfs@uniklinik-freiburg.de, michael.luebbert@uniklinik-freiburg.de

In acute myeloid leukemia the mechanism of action of hypomethylating agents is still unclear and we set out to identify the DNA target regions of these. During treatment, cell type specific genome-wide DNA methylation profiles are obtained using Infinium Human Methylation 450 BeadChip arrays, which cover more than 485.000 cytosine–phosphate–guanine sites (CpGs). We analyze up to four primary blood samples per treatment course. This results in a high-dimensional longitudinal data structure with correlation appearing intrapatient- and inter CpG-wise.

We first identify a response scheme of each individual locus patient-wise as the maximal alteration statistic, which is defined as the maximal deviation from pre treatment and approximately normally distributed. We compare this approach in the context of sparse sequential measurements to more complex approaches, which model the methylation level by predesigned nonlinear functions or general fractional polynomials of degree two in mixed effects models.

In a second stage, we explain the dependence structure between these statistics by incorporating exogenous information. We adapt a procedure proposed for standard differential methylation by Kuan and Chiang [1]. By this we model the maximal alteration statistics from each locus as observations of a non-homogeneous hidden Markov process with hidden states indicating loci responding to the therapy. We further show the existence of inter CpG-wise correlation and compare our approach to standard independent multiple testing step-up procedures and earlier approaches for dependent data by Sun and Cai [2], who apply a homogeneous hidden Markov model.

Statistical testing is validated by a simulation study revealing a high discriminative power even with low methylation dynamics.

Although the testing procedure is presented in a methylation context, it is directly applicable to other high dimensional longitudinal datasets showing an inherent correlation structure.

References:

[1] P. Kuan and D.Y. Chiang. Integrating prior knowledge in multiple testing under dependence with applications to detecting differential DNA methylation. Biometrics, 68(3):774–783, 2012.
[2] W. Sun and T.T. Cai. Large-scale multiple testing under dependence. Journal of the Royal Statistical Society: Series B, 71(2):393–424, 2009.

AUC-based splitting criteria for random survival forests

Schmid, Matthias [1] Wright, Marvin [2] Eifler, Fabian [3]
Ziegler, Andreas [2]

[1] Universität Bonn, Deutschland
schmid@imbie.meb.uni-bonn.de
[2] Universität zu Lübeck, Deutschland
wright@imbs.uni-luebeck.de, ziegler@imbs.uni-luebeck.de
[3] Universität München, Deutschland
fabi.eifler@googlemail.com

Since their introduction in 2001, random forests have become a successful technique for statistical learning and prediction. Ishwaran et al. (2008) extended the original method for classification and regression by proposing a random forest technique for right-censored time-to-event outcomes ("random survival forests").

We present a new AUC-based splitting criterion for random survival forests that is inspired by the concordance index ("Harrell's C") for survival data. Using simulation studies and real-world data, the proposed splitting criterion is compared to traditional methods such as logrank splitting. The performance of the new criterion is evaluated with regard to sample size and censoring rate, and also w.r.t. various tuning parameters such as the forest size and the number of predictor variables selected in each split.

AUC-based splitting criteria are implemented in the R package "ranger" (Wright et al. 2014), which represents a versatile software environment for random survival forests in both high- and low-dimensional settings.

How is dynamic prediction affected if measurement patterns of covariates and outcome are dependent?

Schmidtmann, Irene [1] Weinmann, Arndt [2] Schritz, Anna [1,3]
Zöller, Daniela [1] Binder, Harald [1]

[1] Institut für Medizinische Biometrie, Epidemiologie und Informatik, Universitätsmedizin der Johannes Gutenberg-Universität Mainz, Deutschland
Irene.Schmidtmann@uni-mainz.de,anna.schritz@gmail.com
Daniela.Zoeller@uni-mainz.de,Harald.Binder@uni-mainz.de
[2] I. Medizinische Klinik, Universitätsmedizin der Johannes Gutenberg-Universität Mainz, Deutschland
Weinmann@uni-mainz.de
[3] Hochschule Koblenz, Rhein-Ahr-Campus
fabi.eifler@googlemail.com

In clinical cancer registries, information on initial diagnosis and treatment, but also information on observations at follow-up visits is available. Therefore, dynamic prediction including up to date information on disease recurrence, treatment change, biomarkers and other clinical findings is possible. However, unlike the observation patterns in clinical studies the frequency of follow-up visits and measurement of biomarkers and laboratory values does not follow a regular and uniform pattern. Patients are likely to be seen more frequently if known risk factors are present, if the treatment requires this, but also if treating doctors deem it necessary or patients request it. The latter two reasons may capture unmeasured latent risk factors.

Therefore, laboratory values and biomarkers are more up to date in frequently seen patients who may be at higher risk. In this situation, using the most recent values of covariates may lead to bias.

We suggest to consider the measurement as recurrent events and to model them in an Andersen Gill model. The residual from the Andersen Gill model captures influences on the frequency of measurement updates that are not explained by the measured covariates. When obtaining dynamic prediction, this residual is included as additional covariate. We illustrate the bias reduction that is achieved by this method in a simulation.

We apply the suggested method to data from the Mainz hepatocellular carcinoma registry and show how the estimation of effects is affected.

References:

1. Tan, K. S., et al. (2014). "Regression modeling of longitudinal data with outcome-dependent observation times: extensions and comparative evaluation." Stat Med.

2. Therneau, T. M., et al. (1990). "Martingale-Based Residuals for Survival Models." Biometrika 77(1): 147-160.

Expectile smoothing for big data and visualization of linkage disequilibrium decay

Schnabel, Sabine [1] van Eeuwijk, Fred [1] Eilers, Paul [1,2]

[1] Biometris, Wageningen University and Research Centre, Wageningen, Niederlande
sabine.schnabel@wur.nl, fred.vaneeuwijk@wur.nl
[2] Erasmus Medical Center, Rotterdam, Niederlande
p.eilers@erasmusmc.nl

In Statistical Genetics we are confronted with increasing amounts of genotyping data. To analyze those data, we need appropriate statistical methods as well as ways to visualize them. Here we focus on linkage disequilibrium decay (LD decay). Genetic markers on the same chromosome are not independent. The strength of these correlations decreases with increasing genetic distances between the markers: LD decay. LD decay is typically displayed in a scatterplot of pairwise LD between markers versus marker distance. When the number of markers increases, the interpretation of such scatter plots becomes problematic. With thousands of markers, we get millions of comparisons in terms of correlations. The scatter plot can be improved in several ways: one can fit a (non-) parametric curve to the data cloud to describe the mean relation between LD and marker distance. As an example of the extension of this approach we fit expectile curves. Expectiles give insight into the trend as well as the spread of the data. Computing times are considerably shortened by summarizing the data on a two-dimensional grid in the domain of distance and correlation. For a 100 by 100 grid we get a maximum of 10^4 pseudo-observations, independent of the number of initial data pairs. A scatter plot smoother also tackles this problem. The resulting information about the midpoints of the bins and the weights can be used for further analysis, e.g. with the expectile or quantile curves. The methods will be illustrated using data from plant genetics.

Optimization of dose-finding in phase II/III development programs

Schnaidt, Sven Kieser, Meinhard

University of Heidelberg, Germany
schnaidt@imbi.uni-heidelberg.de, meinhard.kieser@imbi.uni-heidelberg.de

Around 50% of pharmaceutical products in phase III do not make it to final approval [1]. This high attrition rate is in part due to the selection of suboptimal doses, resulting in lack of efficacy and/or intolerable safety-profiles. Thus, strong efforts are made to improve dose selection strategies in phase II/III development programs [2].

Assuming a sigmoidal Emax model for the dose response relationship, Lisovskaja and Burman [3] modeled the probability of success (POS) in phase III trials considering efficacy and safety simultaneously. To increase POS, they recommend the inclusion of two active doses in phase III in case of moderate uncertainty regarding parameters of the Emax model.

We relate this uncertainty, expressed by a prior distribution for ED50, to the size and design of the preceding phase II trial, which allows for an optimization of the POS of phase II/III programs. Among others, the effect of different phase II/III sample size allocations on the POS as well as the optimal number of doses included in phase III in dependence of the phase II sample size is analyzed. Results are discussed to give recommendations for planning of phase II/III development programs.

References:

[1] Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J (2014). Clinical development success rates for investigational drugs. Nature Biotechnology 32(1): 40-51.

[2] Pinheiro J, Sax F, Antonijevic Z, Bornkamp B, Bretz F, Chuang-Stein C, Dragalin V, Fardipour P, Gallo P, Gillespie W, Hsu C-H, Miller F, Padmanabhan SK, Patel N, Perevozskaya I, Roy A, Sanil A, Smith JR (2010). Adaptive and model-based dose-ranging trials: Quantitative evaluation and recommendations. White paper of the PhRMA working group on adaptive dose-ranging studies. Statistics in Biopharmaceutical Research 2(4): 435-454.

[3]. Lisovskaja V, Burman C-F (2013). On the choice of doses for phase III clinical trials. Statistics in Medicine 32: 1661-1676.

Methoden zur Untersuchung der Wechselwirkung von Lebensalter und Multimorbidität auf die körperliche Leistungs- und Koordinationsfähigkeit im Alter

Schneider, Michael Luksche, Nicole Zehnder, Andreas
Lieberoth-Lenden, Dominik Genest, Franca Seefried, Lothar

Orthopädisches Zentrum für Muskuloskelettale Forschung der Universität Würzburg
m-schneider.klh@uni-wuerzburg.de, n-luksche.klh@uni-wuerzburg.de,
andreas.zehnder@kitzingen.info, d.lieberothleden@gmail.com,
f-genest.klh@uni-wuerzburg.de,
l-seefried.klh@uni-wuerzburg.de

Im Zuge einer alternden Bevölkerung gewinnen geriatrische Studien zunehmend an Bedeutung, wobei Lebensalter, Multimorbidität und körperliche Leistungsfähigkeit von besonderem Interesse sind. Eine bedeutsame, aber bisher wenig untersuchte Fragestellung ist die Wechselwirkung von Multimorbidität und Alter auf die altersassoziierte Reduktion der körperlichen Leistungsfähigkeit und des Koordinationsvermögens.

In den Jahre 2013 und 2014 führten wir eine epidemiologische Erhebungsstudie an 507 gehfähigen Männern der Würzburger Wohnbevölkerung im Alter zwischen 65 und 90 Jahren durch, wobei die Erhebung von Messdaten der körperlichen Leistungsfähigkeit und die Analyse von assoziierter Faktoren die primärem Endpunkte waren.

In der vorliegenden Teilstudie untersuchen wir den Einfluss der Multimorbidität auf die körperliche Leistungsfähigkeit und auf das Koordinationsvermögen von Probanden zweier Altersgruppen. Die Klassifikation der Morbidität wurden in Anlehnung an die Definitionen der MultiCare-Studie [1] vorgenommen. Vorteile dieses Verfahrens ist die Abbildung des breiten Diagnosespektrums auf eine Skala von 46 organ- und funktionsbezogenen Diagnosegruppen. In den Analysen sind die Ergebnisse von verschiedenen gerontologischen Leistungs- und Koordinationstests als abhängige Variablen definiert. Unabhängige Variablen sind die Altersgruppe (65-75 Jahre, 76-90 Jahre), die Anzahl der Diagnosegruppen sowie ausgewählte häufige Erkrankungsgruppen (Hypertonie, Hyperlipoproteinämie, Hyperurikämie, Diabetes mellitus u.a.) und deren Komorbidität. Die Analysen befinden sich derzeit in der Auswertungs- und Modellierungsphase.

In dem vorliegenden Beitrag sollen die Methoden der Datenmodellierung und -analyse sowie exemplarisch einige Ergebnisse vorgestellt werden. Erste Analysen weisen darauf hin, dass der Effekt der Multimorbidität auf die körperliche Leistungsfähigkeit in der Gruppe im fortgeschrittenen Alter (76-90 Jahre) nur geringgradig ausgeprägter ist als in der jüngeren Vergleichsgruppe (65 bis 75 Jahre). Diese Beobachtung steht im Einklang mit Voruntersuchungen [2]. Deutlichere Heterogenität beobachten wir in der Untersuchung des Koordinationsvermögens beider Altersgruppen.

Literatur:

[1] Schäfer I, Hansen H, Schön G et al.,The German MultiCare-study: Patterns of multimorbidity in primary health care – protocol of a prospective cohort study; BMC Health Services Research 2009, 9:145

[2] Welmer AK, Kåreholt I, Angleman S, Rydwik E, Fratiglioni L.;Can chronic

multimorbidity the age-related differences in strength, speed and balance in older adults?; Clin Exp Res. 2012 Oct;24(5):480-9

Optimal designs for comparing curves

Schorning, Kirsten Dette, Holger

Ruhr-Universität Bochum, Deutschland
kirsten.schorning@rub.de, holger.dette@rub.de

For a comparison of two regression curves, which is used to establish the similarity between the dose response relationships of two groups, the choice of the two designs (one for each regression curve) is crucial. Therefore an optimal pair of designs should minimize the width of the confidence band for the difference between the two regression functions.

During the talk optimal design theory (equivalence theorems and efficiency bounds) will be presented for this non standard design problem. The results will be illustrated in several examples modeling dose response relationships. For instance, the optimal design pairs for combinations of the EMAX model, the loglinear model and the exponential model will be considered. More precisely, it will be demonstrated that the optimal pair of designs for the comparison of these regression curves is not the pair of the optimal designs for the individual models. Moreover, it will be shown that the use of the optimal designs proposed instead of commonly used ""non-optimal"" designs yield a reduction of the width of the confidence band by more than 50%.

Dealing with recurrent events in the analysis of composite time-to-event endpoints

Schüler, Svenja Kieser, Meinhard Rauch, Geraldine

Universität Heidelberg, Deutschland
schueler@imbi.uni-heidelberg.de, meinhard.kieser@imbi.uni-heidelberg.de, rauch@imbi.uni-heidelberg.de

Composite endpoints combine several time-to-event variables of interest within a single outcome measure. Thereby, the expected number of events is increased compared to a single time-to-event endpoint and the power of the trial is enlarged.

Composite endpoints are usually defined as time-to-first-event variables were only the first occurring event is counted and subsequent events are ignored. By using this approach, standard survival analysis techniques such as the Cox-model can be applied. However, considering only the first occurring event clearly defines a loss in information. The difficulty in modeling subsequent events is that the risk for further events is usually changed whenever a first event has been observed. Therefore, the underlying hazards of the corresponding multi-state model must be estimated separately for each transition from one event to the other.

In this talk, we will show how multi-state models can be used as an adequate framework for the analysis of composite endpoints which takes subsequent events into account. Moreover, other approaches to deal with subsequent events proposed in the literature will be evaluated and compared for the application to composite endpoints. All methods are illustrated by a real clinical trial example.

References:

[1] Beyersmann J, Schumacher M, Allignol A: Competing risks and multistage models with R; Springer, 2012;

Patient-oriented randomisation versus classical block randomisation

Schulz, Constanze Timm, Jürgen

KKS Bremen, Universität Bremen, Deutschland
conschul@math.uni-bremen.de

The "gold standard" for clinical trials is a controlled, randomised and double-blinded trial usually comparing specific treatments. In case of homogeneous patient collectives this is highly efficient. But what happens, if not all patients present superior results for the same drug? In this case some patients will receive the "wrong" drug by randomisation and the trial result becomes sensitive to case selection. Discrepancy between the daily clinical perception and study results may occur and this should always raise the question of the adequacy of the methodological approach. From an ethical point of view, the physician should be involved in the decision concerning treatment, taking risks and healing opportunities of each patient into account. Based on this problem especially in the psychiatric community we created a new clinical design for the NeSSy [1] study, which combines the scientific demand of randomisation with patient-oriented decisions. This design is focussed on strategy comparison instead of comparing specific treatments. The idea is to randomise the strategies and let the physician decide between treatments within these strategies.

The new idea of such patient-oriented randomisation (POR) is generalised to the case of two different strategies with an arbitrary number of treatments within each strategy. The implementation of the randomisation and the practical realisation as well as some theoretical properties will be described. The impact of the physician's decision which specific drug within a strategy he prefers for the respective patient (clear patient-oriented decisions) will be displayed. Main results compare the innovative design POR with a similar block randomisation with regard to balance behaviour between strategies, allocation probability of a specific treatment and efficacy distribution in different strategies for a given heterogeneous population. Results have been generated by theoretical consideration as well as simulation studies.

References:

[1] The Neuroleptic Strategy Study - NeSSy, funded under the program of clinical studies of the Federal Ministry of Education and Research

Meta-Analysis and the Surgeon General's Report on Smoking and Health

Schumacher, Martin Rücker, Gerta Schwarzer, Guido

Universitätsklinikum Freiburg, Deutschland – Institut für Med. Biometrie und Statistik
ms@imbi.uni-freiburg.de, ruecker@imbi.uni-freiburg.de,
sc@imbi.uni-freiburg.de

Although first meta-analyses in medicine have already been conducted at the beginning of the 20th century, its major breakthrough came with the activities of the Cochrane Collaboration during the 1990s. It is less known that the landmark report on "Smoking and Health" to the Surgeon General of the Public Health Service, published in 1964, makes substantial use of meta-analyses performed by statistician William G. Cochran.

Based on summary data given in the report we reconstructed meta-analyses of seven large, prospective studies that were initiated in the 1950s by concentrating on overall and lung cancer mortality [1]. While visualization of results including confidence intervals was largely neglected in the report, we are able to give a vivid impression of the overwhelming evidence on the harmful effects of cigarette smoking. We will put William G. Cochran's contribution into the context of the so-called lung cancer controversy in which other prominent statisticians, e.g. Sir Ronald Fisher, played a major role. In contrast to the latter, who selected a specific study that supported his personal view, Cochran took an impartial, systematic approach for evaluation and followed the major steps of a modern systematic review, including an appraisal of risk of bias based on sensitivity analysis. For that he used statistical methodology that still is state-of-the-art today. Although substantially contributing to an important public policy issue, this work, that had a large impact on the public perception of health risks due to smoking, is often overlooked and deserves much more attention.

Reference:

[1] Schumacher M, Rücker G, Schwarzer G. Meta-Analysis and the Surgeon General's Report on Smoking and Health. N Engl J Med 2014; 370: 186-188.

Regressionsverfahren als Ersatz für bivariate Modelle zur Bestimmung des Surrogatschwellenwerts bei korrelationsbasierten Validierungsverfahren

Schürmann, Christoph Sieben, Wiebke

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)
christoph.schuermann@iqwig.de, wiebke.sieben@iqwig.de

Für die Validierung eines Surrogatendpunkts liegt mit den erstmals von Buyse et al. [1] vorgeschlagenen meta-analytischen, korrelationsbasierten Verfahren ein gut untersuchter Modellierungsansatz vor. Mit dem daraus ableitbaren Surrogatschwellenwert (surrogate threshold effect, STE, [2]) kann für einen gegebenen Effekt auf ein Surrogat eine Vorhersage für den Effekt auf den eigentlich interessierenden Endpunkt erfolgen. Das ursprüngliche bivariate Modell basiert auf individuellen Patientendaten und ist daher nicht anwendbar, falls nur aggregierte Daten zur Verfügung stehen. Wir untersuchen, ob und welche alternativen Modelle in dieser Situation einen adäquaten Ersatz darstellen, wobei wir uns auf Regressionsmodelle (einfache und gewichtete lineare Regression, Meta-Regression mit zufälligen Effekten) beschränken, die bereits in der Literatur ersatzweise verwendet worden sind. Unter Variation möglicher Einflussfaktoren (wie z. B. Studienanzahl, Variation und Korrelation der Effekte) berechnen wir in Simulationen die Abweichungen der STEs aus den vereinfachten Modellen zu dem Resultat aus dem ursprünglichen Ansatz.

Literatur:

[1] Buyse, M., Molenberghs, G., et al., 2000. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics 1, 49-67.

[2] Burzykowski, T., Buyse, M., 2006. Surrogate threshold effect: An alternative measure for meta-analytic surrogate endpoint validation. Pharmaceutical Statistics 5, 173-186.

Ist der gruppenbasierte Trajektorie-Modell-Ansatz übertragbar auf chronische Stoffwechselerkrankungen? Eine Analyse von Patienten mit Typ-1-Diabetes mellitus

Schwandt, Anke [1] Hermann, Julia M [1] Bollow, Esther [1]
Rosenbauer, Joachim [2] Holl, Reinhard W. [1]

[1] Universität Ulm, Deutschland
anke.schwandt@uni-ulm.de, julia.hermann@uni-ulm.de, esther.bollow@uni-ulm.de, reinhard.holl@uni-ulm.de
[2] Deutsches Diabetes-Zentrum (DDZ), Leibniz-Zentrum für Diabetesforschung an der Heinrich-Heine-Universität Düsseldorf, Deutschland
Joachim.Rosenbauer@ddz.uni-duesseldorf.de

Die Diabetes-Patienten-Verlaufsdokumentation (DPV) ist ein multizentrisches, computerbasiertes Diabetes-Patienten-Register. Eine halbjährliche Vergleichsauswertung (Benchmarking) dient den teilnehmenden Zentren zur Qualitätssicherung. Die dabei betrachteten aggregierten Mittelwerte können jedoch den longitudinalen Verlauf der Daten nur teilweise repräsentieren. Im Folgenden sollen daher aus dem sozialwissenschaftlichen Bereich bekannte gruppenbasierte Trajektorie-Modelle nach Nagin et.al. (1) angepasst werden. Diese basieren auf einem Finite-Mixture-Modeling-Ansatz, wobei die Parameter mit der ML-Methode geschätzt werden. Mittels dieser Modelle können verschiedene Gruppen mit spezifischen Trendcharakteristika in Parametern innerhalb eines Zeitintervalls identifiziert werden.

In einem ersten Schritt wurde der HbA1c-Wert, eine der wichtigsten Verlaufsvariablen der Therapie des Typ-1-Diabetes mellitus, zwischen den Jahren 2004 und 2013 auf Zentrumsebene (n=176 Zentren) analysiert. Die Anpassung gruppenbasierter Trajektorie-Modelle erfolgte mit der SAS-Prozedur "Proc Traj" (SAS 9.4) (2). Die Modellwahl hinsichtlich Anzahl der Gruppen und polynomialem Grad des Zeittrends basierte auf dem Bayesian Information Criterion (BIC).

Pro Zentrum und Jahr wurde der HbA1c-Median berechnet. Das nach BIC optimale Modell teilte die Zentren in vier Gruppen mit linearem Trend ein. Drei Gruppen (n1=33, n2=52, n3=84) zeigten stabile Trajektorien, jedoch mit unterschiedlichen HbA1c-Niveaus. Dies spiegelte die Ergebnisse des Benchmarkings wider, welches eine Einteilung in Zentren mit niedrigen, mittleren und hohen HbA1c-Werte vornimmt. Die Trajektorie einer kleinen vierten Gruppe (n4=7) wies über die Jahre eine kontinuierliche Verschlechterung des HbA1c-Wertes auf. Mögliche Erklärungsansätze lagen im strukturellen Bereich der Zentren (kleine Patientenzahl pro Zentrum, änderungen in Versorgungsaufträgen, personelle änderungen).

Die gruppenbasierten Trajektorie-Modelle scheinen geeignet zu sein, zeitliche HbA1c-Verläufe auf Zentrumsebene zu gruppieren. In weiteren Analysen sollen die Modelle auf Patientenebene erweitert werden, um individuelle longitudinale Daten von Typ-1-Diabetes-Patienten in charakteristische Verläufe zu klassieren.

References:

(1) Daniel S. Nagin, Group-Based Modeling of Development, Harvard Univ Pr, 2005

(2) Bobby L Jones, Daniel S. Nagin, Kathryn Roede, A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories, SOCIOLOGICAL METHODS&RESEARCH 2001 Feb 3, 29(3): 374-393

Subgruppenanalysen in der frühen Nutzenbewertung (AMNOG): Fluch oder Segen?

Schwenke, Carsten

SCO:SSiS, Deutschland ohneberg@imbi.uni-freiburg.de, wolke@imbi.uni-freiburg.de, ms@imbi.uni-freiburg.de

Subgruppenanalysen werden derzeit weitreichend diskutiert. Die CHMP hat in diesem Jahr eine Draft Guideline zu Subgruppenanalysen in konfirmatorischen Studien veröffentlicht [1]. Diese hat einen direkten Einfluss auf die frühe Nutzenbewertung nach AMNOG, da die im Studienprotokoll präspezifizierten Subgruppenanalysen auch im Nutzendossier dargestellt und diskutiert werden sollen. Die Verfahrensordnung zum AMNOG verlangt zudem Subgruppenanalysen zumindest nach Alter, Geschlecht, Land/Zentrum und Schweregrad der Erkrankung [2]. Zudem sollen alle präspezifizierten Subgruppenanalysen für alle zur Herleitung des Zusatznutzen verwendeten Endpunkte durchgeführt und dargestellt werden. Auch das IQWiG widmet sich der Interpretation der Subgruppenanalysen in ihrem Methodenpapier [3].

Anhand von Beispielen werden die Vor- und Nachteile der Subgruppenanalysen in der frühen Nutzenbewertung diskutiert und methodische wie interpretatorische Herausforderungen vorgestellt.

Referenzen:

[1] CHMP: Guideline on the investigation of subgroups in confirmatory clinical trials. EMA/CHMP/539146/2013

[2] Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses [online]. 08.05.2014 [Zugriff: 15.12.2014]. URL: https://www.g-ba.de/downloads/62-492-873/VerfO_2014-03-20.pdf

[3] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen: Allgemeine Methoden Version 4.1 [online]. 28.11.2013 [Zugriff: 15.12.2014]. URL: https://www.iqwig.de/download/IQWiG_Methoden_Version_4-1.pdf

Exploration of geriatric mobility scores using semiparametric quantile regression

Sobotka, Fabian [1] Dasenbrock, Lena [2] Bauer, Jürgen M. [2]
Timmer, Antje [1]

[1] Department für Versorgungsforschung, Fakultät VI, Universität Oldenburg, Deutschland
fabian.sobotka@uni-oldenburg.de, antje.timmer@uni-oldenburg.de
[2] Universitätsklinik für Geriatrie, Klinikum Oldenburg, Deutschland
Dasenbrock.Lena@klinikum-oldenburg.de, bauer.juergen@klinikum-oldenburg.de

In geriatrics there are several competing tools to measure a senior patient's mobility in acute care. The Tinetti Performance-Oriented-Mobility Assessment (Tinetti POMA) assesses walking and balancing abilities while the de Morton Mobility Index (DEMMI) consists of a wider range of a patient's movements. For both scores possible influential covariates remain unclear. In the university hospital in Oldenburg, Germany, both mobility tests were conducted for 144 patients between the ages of 70 and 100 with 2/3 being women. We find that the distributions of the mobility indices within the sample are skewed and heteroscedastic along the age. Especially at the beginning of the rehabilitation most patients have rather low mobility scores. We aim to uncover the complete conditional distributions of Tinetti and DEMMI and compare the information gain to mean regression.

We evaluate the effects of age, sex, independency and diagnosis on the values of DEMMI and Tinetti. We propose the use of quantile regression in order to model the complete unknown (and probably different) distributions of the two scores. For possibly nonlinear effects of the continuous covariates we construct p-spline bases for a flexible and smooth model. However, the estimation of frequentist quantiles would be unnecessarily difficult if we added the quadratic penalty of a p-spline basis. A reparameterisation of the spline basis allows for the use of a LASSO penalty which is a natural partner for quantile regression.

This method returns continuously comparable results for Tinetti and DEMMI and improves our understanding of these mobility scores. While the effects of age and independency can appear linear in mean regression, we clearly find nonlinearity in the tails of the score distributions. The effect of sex also increases from the lower towards the upper tail. Hence, we find that mean regression offers too little information for a complete analysis of these mobility scores.

The cure-death-model – A new approach for a randomised clinical trial design to tackle antimicrobial resistance

Sommer, Harriet [1] Wolkewitz, Martin [1] Timsit, Jean-Francois [2]
Schumacher, Martin [1]

[1] Institute for Medical Biometry and Statistics, Medical Center - University of Freiburg (Germany)
sommer@imbi.uni-freiburg.de, wolke@imbi.uni-freiburg.de, ms@imbi.uni-freiburg.de
[2] Bichat Hospital, Paris Diderot University (France)
jean-francois.timsit@bch.aphp.fr

Antimicrobial resistance (AMR) is a growing problem worldwide and with few new drugs making it to the market there is an urgent need for new medicines to treat resistant infections. There is a variety of primary endpoints used in studies dealing with severe infectious diseases, recommendations given by the existing guidelines are not always consistent nor is their practical application. Usually patients' cure rates are compared in trials dealing with AMR but they often suffer from severe diseases besides their infection, so mortality shall not be disregarded. A mortality rate of about 10% until 30% can be assumed within 30 days.

To understand the etiological process how the new treatment influences the cure process, we propose to perform a joint model with two primary endpoints – a combination of non-inferiority study regarding mortality and superiority study concerning cure using a multistate model where death without previous cure acts as competing event for cure and vice versa. Also other authors used a similar combination to deal with the fact that two primary endpoints may be necessary [Röhmel et al.; Biometrical Journal 2006; 48(6):916-933 and Bloch et al.; Biometrics 2001; 57:1039-1047].

Mostly, patients die due to the underlying disease and even if the infection can be considered as cured, patients can die nevertheless. By means of analogies coming from oncology, the model has to be extended to an illness-death-model (here referred to as cure-death-model), a special case of a multistate model [Schmoor et al.; Clin Cancer Research 2013; 9(1):12-21]. Based on real data examples, we simulate different scenarios, compare the simple competing risks model with the cure-death-model and thus show that mortality after being cured cannot be ignored either.

COMBACTE is supported by IMI/EU and EFPIA.

Model Selection in Expectile Regression

Spiegel, Elmar [1] Kneib, Thomas [1] Sobotka, Fabian [2]

[1] Georg-August-Universität Göttingen, Germany
espiege@uni-goettingen.de, tkneib@uni-goettingen.de
[2] Carl von Ossietzky Universität Oldenburg, Germany
fabian.sobotka@uni-oldenburg.de

Expectile regression can be seen as a compromise between quantile regression and normal least squares regression. Expectile regression estimates, similar to results from quantile regression, reflect the influence of covariates on the complete conditional distribution of the response while avoiding the assumption of a specific distribution. Since expectiles depend on the weighted L2 norm of residual deviations, they are a natural generalization of least squares estimation and therefore inherit its advantages of easily including nonlinear covariate effects based on penalised splines or spatial effects for both regional and coordinate data. Model selection on expectiles is especially interesting, as the influence of covariates depends on the asymmetry. On the one hand, the selection can be done for every asymmetry parameter separately if we wish to analyse which covariate has more influence in the upper and lower tail, respectively. On the other hand, the model can be selected for all asymmetry parameters together so that in the end one model specification is valid for all asymmetries. We will show different approaches of model selection including stepwise seletion and shrinkage methods for both separate and joint selection.

This methodology can be applied to many biometric problems when regression effects beyond the mean are estimated. We will illustrate expectile model seletion for an analysis of childhood malnutrition data. It is especially necessary to find an appropriate model for the lower tail of the nutritional score. Generally there are several covariates like gender, age (of child and mother) or location which might have an influence on the child's nutritional status. However, the accuracy of the model strongly depends on an appropriate specification of the model and thus we need model selection tools. Our tools remove improper or too small effects and could also exclude an unnecessary spatial effect, which might bias the estimation.

On the correlation structure of test statistics in genetic case-control association studies

Stange, Jens [1] Dickhaus, Thorsten [1] Demirhan, Haydar [2]

[1] WIAS Berlin, Germany
Jens.Stange@wias-berlin.de, Thorsten.Dickhaus@wias-berlin.de
[2] Hacettepe University, Ankara, Turkey
haydarde@hacettepe.edu.tr

For simultaneous statistical inference of many 2xC contingency tables it is useful to incorporate the correlation structure of the involved test statistics. In genetic case-control association studies with C equal to 2 (allelic association) or 3 (genotypic association) asymptotic normality of the categorical data follows from appropriate multivariate Central Limit Theorems. Therefore the asymptotic correlation structure between the contingency tables is described by Pearson's haplotypic linkage disequilibrium coefficient. Application of the multivariate Delta method reveals that this asymptotic correlation structure is preserved for a wide variety of test statistics. As examples, in which our methodology applies, we consider chi-squared statistics, logarithmic odds ratios and linear trend tests.

Reference:

T. Dickhaus, J. Stange, H. Demirhan: "On an extended interpretation of linkage disequilibrium in genetic case-control association studies", WIAS Preprint No. 2029 (2014),
http://www.wias-berlin.de/preprint/2029/wias_preprints_2029.pdf

Der Einfluss verschiedener Verfahren zum Umgang mit fehlenden Werten und zur Variablenselektion auf die Zusammensetzung und Güte logistischer Regressionsmodelle

Stierlin, Annabel Mayer, Benjamin Muche, Rainer

Universität Ulm, Institut für Epidemiologie und Medizinische Biometrie, Deutschland
Annabel.Stierlin@Uni-Ulm.de, Benjamin.Mayer@Uni-Ulm.de, Rainer.Muche@Uni-Ulm.de

Bei der Entwicklung logistischer Regressionsmodelle wird man – wie auch bei anderen Prognosemodellierungen - vor zwei Herausforderungen gestellt: die Trennung der wesentlichen Einflussvariablen von weiteren sogenannten Störvariablen und die Handhabung fehlender Werte. Viele Methoden zur Variablenselektion und auch zum Umgang mit fehlenden Werten sind bekannt aber jede zeichnet sich durch spezifische Vor- und auch Nachteile aus. Im Falle der multiplen Imputation stellt insbesondere die Umsetzung von Variablenselektionsverfahren ein besonderes Problem dar, welches auch in der Literatur nur wenig diskutiert wird.

Das Ziel dieser Arbeit war es zu demonstrieren, wie verschiedene Verfahren zum Umgang mit fehlenden Werten sowie zur Variablenselektion kombiniert werden können und wie sich dies auf die Modellzusammensetzung sowie die Leistungsfähigkeit der Modelle auswirkt. Das Spektrum der Analysemethoden umfasste Complete Case Analyse, einfache Imputation mit EM-Algorithmus und multiple Imputation mit FCS-Algorithmus zum Umgang mit fehlenden Werten sowie Rückwärtselimination, schrittweise Selektion, LASSO penalisierte Regression und Bootstrap Selektion zur Variablenselektion und wurde auf simulierte sowie auf reale Datensätze angewandt. Die prädiktive Leistungsfähigkeit der Modelle wurde mittels Bootstrapvalidierung gemessen und die Zusammensetzung der finalen Modelle wurde untersucht.

Im Rahmen dieser Auswertung hat sich gezeigt, dass die verschiedenen Modelle eine vergleichbare Prädiktionsfähigkeit aufweisen, aber die Auswahl an Kovariaten konnte in der Regel nicht durch verschiedene Analyseprozeduren repliziert werden. Deswegen ist es wichtig, insbesondere in Anbetracht der unglaublichen Flexibilität in der Modellauswahl und der Auswahl der kritischen Werte, die erfolgten Sensitivitätsanalysen bei der Veröffentlichung eines neuen Modells stets zu kommentieren.

Integrative analysis of case-control data on multiple cancer subtypes

Stöhlker, Anne-Sophie [1] Nieters, Alexandra [2] Binder, Harald [3]
Schumacher, Martin [1]

[1] Institute for Medical Biometry and Statistics, Medical Center - University of Freiburg (Germany)
stoehlker@imbi.uni-freiburg.de, ms@imbi.uni-freiburg.de
[2] Center for Chronic Immunodeficiency, Medical Center - University of Freiburg (Germany)
alexandra.nieters@uniklinik-freiburg.de
[3] Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center Johannes Gutenberg University Mainz (Germany)
binderh@uni-mainz.de

In general, one cancer entity does not have a simple structure but usually breaks down into several (heterogeneous) subtypes. When investigating the associations between genes or, respectively, SNPs and one cancer, this diversity turns into a challenge for the analysis: Some genes might be associated with the respective cancer in general, while some other genes could be related to specific subtypes. However, subgroup analysis might overlook shared genes which determine central characteristics of the cancer whereas an overall analysis could miss type-specific genes representing the singularity of subtypes. Thus, an analysis combining those aspects would be beneficial to understand relations and differences of the subtypes. Data on several cancer subtypes is mostly investigated by comparing analysis results of single subtypes and therefore might suffer from an insufficient amount of data per subtype.

We consider an approach for integrative analysis that analyzes the data on all subtypes simultaneously (1). While this approach was developed for prognosis data, we modify it for case-control settings. It is based on the heterogeneity model allowing a gene to be associated with all, only some, or none of the subtypes, respectively. Building upon this, a tailored compound penalization method is applied to actually find out whether a gene is associated with any of the present subtypes and if so, with which of them. In this context, genes are selected if they contain important SNPs associated with any subtype. The proposed method uses an iterative algorithm to identify interesting genes. To contrast the above-mentioned approach, we also investigate a componentwise boosting approach. Both procedures are applied to real genotyping data derived from a case-control study on the etiology of various subtypes of lymphoma.

References:

1. Liu, J., Huang, J. et al. (2014), Integrative analysis of prognosis data on multiple cancer subtypes. Biometrics, 70:480–488.

Recherchewerkzeug OperationsExplorer: Wie Datenjournalisten mit Daten und Statistik ringen

Stollorz, Volker

Freier Journalist
stollorz@googlemail.com

Im deutschen Gesundheitswesen werden viele Daten gesammelt. Leider mangelt es nicht nur systematisch an Datentransparenz, sondern auch an der Kompetenz von Journalisten, verfügbare Datensätze für eigene Recherchen zu erschließen. Informatiker, Statistiker und Biometriker könnten mit ihren methodischen Kompetenzen helfen, klügeren Journalismus im öffentlichen Interesse zu ermöglichen. Um das Feld des „Data-Driven-Journalismus“ zu stärken und Qualitätsstandards der Datenverarbeitung zu entwickeln wurde in Zusammenarbeit zwischen Wissenschaftsjournalisten und Informatikern am Heidelberger Institut für Theoretische Studien das Recherchewerkzeug „OperationsExplorer“ entwickelt. In der webbasierten Datenbankanwendung können registrierte Journalisten intuitiv nach allen stationären ICD-Diagnosefällen sowie allen Operationen und Prozeduren (OPS-Viersteller) aller meldepflichtigen Krankenhäuser in Deutschland suchen. Der Volldatensatz kann beim Statistischen Bundesamt erworben werden und enthält pro Jahr 18 Millionen stationäre Krankenhausfälle codiert nach Wohnort des Patienten auf Kreisebene. Im OperationsExplorer können sich Journalisten den Datensatz mit einfachen Suchmasken erschließen: nach der Häufigkeit der Fälle, nach dem Wohnort des Patienten auf Kreisebene, nach Alter, Geschlecht und Jahr (2009-2013). Mit dem neuen Recherchewerkzeug werden regionale und zeitliche Auffälligkeiten vollautomatisch altersstandardisiert und in Kartenform auf Kreisebene visualisiert.

Der Vortrag präsentiert erste Ergebnisse und fokussiert dabei auf statistische Herausforderungen, die beim Umgang mit der mächtigen Recherchemaschine zu bewältigen sind. Als Ergebnis steht ein Plädoyer für innovative Formen der Zusammenarbeit zwischen datengetriebenen Wissenschaftlern und Datenjournalisten. Eine statistische Qualitätssicherung für datengetriebenen Journalismus hilft Recherchen im Gesundheitswesen im öffentlichen Interesse zu verbessern.

Mathematical models for anisotropic glioma invasion: a multiscale approach

Surulescu, Christina

TU Kaiserslautern, Deutschland
surulescu@mathematik.uni-kl.de

Glioma is a broad class of brain and spinal cord tumors arising from glia cells, which are the main brain cells that can develop into neoplasms. Since they are highly invasive they are hard to remove by surgery, as the tumor margin it most often not precisely enough identifiable. The understanding of glioma spread patterns is hence essential for both radiological therapy as well as surgical treatment.

We propose a multiscale framework for glioma growth including interactions of the cells with the underlying tissue network, along with proliferative effects. Relying on experimental findings, we assume that cancer cells use neuronal fibre tracts as invasive pathways. Hence, the individual structure of brain tissue seems to be decisive for the tumor spread. Diffusion tensor imaging (DTI) is able to provide such information, thus opening the way for patient specific modeling of glioma invasion. Starting from a multiscale model involving subcellular (microscopic) and individual (mesoscale) cell dynamics, we deduce on the macroscale effective equations characterizing the evolution of the tumor cell population and perform numerical simulations based on DTI data.

Biplots: A graphical method for the detection of infectious disease outbreaks

Unkel, Steffen

Universitätsmedizin, Georg-August-Universität Göttingen, Deutschland
steffen.unkel@med.uni-goettingen.de

The past decade has witnessed a large increase in research activity on the statistical issues that are related to the detection of outbreaks of infectious diseases. This growth in interest has given rise to much new methodological work, ranging across the spectrum of statistical methods (Unkel et al., 2012). In this talk, a new graphical method based on biplots (Gower et al., 2011) is proposed for the detection of infectious disease outbreaks. Biplots, of which various forms do exist, are a graphical tool for simultaneously displaying two kinds of information; typically, the variables and sample units described by a multivariate data matrix. A novel application to multivariate infection surveillance data from the United Kingdom is provided to illustrate the method.

References:

[1] Unkel, S., Farrington, C. P., Garthwaite, P. H., Robertson, C. and Andrews, N. (2012): Statistical methods for the prospective detection of infectious disease outbreaks: a review, Journal of the Royal Statistical Society Series A, Vol. 175, Part 1, pp. 49–82.

[2] Gower, J. C., Lubbe, S. and Le Roux, N. (2011): Understanding Biplots, Wiley.

The need for a third dimension in the external valdiation of clinical prediction rules

Vach, Werner

Medical Center - University of Freiburg, Deutschland
wv@imbi.uni-freiburg.de

When clincial prediction rules have to be validated in an external data set, the focus is often on two dimensions: calibration and discrimination. However, these two dimensions do not cover the whole information about the discrepany between the true event probabilities and the suggested probabilities according to the clinical prediciton rule.

We present some (theoretical) examples with varying degree of agreement between true and suggested event probabilities, which give identical calibration scope, AUC and Brier score.

To overcome the problem, we can consider to estimate directly some measures of the agreement between true and suggested event probabilities, like the euclidian distance. However, such measures may be hard to interprete. As an alternative, we suggest to estimate the inverse calibration slope, i.e. the slope of a regression of the suggested vs. the true event probablities. The joint interpretation of the inverse calibration slope and the ordinary calibration slope is simple: If both are 1, then we have perfect agreement. We demonstrate that the inverse calibration slope can be estimated by a boostrap bias correction of the naive estimate based on a flexible estimate of the true event probabilities.

Quantifizierung des Ausmaßes des Zusatznutzens von neuen Arzneimitteln: “gering” – “beträchtlich” – “erheblich”?

Vach, Werner

Medical Center - University of Freiburg, Deutschland
wv@imbi.uni-freiburg.de

In Deutschland sieht das Arzneimittelmarktneuordnungsgesetz (AMNOG) eine Quantifizierung des Zusatznutzens neuer Arzneimittel vor. In der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) hat das Bundesministerium für Gesundheit in §5 Absatz (7) ausgeführt, das die Quantifizierung des Ausmaßes des Zusatznutzens anhand der Begrilicher “erheblicher Zusatznutzen”, “beträchtlicher Zusatznutzen” und “geringer Zusatznutzen” erfolgen soll. Das IQWiG hat hierzu im Jahre 2011 einen Vorschlag zur Operationalisierung gemacht, der bereits mehrfach angewandt wurde, und mittlerweile auch Eingang in das Methodenpapier des IQWiGs gefunden hat.

In meinem Beitrag gehe ich auf einige grundsätzliche Fragen ein, die sich aus den gesetztlichen Vorgaben und der Operationalisierung des IQWiGs ergeben: Lässt sich mit statistischen Methoden ein “beträchtlicher Zusatznutzen” definieren? Kann eine Klassifikation des Zusatznutzens auf Grund eines geschätzten RR verlässlich sein? Welche grundsätzlichen Möglichkeiten der statistischen Operationalisierung gibt es? Gibt es “das” relative Risiko für den Endpunkt Mortalität? Ist ein relatives Risiko ein sinnvolles Maß für den Zusatznutzen? Sollen in Zukunft Studien so geplant werden, dass ein wahrer “beträchtlicher” Effect gezeigt werden kann?

Der Vortrag berührt die grundsätzliche Frage, wie sich die Biometrie verhalten soll, wenn von der Politik die Umsetzung gesetztlicher Vorgaben erwartet wird, aber diese Umsetzung nicht eindeutig erfolgen kann. Reicht es, eine mögliche Lösung aufzuzeigen, oder sind wir verpflichtet, die dabei getroffenen Entscheidungen in nichteindeutigen Entscheidungssituationen transparent zu machen und mögliche Konseuenzen darzulegen?

Methods for statistical inference in random effects meta-analysis for use in systematic reviews of medical interventions

Veroniki, Areti Angeliki

St. Michael's Hospital, Canada
veronikia@smh.ca

Background
Meta-analyses (Mas) are typically used to combine the results of studies that address the same clinical question and estimate the overall treatment effect of an outcome of interest. An additional aim is to make inferences about the between-study variability, termed heterogeneity, as its presence can have a considerable impact on conclusions from MAs. Several methods have been proposed to estimate heterogeneity including the widely used method proposed by DerSimonian and Laird. However, this method has long been challenged and simulation studies suggest it underestimates heterogeneity when its true level is high.Another important issue in MA is to select the 'best' way to compute the confidence interval (CI) for the overall mean effect among various suggested procedures.While several published papers compare some of the heterogeneity estimators, heterogeneity's CIs, or CIs for the overall mean effect, the literature lacks a full review presenting all alternative estimation options.

Aims
1. To provide a comprehensive overview of methods used for estimating heterogeneity, its uncertainty, and the CI around the overall mean effect in a MA
2. to make clear recommendations for MAs about the selection of these methods based on existing simulation and empirical studies.

Methods/Results
We searched in PubMed for research articles that describe or compare the methods for estimating heterogeneity in MA in simulation or empirical studies, scanned references from included studies, and contacted methodological experts in the field for additional relevant articles. We identified 16 heterogeneity estimators, 7 methods to calculate heterogeneity's CI,and 8 approaches to compute the overall mean effect's CI. Published studies suggested that different methods for heterogeneity, its CI, or the overall effect's CI can provide different or conflicting results and their performance might vary in different MA scenarios. Selection of the most appropriate heterogeneity estimator depends on (i) whether a zero value of heterogeneity is possible, (ii) the properties of the estimators in terms of bias and efficiency, which may depend on the number of studies included and the magnitude of the true heterogeneity, and (iii) the ease of application. The most appropriate CI method for both heterogeneity and overall mean effect should be chosen according to its (i) accuracy, (ii) precision, and (iii) ease of computation.

Conclusions
Our recommendations are based on the scenarios and results presented in published studies. Overall, the estimator suggested by Paule and Mandel and

the restricted maximum likelihood method are better alternatives to estimate heterogeneity, and we recommend the Q-profile method and the alternative approaches based on 'generalized Cochran heterogeneity statistics' to compute the corresponding CI around heterogeneity. The Q-Profile approach has the advantage that is simple to compute, but the method suggested by Jackson with weights equal to the reciprocal of the within-study standard errors outperforms the Q-Profile method for small heterogeneity.We also suggest the method described by Knapp and Hartungto estimate the uncertainty around the overall effect as it is the only method insensitive to the heterogeneity magnitude and estimator, and outperforms other suggested alternatives.When there are few studies included in the MA a sensitivity analysis is advisable using a variety of methods for estimating heterogeneity, its uncertainty,and the uncertainty for the overall effect.

A case-cohort approach for extended illness-death models in hospital epidemiology

von Cube, M. [1] Schumacher, M. [1] Palomar-Martinez, M. [3,4]
Olaechea-Astigarraga, P. [5] Alvarez-Lerma, F. [6] Wolkewitz, M. [1,2]

[1] Institute for Medical Biometry and Statistics, Medical Center - University of Freiburg, Germany
majavoncube@gmail.com, ms@imbi.uni-freiburg.de, wolke@imbi.uni-freiburg.de
[2] Freiburg Center for Data Analysis and Modelling, University of Freiburg, Germany
wolke@imbi.uni-freiburg.de
[3] Hospital Universitari Arnau de Vilanova, Lleida, Spain
mpalomarmartinez@gmail.com
[4] Universitat Autonoma de Barcelona, Spain
mpalomarmartinez@gmail.com
[5] Service of Intensive Care Medicine, Hospital de Galdakao-Usansolo, Bizkaia, Spain
PEDROMARIA.OLAECHEAASTIGARRAG@osakidetza.net
[6] Service of Intensive Care Medicine, Parc de Salut Mar, Barcelona, Spain
FAlvarez@parcdesalutmar.cat

Analyzing the determinants and consequences of hospital acquired infections involves the evaluation of large cohorts. Infected individuals in the cohort are often rare for specific sites or pathogens, since most of the patients admitted to the hospital are discharged or die without such an infection. Death/discharge is a competing event to acquiring an infection since these individuals are no longer at risk to get infected. Therefore the data is best analyzed with an extended survival model - the so called illness-death model. A common problem in cohort studies is the costly collection of covariate values. To provide efficient use of data from infected as well as uninfected patients, we propose a tailored case-cohort approach for the illness-death model . The basic idea of the case-cohort is to use only a random sample of the full cohort, referred to as subcohort, and all cases – the infected individuals. Thus covariate values are only obtained for a small part of the full cohort. The method is based on existing and established methods and is used to perform regression analysis in adapted Cox proportional hazards models. We propose estimation of the cause-specific cumulative hazards and transition probabilities in an illness-death model based on case-cohort sampling. As an example we apply the methodology to infection with a specific pathogen using a large cohort from Spanish hospital data. The obtained results of the case-cohort design are compared to the results in the full cohort to investigate how well the case-cohort design performs.

Fortentwicklung von Diagrammen für Fortgeschrittene

Vonthein, Reinhard

Universität zu Lübeck, Deutschland
Reinhard.Vonthein@imbs.uni-luebeck.de

Oft werden Diagramme im Anfängerunterricht bis zu den Lernzielen Kennen, Verstehen und Anwenden unterrichtet. Anfänger lehrt man zu verstehen, warum alle Daten gezeigt werden sollen, und warum eine Darstellung dem Skalentyp angemessenen ist. Weitere Merkmale werden durch Vervielfachen, durch Farben, Linientypen und Symbole berücksichtigt. Der hier vorgeschlagene Fortgeschrittenenunterricht, etwa während eines praktischen Kurses in einem Masterstudiengang, hat die Lernziele Analysieren, Bewerten und Fortentwickeln grafischer Darstellungen.

Für die Darstellung einer metrischen Variablen je nach Ausprägung einer kategorialen Variablen zum Beispiel stehen mehrere gute Möglichkeiten zur Verfügung. Während eines praktischen Kurses werden die Teilnehmer ihre Daten grafisch explorieren. Die Entscheidung für die eine oder andere Darstellung wird man dann diskutieren und weitere Kriterien vorschlagen.

Die Analyse beginnt damit, dass ein Diagramm aus mehreren Graphemen besteht, welche Information tragende Eigenschaften, sog. aethetics, haben. Der Kern des Ansatzes besteht darin, zu fragen: Werden Grapheme für Daten, Schätzergebnisse oder beides gezeigt? Passt das Schätzungs-Graphem zur parametrischen, robusten oder nichtparametrischen Schätzung in Text und Tabellen? Wie groß darf die Zahl der Beobachtungen mindestens und höchstens sein? In Lehrgespräch oder Impulsvortrag zur gemeinsamen Verteilung eines metrischen und eines kategorialen Merkmals kann man Grapheme Boxplot, zentriertes Histogramm, Violin-Plot, Battleship-Plot und Fehlerbalken nach der Stärke der Voraussetzungen diskutieren. Die übersicht darüber, was wann machbar ist, erklärt dann, warum diese Vielfalt kaum genutzt wird.

Das Beispiel wurde so gewählt, dass es eine oft bestehende Lücke füllt und doch eine lösbare Aufgabe bleibt. Die übertragung auf unterschiedliche Arten von Regression (linear, logistisch, additiv, Splines, ...) bleibt Teil der Teilnehmerprojekte.

Bisher liegen Erfahrungen mit diesem Ansatz nur aus der nicht ganz vergleichbaren Biometrischen Beratung in Tübingen vor. Die Informationsdichte der Diagramme stieg, insbesondere wegen der Darstellung der Variabilität der Messungen.

Assessing predictive performance in multiply imputed data using resampling strategies

Wahl, Simone [1,2] Boulesteix, Anne-Laure [3]

[1] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Germany
simone.wahl@helmholtz-muenchen.de
[2] Institute of Epidemiology II, Helmholtz Zentrum München - German Research Centre for Environmental Health, Germany
simone.wahl@helmholtz-muenchen.de
[3] Department of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany
boulesteix@ibe.med.uni-muenchen.de

Missing values are frequent in human studies, and occur for different reasons including lack of biological samples, laboratory detection limits, data entry errors and incomplete response to a questionnaire. In many situations, multiple imputation is an appropriate strategy to deal with missing values. It involves three steps: (1) generation of M sets of imputed data sets using imputation techniques, (2) application of the statistical method of choice to each imputed data set, and (3) combination of the M obtained parameter estimates Q_m, $m = 1, \ldots, M$, and their variances, thereby accounting for the uncertainty about the imputed values. A well-established way to achieve step (3) are the combination rules by Rubin (1987) [1]. They rely on the assumption that the estimate Q is normally distributed, and on the availability of an estimate for the variance of Q.

When the interest is in estimating the predictive ability of a model, or the added predictive ability of a set of variables, by means of performance criteria such as AUC (i.e., the area under the receiver operating characteristic curve) and pseudo R^2 measures or changes in these criteria (ΔAUC, ΔR^2), respectively, it is less well established how the performance criteria obtained from the M imputed data sets should be combined, and how significance in (added) predictive performance might be assessed.

Bootstrap methods are useful to obtain non-parametric variance estimates and confidence intervals, and to internally validate a model by evaluating it on the out-of-bag observations not included in the bootstrap sample. It has not been well explored how these approaches should be used appropriately in the context of multiple imputation. In extensive simulations, we compare different resampling strategies in terms of unbiased estimation of different performance criteria and confidence interval coverage.

References:

[1] Rubin DB (1987). Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons.

Abwandlung eines Maßes der Erklärten Variation für Überlebenszeitdaten

Weiß, Verena Hellmich, Martin

IMSIE, Uniklinik Köln, Deutschland
verena.weiss@uni-koeln.de, martin.hellmich@uni-koeln.de

In der linearen Regression wird häufig das Bestimmtheitsmaß R^2 verwendet, um mithilfe einer Maßzahl anzugeben, wie gut das Modell die vorliegenden Daten beschreibt. Aufgrund von Zensierungen und einer möglichen Schiefe der Daten lässt sich dieses Maß jedoch nicht bei Daten der überlebenszeitanalyse anwenden. Daher wurden in den letzten Jahren verschiedene Maße der Erklärten Variation für überlebenszeitdaten vorgeschlagen. Eines davon ist das Maß V_1 von Schemper [1, 2].

Ausgehend von dem Maß V_1 wird ein neues Maß der Erklärten Variation entwickelt und vorgestellt. In dem neuen Maß V_1^* wird einerseits eine andere Abstandsdefinition der einzelnen Personenkurve zu dem Kaplan-Meier-Schätzer gewählt und andererseits wird eine kategorielle Kovariate berücksichtigt, indem für eine Person der Abstand der einzelnen Personenkurve zu dem Wert des Kaplan-Meier-Schätzers der Gruppe, zu welcher die Person gemäß der Kovariate gehört, bestimmt wird. Dadurch berücksichtigt das neue Maß eine Kovariate lediglich über den Kaplan-Meier-Schätzer und ist somit vollständig verteilungsfrei. Für die Bestimmung des Maßes V_1 muss dagegen die Annahme proportionaler Hazards erfüllt sein, da das Cox-Modell in dieses Maß eingeht.

Weiterhin wird für das neue Maß V_1^* ein zugehöriger statistischer Signifikanztest vorgeschlagen, mit dem für zwei Kovariaten untersucht werden kann, ob eine Kovariate statistisch signifikant mehr Variation aufklärt als eine andere Kovariate. Angelehnt an die verschiedenen Gewichtungsmethoden des Logrank-Tests werden für das neue Maß verschiedene Möglichkeiten der Gewichtung der Abstände angegeben, so dass sich die Abstände an frühen Zeitpunkten stärker gewichten lassen als die Abstände an späten Zeitpunkten.

References:

[1] Schemper M (1990). The explained variation in proportional hazards regression. Biometrika. 77(1): 216-218

[2] Schemper M (1994). The explained variation in proportional hazards regression - correction. Biometrika. 81(3): 631

Benefits of an integrative strategy for extracting sparse and non-sparse prognostic information from molecular measurements

Weyer, Veronika Binder, Harald

Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center Mainz weyer@uni-mainz.de, binderh@uni-mainz.de

Regularized regression approaches can be used for extracting sparse prognostic signatures. Furthermore, non-sparse hierarchical clustering, based on distance measures, is often used for identifying prognostic groups from molecular measurements. As both, a sparse and a non-sparse approach might be useful for extracting different kinds of information, we investigate the potential advantages of combining both strategies. Specifically we use all non-sparse information detected via cluster analysis in a stratified Cox regression by combining it with a weighting approach via componentwise boosting for automated variable selection for integrating sparse gene information. As an application we consider RNA-Seq gene expression data from acute myeloid leukemia (AML) and kidney renal cell carcinoma (KIRC) patients for examining our stratified, weighted regression approach. The KIRC data were also used for integrating the information of two biological levels (methylation and gene expression). When focusing on risk prediction development in a specific cluster, the observations of all clusters are included in the analysis by down-weighting the observations of the not-analyzed clusters in a weighted partial log-likelihood with each cluster getting its own baseline hazard. Variable selection stability is evaluated based on resampling inclusion frequencies and presented via two visualization tools. We specifically investigate to what extent non-sparse and sparse components of the modeling strategy respectively provide improvement in terms of prediction performance. In both applications prediction performance can be improved adding the non-sparse information via cluster based stratification. Further improvement in prediction performance via the weighted sparse information from the not-analyzed cluster is marginal but weighting has a positive effect on the selection stability. Overall, the combination of sparse and non-sparse information is seen to be useful for gene signature development. Subgroup specific sparse signatures provide only a little benefit in terms of prediction, while they still might be useful for more stably identifying important genes.

Choosing the shrinkage factor in Bayesian logistic regression with variable selection

Wiesenfarth, Manuel [1] Zucknick, Manuela [1] Corberán-Vallet, Ana [2]

[1] DKFZ, Deutschland
m.wiesenfarth@dkfz.de, m.zucknick@dkfz.de
[2] Universidad de Valencia, Spanien
ana.corberan@gmail.com

Bayesian variable selection can be formulated in terms of a spike-and-slab prior on the regression coefficients. Thereby, selected variables are assumed to a priori follow a slab Gaussian distribution with relatively spread support while a point-mass at zero is assigned to the non-selected ones (e.g. [1]). It is well known that a small slab variance shrinks the selected coefficients to the prior mean while on the other hand, posterior selection probabilities of the covariates tend to zero with increasing slab variance leading to the Bartlett-Lindley paradox. This is further aggravated in logistic regression [2] where a large prior variance may induce an implausible prior distribution on the response variable. Therefore, current proposals for hyperpriors assigned to this variance may not be optimal for binary outcomes. We explore various approaches to hyperpriors on the slab variance in logistic regression putting special emphasis on models with high-dimensional molecular data which adds an additional layer of complexity.

References:

[1] George, E. I., & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. Statistica sinica, 7(2), 339-373.

[2] Holmes, C. C., & Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. Bayesian Analysis, 1(1), 145-168.

„Spielen sie schon oder is' grad piano?" – Zur Wahrnehmung der Medizinischen Biometrie

Windeler, Jürgen

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln

Der Vortrag thematisiert das Verhältnis von Evidenzbasierter Medizin und Medizinischer Biometrie.

The covariance between genotypic effects and its use for genetic evaluations in large half-sib families

Wittenburg, Dörte Teuscher, Friedrich Reinsch, Norbert

Leibniz-Institut für Nutztierbiologie, Deutschland
wittenburg@fbn-dummerstorf.de, teuscher@fbn-dummerstorf.de, reinsch@fbn-dummerstorf.de

In livestock, current statistical approaches involve extensive molecular data (e.g. SNPs) to reach an improved genetic evaluation. As the number of model parameters increases with a still growing number of SNPs, multicollinearity between covariates can affect the results of whole genome regression methods. The objective of this study is to additionally incorporate dependencies on the level of genome, which are due to the linkage and linkage disequilibrium on chromosome segments, in appropriate methods for the estimation of SNP effects. For this purpose, the covariance among SNP genotypes is investigated for a specific livestock population consisting of half-sib families. Conditional on the SNP haplotypes of the common parent, the covariance can be theoretically derived knowing the recombination fraction and haplotype frequencies in the population the individual parent comes from. The resulting covariance matrix is included in a statistical model for some trait of interest. From a likelihood perspective, such a model is similar to a first-order autoregressive process but the parameter of autocorrelation is explicitly approximated from genetic theory and depends on the genetic distance between SNPs. In a Bayesian framework, the covariance matrix can be used for the specification of prior assumptions of SNP effects. The approach is applied to realistically simulated data resembling a half-sib family in dairy cattle to identify genome segments with impact on performance or health traits. The results are compared with methods which do not explicitly account for any relationship among predictor variables.

Investigating the effects of treatment for community-acquired infections on mortality in hospital data: a simulation study for the impact of three types of survival bias

Wolkewitz, Martin Schumacher, Martin

Department für Medizinische Biometrie und Medizinische Informatik, Deutschland
wolke@imbi.uni-freiburg.de, ms@imbi.uni-freiburg.de

The current debate whether oseltamivir (Tamiflu) is an effective antiviral drug for treating H1N1 influenza has been stirred up by a meta-analysis of observational studies claiming that Tamiflu reduced mortality [1]. In the analysis, a survival model was applied up to 30 days from influenza onset (time origin); within this time frame, the treatment was handled as a time-dependent covariate in order to avoid time-dependent bias. This finding was based on hospital data.

In addition to time-dependent bias, we consider two other types of survival bias which might occur when using hospital data. First, hospital admission is usually a few days after influenza onset; ignoring this in the analysis may lead to length bias. Second, discharged patients cannot simply be handled as censored since they are usually in a better health condition than hospitalized patients; they should be handled as competing events. Classical survival models such as Kaplan-Meier curves fail to address these issues. We propose a multi-state model (onset, admission, treatment, discharge and death) and use a simulation study to investigate the impact of bias due to ignoring the time-dependency of treatment, to left-truncation and to competing events. The impact differs in magnitude and direction and will be displayed in isolation as well as in combination.

References:

[1] Muthuri SG, et al. Effectiveness of neuraminidase inhibitors in reducing mortality in patients admitted to hospital with influenza A H1N1pdm09 virus infection: a meta-analysis of individual participant data. Lancet Resp Med 2014;2:395-404.

Qualitätsveränderungen der Wissenschaftskommunikation am Beispiel medizinischer Themen

Wormer, Holger Anhäuser, Marcus Serong, Julia

TU Dortmund, Deutschland
holger.wormer@tu-dortmund.de, marcus.anhaeuser@tu-dortmund.de, julia.serong@tu-dortmund.de

Im Zuge der ökonomisierung und Medialisierung von Wissenschaft orientiert sich die Wissenschaftskommunikation zunehmend an den Mechanismen einer medialen Aufmerksamkeitsökonomie (vgl. Franzen/Rödder/Weingart 2012). Der Wissenschaftsjournalismus ist gefragt, diese Entwicklungen kritisch zu beobachten, gerät jedoch seinerseits unter ökonomischen Druck. Empirische Studien geben Hinweise darauf, dass der Journalismus in erheblichem Maße von der Wissenschaftskommunikation der Forschungseinrichtungen und Fachzeitschriften abhängt (vgl. Stryker 2002, Schwartz/Woloshin et al. 2012, Yavchitz et al. 2012). Da Studienabstracts sowie Pressemitteilungen zudem nicht mehr nur Fachleuten und professionellen Journalisten zur Verfügung stehen, sondern über das Internet einem breiten Laienpublikum direkt zugänglich sind, stellt sich die Frage nach den normativen Qualitätsansprüchen an Wissenschaftskommunikation mit wachsender Dringlichkeit.

Im Rahmen eines Forschungsprojektes haben wir untersucht, wie sich die Informationsqualität im Prozess der Wissenskommunikation von der Fachpublikation über die Pressemitteilung bis zum journalistischen Beitrag verändert. Medizinische und wissenschaftsjournalistische Experten bewerten die Abstracts der Studien, Pressemitteilungen und journalistischen Beiträge mithilfe eines Kriterienkatalogs, der wissenschaftliche und journalistische Normen und Standards berücksichtigt. Auf diese Weise soll ermittelt werden, ob und inwiefern sich die im Medizinjournalismus erprobten Qualitätskriterien auf andere Stufen der Wissenschaftskommunikation übertragen lassen.

Literatur:

Rödder, S. / Franzen, M. / Weingart, P. (Hrsg.) (2012): The Sciences' Media Connection: Public Com-munication and Its Repercussions. Bd. 28. Dordrecht: Springer.

Schwartz, L. / Woloshin, S. et al. (2012).; Influence of medical journal press releases on the quality of associated newspaper coverage: retrospective cohort study. In: British Medical Journal 2012: 344: d8164. doi: 10.1136/bmj.d8164.

Stryker, J. (2002): Reporting Medical Information: Effects of Press Releases and Newsworthiness on Medical Journal Articles' Visibility in the News Media. In: Preventive Medicine, Jg. 35: 519-530.

Yavchitz, A. et al. (2012): Misrepresentation of Randomized Controlled Trials in Press Releases and News Coverage: A Cohort Study. In: PLOS Medicine, Jg. 9, Nr. 9: e1001308.

Ranger: A Fast Implementation of Random Forests for High Dimensional Data

Wright, Marvin N. [1] Ziegler, Andreas [1,2,3]

[1] Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Germany
wright@imbs.uni-luebeck.de,ziegler@imbs.uni-luebeck.de
[2] Zentrum für Klinische Studien, Universität zu Lübeck, Germany
ziegler@imbs.uni-luebeck.de [3] School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa
ziegler@imbs.uni-luebeck.de

Random Forests are widely used in biometric applications, such as gene expression analysis or genome-wide associations studies (GWAS). With currently available software, the analysis of high dimensional data is time-consuming or even impossible for very large datasets. We therefore introduce Ranger, a fast implementation of Random Forests, which is particularly suited for high dimensional data. We describe the implementation, illustrate the usage with examples and compare runtime and memory usage with other implementations. Ranger is available as standalone C++ application and R package. It is platform independent and designed in a modular fashion. Due to efficient memory management, datasets on genome-wide scale can be handled on a standard personal computer. We illustrate this by application to a real GWAS dataset. Compared with other implementations, the runtime of Ranger proves to scale best with the number of features, samples, trees, and features tried for splitting.

Nonparametric meta-analysis for diagnostic accuracy studies

Zapf, Antonia [1] Hoyer, Annika [2] Kramer, Katharina [1]
Kuss, Oliver [2]

[1] Universitätsmedizin Göttingen, Deutschland
antonia.zapf@med.uni-goettingen.de, katharina.kramer@med.uni-goettingen.de
[2] Deutsches Diabeteszentrum Düsseldorf, Deutschland
annika.hoyer@ddz.uni-duesseldorf.de, oliver.kuss@ddz.uni-duesseldorf.de

Summarizing the information of many studies using a meta-analysis becomes more and more important, also in the field of diagnostic studies. The special challenge in meta-analysis of diagnostic accuracy studies is that in general sensitivity (as true positive fraction) and specificity (as true negative fraction) are co-primary endpoints. Within one study both endpoints are negatively correlated, because of the threshold effect [1]. This leads to the challenge of a bivariate meta-analysis.

In the talk we will present a new, fully nonparametric approach for the meta-analysis of diagnostic accuracy studies. The method borrows from the analysis of diagnostic studies with several units per individual [2], where analogue dependency structures are present. The nonparametric model has the advantage that compared to competing approaches less assumptions regarding the distribution functions and the correlation structure have to be made. Furthermore the approach is expected to be numerically more stable than competing approaches, can deal with studies with only sensitivity or specificity and is also applicable to diagnostic studies with factorial design.

In a simulation study it becomes apparent that results of the nonparametric meta-analyses are in general unbiased, the type I error is kept, and the mean squared error (MSE) is quite small. The approach is also applied to an example meta-analysis regarding the diagnostic accuracy of telomerase for the diagnosis of primary bladder cancer.

References:

[1] Leeflang, Deeks, Gatsonis, Bossuyt - on behalf of the Cochrane Diagnostic Test Accuracy Working Group (2008). Systematic reviews of diagnostic test accuracy. Annals of Internal Medicine 149, 889–897.

[2] Lange (2011). Nichtparametrische Analyse diagnostischer Gütemaße bei Clusterdaten. Doctoral thesis, Georg-August-University Göttingen.

Didaktische Umstrukturierung der Grundvorlesung Biometrie und Epidemiologie – Erfahrungen aus der Veterinärmedizin

Zeimet, Ramona [1] Doherr, Marcus G. [2] Kreienbrock, Lothar [1]

[1] Stiftung Tierärztliche Hochschule Hannover, Deutschland
Ramona.Zeimet@tiho-hannover.de, Lothar.Kreienbrock@tiho-hannover.de
[2] Freie Universität Berlin, Deutschland
Marcus.Doherr@fu-berlin.de

Die Lehre statistischer Konzepte in der Veterinärmedizin (wie auch in der Humanmedizin) stellt eine Herausforderung für Dozierende dar, da die Studierenden gewöhnlich einen sehr heterogenen mathematischen Grundkenntnisstand aufweisen. Zudem liegt das primäre Interesse der Studierenden darin, Patienten zu untersuchen und zu behandeln. Infolgedessen beschäftigen wir uns an den veterinärmedizinischen Bildungsstätten vermehrt mit der Frage, welche biometrischen und epidemiologischen Inhalte für die Studierenden von zentraler Relevanz sind und wie diese Themen vermittelt werden sollten, um sie attraktiv, nachhaltig verfügbar und anwendbar zu machen.

Aus einer Online-Befragung, die an allen acht veterinärmedizinischen Bildungsstätten in Deutschland (5), österreich (1) und der Schweiz (2) durchgeführt wurde, konnte ein zentraler Themenkatalog für die Grundvorlesung Biometrie und Epidemiologie entwickelt werden. Dieser umfasst jene Kenntnisse, die Studierende der Veterinärmedizin im Rahmen der Grundvorlesung erlangen sollten, um für die anschließenden Fachvorlesungen gerüstet zu sein. Basierend auf diesem Katalog, wurden konkrete operationalisierte Lernziele für die Grundvorlesung Biometrie und Epidemiologie formuliert und die Veranstaltungsstruktur entsprechend didaktisch umstrukturiert.

Nachdem im Sommersemester 2014 an der Tierärztlichen Hochschule Hannover und an der Freien Universität Berlin die Grundvorlesung Biometrie und Epidemiologie in neuer Struktur und unter Einbeziehung kreativer und aktivierender Lehrmethoden und -materialien erfolgt ist, können wir durch den Vergleich zum Vorjahr und zwischen den Hochschulstandorten ein Fazit hinsichtlich des Nutzens und Erfolgs didaktischer Methoden in der Biometrie- und Epidemiologie- Lehre im Rahmen des Kolloquiums präsentieren.

Tree and Forest Approaches to Identification of Genetic and Environmental Factors for Complex Diseases

Zhang, Heping

Yale School of Public Health, United States of America
heping.zhang@yale.edu

Multiple genes, gene-by-gene interactions, and gene-by-environment interactions are believed to underlie most complex diseases. However, such interactions are difficult to identify. While there have been recent successes in identifying genetic variants for complex diseases, it still remains difficult to identify gene-gene and gene-environment interactions. To overcome this difficulty, we propose forest-based approaches and new concepts of variable importance, as well as other implications in the construction of trees an forests. The proposed approaches are demonstrated by simulation study for its validity and illustrated by real data analysis for its applications. Analyses of both real data and simulated data based on published genetic models have revealed the potential of our approaches to identifying candidate genetic variants for complex diseases.

Smoothing for pseudo-value regression in competing risks settings

Zöller, Daniela [1] Schmidtmann, Irene [1] Weinmann, Arndt [2]
Gerds, Thomas [3] Binder, Harald [1]

[1] Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center Mainz, Germany
daniela.zoeller@uni-mainz.de, ischmidt@uni-mainz.de,binderh@uni-mainz.de
[2] First department of Medicine, University Medical Center Mainz, Germany
weinmann@mail.uni-mainz.de
[3] Department of Public Health, University Copenhagen, Denmark
tag@biostat.ku.dk

In a competing risks setting for time-to-event data the cumulative incidence, i.e. the proportion of an event type observed up to a certain time, is an easily interpretable quantity and patients can be compared with respect to their cumulative incidence for different competing risks. One way to analyze effects on the cumulative incidence is by using a pseudo-value approach, which allows, among others, the application of regression with time-varying effects. As the resulting effect estimates are often not smooth over time, which is not reasonable for a cumulative quantity, we present a smoothing algorithm based on pairwise switches of the first observed event times and its impact on the estimates. In order to improve the interpretability we additionally present confidence bands for the separate estimates based on the resampling method.

These methods are specifically illustrated for a stagewise regression algorithm based on pseudo-values for the cumulative incidence estimated at a grid of time points for a time-to-event setting with competing risks and right censoring. This algorithm provides an approach for estimating the effect of covariates in the course of time by coupling variable selection across time points but allowing for separate estimates. We apply the algorithm to clinical cancer registry data from hepatocellular carcinoma patients.

The use of smoothing is seen to improve interpretability of results from pseudo-value regression which enables the estimation of model parameters that have a straightforward interpretation in particular in combination with smooth confidence bands. Additionally time-varying effects on the cumulative incidence can be judged with the help of the confidence bands, as illustrated for the application.

Posters

Antibiotic Prophylaxis in Inguinal Hernia Repair - Results of the Herniamed Registry

Adolf, Daniela [1] Köckerling, Ferdinand [2]

[1] StatConsult GmbH
daniela.adolf@statconsult.de
[2] Department of Surgery and Center for Minimally Invasive Surgery, Vivantes Hospital Berlin
ferdinand.koeckerling@vivantes.de

"The Herniamed quality assurance study is a multi-center, internet-based hernia registry into which 383 participating hospitals and surgeons in private practice in Germany, Austria and Switzerland enter data prospectively on their patients who had undergone hernia surgery.

Previous prospective randomized trials have identified postoperative infection rates of between 0 and 8.9 % in the absence of antibiotic prophylaxis, and of between 0 and 8.8 % on administration of antibiotic prophylaxis [1]. So since the use of antibiotic prophylaxis in inguinal hernia surgical repair is a controversial issue, we used data from the Herniamed Registry to analyze the influence of antibiotic prophylaxis on the postoperative outcome. Furthermore, we explored the influence that the use of an endoscopic technique for repair of inguinal hernia had on the rate of impaired wound healing and on deep infections with mesh involvement compared with open surgery.

The results being presented here are based on the prospectively collected data on all patients who had undergone unilateral, bilateral or recurrent repair of inguinal hernia using either endoscopic or open techniques between 1 September, 2009 and 5 March, 2014. In addition, formal as well as statistical challenges in analyzing these registry data will be discussed.

References:

[1] Bittner R, Arregui ME, Bisgaard T, Dudai M, Ferzli GS, Fitzgibbons RJ, Fortelny RH, Klinge U, Koeckerling F, Kuhry E, Kukleta J, Lomanto D, Misra MC, Montgomery A, Morales-Condes S, Reinpold W, Rosenberg J, Sauerland S, Schug-Pass C, Singh K, Timoney M, Weyhe D, Chowbey P (2011). Guidelines for laparoscopic (TAPP) and endoscopic (TEP) treatment of inguinal hernia. Surg Endosc; Sep; 25 (9): 2773-843.

Identification of patient subgroups in high-throughput data

Ahrens, Maike [1,2] Turewicz, Michael [2] Marcus, Katrin [2]
Meyer, Helmut E. [3] Eisenacher, Martin [2] Rahnenführer, Jörg [1]

[1] Department of Statistics, TU Dortmund University, Germany
maike.ahrens@tu-dortmund.de, rahnenfuehrer@statistik.tu-dortmund.de
[2] Medizinisches Proteom-Center, Ruhr University Bochum, Germany
maike.ahrens@tu-dortmund.de, michael.turewicz@rub.de, katrin.marcus@rub.de, martin.eisenacher@rub.de
[3] Leibniz-Institut für Analytische Wissenschaften - ISAS - e.V., Dortmund, Germany
helmut.e.meyer@isas.de

We present methods for the detection of patient subgroups in molecular data. A subgroup is a subset of patients with differential gene or protein expression in a limited number of features compared to all other patients. The goal is to identify the subgroup, but also the subgroup-specific set of features. Applications range from cancer biology over developmental biology to toxicology. In cancer, different oncogenes initiating the same tumor type can cause heterogeneous expression patterns in tumor samples. The identification of subgroups is a step towards personalized medicine. Different subgroups can indicate different molecular subtypes of a disease which might correlate with disease progression, prognosis or therapy response, and the subgroup-specific genes or proteins are potential drug targets.

We present univariate methods that score features for subgroup-specific expression one by one, as well as multivariate methods that group features which consistently indicate the same subgroup. We conducted a comprehensive simulation study to compare several univariate methods [1]. For many parameter settings, our new method FisherSum (FS) outperforms previously proposed methods like the outlier sum or PAK (Profile Analysis using Kurtosis). We also show that Student's t-test is often the best choice for large subgroups or small amounts of deviation for the subgroup. For the multivariate case we compare classical biclustering to a new method called FSOL. This algorithm combines the scoring of single features with FS and the comparison of features regarding their induced subgroups with the OrderedList method [2]. Results from simulations as well as from a real data example are presented.

References:

[1] Ahrens M, Turewicz M, Casjens S, May C, Pesch B, Stephan C, Woitalla D, Gold R, Brüning T, Meyer HE, Rahnenführer J, Eisenacher M. (2013). PLOS ONE 8 (11).

[2] Yang X, Scheid S, Lottaz C (2008). OrderedList: Similarities of Ordered Gene Lists. R package version 1.38.0.

Estimation of Total Cholesterol Level by Kriging Metamodeling

Aydin, Olgun

Mimar Sinan University, Turkey
olgunaydinn@gmail.com

Cholesterol levels should be measured at least once every five years in everyone who is over age 20. The tests which are usually performed is a blood test called a lipid profile. Experts recommend that men are 35 years old or older and women are 45 years old and older should be frequently screened for lipoprotein profile includes; total cholesterol, LDL (low-density lipoprotein cholesterol), HDL (high-density lipoprotein cholesterol), triglycerides (triglycerides are the form in which fat exists in food and the body).[1]

In this study, aimed to estimate total cholesterol level by the aid of Kriging metamodeling. In this model; systolic, diastolic blood pressure, waist circumference, hip size, height, weight and age were used as independent variables and total cholesterol level was used as dependent variable. This data were provided from 1956 people who are in 17 - 68 age range. All the data were collected by nurses from the people who applied to a public health care center in three months time period within the scope of general health screening.

By this study total cholesterol level was modeled by using Kriging. For verifying the model, differences were calculated by dividing observed total cholesterol levels to their estimated values predicted with excluding related data points. Mean of the differences have been found 1.016, and standart deviation of differences have been found 0.1529. It means, total cholesterol level was estimated averagely %1.6 higher than observed total cholesterol level.

By the agency of this model, people would have had an idea about their total cholesterol levels not to need blood tests. They could put their age, weight, height, systolic and diastolic blood pressure, waist circumference, hip size into this model and their total cholesterol level would have been estimated. By this means, this model would have been an early warning system for too high or too low total cholesterol levels.

References:

[1] What is cholesterol? National Heart, Lung, and Blood Institute. http://www.nhlbi.nih.gov/health/health-topics/topics/hbc/.

A weighted polygenic risk score to predict relapse-free survival in bladder cancer cases

Bürger, Hannah [1,2] Golka, Klaus [1] Blaszkewicz, Meinolf [1]
Hengstler, Jan G. [1] Selinski, Silvia [1]

[1] Leibniz-Institut für Arbeitsforschung an der TU Dortmund (IfADo), Deutschland
hannah.buerger@tu-dortmund.de, golka@ifado.de, blaszkewicz@ifado.de, hengstler@ifado.de, selinski@ifado.de
[2] Fakultät Statistik, TU Dortmund, Deutschland
hannah.buerger@tu-dortmund.de

Bladder cancer (BC) is a smoking and occupational related disease showing a substantial genetic component. Though the prognosis is generally good, a major problem are the frequent relapses affecting about half of the patients. Meanwhile a panel of susceptibility SNPs for BC development has been discovered and validated by GWAS (1). We aim to investigate the common impact of these BC SNPs on relapse-free survival using a weighted polygenic risk score (PRS).

We used follow-ups of three case-control studies from Lutherstadt Wittenberg (n=205), Dortmund (n=167) and Neuss (n=258) without missing values for age, gender, smoking habits and 12 polymorphisms (GSTM1 deletion, rs1014971, rs1058396, rs11892031, rs1495741, rs17674580, rs2294008, rs2978974, rs710521, rs798766, rs8102137, rs9642880). Adjusted (age, gender, smoking habits, study group) Cox proportional hazards models predicting relapse-free survival were fitted for each polymorphism assuming an additive mode of inheritance. A PRS was calculated as the weighted sum of risk alleles across all polymorphisms. Weights for the individual polymorphisms were their adjusted log HRs. Association of the PRS (as linear predictor) with relapse-free survival was tested adjusted for age, gender, smoking habits and study group. Quartiles of the PRS were used for Kaplan-Meier curves.

For 349 (55%) of the cases at least one relapse was confirmed. Higher PRS was associated significantly with shorter recurrence-free time (adjusted HR=2.093, p=0.000098). Kaplan-Meier curves of PRS quartiles showed shortest relapse-free survival times for the upper quartile, especially in the first 5 years after first diagnosis, whereas the lowest 75% PRS showed quite similar survival times.

Weighted PRS that are common in case-control studies (1) are also useful to investigate common effects of polymorphism panels in survival analysis.

References:

(1) Garcia-Closas et al. (2013). Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer. Cancer Res 73:2211-20.

Variable selection under multiple imputation and bootstrapping – an application to self-reported quality of life in adolescents with cerebral palsy

Eisemann, Nora

Universität zu Lübeck, Deutschland
nora.eisemann@krebsregister-sh.de

The exclusion of cases with missing data can lead to biased results. An alternative to complete case analysis is multiple imputation, where each missing value is replaced by multiple predicted values, resulting in several completed data sets. When conducting variable selection in a regression analysis, backward elimination is not straightforward in such a situation because it can produce different sets of selected variables on each of the multiple data sets. A final set of variables can then be chosen depending on their inclusion frequency, and the results from the regression analyses can be combined. When multiple data sets are already generated because of missing data, the framework can easily be extended to include further data sets derived by bootstrapping [1]. This reduces overfitting of the regression model which may otherwise occur due to sampling error.

This procedure is illustrated by an application to a data set on self-reported quality of life in adolescents with cerebral palsy with missing data in the possible predictors of quality of life [2]. Missing values were imputed ten times by multiple imputation with chained equations. From each completed data set, 10 bootstrap data sets were generated with replacement. The inclusion frequency of the predictors was counted after backward elimination, and predictors selected at least in 60% of the data sets were chosed for the final set of predictors.

Although the combination of multiple imputation and bootstrapping for variable selection is attractive and does not require much additional effort, it is not widely applied in practice yet.

References:

[1] Heymans, van Buuren et al: Variable selection under multiple imputation using the bootstrap in a prognostic study. BMC MedResMethodol 2007, 7:33.

[2] Colver, Rapp et al: Self-reported quality of life of adolescents with cerebral palsy: a cross-sectional and longitudinal analysis. Lancet 2014.

Impact of a clinical hold with treatment suspension on primary objective evaluation

Erhardt, Elvira [1] Beyersmann, Jan [1] Loos, Anja-Helena [2]

[1] Universität Ulm, Deutschland
elvira.erhardt@uni-ulm.de, jan.beyersmann@uni-ulm.de
[2] Merck KgaA
Anja-Helena.Loos@merckgroup.com

Abstract: This work is motivated by a real situation faced during the phase 3 START study where the FDA issued a clinical hold (CH). No new patients could be recruited and investigational treatment had to be stopped.[1,2] Primary endpoint was overall survival time. After the CH was lifted, all subjects randomized 6 months prior to the CH were excluded to avoid potential bias. Additional patients were recruited to reach the pre-specified number of events for final analysis.[3]

Although CHs do not seem to be rare, there is no specific statistical guidance on how to deal with it. Not accounting for the CH adequately as an external time-dependent exposure in trials with time-to-event variables, can lead to "time-dependent bias"/"immortal time bias".[4,5]

While START did not show improvement in overall survival time, one might ask whether the CH impacted trial results or an analysis method accounting adequately for the treatment suspension could have been used.

Results of a simulation study, close to the START setting, will be shared. The study has explored the impact of the treatment suspension and the evaluation of a multi-state illness-death model as alternative statistical method to account for the CH.

References:

[1] FDA (2000). Guidance for Industry. Submitting and Reviewing Complete Responses to Clinical Holds.
http://www.fda.gov/downloads/Drugs/GuidanceComplinceRegulatory Information/Guidances/UCM078764.pdf. Accessed November 2014

[2] FDA (2013). IND Application Procedures: Clinical Hold.
http://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsare DevelopedandApproved/ApprovalApplications/InvestigationalNewDrug INDApplication/ucm362971.htm. Accessed November 2014

[3] Butts C et al (2014). Tecemotide (L-BLP25) versus placebo after chemoradiotherapy for stage III non-small-cell lung cancer (START): a randomised, double-blind, phase 3 trial. Lancet Oncol., 15:59–68.

[4] Wolkewitz M et al (2010).Two Pitfalls in Survival Analyses of Time- Dependent Exposure: A Case Study in a Cohort of Oscar Nominees. Am. Stat., 64:205-211

[5] Sylvestre M et al (2006). Do OSCAR Winners Live Longer Than Less Successful Peers? A Reanalysis of the Evidence. Ann. Intern. Med., 145:361-363.

Optimizing graphical multiple testing procedures

Görlich, Dennis Faldum, Andreas Kwiecien, Robert

Westfälische Wilhelms-Universität Münster, Deutschland
dennis.goerlich@uni-muenster.de, andreas.faldum@ukmuenster.de,
robert.kwiecien@ukmuenster.de

For clinical trials with multiple endpoints a large variety of multiple testing procedures are available. In certain situations the sequential rejective testing procedures (SRTP) proposed by Bretz et al [1,2] are an appealing tool to plan clinical trials controlling the family-wise-error-rate. In general, the structure of the testing procedure will be developed by the study statistican using information about the importance of hypotheses and other constraints. Nevertheless, often it is not completely clear whether there might exist a better design with the same statistical properties answering the trial hypotheses. We, here, propose an evolutionary algorithm to optimize graphical statistical trial designs with respect to the procedures power. In particular, we optimize the node and edge weights of the underlying graph. Different measures of power can be used for optimization, i.e., at-least-one rejection, rejection of selected hypotheses, and rejection of all hypotheses. The evolutionary algorithm is a simple evolutionary (3,1)-strategy without recombination. In each generation the best (= highest power) individual (=SRTP) replaces the worst (=lowest power) and is mutated randomly. Fitness is evaluated by simulations of 10000 clinical trials per generation. To evaluate a SRTP the gMCP package [3] for R is used. We will discuss the potential to optimize SRTPs by our algorithm

References:

[1] Bretz F, Maurer W, Brannath W, Posch M (2009), A graphical approach to sequentially rejective multiple test procedures. Stat. Med. 28(4):586-604.

[2] Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K (2011), Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes or parametric tests. Biometrical Journal 53 (6):894-913.

[3] Rohmeyer K, Klinglmueller F (2014). gMCP: Graph Based Multiple Test Procedures. R package version 0.8-6. http://CRAN.R-project.org/package=gMCP.

Toxicogenomics directory of chemically exposed human hepatocytes

Grinberg, Marianna [1] Stöber, Regina [2] Rempel, Eugen [1]
Rahnenführer, Jörg [1] Hengstler, Jan [2]

[1] Technische Universität Dortmund, Deutschland
grinberg@statistik.tu-dortmund.de, rempel@statistik.tu-dortmund.de, rahnenfuehrer@statistik.tu-dortmund.de
[2] Leibniz-Institut für Arbeitsforschung an der TU Dortmund (IfADo), Dortmund, Deutschland
Stoeber@ifado.de, Hengstler@ifado.de

Understanding chemically-induced toxicity is important for the development of drugs and for the identification of biomarkers. Recently, large-scale gene expression data sets have been generated to understand molecular changes on a genome-wide scale. The Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system (TG-GATEs) is an open-source project in Japan (`http://toxico.nibio.go.jp`). The database contains data for more than 150 compounds applied to cells from rats (liver, kidney) and to human hepatocytes. For many compounds, incubations with different concentrations and for different time periods are available.

Our goal is to identify general principles of the toxicotranscriptome in human hepatocytes with the help of statistical methods. Besides the curse of dimensionality (many more variables than observations) the greatest challenges here are the small number of replicates and several batch effects. However, two advantages of the extensive database can be exploited. Firstly, the different concentrations of the compounds can be used as additional information. We present a curated database where compounds with unusual concentration progression profiles of gene expression alterations are detected and excluded. Secondly, due to the high number of tested compounds we can differentiate between stereotypical and compound-specific expression responses.

This resulted in a toxicotranscriptomics directory that is now publicly available under `http://wiki.toxbank.net/toxicogenomics-map/`. The directory provides information for all genes about the following key features: (1) is the gene deregulated by chemicals and, if yes, by how many and which compounds; (2) is the change in gene expression a stereotypical response that is typically observed when hepatocytes are exposed to high concentrations of chemicals; (3) is the gene also deregulated in human liver disease; (4) does the gene belong to a group of 'unstable baseline genes' that are up- or downregulated due to the culture conditions without exposure to chemicals?

The impact of number of patients lost to follow-up in cardiovascular outcome trials

Hantel, Stefan

Boehringer Ingelheim Pharma GmbH & Co. KG, Deutschland
stefan.hantel@boehringer-ingelheim.com

In all long term clinical trials, it is challenging to keep almost all patients in the trial. In cardiovascular outcome trials (CVOT), patients are asked to stay in the trial and to attend the clinical visits even after stopping the interventional treatment. However, this is not achieved for all patients, who terminate their treatment prematurely without having experienced an event of interest (primary endpoint). Since in most CVOTs in diabetes patients, the primary endpoint is the composite endpoint of cardiovascular death (CV death), non-fatal myocardial infarction and stroke, which is called major adverse cardiovascular event (MACE), it is not sufficient, to know at the end of the trial, whether the patient is still alive or dead. However, this information is still helpful to assess the impact of missing data.

In the discussion about patient retention, it is often mentioned that not more than 1% of the patients may be lost to follow-up for vital status, i.e., for death of any cause. This poster assess the impact of this rate on the primary analyses and suggests sensitivity analysis by imputing missing data based on the detected event rate in the follow-up of patients who terminated the trial prematurely but could be followed-up for their vital status. This assessment is based on simulations in a framework of a Cox regression model.

Classification and diagnosis of diseases based on breath sample analysis

Horsch, Salome [1] Baumbach, Jörg Ingo [2] Rahnenführer, Jörg [1]

[1] TU Dortmund, Deutschland
horsch@statistik.tu-dortmund.de, rahnenfuehrer@statistik.tu-dortmund.de
[2] B & S Analytik, BioMedizinZentrum Dortmund
baumbach@bs-analytik.de

In the past few decades, breath sample analysis has become a new research field for the diagnosis of diseases. MCC-IMS (multi-capillary column-ion mobility spectrometry) is a technology that allows analyzing volatile organic compounds in an air sample in approximately 10 minutes using comparatively small devices. In the past IMS has therefore been used for example at airports for the detection of explosives.

A current research topic is to find out if breath analysis could help medical practitioners with diagnosing diseases concerning the lungs, like COPD or lung carcinoma. Also other diseases might cause differences in the composition of human breath. Some gold standard diagnostic tools like growing a blood culture in order to diagnose a sepsis are time consuming and expensive. Breath analysis is comparatively both fast and cheap and could therefore amend the existing techniques.

The statistical challenge lies in finding appropriate methods for discriminating between two groups depending on the observed data. There is a large variety of classification algorithms available in the literature, but it is unclear which ones work best in a given scenario. Furthermore, most classification tasks in this field face the curse of dimensionality, with a small sample size compared to a large number of measured variables. Mostly there is no or just little information available on which features contain relevant information, such that variable selection is required.

We present a comprehensive comparison of various classification algorithms on data sets with IMS measurements. For different lung diseases basic linear methods are compared to SVMs, Random Forests and Boosting approaches.

Integrative analysis of genome-wide gene copy number changes, gene expression and its clinical correlates in human non-small cell lung cancer

Jabs, Verena

TU Dortmund, Deutschland
jabs@statistik.tu-dortmund.de

In genomics, the aim is often to find oncogenes and tumour suppressor genes. Two approaches to that end are gene expression analysis and the analysis of copy number variations. Usually, only one type of analysis is chosen.

We deal with non-small cell lung cancer (NSCLC), which presents a notoriously genomically unstable cancer type with numerous large scale and focal genomic aberrations. This leads to varying gene copy numbers across the whole genome. The relations of these gene copy number variations to subsequent mRNA levels are only fragmentarily understood. To gain an insight into the relationship, we perform an integrated analysis of gene copy number variations and corresponding gene expressions in a large clinically annotated NSCLC patient cohort.

The main goal is to understand the link between copy number variations and gene expression. To that end, we apply multiple genome-wide methods, for example a correlation analysis using an externally centered correlation coefficient and a survival analysis using Cox models. Our results indicate that gene copy number variations in lung cancer represent a central mechanism to influence gene expression. The copy number variations are also associated with specific gene function and tumorigenic pathways. From the analyses we obtain a list of conspicuous genes. This list can serve as a starting point for further functional studies to identify druggable cancer drivers in NSCLC.

„Beyond CONSORT“ - Illustration von Limitationen des CONSORT Statements aufgrund fehlender Berücksichtigung der methodischen Belastbarkeit von Angaben zu CONSORT-Kriterien am Beispiel einer RCT der zahnärztlichen Implantologie

Knippschild, Stephanie Hirsch, Jessica Baulig, Christine
Krummenauer, Frank

Institut für Medizinische Biometrie und Epidemiologie (IMBE), Universität Witten / Herdecke, Deutschland
stephanie.knippschild@uni-wh.de, jessica.hirsch@uni-wh.de, christine.baulig@uni-wh.de, frank.krummenauer@uni-wh.de

Zielsetzung: Durch das CONSORT Statement [1] soll Transparenz in Publikationen Klinischer Prüfungen gesichert werden. Hierzu soll demonstriert werden, dass eine konsequente Berücksichtigung aller Kriterien der Checkliste durch Autoren nicht zwingend eine valide Studiendarstellung bedingt.

Methode: Das CONSORT Statement wurde auf eine aktuell publizierte 1:1 randomisierte Klinische Prüfung [2] der Zahnärztlichen Implantologie angewendet. Hierbei wurde zum einen das reine Erfülltsein der Kriterien der Checkliste in der Publikation überprüft, zum anderen deren jeweils inhaltliche Korrektheit.

Ergebnis: Von den bearbeitbaren Kriterien der Checkliste inclusive aller Unterpunkte wurden 88% in der Publikation dokumentiert. Werden die Kriterien der CONSORT-Checkliste allerdings auch hinsichtlich ihrer inhaltlich korrekten Umsetzung überprüft, konnte nur in 41% eine korrekte Umsetzung attestiert werden. Exemplarisch sei eine korrekt in der Publikation ausformulierte Berichtung der dieser Studie zugrunde liegenden Fallzahlplanung benannt, die jedoch inhaltlich falsch ist aufgrund einer Nicht-Eignung zum faktischen Studienziel.

Schlussfolgerung: Bei aller Berechtigung des CONSORT Statements zur Sicherstellung transparenter Berichterstellung klinischer Studien und zur Dokumentation ihrer externen Validität muss stets neben dem reinen Vorhandensein von Angaben auch deren inhaltliche Belastbarkeit überprüft werden. Folglich sind auch Gutachter von Publikationen zu Klinischen Studien stets in der Verantwortung, nicht nur das Vorliegen der laut CONSORT zu machenden Angaben zu prüfen, sondern zumindest stichprobenhaft auch deren Sinnhaftigkeit im konkreten Kontext [3].

References:

[1] Schulz, K.F., et al., CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med, 2010. 8:18

[2] Kim, Y.K., et al., A randomized controlled clinical trial of two types of tapered implants on immediate loading in the posterior maxilla and mandible. Int J Oral Maxillofac Implants, 2013. 28:1602-11.

[3] Knippschild S., et al., Das CONSORT-Statement zur standardisierten Berichterstattung Randomisierter Klinischer Prüfungen - Evidenz durch Transparenz. ZZI, 2015 (in press).

Gene set analysis in gene expression series

König, André Rahnenführer, Jörg

TU Dortmund, Department of Statistics, Germany
akoenig@statistik.tu-dortmund.de, rahnenfuehrer@statistik.tu-dortmund.de

Gene expression time series and gene expression multi-dose experiments challenge the investigator with a large number of genes in combination with a small number of replicates per sample in the series. The analysis of gene sets instead of single genes is a common strategy to face the high dimensionality in the data.

There are different competing approaches in literature available. The conference contribution gives an overview about the wide spectrum of approaches, their analysis steps and their implementation in R. A simulation study illustrates the computational effort, the robustness and the similarities of the included methods (e.g. [1], [2], and [3]).

References:

[1] König, André (2014). Temporal Activation Profiles of Gene Sets for the Analysis of Gene Expression Time Series. Dissertation, TU Dortmund, URL: http://hdl.handle.net/2003/32846.

[2] Ernst, Jason, Nau, Gerard J. and Bar-Joseph, Ziv, Clustering short time series gene expression data, 2005, Bioinformatics 21(Suppl 1), URL: http://bioinformatics.oxfordjournals.org/content/21/suppl_1/i159.abstract.

[3] Nueda, Maria, Sebastian, Patricia, Tarazona, Sonia, Garcia-Garcia, Francisco, Dopazo, Joaquin, Ferrer, Alberto and Conesa, Ana (2009). Functional assessment of time course microarray data. BMC Bioinformatics 10.Suppl 6, URL: http://www.biomedcentral.com/1471-2105/10/S6/S9.

Beyond Statistical Consulting“ - Clusterung von Ursachen des Einschaltens der „Vertrauensstelle für Promotionsbelange“ an der Fakultät für Gesundheit der Universität Witten/Herdecke

Krummenauer, Frank

Universität Witten/Herdecke, Deutschland
frank.krummenauer@uni-wh.de

Zielsetzung: „Klassische“ Aufgabe universitärer Abteilungen für Medizinische Biometrie ist die methodische Unterstützung des akademischen Nachwuchses der Fakultäten. Zunehmend erweitert sich dies auf Methoden des Projektmanagement und der akademischen Qualitätsentwicklung: Gerade Promotionsvorhaben zum „Dr. med. (dent.)“ haben durch ihre frühe Ansiedelung im Studium und die Betreuung durch klinisch tätige Ansprechpartner suboptimale Erfolgsaussichten. Vor diesem Hintergrund hat die Fakultät für Gesundheit 2010 eine „Vertrauensstelle für Promotionsbelange“ eingerichtet. Nun soll bewertet werden, welche Anfragen an eine solche Vertrauensstelle gerichtet werden, und ob der Ansatz als Modellvorschlag gelten kann.

Methode: Zum Studium der Human- und Zahnmedizin an der Wittener Fakultät für Gesundheit sind jährlich insgesamt circa 150 Immatrikulationen zu verzeichnen. Im Zeitraum 04/2010 – 09/2014 seit Einrichtung der Vertrauensstelle wurden daraus insgesamt 53 Konflikt-Fälle betreut, deren Situation nicht mit maximal zwei moderierenden Gesprächen in eine reguläre Zusammenarbeit zurück geführt werden konnte. Nur in zwei Fällen wurde die Vertrauensstelle Betreuer-seitig angerufen. Ursachen der Konflikt-Konstellationen wurden aus der Dokumentation der Vertrauensstelle klassiert in Cluster, die in mehr als 10% der moderierten Fälle auslösende Rolle hatten; Mehrfachnennungen waren möglich.

Ergebnis: Häufigste Konflikt-Konstellationen waren „mangelnde Projektmanagement-Fixierung“ (z.B. ein Exposee für das Promotionsvorhaben) in 91% der Fälle sowie „nicht Fach-sichere Delegation der Betreuungsverantwortung“ (z.B. durch die Klinikleitung an Betreuungs-unerfahrene Assistenzärzte respektive durch nicht-habilitierte Themensteller an nicht aktiv integrierte Habilitierte) in 60% der Fälle. In beiden Konstellationen konnten mit Methoden des Projektmanagement jenseits bestehender Verwerfungen gangbare Abläufe zur Rettung des Promotionsprojektes moderiert werden. Die hierfür notwendige Sprechstunden-Auslastung bedingte mehr als einem Tag pro Woche.

Schlussfolgerung: Die Vorhaltung einer „Vertrauensstelle für Promotionsbelange“ bedingt eine Ressourcen-intensive Investition für eine Fakultät; gleichzeitig ist durch „Rettung“ entsprechender Projekte und daraus möglicherweise generierbarer Publikationen ein Return on Investment erwartbar. Insofern ist die Einrichtung einer Vertrauensstelle als Modell für Medizinische Fakultäten anzuraten; daneben müssen bei Häufung obiger Konfliktquellen Maßnahmen zur Steigerung der Betreuer-Kompetenz und –Verantwortung implementiert werden.

„Pricing Biometrical Service“ - Vorschlag einer Kalkulationsmatrix für biometrische Dienstleistungen in regulatorischen Klinischen Prüfungen

Krummenauer, Frank Hirsch, Jessica

Universität Witten/Herdecke, Deutschland
frank.krummenauer@uni-wh.de, jessica.hirsch@uni-wh.de

Zielsetzung: 2012 wurde für eine zweiarmige zahnmedizinische RCT (zwölfseitiger CRF, keine methodischen Besonderheiten, GCP-Konformität) parallel von sechs universitären Abteilungen für Medizinische Biometrie ein Kostenvoranschlag zur vollständigen biometrischen Betreuung der Studie eingeholt. Die erhaltenen Kostenvoranschläge rangierten zwischen summarisch mininal 3.000 €und maximal 62.000 €, im Median wurden 32.000 €veranschlagt (Angaben auf 1.000 €gerundet); nur drei wiesen Teil-Leistungen mit stratifizierter Bepreisung aus. Ein Erklärungsansatz für diese Spanne mag sein, dass unter Maßgaben der GCP für biometrische Dienstleistungen spezifische Kosten anstehen bei Führung der Prozesse entlang eines Qualitätsmanagement-Systems [1]. Vorgestellt werden soll eine Kalkulationsmatrix für biometrische Dienstleistungen in regulatorischen Studien.

Methode: Entlang des biometrischen SOP-Katalogs der Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (TMF) respektive des KKS-Netzwerks (KKSN) sowie eigener institutioneller SOPs wurde eine Maximal-Listung der aus biometrischer Perspektive anbietbaren Teil-Dienstleistungen zusammen gestellt, welche dann mit TVL-Kalkulationsannahmen für die Leistungen entsprechend der anstehenden Mindest-Qualifikation des einzubeziehenden Personals unterlegt wurde (exemplarisch sei für den „Qualified Statistician“ die Kalkulationsannahme „mindestens TVL-14 in Stufe 3-4“ mit Blick auf mindestens fünfjährige Berufserfahrung nach ICH Guideline E-9 benannt).

Ergebnis: Als biometrische Dienstleistungen sind mindestens folgende Tätigkeiten zu bepreisen: Planung und Parametrisierung des Studiendesigns; Randomisierung; Erstellung des Statistischen Analyse-Plans; Programmierung der Auswertung (ggf. parallel!); Entblindung inklusive Blind Data Review Meeting; Auswertung inklusive Tabellen- und Abbildungserstellung; Erstellung des biometrischen Ergebnisberichts. Optional sind z.B. orale Präsentationen vor Auftraggebern oder die Durchführung geplanter Interimanalysen zu bepreisen. Werden TVL-Verdienstsätze unterstellt zuzüglich Arbeitgeber-seitiger Mehrkosten um pauschal 20%, so ergibt sich zur eingangs erwähnten „einfachen“ Studie eine summarische Kalkulation von 38.000 €aus Arbeitgeber-Perspektive.

Schlussfolgerung: Grundsätzlich sind einheitliche Preis/Leistungs-Relationen biometrischer Dienstleistungen in regulatorischen Klinischen Prüfungen anzustreben; die hier vorgeschlagene Kalkulationsmatrix mag hierzu Hilfestellung bieten.

References:

[1] Krummenauer F et al. (2011): Studiendesigns in der Implantologie (VI): Budgetierung Klinischer Studien - was kostet das Ganze? Zeitschrift für Zahnärztliche Implantologie 27: 354-61

Subgroup-specific survival analysis in high-dimensional breast and lung cancer datasets

Madjar, Katrin Rahnenführer, Jörg

TU Dortmund, Deutschland
madjar@statistik.tu-dortmund.de, rahnenfuehrer@statistik.tu-dortmund.de

Survival analysis is a central aspect in cancer research with the aim of predicting the survival time of a patient on the basis of his features as precisely as possible. Often it can be assumed that specific links between covariables and the survival time not equally exist in the entire cohort but only in a subset of patients. The aim is to detect subgroup-specific effects by an appropriate model design. A survival model based on all patients might not find a variable that has a predictive value within one subgroup. On the other hand fitting separate models for each subgroup without considering the remaining patient data reduces the sample size. Especially in small subsets it can lead to instable results with high variance. As an alternative we propose a model that uses all patients but assigns them individual weights. The weights correspond to the probability of belonging to a certain subgroup. Patients whose features fit well to one subgroup are given higher weights in the subgroup-specific model. Since censoring in survival data deteriorates the model fit we seek an improvement in the estimation of the subgroup-weights.

10 independent non-small cell lung cancer and 36 breast cancer cohorts were downloaded from a publicly available database and preprocessed. We define them as subgroups and further divide them according to histology, tumor stage or treatment in order to obtain as homogeneous subgroups as possible. We apply and evaluate our model approach using these datasets with some clinical variables and high-dimensional Affymetrix gene expression data and compare it to a survival model based on all patients as well as separate models for each subgroup.

Steuerung und Regelung der Dynamik physiologischer Wirkungsflüsse

Mau, Jochen

iqmeth Privates Institut für Quantitative Methodik, Deutschland
iqmeth@t-online.de

Das biologische System des menschlichen Körpers ist ein entscheidungsautonomer Automat, der in Ver-/Entsorgung seiner inneren Verbrennungsvorgänge nicht autark ist. Energiebedarf, Zugang zu Energieressourcen im ökosystem und darauf gerichtetes Verhalten wird vom übergeordneten System des sozioökonomischen Umfelds und der politischen Rahmenbedingungen beeinflußt.

Die Ebenen der physiologischen Steuerung des Körpers entsprechen aus funktionaler Sicht der hierarchischen Systematik der Automatisierungstechnik, also Lenkung, Steuerung und Regelung, jedoch sind Ordnung und Lösungen der Realisierung - unabhängig vom Materialaspekt - ganz andere als in der Technik. Die übertragung von Regelungskonzepten und dem modularen Konstruktionsweg der Ingenieurwissenschaften war in der Systembiologie mehrfach erfolgreich[1,2], so daß auch die übernahme von technischen Kriterien der Systemrobustheit naheliegend wurde[3,4,5]. Jedoch tragen die Analogien mit technischen Anlagen nicht weit: Ganz abgesehen vom Reproduktionsauftrag sind Bipolarität und Rhythmus durch physiologischen Antagonismus von technischen Systemen abweichende Prägungen des Systemverhaltens des menschlichen Körpers[6]. Zusätzlich wird die Diskrepanz in der Zusammenlegung von Systemversorgung, -entsorgung und Signalwegen der hormonellen Regelung in einen einzigen übertragungskanal, die extrazelluläre Flüssigkeit, manifest[7].

Ohne hierarchische Steuerungsstrukturen modularer Technik müssen Steuerungszentren verschiedener, sich überlagernder Ebenen in diese vielfältige träge Regelung eingreifen, um die Dynamik von Massen- und Signalübertragung erfolgreich ablaufen zu lassen.

Diese physikalischen Gegebenheiten werfen klinisch relevante Fragen auf:

1) Können Unausgeglichenheiten dieses “Regelungsgeflechts” (i. S. einer “Unwucht”) latent bestehen?

2) Können - bei schnellem Umschalten und träger Signalleitung - notwendigerweise entstehende “Schwingungsbreiten” deutlich größer als normal, aber im Mittel unauffällig bleiben und z.B. durch Resonanzschwingungen andere Regelkreise beeinflussen?

3) Wann und wie lassen sich in einem nichthierarchischen komplexen Regelungsgeflecht Wirkungen auf Ursachen zurückführen, wenn variable Regelungswege den Ursache-Wirkungs-Zusammenhang physiologisch verändern?

References/Literaturhinweise

[1] Csete ME, Doyle JC.:Reverse engineering of biological complexity. Science. 2002;295:1664-9

[2] El-Samad, H., Kurata, H.,Doyle,J.C.,Gross,C.A.,Khammash,M.:Surviving heat shock: Controlstrategies for robustness and performance. Proc Natl Acad Sci USA.2005;102:2736-41.

[3] Kitano H.:Systems biology: a brief overview. Science.2002; 295:1662-4

[4] Ahn AC, Tewari M, Poon ChS, Phillips RS.:The limits of reductionism in medicine: Could systems biology offer an alternative. PLOS Med2006;3(6):e208.

[5] Ahn AC, Tewari M, Poon ChS, Phillips RS.:The clinical applications of a systems approach. PLOS Med2006;3(7):e209.

[6] Sachsse,H.:Einführung in die Kybernetik, Vieweg, Braunschweig, 1971/1974.

[7] Hall,J.E.:Guyton and Hall Textbook of Medical Physiology, 12th ed., Saunders, Philadelphia, 2011.

Kybernetik - Von der Zelle zur Population

Mau, Jochen

iqmeth Privates Institut für Quantitative Methodik, Deutschland
iqmeth@t-online.de

Die ingenieurwissenschaftliche Methodik der Steuerung und Regelung von komplexen dynamischen Systeme hat auf verschiedenen Skalenebenen Bedeutung: Bei der Beschreibung zellulärer Regelkreise in der Systembiologie [1, 2], ohnehin bei physiologischen Regelungs- und Steuerungsvorgängen [3, 4] sowie bei ökologischen und sozioökonomischen Zusammenhängen, hier insbesondere auch in der Epidemiologie [5].

Die einschlägigen Konzepte werden anhand von Literaturbeispielen erläutert:

Für die zelluläre Ebene wird das Verhalten bei Wärmeeinwirkung (heat shock response) gewählt; hier wurde in der Arbeit von El-Samad et al [2] die modulare Konstruktion und Notwendigkeit von komplexer Regelung für optimales Verhalten untersucht.

Auf der physiologischen Ebene gibt es zahlreiche Beispiele; die Komplexität der ineinandergreifenden Systeme und Ebenen wird gut deutlich bei der Regelung des Blutzuckers [3, 4]. Oft handelt es sich um sogenannte Rückkopplungsschleifen (negative feedback control loops); interessant sind aber auch Beispiele von positivem, also beschleunigendem Abweichen vom Normalwert: Die Regelung der Wehentätigkeit und der Sichelzellanämie [4].

Auf der Bevölkerungsebene trifft man auf hochgradig komplexe Interaktionen zwischen sozialen, ökonomischen und politischen Gefügen, die gesamthaft auf das Körpersystem des Individuums physiologisch einwirken und insbesondere die prinzipiell autonomen Lenkungsentscheidungen moderieren. Zur Illustration dient die Aufklärung der Risikofaktoren und von Gegenmaßnahmen bei der Adipositas [5].

Hier sind Grundlagen zur Definition von Regelkreisen bisher nur in sehr reduziertem Ausmaß vorhanden, so daß mit anderen Ansätze flexiblere Modellierungen des Systemverhaltens unternommen werden: Hervorzuheben ist das agent-based modeling (ABM) [5], für das in-silico-Verfahren benutzt werden. Der Ansatz hat interessante Schlußfolgerungen für die statistische Datenanalyse [5]. Grundsätzlich erscheint es aber auch hier möglich, komplizierte "Regelungsgeflechte" - ähnlich wie in der Physiologie des menschlichen Körpers - zugrunde zu legen, sobald hinreichende theoriebildende Vorstellungen bereit gestellt werden.

References/Literatur

[1] Csete M.E., Doyle J.C.: Reverse engineering of biological complexity. Science. 2002;295:1664-9.

[2] El-Samad, H.,Kurata,H.,Doyle,J.C.,Gross,C.A.,Khammash,M.: Surviving heat shock: Control strategies for robustness and performance. Proc Natl Acad Sci USA.2005;102:2736-41.

[3] Sachsse, H.: Einführung in die Kybernetik, Vieweg, Braunschweig,1971/ 1974.

[4] Hall,J.E.: Guyton and Hall Textbook of Medical Physiology, 12th ed., Saunders, Philadelphia, 2011.

[5] Galea,S., Riddle,M., Kaplan, G.A.: Causal thinking and complex system approaches in epidemiology. Intl J Epid.2010;39:97-106.

Statistical Lunch – A Practical Approach to Increase Statistical Awareness

Müller, Tina Igl, Bernd-Wolfgang Ulbrich, Hannes-Friedrich

Bayer Pharma AG, Deutschland
tina.mueller@bayer.com, bernd-wolfgang.igl@bayer.com, hannes-friedrich.ulbrich@bayer.com

Every statistician who works in a consulting function in an interdisciplinary working environment is often faced with two challenges:

Firstly, the value of statistics is often underestimated. Statistical support might be sought by clients because of some guideline requirements, but the real benefit and the opportunities of a well done statistical analysis are unknown. Secondly, statistics is often poorly understood by the client, which can lead to tedious and inefficient discussions.

However, the client can seldom spare the time or is unwilling to get properly educated in statistics, partly due to the large amount of work they themselves are already faced with in their job.

To meet all this, we regularly organize "Statistical Lunch" meetings around midday hours. This time slot can usually be made by clients. We present relevant statistical topics in small chunks while snacks are available for everyone. The topics are carefully prepared for a non-statisticians audience and invite all participants to lively discussions.

So far, the feedback is overwhelmingly positive, and the Statistical Lunches are on their way to become a permanent institution.

The poster contains the organizational frame as well as the topics we have presented so far.

Vergleich der Schätzung der Standardfehler von Tibshirani und Osborne in der Lasso-Regression

Näher, Katja Luisa Kohlhas, Laura Berres, Manfred

Hochschule Koblenz, RheinAhrCampus, Deutschland
kalunae@gmx.de, laura.kohlhas@web.de, berres@rheinahrcampus.de

Tibshirani [1] entwickelte 1996 einen neuen Ansatz um lineare Modelle zu schätzen, den Least Absolute Shrinkage and Selection Operator (Lasso). Das von Tibshirani vorgestellte Verfahren zur Schätzung der Standardfehler der Regressionskoeffizienten wurde 1999 von Osborne aufgegriffen und modifiziert [2]. Diese beiden Methoden werden in einer Simulationsstudie miteinander verglichen. Dabei werden die Standardfehler empirisch durch die in den Simulationen ermittelten Koeffizienten geschätzt und den Standardfehlern nach Tibshirani und Osborne gegenübergestellt. Neben dem Einfluss der Korrelationsstruktur der Koeffizienten (keine Korrelation, Autokorrelation und tridiagonale Toeplitz-Korrelation) werden auch die Auswirkungen durch die Anzahl der Variablen, die Anzahl der Beobachtungen und den Wert der wahren Koeffizienten untersucht.

Für alle Szenarien liefern die Standardfehler nach Osborne gute Schätzungen. Dagegen werden die Standardfehler nach Tibshirani meist zu klein geschätzt. Bei den Standardfehlern nach Tibshirani fällt auf, dass die Schätzungen von der Korrelation abhängen und davon, ob ein Koeffizient den wahren Wert 0 besitzt. Bei Osborne hat die Tatsache, ob ein Koeffizient tatsächlich 0 ist keinen Einfluss auf die Schätzung der Standardfehler, die Korrelation der Variablen hingegen schon. So zeigte sich bei einer hohen Autokorrelation ein kleinerer Standardfehler für die erste und letzte Variable und für die Toeplitz-Korrelation eine bogenartige Entwicklung der Standardfehler.

References:

[1] Tibshirani R. Regression Shrinkage and Selection via the Lasso. J. R. Statist. Soc. B 1996; 58: 267-288.

[2] Osborne M., Presnell B., Turlach B. On the LASSO and its Dual. Journal of Computational and Graphical Statistics 2000; 9 (2): 319-337.

Zero-inflated models for radiation-induced chromosome aberration data

Oliveira, Maria [1] Ainsbury, Liz [2] Einbeck, Jochen [1]
Rothkamm, Kai [2] Puig, Pedro [3] Higueras, Manuel [2,3]

[1] Durham University, UK
maria.oliveira@usc.es, jochen.einbeck@durham.ac.uk
[2] Public Health England Centre for Radiation, Chemical and Environmental Hazards (PHE CRCE), UK
Liz.Ainsbury@phe.gov.uk, Kai.Rothkamm@phe.gov.uk, Manuel.Higueras-Hernaez@phe.gov.uk
[3] Universitat Autonoma de Barcelona, Spain
ppuig@mat.uab.cat, Manuel.Higueras-Hernaez@phe.gov.uk

The formation of chromosome aberrations is a well-established biological marker of exposure to ionizing irradiation, and they can lead to mutations which cause cancer and other health effects. Accordingly, radiation-induced aberrations are studied extensively to provide data for biological dosimetry, where the measurement of chromosome aberration frequencies in human lymphocyte is used for assessing absorbed doses of ionizing radiation to individuals. That is, d blood samples from one or various healthy donors are irradiated with doses $x_i,\ i = 1, \ldots, d$. Then for each irradiated sample, n_i cells are examined and the number of observed chromosomal aberrations $y_{ij},\ j = 1, \ldots, n_i$ is recorded. For such count data, the Poisson distribution is the most widely recognised and commonly used distribution and constitutes the standard framework for explaining the relationship between the outcome variable and the dose. However, in practice, the assumption of equidispersion implicit in the Poisson model is often violated due to unobserved heterogeneity in the cell population, which will render the variance of observed aberration counts larger than their mean, and/or the frequency of zero counts greater than expected for the Poisson distribution. The goal of this work is to study the performance of zero–inflated models for modelling such data. For that purpose, a substantial analysis is performed, where zero-inflated models are compared with other models already used in biodosimetry, such as Poisson, negative binomial, Neyman type A, Hermite and, in addition, Polya–Aeppli and Poisson–inverse Gaussian. Several real datasets obtained under different scenarios, whole and partial body exposure, and following different types of radiation are considered in the study.

Assessing clinical expertise using information theoretical concepts of evidence and surprise

Ostermann, Thomas [1] Rodrigues Recchia, Daniela [1]
Jesus Enrique, Garcia [2]

[1] University of Witten/Herdecke, Germany
Thomas.Ostermann@uni-wh.de, Daniela.RodriguesRecchia@uni-wh.de
[2] University of Campinas, Department of Statistics, Brazil
jg@ime.unicamp.br

Background
Besides the evidence generated from clinical studies, a subjective evidence (SE) exists referring to the expertise of a physician.

Material and Methods
Based on two examples a method to calculate SE is proposed relying on the basic work on subjective evidence from Palm (1981 & 2012) derived from Shannon's concept of information:

$$E = -\sum_{i}^{n} p(A_i) \log_2 q(A_i).$$

Only those results are considered, which are evident from the expert's point of view, in the formula this is represented for q which is the subjective probability for the expert and p is the true/known probability for the event. Thus, the greater the difference between personal expertise and empirical data, the higher is the expert's „surprise".

Results
The first example is based on a notional pre-post study of a new antihypertensive drug which is supposed to lower the blood pressure. A believes in the new drug whether B does not. Using a triangular probability density function modeling the pre post difference we were able to confirm that SE adequately describes personal expertise.

Moreover, using normally distributed data we were able to show that the discrete model converges against the continuous limit distribution given by

$$I = -\int_{-\infty}^{\infty} f(x) \log_2 f(x) d(x).$$

Discussion
SE contributes to evaluate personalized health care as it helps to quantify expertise in a straightforward manner. Additional basic research on the underlying mathematical models and assumptions should foster the use of this model and help to bring it into line with other models.

References:

(1) Ostermann T, Recchia DR, Garcia JE: Von der klinischen zur personalisierten Evidenz: ein Methodenvorschlag. DZO 2014:46:1-4.

(2) Palm G. Evidence, information, and surprise. Biol Cybern 1981; 42(1): 57–68.

(3) Palm G. Novelty, Information and Surprise. Berlin, Heidelberg: Springer; 2012.

Two-factor experiment with split units constructed cyclic designs and square lattice designs

Ozawa, Kazuhiro [1] Kuriki, Shinji [2]

[1] Gifu College of Nursing, Japan
k-ozawa@gifu-cn.ac.jp
[2] Osaka Prefecture University, Japan
kuriki@ms.osakafu-u.ac.jp

We consider a nested row-column design with split units for a two-factor experiment. The experimental material is divided into b blocks. Each block constitutes a row-column design with k_1 rows and k_2 columns. The units of the row-column design are treated as the whole plots, and the levels of a factor A (called whole plot treatments) are arranged on the whole plots. Additionally, each whole plot is divided into k_3 subplots and the levels of the second factor B (called subplot treatments) are arranged on the subplots.

Kachlicka and Mejza [1] presented a mixed linear model for the observations with fixed treatment combination effects and random block, row, column, whole plot and subplot effects, and the mixed model results from a four-step randomization. This kind of randomization leads us to an experiment with an orthogonal block structure, and the multistratum analysis can be applied to the analysis of data in the experiment.

Kuriki et al. [2] used a Youden square repeatedly for the whole plot treatments and a proper design for the subplot treatments. In this talk, we use a cyclic design for the whole plot treatments and a square lattice for the subplot treatments. We give the stratum efficiency factors for such a nested row-column design with split units, which has the general balance property.

References:

[1] Kachlicka, D. and Mejza, S.: Repeated row-column designs with split units, *Computational Statistics & Data Analysis*, **21**, pp. 293–305, 1996.

[2] Kuriki, S. Mejza, S. Mejza, I. Kachlicka, D.: Repeated Youden squares with subplot treatments in a proper incomplete block design, *Biometrical Letters*, **46**, pp. 153–162, 2009.

Bestimmung der optimalen Penalisierung im Least Absolute Shrinkage and Selection Operator mittels Kreuzvalidierung und generalisierter Kreuzvalidierung

Ozga, Ann-Kathrin Berres, Manfred

Hochschule Koblenz, RheinAhrCampus Remagen, Deutschland
a.ozga@gmx.net, berres@rheinahrcampus.de

Die lineare Regression ist ein zentrales statistisches Verfahren. Meist werden die Koeffizienten nach der Methode der kleinsten Quadrate geschätzt. Allerdings ist die Varianz dieser Schätzer bei korrelierten Kovariablen groß. Um dies zu vermeiden, wurden weitere Methoden vorgeschlagen, wie die Subset Selection oder Ridge Regression, welche Koeffizienten auf Null setzen oder schrumpfen. Diese weisen jedoch ebenfalls Nachteile auf, sodass Tibshirani [1] den Least Absolute Shrinkage and Selection Operator (Lasso) entwickelt hat. Dabei wird das lineare Regressionsmodell unter einer Nebenbedingung über die 1-Norm minimiert. Das führt zu einer Schrumpfung der Koeffizienten, wobei diese auch Null werden können. Ein ähnliches Verfahren ist die Least Angle Regression (Lars) [2], die mit einer kleinen Modifikation einen schnellen Lasso-Schätzer liefert.

Ziel war es nun, mittels k-facher und generalisierter k-facher Kreuzvalidierung den besten in der Nebenbedingung des Lasso-Modells stehenden Penalisierungsparameter zu ermitteln und die Modellauswahl der einzelnen Methoden zu vergleichen. Dazu wurden Datensätze mit unterschiedlichen Korrelationsstrukturen und Stichprobenumfängen simuliert.

Es zeigt sich, dass die mittels k-facher Kreuzvalidierung, generalisierter Kreuzvalidierung und die für das Lars-Modell bestimmten mittleren Penalisierungsparameter in den betrachteten Szenarien sich sehr ähneln. Erkennbare Unterschiede zwischen den Schätzungen der Methoden bestehen nur bei kleinerem Stichprobenumfang. Die Methoden führen ebenfalls zu ähnlichen Ergebnissen für die Koeffizientenschätzung und bei der Modellauswahl.

References:

[1] Tibshirani, Robert (1996). Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 58, No. 1 267-288.

[2] Efron, Bradley; Hastie ,Trevor; Johnstone, Iain; Tibshirani, Robert (2004). Least Angle Regression. The Annals of Statistics, Vol. 32, No. 2, 407-499.

Zeitlich-räumliche Modellierung mit der verallgemeinerten Gamma-Poisson Mischung

Rausch, Tanja K. [1,2] Müller, Christine H. [2]

[1] IMBS, Universität zu Lübeck, Deutschland
tanja.rausch@imbs.uni-luebeck.de
[2] Fakultät Statistik, TU Dortmund, Deutschland
tanja.rausch@imbs.uni-luebeck.de, cmueller@statistik.tu-dortmund.de

Um Korrelationen zwischen benachbarten Zellen im 2- und 3-dimensionalen Raum in verschiedene Richtungen modellieren und schätzen zu können, wurde das verallgemeinerte Gamma-Poisson-Mischungsmodell weiterentwickelt, welches von [1] beschrieben wurde. Dieses Modell wurde ursprünglich vorgeschlagen, um in Presence-Absence-Maps Gesamtanzahlen von Individuen auf einem Gebiet zu schätzen. Es hat den Vorteil, dass Korrelationen mit kleineren Parameterwerten ebenfalls geschätzt werden können. Das Verhalten des weiterentwickelten Modells wurde mittels Simulationen untersucht. Als Anwendungsbeispiel wurden Anzahlen von Giraffen mit genauer Lokalisation im Krüger-Nationalpark in Südafrika aus den Jahren 1977 - 97 [2] betrachtet. Weitere Anwendungsmöglichkeiten des Modells wie die Vorhersage der Lebensdauer von Prothesen bei täglicher Belastung werden diskutiert.

References:

[1] Müller, Ch. H., Huggins, R. und Hwang, W.-H. (2011) Consistent estimation of species abundance from a presence-absence map. Statistics & Probability Letters 81(9), 1449-57.

[2] http://dataknp.sanparks.org/sanparks/metacat?action=read&qformat=sanparks&sessionid=0&docid=judithk.739.17 (24.11.2014).

A boosting technique for longitudinal microarray data in multivariable time-to-event models

Reiser, Veronika [1] Vach, Werner [1] Binder, Harald [2]
Schumacher, Martin [1]

[1] Medical Center - University of Freiburg, Deutschland
reiser@imbi.uni-freiburg.de, wv@imbi.uni-freiburg.de, ms@imbi.uni-freiburg.de
[2] Medical Biostatistics - University Medical Center of Mainz, Deutschland
binderh@uni-mainz.de

There are several methods for linking gene expression data to clinical outcomes, such as the time to an event of interest. However, current methods for considering repeated gene expression measurements, which are increasingly becoming available, are mostly univariate [1], i.e., only the effect of one gene at a time is evaluated. For obtaining a parsimonious signature that can predict survival based on longitudinal microarray data, modeling techniques are needed that can simultaneously incorporate the time course of all genes and perform variable selection.

We present a boosting approach for linking longitudinal high-dimensional microarray data to a censored event time, adapting a multivariable componentwise likelihood-based boosting approach, which was originally developed for standard time-to-event settings [2]. The employed approach identifies a small subset of genes by providing sparse model fits, where most of the estimated regression parameters are equal to zero. Incorporating also baseline measurements, our approach can distinguish between genes that are informative at baseline and other genes where measurements in the course of time are important.

For evaluation in a simulation study and an application example, we consider bootstrap .632+ prediction error curve estimates, adapted for longitudinal covariates.

References:

[1] Natasa Rajicic, Dianne M. Finkelstein, and David A. Schoenfeld. (2006). Survival analysis of longitudinal microarrays. Bioinformatics.

[2] Binder, H. and Schumacher, M. (2008). Allowing for mandatory covariates in boosting estimation of sparse high-dimensional survival models. BMC Bioinformatics. 9:14.

Proteome-wide phylogenetic analysis by similarity search in tandem mass spectra

Rieder, Vera Rahnenführer, Jörg

TU Dortmund, Deutschland
rieder@statistik.tu-dortmund.de, rahnenfuehrer@statistik.tu-dortmund.de

Phylogenetic analysis has the goal to identify evolutionary relationships between organisms. Modern molecular approaches are based on the alignment of DNA sequences or amico acid sequences. The analysis is often based on selected parts of the genome, but also genome-wide comparisons of sequences are performed. Recently the construction of phylogenetic trees based on proteomics has become popular, supported by the progress in the technical development of mass spectrometers. A typical approach is to first estimate protein sequences from tandem mass spectra with the help of data base searches, given species with known genome-wide sequence information, and then apply sequence based methods.

A new approach is to directly compare tandem mass spectra without using sequence information. One advantage is that also experiments from species without genome annotation can be compared. Furthermore the procedure does not depend on the quality of databases and database search algorithms. A first basic algorithm was introduced by Palmblad and Deelder [2] who demonstrate the reconstruction of the unique correct phylogenetic tree for the great apes and other primates.

First we review distance and similarity measures for the direct comparison of mass spectra, especially the cosine distance and the Hausdorff distance. Then we introduce measures and algorithms for the proteome-wide comparison. Finally the algorithms are applied to selected examples, including technical replicated mass spectra of human platelet [1] as well as mass spectra of different tissues of rats.

References:

[1] Burkhart et al. (2012): The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways, Blood, 120(15), 73-82.

[2] Palmblad M, Deelder AM (2012): Molecular phylogenetics by direct comparison of tandem mass spectra, Rapid Commun Mass Spectrom, 26(7), 728-32.

Konzeption und Implementierung einer Standard Operation Procedure (SOP) zur Funktionsvalidierung von statistischer Auswertungs-Software zur GCP-konformen Durchführung von Klinischen Prüfungen

Schaper, Katharina [1] Hirsch, Jessica [1] Eglmeier, Wolfgang [2]
Huber, Peter [3] Krummenauer, Frank [1]

[1] Institut für Medizinische Biometrie und Epidemiologie (IMBE), Universität Witten/Herdecke
katharina.schaper@uni-wh.de, jessica.hirsch@uni-wh.de, frank.krummenauer@uni-wh.de
[2] Zentrum für Klinische Studien (ZKS-UW/H), Universität Witten/Herdecke
wolfgang.eglmeier@uni-wh.de
[3] Bereich Informationstechnologie (BIT), Universität Witten/Herdecke
peter.huber@uni-wh.de

Zielsetzung:
Die Durchführung Klinischer Prüfungen muss den Maßgaben der GCP (Good Clinical Practice) folgen. Dazu zählt, dass jegliche Software, die für die Durchführung einer Klinischen Prüfung benutzt wird, belegbar funktionsvalidiert sein muss. Während die meisten GCP-Umsetzungen von Biometrie- und Data-Management-bezogenen Prozessen an Klinischen Studienzentren (KKS, ZKS) expliziten SOPs unterliegen, oft angelehnt an die SOP-Bibliothek der Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (TMF) respektive des KKS-Netzwerks (KKSN), steht derzeit keine SOP zur Funktionsvalidierung von Studiensoftware zur Verfügung. Für an KKSN-Standorten üblicherweise im Gebrauch befindliche Anwendungs-Software (wie Microsoft Office ®, SAS ®, IBM SPSS Statistics ®, OpenClinica ®) kann ab dem Moment des öffnens der CD-Hülle vom Hersteller keine Garantie für die Funktionalität der Software übernommen werden, sodass eine lokale Funktionsvalidierung für jede Software vor deren Nutzung in Klinischen Prüfungen vorzunehmen ist.

Methode:
Aus dem beschriebenen Handlungsbedarf heraus wurde eine SOP zur Funktionsvalidierung generiert und implementiert, welche eine explizite Methodik zur Prüfung der Funktionsvalidität entlang bestehender Beispiel-Datensätze und standardisierter Analyse-Routinen vorgibt. Insbesondere wurde eine Check-Liste zur funktionellen als auch zur im Vorfeld nötigen technischen Validierung der Software bezogen auf Installationsprozesse abgeleitet zur Dokumentation der Validierungsprüfung.

Ergebnis:
Ein Audit-belastbares und GCP-konformes SOP-Paket, bestehend aus einer Kern-SOP, zwei Checklisten als SOP-Anhang zur technischen und zur funktionellen Validierung, drei Programmcodes zur funktionellen Validierung und einem Programmcode zur Erstellung eines Testdatensatzes wurden generiert und implementiert. Das SOP-Paket kommt zum Einsatz, sowohl wenn die Lizenz für eine Software erneuert als auch wenn die komplette Software auf einem Rechner neu installiert wird.

Schlussfolgerung:
Um Klinische Prüfungen GCP-konform durchführen zu können, muss eine Systematik für die technische und funktionelle Validierung von Studiensoft-

ware bereitgestellt werden. Die vorliegende SOP mit ihren Anlagen ist eine Möglichkeit der konkreten Umsetzung und kann zeitnah lokal von Klinischen Studienzentren in Zusammenarbeit mit den zuständigen Abteilungen für Medizinische Biometrie sowie Informations-Technologie eingesetzt werden.

Author index