CEN 2011

Bridging Biostatistical Theory and Application

2nd Conference of the Central European Network

CEN 2011

Bridging Theory and Applications

September 12-16, Zurich, Switzerland

2nd conference of the Central European Network
27. ROeS Seminar
57. Biomtrisches Kolloquium

Abstracts

Joint Biometric Conference of the Region Austria-Switzerland (ROeS), the German Region (DR) and the National Group Poland of the International Biometric Society (IBS)

Editors:

Norbert Neumann (Basel), Hans Ulrich Burger (Basel), Katja Ickstadt (Dortmund), Stanislaw Mejza (Posznan), Dominik Heinzmann (Basel)

CEN 2011
Bridging Theory and Application

- Abstracts booklet –

2nd Conference of the Central European Network
27. ROeS Seminar
57. Biometrisches Kolloquium Deutsche Region
September 12-16, Zürich, Switzerland

Universität Zürich, Switzerland

Editors: Norbert Neumann (Basel), Hans Ulrich Burger (Basel), Katja Ickstadt (Dortmund), Stanislaw Mejza (Posznan), Dominik Heinzmann (Basel)

ISBN 978-3-033-03083-1
Coverphoto: Copyright © Zürich Tourism / Rubiano Soto

Content

Forword

Dear Participant, dear Reader,

We welcome you to the Second Conference of the Central European Network at the University Zurich, Switzerland – CEN2011. Our motto "Bridging Biostatistical Theory and Application" relates to challenges and perspectives in the development of methods and applications in the field of biometry, biostatistics and bioinformatics.

The Central European Network of the International Biometrics Society (IBS) was established in 2007 and consists currently of the German Region (DR), the Region Austria-Switzerland (ROeS) and the Polish Region (PR) of the IBS. One of the main purposes of the network is to facilitate scientific exchange in the area of biometry among the member groups and regions of the IBS and to organize joint meetings in which scientific exchange can take place.

The conference CEN2011 is the second joint meeting of this network with more than 150 paper and poster submissions from the three regions and groups. Taking place in Zurich Switzerland the Region Austria-Switzerland took the lead in organizing this event with the strong support of the German Region and the Polish Region. We hope it is notable already from the program and this abstract book that this conference goes beyond a traditional ROeS Meeting held by the Region Austria - Switzerland every second year and the Biometrics Colloquium held by the German Region every year. The creativity within the whole network really enriches our scientific perspectives in valuable presentations and interesting discussions with many colleagues from all member regions of the network.

For this conference we organized a number of sessions on interesting and hot topics of biometry. The conference program starts with a series of introductory sessions on topics as Health Economics, Statistics in Molecular Biology and Modern Time to Event Methods for Registry Data. We are happy that we have been able to gain support in organizing these sessions by our well known colleagues Andreas Ziegler (Luebeck), Harald Binder (Freiburg) and Hanno Ulmer (Innsbruck). At the end of the conference there is also a full day workshop on "Probability Estimation in Prognostics" providing a comprehensive overview on machine learning approaches – thanks to Andreas Ziegler for organizing this workshop.

In our conference we have a special focus on our younger colleagues by giving an opportunity to present in the "Young Statisticians" session –thanks to Andrea Berghold (Graz) and Stanislaw Mejza (Poznan). As started in 2010 there are again "Statistics in Practice" sessions presenting recent theoretical research put into practice. Another opportunity to get valuable insight from experienced colleagues – thanks to the working group "Nachwuchs".

Other sessions were organized by the working groups (AGs, Sektionen) of our society and reflect very well the diversity and the quality of biometry in our regions and groups: For example the working groups on 'Adaptive und multiple Verfahren', on 'Landwirtschaftliches Versuchswesen', on 'Populationsgenetik-Genomanalyse', on 'Bayes Methodik', on 'Nichtparametrische Methoden' and others have organized sessions reflecting the present work of many active colleagues.

We want to make the CEN2011 conference interesting and attractive for all of our colleagues and at least maintain the standards set in the first joint network meeting. We will certainly monitor the result of this meeting to find out what we can improve for future meetings.

Last but not least we would like to thank Leo Held with the Institute for Social and Preventive Medicine at the Zurich University (ISPM) together with the University of Zurich for hosting the conference and the local organizing committee for all their dedication and the many hours they have put into the preparation of this meeting to make it a success. Only those who have organized such a congress once really know how much work is really associated with it. We would also like to thank the session chairs and all the presenters for their contributions to this conference. Finally we hope that every participant will enjoy the conference and will be able to attend many interesting sessions and use this great opportunity for social and scientific interaction.

Finally, don't forget that Zurich as the largest city in Switzerland has a lot to offer. Beside its vibrant nightlife the opportunities range from the renowned Opera House and concert hall to the city theatre and around 50 museums.

Zurich

September 2011

Norbert Neumann (President ROeS)

Katja Ickstadt (President DR)

Richardus Vonk (past President DR)

Stanislaw Mejza (President PR)

Greetings of the local Organization committee

Dear friends and colleagues,

The conference in Zurich will be the second conference organized by the CEN, the Central European Network, organized by the German region (DR) , the Austrian-Swiss region (ROeS) and the Polish group of the International Biometric Society. The CEN is open for other biometric groups in central Europe and surrounding regions. The goal of the network is to provide a forum for scientific exchange in the area of biometry among members of the region and to foster international collaboration. The first CEN conference was held in Munich 2008 and this year's conference in Zurich is the second forum in which biometricians from all these countries and surrounding regions will come together.

The topic of this conference is bridging biostatistical theory and application. Linking biometry research and methodology to real world application is important to move forward biometry as a science. The duty of biometricians is not only the development of complex theories and models, but also formulating, summarizing and communicating conclusions when applying those theories and models to practical problems. This conference thus will provide a forum for stimulating discussions and enable sharing of experiences and learnings, covering theoretical areas as well as applications in the field of agriculture, environmental science, ecology, biology, epidemiology and health care.

The dialog with scientists from other disciplines is one of the important parts of the work of many statisticians as their success often depends on a proper understanding of their colleagues and their associated disciplines. On the other hand, arising questions may find answers in scientific discussions with other statisticians including those working more theoretically on the development of new methods or working in other fields. That is why cooperation between all different areas of biometry is specifically important. We hope that this conference will provide a good forum and opportunity for exchange of problems and ideas and will help finding solutions to burning questions in our work.

The conference will comprise more than 30 scientific session including around 150 oral and many poster presentations. Sessions will cover basically all major areas in which statisticians are currently active. The integration of colleagues from different countries is increasingly important as biometry becomes more and more part of a global environment. This is especially important for colleagues from Eastern Central Europe. However, when working in a global multi-disciplinary environment European statisticians must position themselves to deserve a valuable place at the global table. Hence working together and exchanging information will become a key to success. Europe as an essential place for rigorous research and manifold applications in biometry will be sustainable only if a lively biometric community can be maintained across all Europe with frequent opportunities for exchange and bridging research and applications. This is a requisite to retain Europe as an attractive and competitive area for biometry. The Central

European Network and its conference every three years, national and regional frequent biometric conferences and other forums in which biometricians can come together will play an essential role in this. Last but not least we have to keep in mind that not only biometry as a discipline but also our partners increasingly globalize their working environment.

In line with previous ROeS conferences and Biometric Colloquiums of the German Region, this conference will put specific focus on the next generation of biometricians. There are dedicated sessions for "young" statisticians from all parts of CEN, in which the next generation can present their work, and these sessions have turned out to be of high quality and stimulating in the past. In addition there are special sessions for "statisticians in action" which should help providing an overview of today's environment in which statisticians are working. Both elements should help to prepare and integrate the younger generation into the community, a major task to secure our future.

Last but not least we need to thank to a number of people in Basel and Zurich for their great help in making this conference possible. There is Leonhard Held and his continuous support in organizing the facilities at the University of Zurich and his stimulating interest in all aspects of the conference. There is to thank all the staff from his department, the Institute for Social and Preventive Medicine at the Zurich University (ISPM), especially Daniel Sabanés Bové and Sinikka Kohler for making a lot of arrangements at the University feasible. And finally, we need to thank the University of Zurich for hosting the conference and the Dekan of the Medical Faculty for his support. Furthermore we would like to thanks Barbora Martinec, David Burger, Eliane Imfeld and Friedrich Burger for their administrative support without whom this event would have been impossible.

We are truly looking forward to this conference. We hope the conference including all presentations and posters as a reflection of our work in the past will stimulate the scientific discussions and will help to manage our future work. We also hope that the supporting social elements of the conference, the essential coffee and lunch breaks, the reception on Tuesday, the excursion and the joint dinner on Wednesday will help bringing us together and will provide a friendly atmosphere facilitating networking and the very important coming together. We hope as well that you will get a flavor of our hosting city and the University of Zurich, a major fascinating center of science and commerce in Switzerland.

We wish you an enjoyable and stimulating time at this conference in Zurich!

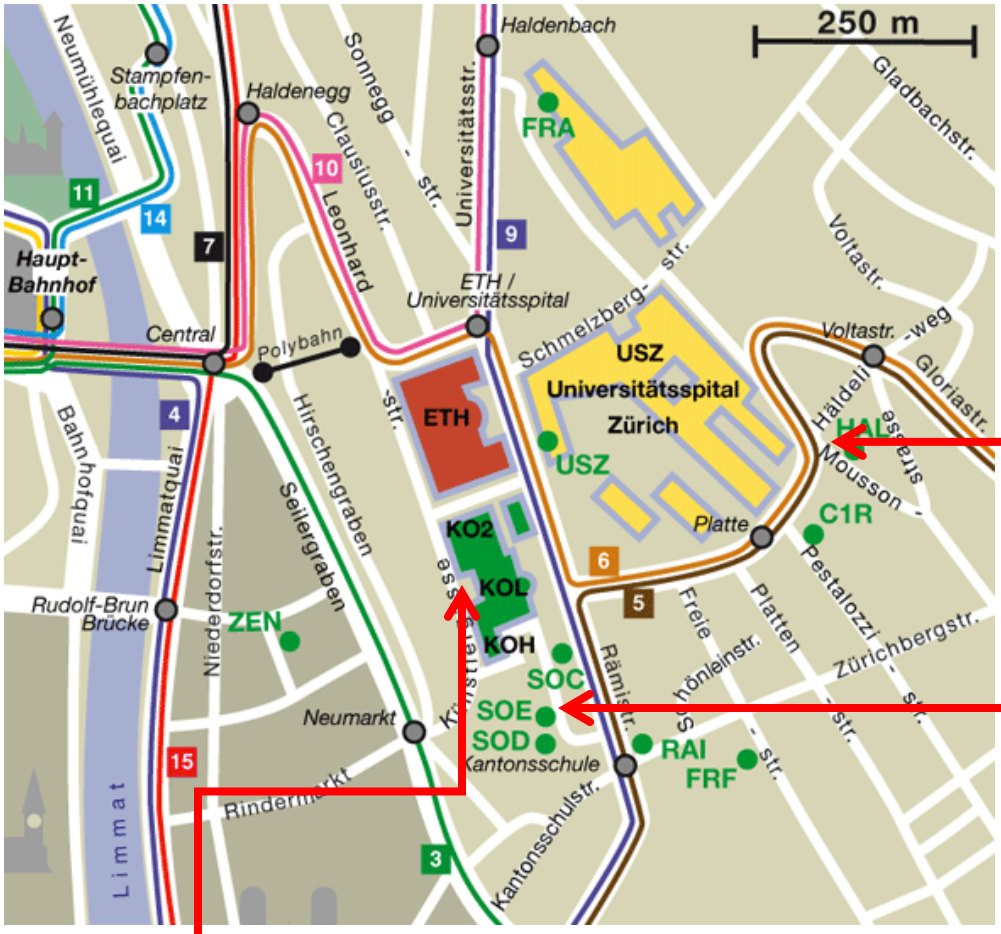
Basel and Zurich, August 2011

Local Organizing Committee

Mike Branson, Frank Bretz, Uli Burger, Dominik Heinzmann and Willi Maurer

Locations

Location of the University

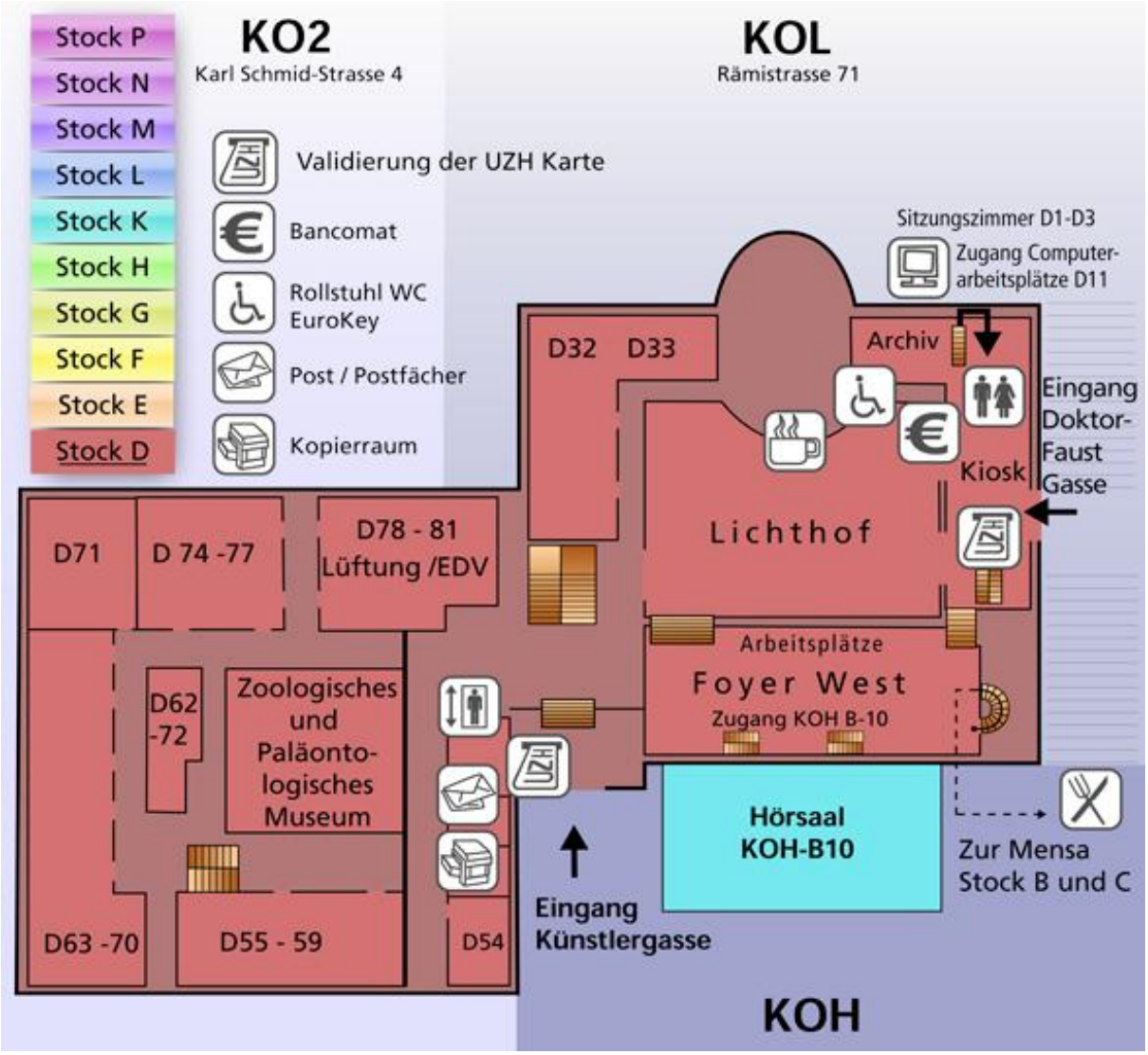


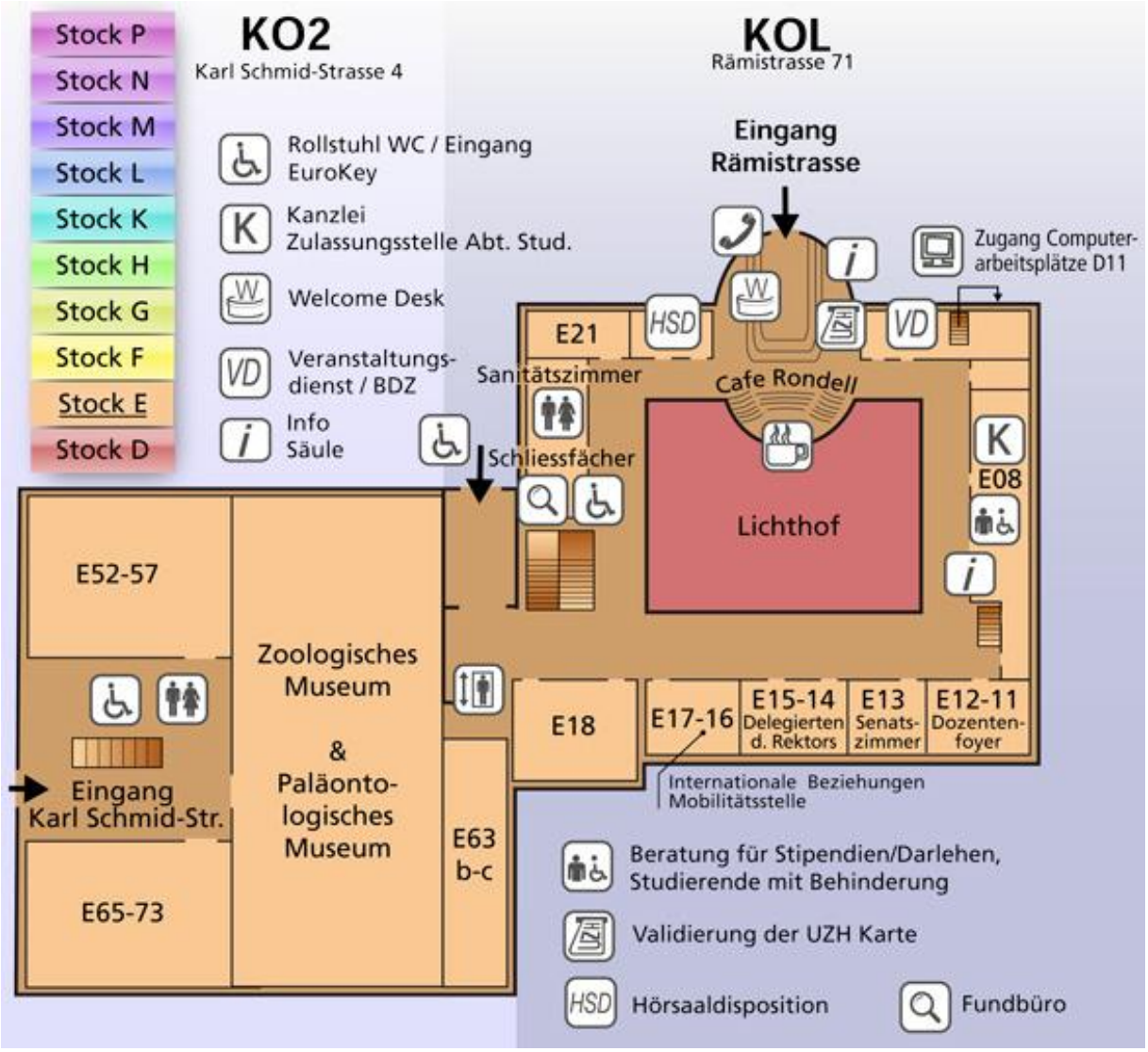
15. Sept.
Häldeliweg 2

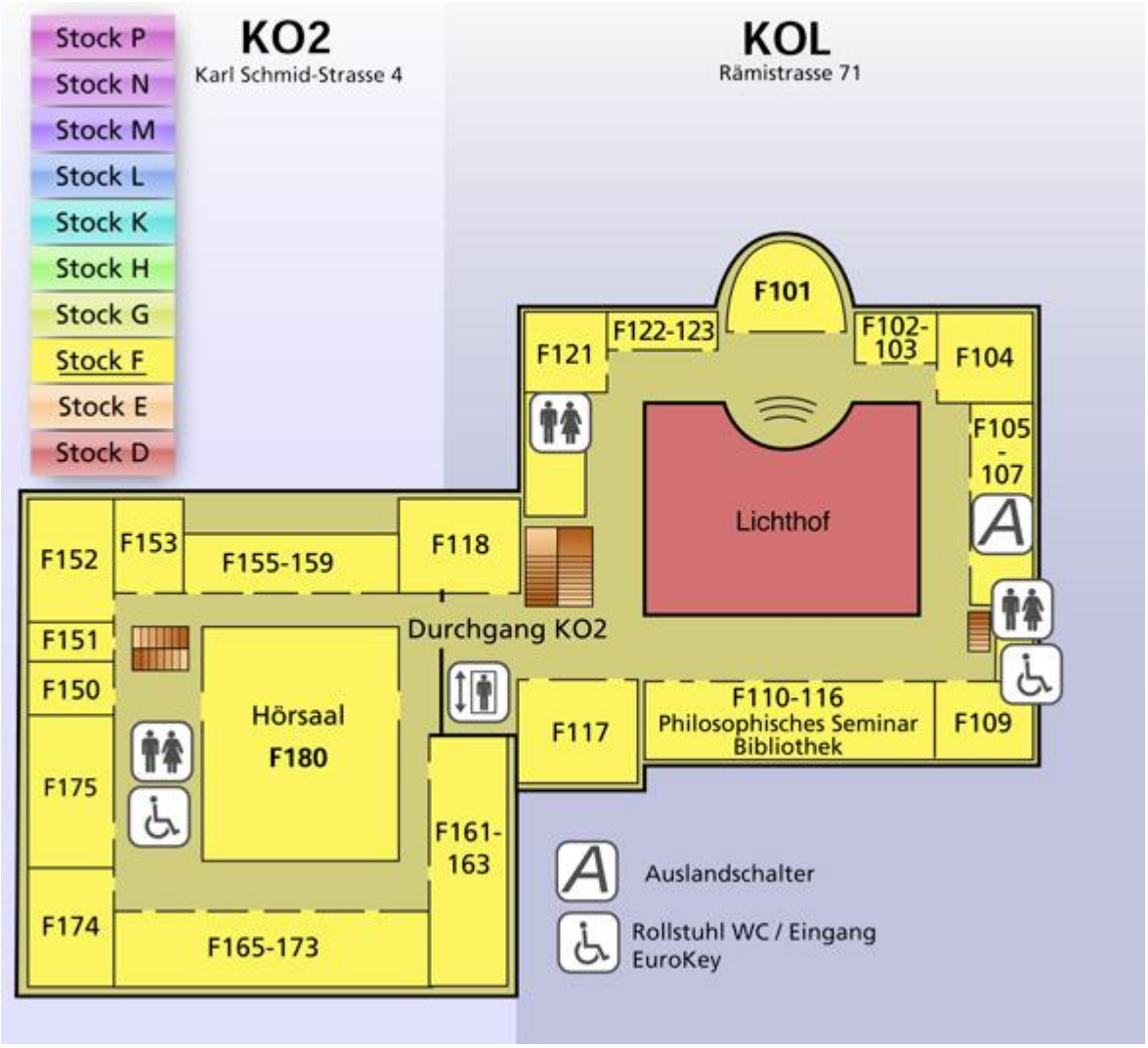
16. Sept.
Schönberggasse 11

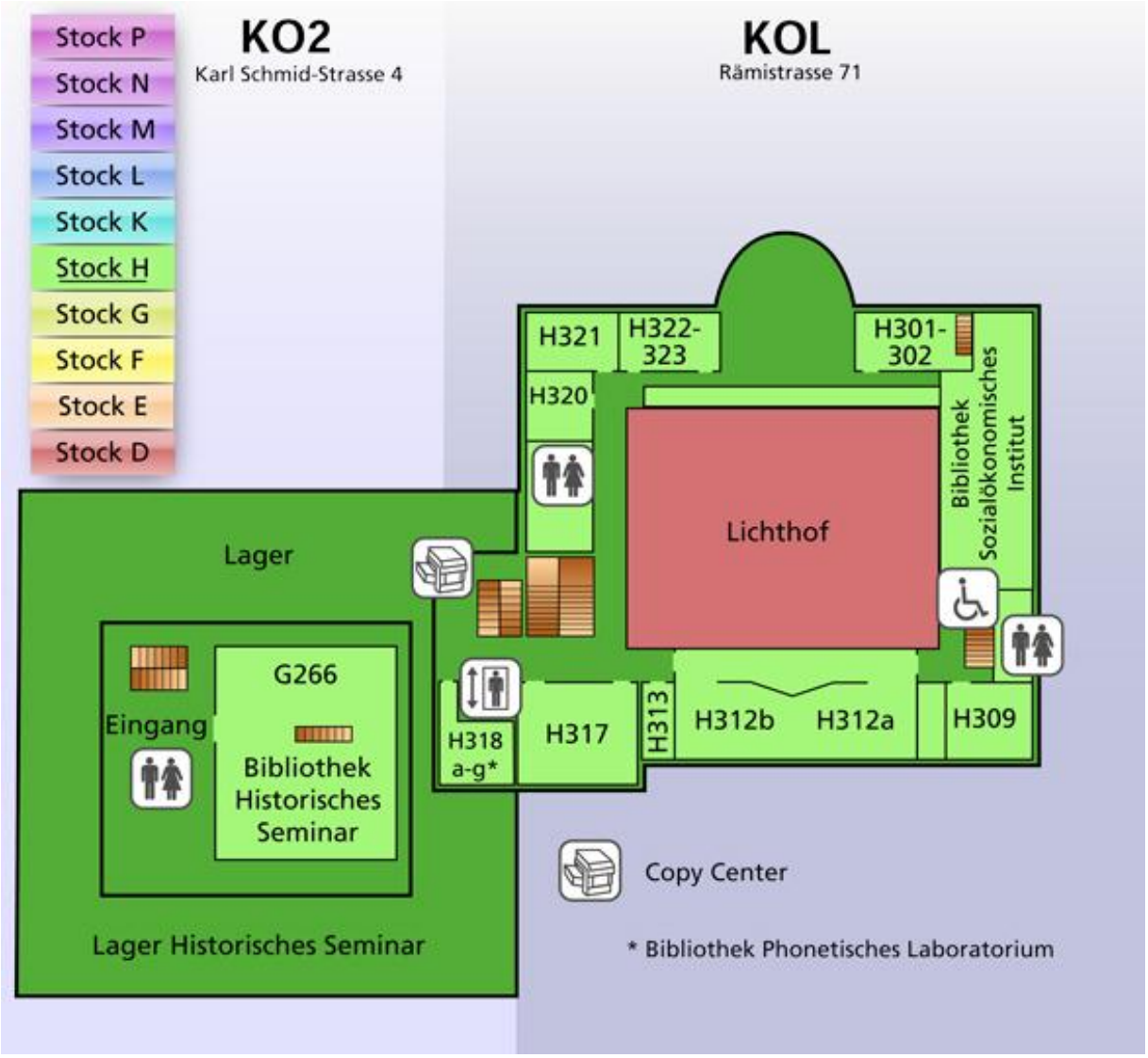
12. – 14. September
Main Building
Rämistr. 71

Location of lecture rooms

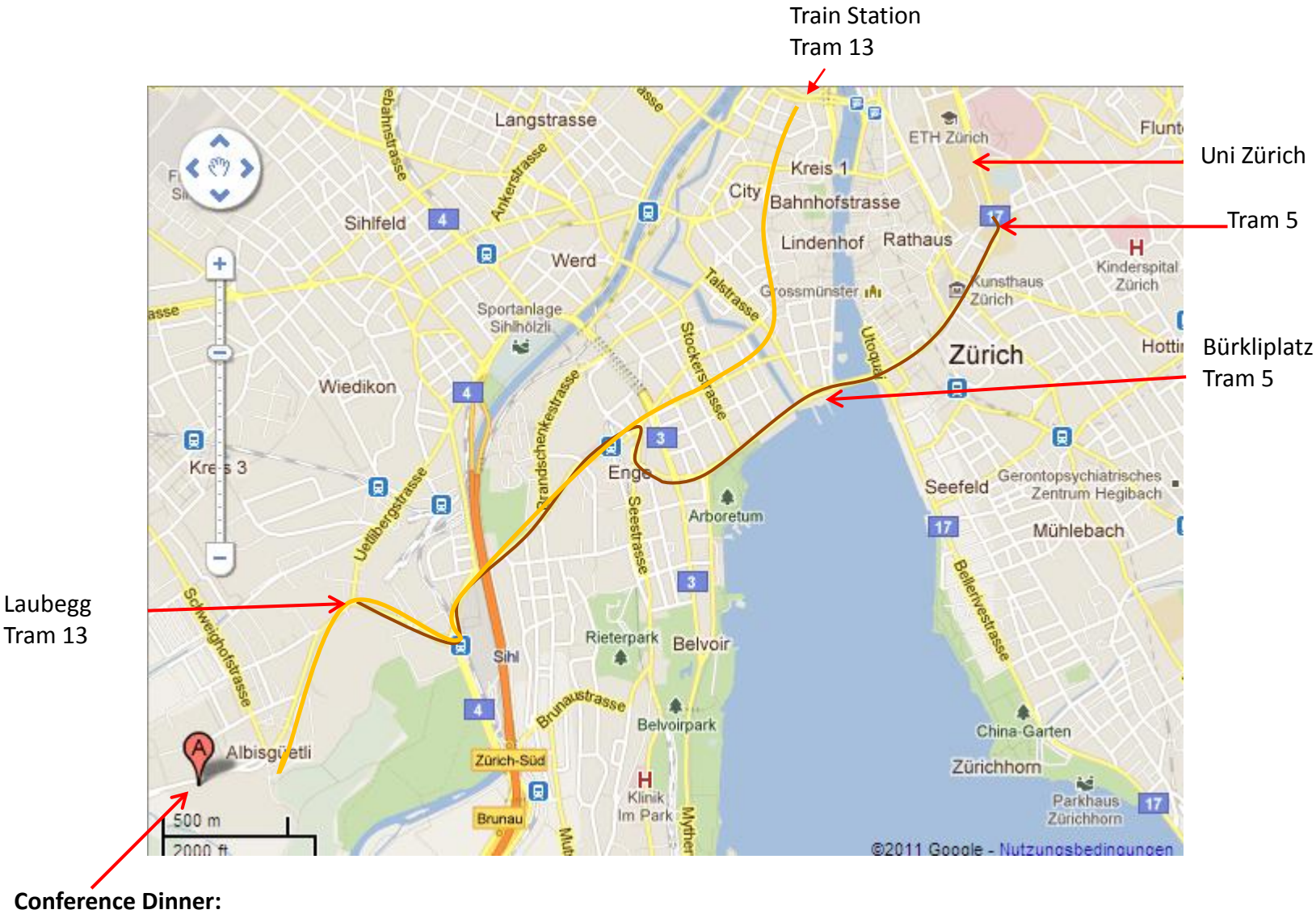



Stock P
Stock N
Stock M
Stock L
Stock K
Stock H
Stock G
Stock F
Stock E
Stock D
KO2
Karl Schmid-Strasse 4
KOL
Rämistrasse 71
Rollstuhl WC / Eingang EuroKey
Kanzlei Zulassungsstelle Abt. Stud.
Welcome Desk
Veranstaltungs-dienst / BDZ
Info Säule
Eingang Rämistrasse
Zugang Computer-arbeitsplätze D11
E21
HSD
VD
Sanitätszimmer
Cafe Rondell
Schliessfächer
Lichthof
E08
E52-57
Zoologisches Museum & Paläonto-logisches Museum
E18
E17-16
E15-14 Delegierten d. Rektors
E13 Senats-zimmer
E12-11 Dozenten-foyer
Internationale Beziehungen Mobilitätsstelle
Eingang Karl Schmid-Str.
E63 b-c
E65-73
Beratung für Stipendien/Darlehen, Studierende mit Behinderung
Validierung der UZH Karte
Hörsaaldisposition
Fundbüro


Stock P
Stock N
Stock M
Stock L
Stock K
Stock H
Stock G
Stock F
Stock E
Stock D
KO2
Karl Schmid-Strasse 4
KOL
Rämistrasse 71
F101
F121
F122-123
F102-103
F104
F105-107
Lichthof
F152
F153
F155-159
F118
Durchgang KO2
F151
F150
F175
Hörsaal
F180
F117
F110-116
Philosophisches Seminar
Bibliothek
F109
F161-163
F174
F165-173
Auslandschalter
Rollstuhl WC / Eingang
EuroKey

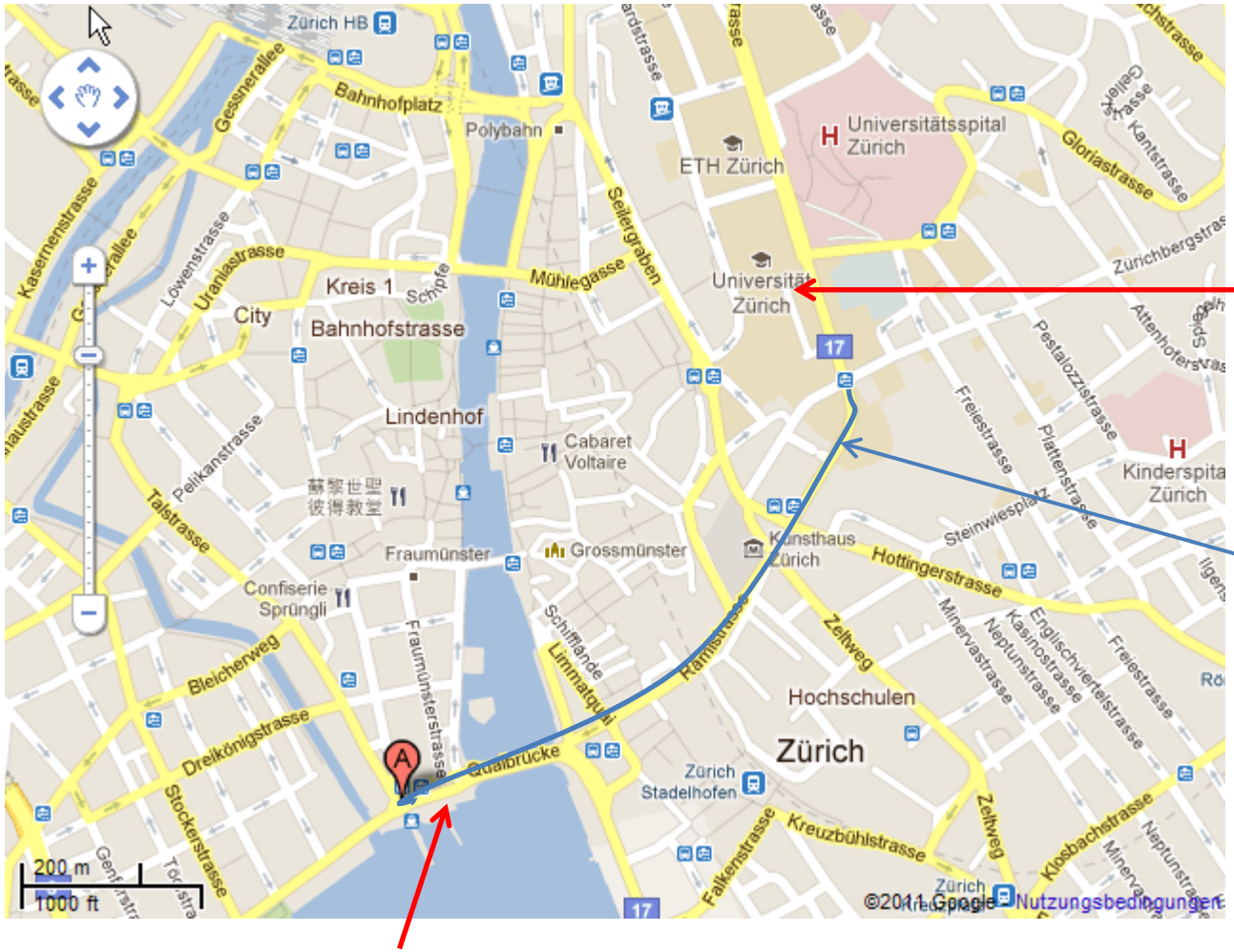

Stock P
Stock N
Stock M
Stock L
Stock K
Stock H
Stock G
Stock F
Stock E
Stock D
KO2
Karl Schmid-Strasse 4
KOL
Rämistrasse 71
H321
H322-323
H301-302
H320
Lichthof
Bibliothek Sozialökonomisches Institut
Lager
G266
Eingang
Bibliothek Historisches Seminar
H318 a-g*
H317
H313
H312b
H312a
H309
Copy Center
Lager Historisches Seminar
* Bibliothek Phonetisches Laboratorium

Location of the Conference Dinner



Conference Dinner:
Meet between 18:30 – 19:30 at Schützenhaus Albisgütli Uetlibergstrasse 341, Zuerich

Location of the Excursion (ROeS)



ROeS Excursion 15:00 Tramstation «Bürkliplatz»
Meeting point: Limmatbootsteg 6 (towards lake)

Session Overview

and

Abstracts for Presentations and Posters

Session Overview

Monday 12. Sep 2011					
Time	**Lecture Hall KOH-B10**	**Lecture Hall KO2-F180**	**Lecture Hall KOL-F101**	**Lecture Hall KOL-F104**	**Lecture Hall KOL-F121**
9:00-10:30			**Tutorial** Introduction to health economics (1) U. Siebert	**Tutorial** Introduction to molecular biology for statisticians (1) T. Zeller	**Tutorial** Modern Time to Event Methods for Registry Data (1) M. Cvancarova Smastuen
10:30-11:00	**Break**				
11:00-12:30			**Tutorial** Introduction to health economics (2) U. Siebert	**Tutorial** Introduction to molecular biology for statisticians (2) T. Zeller	**Tutorial** Modern Time to Event Methods for Registry Data (2) M. Cvancarova Smastuen
12:30-14:00	**Lunch**				
14:00-15.30	**Opening Session**				
15:30-16:00	**Break**				
16:00-17:30	**Adaptive and sequential designs (1)** M. Vandemeulebroecke	**Non-parametric procedures in biometry (1)** F. Konietschke	**Spatial Statistics** L. Held and J. Dreesman	**Longitudinal Data** H. Binder	**Missing Values** N. Neumann
Evening	**City tour**				

Tuesday 13. Sep 2011					
Time	**Lecture Hall KOH-B10**	**Lecture Hall KO2-F180**	**Lecture Hall KOL-F101**	**Lecture Hall KOL-F104**	**Lecture Hall KOL-F121**
9:00-10:30	**Adaptive and Sequential Designs (2)** M. Kieser	**Nonparametric Procedures in Biometry (2)** M. Neuhäuser/ F. Konietschke	**Health Economics** C. Schwenke and R. Nixon	**Challenges in Agricultural Science (1)** J. Spilke	**Multiple Testing (1)** L. Hothorn
10:30-11:00	**Break**				
11:00-12:30	**Bayesian Methods and Decision Theory (1)** K. Ickstadt	**Young Statisticians (1)** A. Berghold	**Time to Event Analysis (1)** J. Beyersmann	**Challenges in Agricultural Science (2)** C. Richter	**Multiple Testing (2)** S. Zehetmayer
12:30-14:00	**Lunch**				
14:00-15.30	**Bayesian Methods and Decision Theory (2)** K. Ickstadt	**Clinical Trial Application** A. Berghold and J. Roehmel	**Time to Event Analysis (2)** H. Binder	**Diagnostics 1: ROC-Analyses - Sensitivity and Specificity** C. Schwenke	**Multiple Testing (3)** W. Maurer
15:30-16:00	**Break**				
16:00-18:00	**Awards Session DR and ROeS**		**Time to Event Analysis (3)** A. Wienke	**Diagnostics 2: Biomarkers and Clustered Data** C. Schwenke	**Miscellaneous** W. Maurer
Evening	**Reception by the City and Kanton Zürich / Poster and Wine: Lichthof (atrium) of University (KOL-E)**				

Wednesday 14. Sep 2011					
Time	**Lecture Hall KOH-B10**	**Lecture Hall KO2-F180**	**Lecture Hall KOL-F101**	**Lecture Hall KOL-F104**	**Lecture Hall KOL-F121**
9:00-10:30	**Young Statisticians (2)** S. Mejza	**Bayesian Methods in Translational Medicine (1)** M. Branson and L. Colin	**Multivariable model-building with continuous variables – a comparison of flexible regression approaches** W. Sauerbrei	**Biometrical Journal: Editor's Choice** L. Held	**Epidemiology and Statistics (1)** H. Ulmer
10:30-11:00	**Break**				
11:00-12:30	**Statistics in practice (1) Double Session (for Young Statisticians) - Meta-Analysis & Practical Experiences of Senior Statisticians** T. Mueller	**Bayesian Methods in Translational Medicine (2)** M. Branson and L. Colin	**Observational Studies** H.-U. Burger and A. Berghold	**Genetics and Biomarker (1)** H. Binder	**Epidemiology and Statistics (2)** H. Ulmer and J. Wellmann
12:30-13:00	**Lunch**				
13:00-14:00			**ROeS General Assembly**		
14:00-15.00	**DR General Assembly**				
15:00-16:00			**ROeS Excursion**		
16:00-18:00					
Evening	**Conference Dinner**				

Thursday 15. Sep 2011				
Time	**Lecture Hall HAH-E-3**	**Lecture Hall HAH-E-11**	**Lecture Hall HAH-F-1**	**Lecture Hall HAH-E-10**
9:00-10:30	**Clinical Enpoints in Oncology (1)** H.-U. Burger and A. Berghold	**Statistics in practice (2)** AG Nachwuchs, AG Weiterbildung, S. Roll	**Genetics and Biomarker (2)** P. Schlattmann	**Modeling and Simulation** M. Vandemeulebroecke
10:30-11:00	**Break**			
11:00-12:30	**Clinical Endpoints in Oncology (2)** H.-U. Burger and A. Berghold	**Statistics in practice (3)** AG Nachwuchs, AG Weiterbildung, S. Roll	**Genetics and Biomarker (3)** A. Ziegler	**Benefit Risk** Günter Heimann
12:30-13:00	**Lunch**			
13:00-14.30	**Closing Session**			
15:30-16:00	**Break**			

Friday 16. Sep 2011					
Time	**Lecture Hall SOE-F-02**				
9:00-16:00	**Workshop** Probability estimation in Prognostics				

Detailed Session Overview

Monday 12. Sep 2011			
Time	**Lecture Hall KOH-B10**	**Lecture Hall KO2-F180**	**Lecture Hall KOL-F101**
9:00-10:30	**Tutorial** Introduction to health economics (1) U. Siebert	**Tutorial** Introduction to molecular biology for statisticians (1) T. Zeller	**Tutorial** Modern Time to Event Methods for Registry Data (1) M. Cvancarova Smastuen
10:30-11:00	**Break**		
11:00-12:30	**Tutorial** Introduction to health economics (2) U. Siebert	**Tutorial** Introduction to molecular biology for statisticians (2) T. Zeller	**Tutorial** Modern Time to Event Methods for Registry Data (2) M. Cvancarova Smastuen
12:30-14:00	**Lunch**		
14:00-15.30	**Opening Session** Dekan Medical Faculty, Universitaet Zuerich Presidents of ROeS, IBS Germany and Poland Greeting 50th anniversary ROeS: Prof. J. Roehmel Keynote talk: Prof. A. Grieve		
15:30-16:00	**Break**		
16:00-17:30	**Adaptive and sequential designs (1)** M. Vandemeulebroecke **Adaptive enrichment designs, their methodological approaches and utility – what are regulatory implications** Wang, Sue-Jane **Adaptive population enrichment: switching to a sub-population in response to interim trial data** Jennison, Christopher **Group sequential enrichment design incorporating subgroup selection** Magnusson, Baldur P., Turnbull, Bruce W **Statistical Methodology for Adaptive Patient Enrichment Designs** Wassmer, Gernot	**Non-parametric Procedures in Biometry (1)** F. Konietschke **Concordance Regression** Heinze, Georg **Unimodal regression using B-splines** Köllmann, Claudia; Bornkamp, Björn; Grecco, Hernan Edgardo; Fried, Roland; Ickstadt, Katja **Nonparametric inference on the cumulative abortion probability in a competing risk model** Di Termini, Susanna; Hieke, Stefanie; Schumacher, Martin; Beyersmann, Jan	**Spatial Statistics** L. Held and J. Dreesman **Spatial distribution of notified infectious diseases in Germany 2001-2010** Zoellner, Iris; Seibold-Krämer, Erik **Influenzanet: An Internet-based System for Monitoring Influenza-like Illness** Schwehm, Markus **Spatio-temporal modelling of infectious disease surveillance counts** Paul, Michaela; Held, Leonhard
Evening	**City tour**		

Monday 12. Sep 2011		
Time	**Lecture Hall KOL-F104**	**Lecture Hall KOL-F121**
9:00-10:30		
10:30-11:00	**Break**	
11:00-12:30		
12:30-14:00	**Lunch**	
14:00-15.30		
15:30-16:00	**Break**	
16:00-17:30	**Longitudinal Data** H. Binder **Selection of variables in high-dimensional time-series analysis** Läuter, Jürgen; Rosolowski, Maciej **Variable Selection for Generalized Linear Mixed Models by L1-Penalized Estimation** Groll, Andreas; Tutz, Gerhard **Linear mixed models with a penalized normal mixture as random effects distribution** Heinzl, Felix; Tutz, Gerhard **How to analyse multiple ordinal scores in a clinical trial?** Laffont, Celine Marielle; Concordet, Didier	**Missing Values** N. Neumann **Estimation of the Treatment Effect in the Presence of Noncompliance and Missing Data** Leuchs, Ann-Kristin; Benda, Norbert; Zinserling, Jörg; Neuhäuser, Markus; Berres, Manfred **Orthogonal regression and related methods for bivariate data** Berres, Manfred; Volland, Ruth; Beck, Irene; Monsch, Andreas U. **The Value of Pattern-Mixture Models for Sensitivity Analysis** Akacha, Mouna **Effect of genotype imputation errors on the accuracy of genomic selection** Neugebauer, Nadine; Wellmann, Robin; Bennewitz, Jörn
Evening	**City tour**	

Tuesday 13. Sep 2011 Morning			
Time	**Lecture Hall KOH-B10**	**Lecture Hall KO2-F180**	**Lecture Hall KOL-F101**
9:00-10:30	**Adaptive and Sequential Designs (2)** M. Kieser **Improving adaptive group sequential designs with discrete outcomes** Englert, Stefan; Kieser, Meinhard **Estimation of the Hazard Ratio in Adaptive Designs with Sample Size Readjustment** Ligges, Sandra, Wassmer, Gernot, Müller, Christine **A graphical approach for adaptive clinical trials testing multiple hypotheses** Klinglmueller, Florian , Koenig, Franz, Posch, Martin **Maximum type 1 error rate inflation of conventional tests applied in trials with (inbalanced) sample size reassessment and treatment selection.** Graf, Alexandra; Bauer, Peter	**Non-parametric Procedures in Biometry (2)** M. Neuhäuser/ F. Konietschke **High-dimensional structured repeated measures under non-normality: An approximation for finite dimensions** Helms, Hans-Joachim; Brunner, Edgar **High-dimensional structured repeated measures for two samples** Becker, Benjamin Markus; Brunner, Edgar **High-dimensional structured repeated measures under non-normality: Asymptotic distribution of quadratic forms** Brunner, Edgar; Ellenberger, David **Ranking procedures for matched pairs with missing data** Konietschke, Frank	**Health Economics** C. Schwenke and R. Nixon **Estimating the costs of hospital-acquired infections: Bernoulli vs. Landmarking** Beyersmann, Jan; Schumacher, Martin **The role of cost effectiveness analysis in NICE Technology Appraisals: current practice, challenges and future developments** Wailoo, Allan **Biometrical Requirements for Dossiers in the Framework of Early Benefit Assessment** Ralf Bender, Yvonne-Beatrice Schüler
10:30-11:00	**Break**		
11:00-12:30	**Bayesian Methods and Decision Theory (1)** K. Ickstadt **Sensitivity analysis in Bayesian generalized linear mixed models for binary data** Roos, Malgorzata; Held, Leonhard **Estimating the number of compartments with reversible jump MCMC** Kärcher, Julia C.; Schmid, Volker J. **Functional uniform prior distributions for nonlinear regression** Bornkamp, Björn **On bayesian and generalized confidence intervals on the variance components in mixed linear models** Michalski, Andrzej	**Young Statisticians (1)** A. Berghold **On generalized Hurwitz--Lerch Zeta distributions occuring in statistical inference** Jankov, Dragana **Estimating the risk of a Down's syndrome term pregnancy using age and serum markers. Comparison of various methods.** Sikolya, Kinga; Baran, Sándor; Veress, Lajos **Hierarchical regression to adjust for multiple comparisons in a case-control study of occupational risk for lung cancer** Corbin, Marine; Vermeulen, Roel; Kromhout, Hans; Peters, Susan; Simonato, Lorenzo; Richiardi, Lorenzo; Merletti, Franco; Pearce, Neil; Maule, Milena **The Extended-Quasi-Likelihood-Function in Generalized Linear Models** Thaler, Thorn **State-of-the art solutions for class-imbalance problem. Why don't they work on high-dimensional class-imbalanced data** Blagus, Rok	**Time to Event Analysis (1)** J. Beyersmann **Quanitifying the correlation of paired survival times under censoring** Schemper, Michael; Kaider, Alexandra; Wakounig, Samo **Copulas and frailty models for clustered survival data** Wienke, Andreas **Software for the Analysis of shared gamma and lognormal Frailty Models in SAS and R** Hirsch, Katharina; Wienke, Andreas **Sample size calculation in cluster randomized trials with a time to event outcome** Jahn-Eimermacher, Antje; Schneider, Astrid

Tuesday 13. Sep 2011 Morning		
Time	**Lecture Hall KOL-F104**	**Lecture Hall KOL-F121**
9:00-10:30	**Challenges in Agricultural Science (1)** J. Spilke **On the dependence of the size of GxE interaction on the number of trials** Pilarczyk, Wieslaw **Analyzing genotypes by environment interaction by curvilinear regression** Mejza, Stanislaw; Mejza, Iwona; Pereira, Dulce; Rodrigues, Paulo; Mexia, Joao **Augmented p-rep designs** Piepho, Hans-Peter; Williams, Emlyn; Whitaker, David	**Multiple Testing (1)** L. Hothorn **On the null-problem in multiple hypotheses testing** Gontscharuk, Veronika; Finner, Helmut **Significance and dependence occurring in gene expression analyses** Landwehr, Sandra; Finner, Helmut; Gontscharuk, Veronika **Expected Number of False Rejections of FDR- and FWER-Controlling Procedures Under Dependence** Scheer, Marsel; Finner, Helmut **Repeated significance tests for high-dimensional data** Zehetmayer, Sonja, Posch, Martin
10:30-11:00	**Break**	
11:00-12:30	**Challenges in Agricultural Science (2)** C. Richter **Methods and Models to analyze ordered categorical data with spatial covariance - a simulation based comparison** Höltl, Karen; Thamm, Katrin; Mielenz, Norbert; Spilke, Joachim **Analysis of count data with repeated measurements – tested by simulation studies** Thamm, Katrin; Höltl, Karen; Mielenz, Norbert; Spilke, Joachim **Bayesian analysis with plant breeding data** Möhring, Jens; Piepho, Hans-Peter **Sparse covariance matrices in random models for quantitative trait locus discovery in F2 populations** Zimmer, Daisy; Reinsch, Norbert	**Multiple Testing (2)** S. Zehetmayer **Calculation of Simultaneous Confidence Intervals by Constraint Propagation** Gutjahr, Georg , Bretz, Frank **Multiple comparisons to both a negative and a positive control** Hasler, Mario **Test Procedures for the Assessment of the Components of Composite Endpoints** Rauch, Geraldine; Kieser, Meinhard

Tuesday 13. Sep 2011 Afternoon			
Time	**Lecture Hall KOH-B10**	**Lecture Hall KO2-F180**	**Lecture Hall KOL-F101**
12:30-14:00	**Lunch**		
14:00-15.30	**Bayesian Methods and Decision Theory (2)** K. Ickstadt **Incorporating utilities in Bayesian models for risk-benefit decision-making** Ashby, Deborah **Benefit-Risk of Multiple Sclerosis Treatments: Lessons Learnt in Multi-Criteria Decision Analysis** Nixon, Richard; Oliveira, Pedro **Use of historical data to support the planning, analyzing and decision making of Proof of concept studies: an example in Irritable Bowel Syndrome** Belleli, Rossella **Integration of Copy Number Variation and Gene Expression Data in Bayesian Regression Models for Prediction and Biomarker Selection** Zucknick, Manuela ; Pfister, Stefan ; Benner, Axel	**Clinical Trial Application** A. Berghold and J. Roehmel **Multiplicity in Confirmatory Clinical Trials: New aspects and emerging principles** Benda, Norbert **A new approach for the simultaneous assessment of statistical significance and clinical relevance** Kieser, Meinhard **Is there a danger of "bio-creep" with non-inferiority trials?** Beryl, Primrose; Vach, Werner **QTc Analysis Employing Continuous 24 h Holter ECGs** Ferber, Georg, Holzgrefe, Henry	**Time to Event Analysis (2)** H. Binder **Cox-based structural equation modelling with latent variables in cardiovascular trials** Herich, Lena; zu Eulenburg, Christine; Wegscheider, Karl **On time-varying effects in high-dimensional survival studies** Buchholz, Anika; Sauerbrei, Willi; Binder, Harald **Nonparametric hazard rate estimation for relative survival models** Frantal, Sophie; Brannath, Werner **Modified Aalen-Johansen Estimator of the Cumulative Incidence Function for Left-Truncated Competing Risks Data** Allignol, Arthur; Schumacher, Martin; Meister, Reinhard; Beyersmann, Jan
15:30-16:00	**Break**		
16:00-17:30	**Awards Session DR and ROeS**		**Time to Event Analysis (3)** A. Wienke **A new insight in the validation of prognostic indices based on frailty models and an application to the validation of a prognostic index in bladder cancer patients.** Legrand, Catherine **The survAUC package: Tools to Evaluate the Prediction Accuracy of Survival Models** **Schmid, Matthias** **The trend renewal process: a useful model for medical recurrence data** Pietzner, Diana
Evening	**Reception by the City and Kanton Zürich / Poster and Wine: Lichthof (atrium) of University (KOL-E)**		

Tuesday 13. Sep 2011 Afternoon		
Time	**Lecture Hall KOL-F104**	**Lecture Hall KOL-F121**
12:30-14:00	**Lunch**	
14:00-15.30	**Diagnostics 1: ROC-Analyses - Sensitivity and Specificity** C. Schwenke **A smooth ROC curve estimator based on log-concave density estimates** Rufibach, Kaspar **Three principles for the joint evaluation of sensitivtiy and specificity in analysing a diagnostic study** Vach, Werner; Hoilund-Carlsen, Poul Flemming; Gerke, Oke **Comparison of two diagnostic tests regarding sensitivity and specificity** Rieck, Daniela; Zapf, Antonia **Evaluation of Reader Heterogeneity in Diagnostic Trials** Haeussler, Katrin; Koch, Armin; Zapf, Antonia	**Multiple Testing (3)** W. Maurer **Multiple Comparisons Problems in Complex Clinical Trial Designs** Hung, H.M. James , Wang, Sue-Jane **Graphical approaches for multiple endpoint problems using weighted parametric tests** Glimm, Ekkehard; Frank, Bretz; Willi, Maurer **Simultaneous confidence intervals for evaluation of multi-arm trials** Hothorn, Ludwig A **Use of modeling approaches to support dose selection at interim in adaptive designs for confirmatory clinical trials** Koenig, Franz , Bretz, Frank, Bornkamp, Bjoern , Graf, Alexandra .
15:30-16:00	**Break**	
16:00-17:30	**Diagnostics 2: Biomarkers and Modeling** C. Schwenke **Study planning for the validation of a prognostic marker** Lanius, Vivian **Subgrouping variables for the genetic characterization of lung cancer** Netzer, Christian; Rahnenführer, Jörg **Comparison of methods for clustered data and meta-analysis in diagnostic studies** Zapf, Antonia; Kuss, Oliver **Nonparametric analysis of diagnostic trials regarding clustered data** Lange, Katharina; Brunner, Edgar	**Miscellaneous** W. Maurer **Multidimensional ordination of classification methods performance for microarray data** Małgorzata Cwiklinska-Jurkowska, Magdalena Wietlicka-Piszcz **Test methods for correlated functional imaging data** Daniela Adolf, Siegfried Kropf **On the normality of the log odds ratio** Karl-Ernst Biebler, Bernd Jäger **Systematische Übersicht zur Therapie analer Inkontinenz bei Erwachsenen mit Biofeedback und Elektrostimulation** Reinhard Vonthein, Tankred Heimerl, Inke R. König, Christiane Wichmann, Claudia Hemmelmann, Thilo Schwandner, Andreas Ziegler
Evening	**Reception by the City and Kanton Zürich / Poster and Wine: Lichthof (atrium) of University (KOL-E)**	

Wednesday 14. Sep 2011 Morning			
Time	**Lecture Hall KOH-B10**	**Lecture Hall KO2-F180**	**Lecture Hall KOL-F101**
9:00-10:30	**Young Statisticians (2)** S. Mejza **Gene selection procedures including correlation between genes** Zyprych-Walczak, Joanna Grazyna; Siatkowski, Idzi **Linear Mixed Model in Gene Selection Problem** Szabelska, Alicja; Siatkowski, Idzi **Greenhouse gases emission in agriculture - measurements for today and prospects for the future** Wójcik-Gront, Elżbieta **Sample size calculation for the three-arm 'gold standard' non-inferiority design** Stucke, Kathrin; Kieser, Meinhard **Sensitivity-Preferred Strategy in Building Classifiers for High-Dimensional Data** Agueusop, Inoncent; Lehr, Stephan; Vonk, Richardus; Ickstadt, Katja	**Bayesian Methods in Translational Medicine (1)** M. Branson and L. Colin **What Can we Learn from 30 Years of Bayesian Methods in Early Drug Development?** Grieve, Andrew Peter **Sample size considerations for proof of concept studies with binary outcomes** Becka, Michael **A respiratory case study in early development: historical data inclusion and level-of-proof decision-making within a Bayesian framework.** Di Scala, Lilla; Kerman, Jouni; Beat, Neuenschwander	**Multivariable model-building with continuous variables – a comparison of flexible regression approaches** W. Sauerbrei **Multivariable model-building with continuous covariates: 1. Performance measures and simulation design** Binder H, Sauerbrei W, Royston P **Multivariable model-building with continuous covariates: 2. Comparison between splines and fractional polynomials** Sauerbrei W, Binder H, Royston P **Penalized splines and fractional polynomials for flexible modelling of the effects of continuous predictor variables: A comparison** Strasak, Alexander **Discussion** Heinze, Georg
10:30-11:00	**Break**		
11:00-12:30	**Statistics in practice (1): Double Session (for Young Statisticians)- Meta-Analysis and Experiences of Senior Statisticians** AG Nachwuchs, AG Weiterbildung, T. Mueller **Introduction to meta-Analysis (~45min)** Berghold, Andrea **Practical Experience of Senior Statisticians (~45min)** with: - Bretz, Frank (Novartis) - Berss, Joachim (University of Muenster) - Kappler, Martin (Novartis) - Wang, Sue-Jane (US Food and Drug Administration)	**Bayesian Methods in Translational Medicine (2)** M. Branson and L. Colin **Comparative Bayesian escalation designs** Lesaffre, Emmanuel; Hamberg, Paul; Verweij, Jaap; Dejardin, David **Combining information from healthy volunteers and patients for dose selection: case studies applying Bayesian statistics in early phase trials** Sauter, Annette; Niggli, Markus; Jordan, Paul **A Bayesian nonlinear mixed effects approach to analyzing data from rectal barostat experiments** Fisch, Roland; Belleli, Rossella	**Observational Studies** H.-U. Burger and A. Berghold **Pharmacoepidemiological databases: Strengths, limitations, methodological challenges** Pigeot, Iris **On observational studies for addressing safety concerns** Thakrar, Bharat **Estimation of survival probabilities for binary time-dependent covariates** Mittlboeck, Martina; Pötschger, Ulrike; Heinzl, Harald
12:30-13:00	**Lunch**		

Wednesday 14. Sep 2011 Morning		
Time	**Lecture Hall KOL-F104**	**Lecture Hall KOL-F121**
9:00-10:30	**Biometrical Journal: Editor's Choice** L. Held **Use of pre-transformation to cope with extreme values in important** Anne-Laure Boulesteix, Vincent Guillemot, Willi Sauerbrei **Adaptive Dose-finding: Proof of Concept with Type I Error Control** Miller, Frank **Assessing inter-rater reliability when the raters are fixed : two concepts and two estimates** Rousson, Valentin	**Epidemiology and Statistics** H. Ulmer and J. Wellmann **The Relative Frailty Variance and Shared Frailty Models** Unkel, Steffen; Farrington, C. Paddy; Anaya-Izquierdo, Karim **Estimating the number of malaria infections in blood samples using high-resolution genotyping data** Ross, Amanda; Köpfli, Cristian; Schöpflin, Sonja; Müller, Ivo; Felger, Ingrid; Smith, Tom **Parameter estimation in a six-state model for partially observable data in chronic kidney disease** Begun, Alexander; Icks, Andrea; Brinks, Ralph; Waldeyer, Regina; Koch, Michael; Giani, Guido **Modeling of continuous covariates in the mean structure of generalized estimating equations** Vens, Maren; Stolpmann, Jördis; Hemmelmann, Claudia; Ziegler, Andreas
10:30-11:00	**Break**	
11:00-12:30	**Genetics and Biomarker (1)** H. Binder **Temporal Activation Profiles of Gene Groups** König, André; Rahnenführer, Jörg **SVM-based models for cancer classification and survival prediction with high-dimensional genomic predictors** Rempel, Eugen; Rahnenführer, Jörg **Prediction of progression and therapy response for cancer patients: Are high-dimensional genomic data a blessing or a curse?** Rahnenführer, Jörg **Analysis of small-sample gene expression and gene interactions via Bayesian hierarchical models** Gasparini, Mauro; Rockstroh, Anja; Wells, Christine; Kennedy, Derek **A Bayesian network approach for pathway analysis using the gene ontology database** Foraita, Ronja; Karl, Janine; Leseberg, Annika; Günther, Frauke	**Miscellaneous** N. Neumann **Using penalized splines in extended Cox-type additive hazard regression to flexibly estimate the effect of time-varying serum uric acid on risk of cancer** Strasak, Alexander **Relation between incidence, prevalence and mortality in terms of a stochastic differential equation – formulation and application to renal replacement therapy** Brinks, Ralph; Landwehr, Sandra; Icks, Andrea; Koch, Michael; Giani, Guido **Statistical modeling of flexible pooling in unbalanced experiments for gene expression data** Pricop-Jeckstadt, Mihaela; Reinsch, Norbert; Rudolf, Henrik **Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions** Ogutu, Joseph O.; Schulz-Streeck, Torben; Piepho, Hans-Peter

<table>
<tr><th colspan="4">Wednesday 14. Sep 2011 Afternoon</th></tr>
<tr><th>Time</th><th>Lecture Hall
KOH-B10</th><th>Lecture Hall
KO2-F180</th><th>Lecture Hall
KOL-F101</th></tr>
<tr><td>12:30-13:00</td><td colspan="3">Lunch</td></tr>
<tr><td>13:00-14:00</td><td></td><td rowspan="2">ROeS General Assembly</td><td></td></tr>
<tr><td>14:00-15.00</td><td rowspan="2">DR General Assembly</td><td></td></tr>
<tr><td>15:00-16:00</td><td rowspan="2">ROeS Excursion</td><td></td></tr>
<tr><td>16:00-18:00</td><td></td><td></td></tr>
<tr><td>Evening</td><td colspan="3">Conference Dinner</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>

Thursday 15. Sep 2011		
Time	**Lecture Hall HAH-E-3**	**Lecture Hall HAH-E-11**
9:00-10:30	**Clinical Enpoints in Oncology (1)** H.-U. Burger and A. Berghold **Biomarkers and Surrogate Endpoints In Clinical Trials** Fleming, Tom **Titel Missing** Burzykowski, Tomasz **Challenges with the Overall Survival Endpoint in Oncology Trials and Investigation of some Alternatives** Neate, Colin James; Tong, Barbara; Burger, Hans Ulrich **Validating surrogate endpoints in breast and colon cancer: A systematic literature review** Schürmann, Christoph; Bender, Ralf; Kaiser, Thomas; Ringsdorf, Susanne; Vervölgyi, Elke; Vervölgyi, Volker; Wieseler, Beate	**Statistics in practice (2)** AG Nachwuchs, AG Weiterbildung, S. Roll Salanti, G
10:30-11:00	**Break**	
11:00-12:30	**Clinical Endpoints in Oncology (2)** H.-U. Burger and A. Berghold **Complex clinical endpoints are present in studies in hematopoietic cell transplantation** Schmoor, Claudia; Beyersmann, Jan **Composite Cancer Endpoints vs Combined Cancer Evidence** Wittkowski, Knut M. **Allogeneic transplants in acute myeloid leukemia (AML) – A case study on the performance of an Andersen-Gill model for a time-dependent intervening event** Zucknick, Manuela; Schlenk, Richard F.; Benner, Axel **Multiple Imputation of Missing Covariates for Multiple Survival Endpoints** Klingbiel, Dirk; Hsu Schmitz, Shu-Fang; Roth, Arnaud; Dietrich, Daniel; Delorenzi, Mauro	**Statistics in practice (3)** AG Nachwuchs, AG Weiterbildung, T. Mueller
12:30-13:00	**Lunch**	
13:00-14.30	**Closing Session** Closing talk: Prof. L. Hothorn	
15:30-16:00	**Break**	

Thursday 15. Sep 2011		
Time	**Lecture Hall HAH-F-1**	**Lecture Hall HAH-E-10**
9:00-10:30	**Genetics and Biomarker (2)** P. Schlattmann **Strategies for integrated analysis of genome wide measurements in risk prediction models** Hieke, Stefanie; Hielscher, Thomas; Schlenk, Richard F.; Schumacher, Martin; Benner, Axel; Bullinger, Lars; Binder, Harald **Improving SNP selection in genome-wide association studies with CAT and CAR scores** Zuber, Verena; Strimmer, Korbinian **Testing and Genetic Model Selection in Genome-Wide Association Studies** Loley, Christina; König, Inke R.; Hothorn, Ludwig; Ziegler, Andreas **Automated investigation of genotype calling using angles and tests for unimodality** Schillert, Arne; Pfützenreuter, Michael; Ziegler, Andreas **Automated Allele Calling** Schlattmann, Peter; Verba, Maryna	**Modeling and Simulation** M. Vandemeulebroecke **Capture-recapture with boosting: An application to estimating the comprehensiveness of literature searches for systematic reviews** Rücker, Gerta; Reiser, Veronika; Motschall, Edith; Binder, Harald; Meerpohl, Joerg J.; Antes, Gerd; Schumacher, Martin **Investigations on non-inferiority: the case of the FDA draft guidance on treatments for Nosocomial Pneumonia** Roehmel, Joachim; Kieser, Meinhard **Regularization and Model Selection with Ordinal Covariates** Gertheiss, Jan; Stelz, Veronika; Tutz, Gerhard
10:30-11:00	**Break**	
11:00-12:30	**Genetics and Biomarker (3)** A. Ziegler **Micronutrient Deficiencies and the Human Plasma Nutriproteome** Ruczinski, Ingo **Two-group comparisons of intensity values in omics experiments** Gleiss, Andreas; Mischak, Harald; Heinze, Georg **Group Effects in miRNA and related Target Gene Set Expression** Artmann, Stephan; Jung, Klaus; Bleckmann, Annalen; Beißbarth, Tim	**Benefit Risk** Günter Heimann **Benefit-Risk Assessment: A Data-Driven Approach.** Sarac, Sinan **What role should formal risk-benefit decision-making play in the regulation of medicines?** Ashby, Deborah **Quantitative Benefit-Risk Assessment and the Role of Statistics – Is there a need for ICH E9+?** Jürgen Kübler **Application of Spatial Weather Generator for Runoff Simulation in River Catchment** Kuchar, Leszek; Iwanski, Slawomir; Jelonek, Leszek; Szalinska, Wiwiana
12:30-13:00	**Lunch**	
13:00-14.30	**Closing Session** Closing talk: Prof. L. Hothorn	
15:30-16:00	**Break**	

Friday 16. Sep 2011					
Time	**Lecture Hall** **SOE-F-02**				
9:00-16:00	**Workshop** Probability estimation and data mining methods				

Schedule of AG meetings of the German Region

Council and Working group (AG) meetings					
Tuesday 13. Sep 2011					
Time	**Lecture Hall KOL-E-21**	**Seminar Room KOL-F-103**	**Seminar Room KOL-H-309**	**Seminar Room KOL-H-320**	**Lecture Hall KOL-F104**
12:30-13:30	AG Statistical Computing	AG Nachwuchs	AG Nichtparametrische Methoden		Editorial Board of the Biometrical Journal
13:30-14:00				AG Landwirt-schaftliches Versuchswesen	
Wednesday 14. Sep 2011					
08:15-09:00		Besprechung der AG-Leiter			
10:30-11:00	AG Räumliche Statistik				
12:30-14:00		Vorstand und Beirat der IBS, DR			
16:00-17:00	AG Statistische Methoden in der Medizin				

Part 1: Abstracts for presentations

****

Adaptive and sequential designs (1)

M. Vandemeulebroecke

Monday 12. September, 16:00 - 17:30, Lecture Hall KOH-B10

Adaptive enrichment designs, their methodological approaches and utility – what are regulatory implications

Sue-Jane Wang
US Food and Drug Administration, United States of America; suejane.wang@fda.hhs.gov

Keywords: Adaptive enrichment, pharmacogenomics, biomarker

Traditional, subgroup analyses are routinely performed in controlled clinical trials to assess the consistency of subgroups relative to the treatment effect observed in the intent-to-treat patients. In regulatory submissions, enrichment strategy has been one type of study designs to assess treatment effect in a relatively narrowly defined patient subgroup based on clinical features thought to increase the study power. With the advent of genomics, the concept of subgroup is gradually elevated to subpopulation due to the belief of potentially more accurately defined molecular targets.

In this paper presentation, a variety of on-trial enrichment strategy will be introduced. The utility of each strategy including design consideration, analysis approaches, case examples and regulatory perspectives will be presented and discussed. Some may be more suitable for learning purpose while others may be considered confirmatory.

Adaptive population enrichment: switching to a sub-population in response to interim trial data

Christopher Jennison
University of Bath, United Kingdom; c.jennison@bath.ac.uk

Keywords: Clinical trials, adaptive, enrichment

I shall describe the use of adaptive methods to change the focus of a clinical trial to a sub-population in which the treatment appears to be most effective. This requires a combination of sequential, adaptive and multiple testing methodologies. I shall assess the potential efficiency gains from these procedures and discuss point estimates and confidence intervals for treatment effects after such an adaptive trial.

Group sequential enrichment design incorporating subgroup selection

Baldur P. Magnusson[1], **Bruce W. Turnbull**[2]
[1]Novartis Pharma AG, Switzerland; [2]Cornell University; bpm36@cornell.edu

Keywords: Enrichment design, group sequential design, seamless Phase IIb/III, selection bias, subgroup analysis

An important component of clinical trials in drug development is the analysis of treatment efficacy in patient subgroups (subpopulations). Due to concerns of multiplicity and of the small sample sizes often involved, such analyses can present substantial statistical challenges and may lead to misleading conclusions. We propose an adaptive enrichment group sequential procedure in which resources can be concentrated on subgroups most likely to respond to treatment. This procedure can be applied as a confirmatory Phase IIb/III design. Stopping boundaries are defined through upper and lower spending functions. The procedure is presented in terms of the efficient score, enabling the analysis of normal, binary, or time-to-event data. The dilution effect is addressed by eliminating populations at the first stage that appear likely to derive no therapeutic benefit, and subsequently proceeding with the definitive assessment of treatment efficacy among the remaining pooled populations using a group sequential design. The procedure provides strong protection of familywise Type I error rate. We give an example to illustrate how the design is used, and draw comparisons with other methods found in the literature.

Statistical Methodology for Adaptive Patient Enrichment Designs

Gernot Wassmer
ADDPLAN An Aptiv Solutions Company, Germany; wassmer@addplan.com

Keywords: Adaptive designs, Patient enrichment, Sequential design

Adaptive patient enrichment designs enable the data-driven selection of one or more pre-specified subpopulations in an interim analysis, and the confirmatory proof of efficacy in the selected subset at the end of the trial. Strong control of the experimentwise Type I error rate is guaranteed by use of the combination testing principle due to Bauer and Koehne (1994) together with the closed testing argument (see, e.g., Wassmer, 2010). Using these principles, the way of how to perform the subset selection and the sample size recalculation needs not to be pre-specified. This procedure can be extended to the multi-stage case including additional sample size reassessment procedures at the interim analyses. It also applies to nested subsets of patients as described in Wang et al. (2009). In this talk the general methodology and designing issues when planning such a design are described. It is shown how to define overall confidence intervals and p-values. Furthermore, the advantage of using the adaptive approach as compared to the classical approach is assessed by simulation.

REFERENCES:

P. Bauer. and K. Koehne (1994). Evaluation of experiments with adaptive interim analyses. Biometrics 50:1029–1041.

S.-J. Wang, H.M.J. Hung, and R.T. O'Neill (2009). Adaptive patient enrichment designs in therapeutic trials. Biometrical J. **51**:358–374.

G. Wassmer (2010). Adaptive Interim Analyses in Clinical Trials. In: Handbook of Adaptive Designs in Pharmaceutical and Clinical Development (Pong, Chow, eds.). CRC Press.

Non-parametric Procedures in Biometry (1)

F. Konietschke

Monday 12. September, 16:00 – 17:30, Lecture Hall KO2-F180

Concordance Regression

Heinze, Georg
Section for Clinical Biometrics, CeMSIIS, Medical University of Vienna,
http://www.muw.ac.at/user/georg.heinze

Keywords: Concordance regression, semiparametric

Although flexible models exist to evaluate the relationship between a continuous covariate and an outcome variable, one often wants to evaluate monotonic relationships using robust nonparametric methods. We present concordance regression, a general regression framework suitable for this task. Here, the summands of the log-likelihood are all possible pairs of observations, and the dependent variable is an indicator of the concordance of each two observations forming a pair. The link function is such that the resulting regression coefficients are readily interpretable as log odds of concordance, and thus can easily be transformed into concordance probabilities. A sandwich estimate of the variance corrects inference for the multiple considerations of individuals in the likelihood. The regression framework is semiparametric in nature, but can be modified to allow for non-parametric estimation. Applied to a two-group comparison problem, concordance regression then simplifies to the Mann-Whitney-U test. Applied to more complex problems, concordance regression may enable, e. g., robust estimation in the presence of outliers or the consideration of ordinal independent or dependent variables. The model does not need to specify any particular functional relationship between independent and dependent variables except monotonocity. After presenting the basic theory behind concordance regression we explore some potential applications: nonparametric ANCOVA, survival analysis with possibly non-proportional hazards, and group comparison of zero-inflated data. The method is implemented in the R package concreg.

Unimodal regression using B-splines

Claudia Köllmann[1], Björn Bornkamp[2], Hernan Edgardo Grecco[3], Roland Fried[1], Katja Ickstadt[1]

[1]Department of Statistics, TU Dortmund University, Dortmund, Germany; [2]Novartis Pharma AG, Basel, Switzerland; [3]Max Planck Institute of Molecular Physiology, Dortmund, Germany; koellmann@statistik.tu-dortmund.de

Keywords: Nonparametric regression, shape constraint, penalized splines, dose finding, systems biology

Since the initial work on unimodal regression by Frisén in 1986, see [1], there has been little research in this field of nonparametric regression. However, there is need for such methods in applications, where a unimodal relationship between a predictor and a response variable is likely, as for example in dose-response studies or systems biology.

The idea of [1] is to iterate through all possible modes, as for a known mode the unimodal regression can be replaced by two monotonic regressions (one increasing, one decreasing). However, isotonic regression methods are based on pointwise estimates, which do not produce smooth curves.

Several authors make use of regression splines today, which are a very flexible tool for curve fitting. In some contexts splines might even be too flexible, but there exists a variety of analyses showing that imposing shape constraints such as monotonicity or convexity on *penalized* splines is very promising, see for example [2].

In this talk we will extend these possible shape constraints by unimodality. We will give application examples from dose-response and systems biology data. The spline regression is implemented taking up the original idea of [1] and making use of B-spline basis functions. In addition we will also refer to extensions of the proposed method using Bayesian and robust techniques.

References

[1] M. Frisén (1986), Unimodal Regression, The Statistician, **35**:479-485.

[2] M. L. Hazelton and B. A. Turlach (2011), Semiparametric Regression with Shape Constrained Penalized Splines, Computational Statistics and Data Analysis, online available at doi:10.1016/j.csda.2011.04.018.

Nonparametric inference on the cumulative abortion probability in a competing risk model

Susanna Di Termini[1,2], Stefanie Hieke[1,2], Martin Schumacher[2], Jan Beyersmann[1,2]
[1]Institute of Medical Biometry and Medical Informatics, University of Freiburg, Germany;
[2]Freiburg Centre for Data Analysis and Modeling, University of Freiburg, Germany;
susanna@fdm.uni-freiburg.de

Keywords: Cumulative incidence function, wild bootstrap, abortion, delayed study entry

The motivating data example is based on a study investigating the impact of coumarin exposure on pregnancy outcome. This is a left-truncated competing risk model where women enter the study several weeks after conception and where, spontaneous abortion is the event of primary interest and live birth and induced abortion are competing events.

Our goal is inference on the cumulative abortion probability, affected by the presence of left-truncation, by means of simultaneous confidence bands providing results both for one-sample and for two-sample situations. A theoretical difficulty in dealing with a direct evaluation of the aymptotic process resulting from a complicated nature of covariance function, suggested the use of a non-parametric wild bootstrap-type resampling [1].

Such approach has sometimes been used in biostatistical applications but has coverage probability problems as recently pointed out in [2].

A first improvement comes from considering a Greenwood-type wild-bootstrap technique whose variance estimator is of Greenwood-type, which is less biased variance type estimator.

The wild bootstrap point of view also allows to use non-normal variates. We investigate whether the use of mean-zero Poisson variates with variance 1 combined with a Greenwood approach may further improve on the coverage probabilities results.

To this aim, simulation studies reflecting the left-truncated pregancy outcome data are provided.

References

[1] DY. Lin (1997). Non-parametric inference for cumulative incidence functions in competing risks studies, Statistics in Medicine **16**:901-910.

[2] R. Bajorunaite, JP. Klein (2007). Two-sample tests of the equality of two cumulative incidence functions, Computational Statistics & Data Analysis; **51**:4269-4281.

Spatial Statistics

L. Held and J. Dreesman

Monday 12. September, 16:00 – 17:30 Lecture Hall KOL-F101

Spatial distribution of notified infectious diseases in Germany 2001-2010

Iris Zoellner[1], Erik Seibold-Krämer[2]

[1]LGA BW, Germany; [2]Universität Mainz; Iris.Zoellner@rps.bwl.de

Keywords: Disease mapping, infectious diseases, notified cases, incidences

Since 2001 the notification of infectious diseases in Germany is managed via electronic data transfer from local health offices to the federal health administration level (Bundeslaender) and to the Robert Koch Institute, Berlin, where infectious disease data for Germany are collected in a database[1]. Disease mapping and data analysis of notified incidence rates are possible using the programme Survstat (RKI Berlin) [2].

Incidence maps of the most frequently notified infectious diseases in administrative units (Stadt-/Landkreise) will be presented for the years 2001 to 2010. For some infections the disease maps show specific areas of higher risk (tick-borne encephalitis, Hanta virus infec-tions, Q fever), other maps reflect regional differences in vaccination coverage, and for some map series differences in notification rates seem to give a plausible explanation of the spatial patterns observed.

Disease mapping and its role in outbreak situations is illustrated using data from a legionellosis outbreak in Ulm 2009/10 and several q fever outbreaks.

References:

[1] D. Faensen, H. Claus, J. Benzler, A. Ammon, T. Pfoch, T. Breuer and G. Krause (2006). Suvnet@RKI - a multistate electronic reporting system for communicable diseases. Eurosurveillance (11).

[2] E. Seibold-Krämer (2010). Spatial distribution of reported infectious diseases in Germany - An example of using epidemiological data from SurvStat. Master Thesis, Johannes-Gutenberg-Universität Mainz.

Influenzanet: An Internet-based System for Monitoring Influenza-like Illness

Markus Schwehm
ExploSYS GmbH, Germany; markus.schwehm@explosys.de

Keywords: Influenza-like illness, syndromic surveillance, spatial epidemiology, pattern-based modelling

The influenza pandemic 2009 has revealed that the existing surveillance of influenza-like illness (ILI) in Europe is not sufficient. The existing sentinel systems via voluntarily participating physicians count only severe cases that seek medical help, thus omitting moderately sick cases that do not need help. Furthermore, Germany, France and Bulgaria only count the less specific cases of acute respiratory illness (ARI). But even for those countries where ILI data is collected, the data is not directly comparable, since the countries use different ILI case definitions. The EU-funded EPIWORK project establishes Influenzanet, a new internet-based system for monitoring ILI activity in European countries. Volunteers can register on national websites to answer a weekly questionnaire about their ILI-related symptoms during the influenza season. The first Influenzanet website was launched 2003/04 in the Netherlands an Belgium (deGroteGriepMeting). Since then websites in Italy (Influweb), Portugal (gripenet) and the United Kingdom (flusurvey) have been established. Today more than 30.000 volunteers participate in the Influenzanet project. In the coming influenza season Sweden (Influenzaskoll) and the German speaking countries (Aktiv-gegen-Grippe) will follow. Initiatives for France and Spain are in preparation. Using a uniform ILI case definition for all countries will provide an overview over the spatial dynamics of ILI activity in Europe. The Influenzanet project is designed to complement the existing sentinel surveillance systems. Aggregated data about ILI activity in Europe is directly accessible via the participating websites. The collected spatial and age-stratified raw data is used for research and modelling in spatial epidemiology.

Spatio-temporal modelling of infectious disease surveillance counts

Michaela Paul, Leonhard Held
Institute for Social and Preventive Medicine, University of Zurich, Switzerland;
michaela.paul@uzh.ch

Keywords: Multivariate time series of counts, infectious diseases, random effects, proper scoring rules

National surveillance systems routinely collect data on various infectious diseases. Typically, surveillance data are available as weekly time series of counts of confirmed new cases, stratified for example by administrative regions. In this talk I will discuss a statistical modelling framework for the analysis of such data [1]. Inherent (spatio-)temporal dependencies are incorporated via an observation-driven formulation. Heterogeneous incidence levels or differences in disease transmission across regions may be due to factors such as age, vaccination status or environmental conditions. They are addressed in the model formulation by region-specific and possibly spatially correlated random effects. Inference is based on penalized likelihood methodology for mixed models. As the use of classical model choice criteria such as AIC or BIC can be problematic in the presence of random effects, models are compared by means of probabilistic one-step-ahead predictions and proper scoring rules. The methods of this talk will be illustrated by surveillance data from Germany.

References

[1] M. Paul and L. Held (2011). Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts, Stats Med, **30**:1118-1136.

Longitudinal Data

H. Binder

Monday 12. September, 16:00 – 17:30, Lecture Hall KOL-F104

Selection of variables in high-dimensional time-series analysis

Jürgen Läuter[1], Maciej Rosolowski[2]

[1]University of Magdeburg, University of Leipzig, Germany; [2]University of Leipzig, Germany; juergen.laeuter@med.ovgu.de

Keywords: Selection of variables, high-dimensional analysis, time series, gene expression analysis

We report investigations from the BMBF project "HaematoSys". We search for sets of co-regulated genes in lymphoma precursor cells under CD40 or BCR stimulation. The given data consist of n=3 independent repetitions with p=20000 genes at t=12 times (over 24 hours).

In the preliminary step, the nxp given time curves are shifted in such a way that the curves vary around zero. Thus, n=3 independent vectors of the dimension txp=12x20000=240000 arise. Our null hypothesis is that the n repetitions have the mean vector $\mathbf{0}$. Deviations from the hypothesis come about if some genes have non-stationary mean-value curves.

Let $\mathbf{Y}$ be the full data matrix with n rows and txp columns. Then weight vectors $\mathbf{d}_1$, $\mathbf{d}_2$,... are derived from $\mathbf{Y}'\mathbf{Y}$ by special principal components methods. Thus, score vectors $\mathbf{z}_1=\mathbf{Y}\mathbf{d}_1$, $\mathbf{z}_2=\mathbf{Y}\mathbf{d}_2$,... are obtained. Each score corresponds to a set of selected genes. To construct a set of similar genes, a minimum correlation of the time curves must be exceeded. The order of the scores is determined from the strength of non-stationarity. The sets are mutually non-overlapping.

We test the scores $\mathbf{z}_1$, $\mathbf{z}_2$,... for non-sphericity of their distributions. The property of non-sphericity shows the non-stationarity of the mean-value curves. The multiple significance level is kept by using the order of the scores. The whole selecting and testing procedure is built in such a way that any false significances, i.e. spherically distributed scores, corresponding to curves without systematic fluctuations, may occur with probability alpha, at most.

In the result, the significant gene-expression curves are presented.

Variable Selection for Generalized Linear Mixed Models by L1-Penalized Estimation

Andreas Groll, Gerhard Tutz
Ludwigs-Maximilians-Univerity, Germany; andreas.groll@stat.uni-muenchen.de

Keywords: Generalized linear mixed model, lasso, gradient ascent, penalty, variable selection

Generalized linear mixed models are a widely used tool for modeling longitudinal data. However, their use is typically restricted to few covariates, because the presence of many predictors yields unstable estimates. The presented approach to the fitting of generalized linear mixed models includes an L_1-penalty term that enforces variable selection and shrinkage simultaneously. A gradient ascent algorithm is proposed that allows to maximize the penalized log-likelihood yielding models with reduced complexity. In contrast to common procedures it can be used in high-dimensional settings where a large number of potentially influential explanatory variables is available. The method is investigated in simulation studies and illustrated by use of real data sets.

Linear mixed models with a penalized normal mixture as random effects distribution

Felix Heinzl, Gerhard Tutz
Ludwig-Maximilians-University Munich, Germany; felix.heinzl@stat.uni-muenchen.de

Keywords: Longitudinal data, linear mixed models, group lasso, EM algorithm

A method is proposed that aims at identifying clusters of individuals that show similar patterns when observed repeatedly. As background model we consider the linear mixed models which is widely used for the modeling of longitudinal data. In contrast to the classical assumption of a normal distribution for the random effects (for example [1] and [2]) a finite mixture of normal distributions used, for example, by [3] is assumed. Typically, the number of mixture components is unknown and has to be chosen, ideally by data driven tools. For this purpose an EM algorithm based approach is considered that uses a penalized normal mixture as random effects distribution. The penalty term shrinks the pairwise distances of cluster centers based on the group lasso method ([4]). The strength of regularization is determined by one penalization parameter. The effect is that individuals with similar time trends are merged into the same cluster.

References:

[1] G. Verbeke, and G. Molenberghs (2000). Linear mixed models for longitudinal data, New York: Springer.

[2] D. Ruppert, M.P. Wand, and R.J. Carroll (2003). Semiparametric Regression, Cambridge: Cambridge University Press.

[3] G. Verbeke and E. Lesaffre (1996). A linear mixed-effects model with heterogeneity in the random-effects population, Journal of the American Statistical Association **91**:217-221.

[4] M. Yuan and Y. Lin (2006). Model selection and estimation in regression with grouped variables, Journal of the Royal Statistical Society B, **68**:49-67.

How to analyse multiple ordinal scores in a clinical trial?

Celine Marielle Laffont[1,2], **Didier Concordet**[1,2]
[1]INRA, UMR 1331, Toxalim, F-31027 Toulouse, France.; [2]Université de Toulouse, INPT, ENVT, UPS, EIP, F-31076 Toulouse, France.; c.rousset@envt.fr

Keywords: Multivariate, ordinal, scores, longitudinal, modeling

Longitudinal measurements of ordinal responses (scores) are very frequent in clinical trials to assess drug efficacy/safety. Usually, several scores are recorded and this multiplicity is an issue for data analysis. Current methods consist in analysing an average/aggregated score or each score separately. These methods ignore the correlations between scores which often document different aspects of a same physio-pathological process (e.g. pain) and in the first case, the actual metric of the scores is not taken into account. We thus propose to generalise the approach described in [1] to properly analyse multiple longitudinal ordinal data and investigate similarities between scores.

We use the concept of latent variables to derive the joint distribution of K ordinal responses, together with mixed effects models. A stochastic EM algorithm was implemented in C for model estimation. Two simulation studies were carried out to assess the applicability of our method using drug dose or plasma concentration as a covariate. In the end, a principal component analysis was performed to summarise correlations between scores.

All model parameters were correctly estimated. As expected, multivariate and univariate analyses produced similar estimation of marginal distributions but gave different results regarding the percentage of subjects within each "crossing" category. We thus show that multivariate analyses can be more appropriate than univariate analyses for the assessment of drug efficacy and safety.

References:

[1] D. Todem, K. Kim and E. Lesaffre (2007). Latent-variable models for longitudinal data with bivariate ordinal outcomes, Stat Med, **26**:1034-1054.

Acknowledgement

We thank Novartis Pharma AG, Switzerland, for funding this project.

Missing Values

N. Neumann

Monday 12. September, 16:00 – 17:30, Lecture Hall KOL-F121

Estimation of the Treatment Effect in the Presence of Noncompliance and Missing Data

Ann-Kristin Leuchs[1,2], Norbert Benda[2], Jörg Zinserling[2], Markus Neuhäuser[1], Manfred Berres[1]

[1]RheinAhrCampus Remagen, Germany; [2]BfArM; aleuchs@rheinahrcampus.de

Keywords: Noncompliance, missing values, clinical trials, depression

Treatment noncompliance and missing data are a frequently emerging problem in clinical trial routine. Noncompliance is a broad term which incorporates any kind of deviation from the assigned treatment protocol (treatment discontinuation, rescue medication ...). Such violations often result in missing values.

Both, missing values and treatment noncompliance may bias the results. It may be useful to observe all patients until the trial's end irrespective of their protocol adherence. We consider the situation that at least for some noncomplying patients observations were collected even after their protocol violation.

We assumed noncompliance as being the discontinuation of treatment and to be the same for all patients. As a result the patient's longitudinal profile is dividable into on- and off-treatment observations.

Within the framework of longitudinal depression trials we compared different analysis strategies to include, both, on- and off-treatment observations to get insight in how the additional use of off-treatment data may affect the trial's outcome. We compared simple strategies that ignore off-treatment data or treat on- and off-treatment data the same with more complex strategies based on two-piece linear mixed models which assume different treatment effects for on- and off-treatment data.

We showed that the simple methods considered could be unacceptable since they tend to overestimate the treatment effects. Therefore it should be put effort in following all patients up to the trial's end irrespective of their compliance and include all available data in the analysis.

Orthogonal regression and related methods for bivariate data

Manfred Berres[1], Ruth Volland[1], Irene Beck[2], Andreas U. Monsch[2]
[1]RheinAhrCampus Remagen, FH Koblenz, Germany; [2]Memory Clinic, Dept. of Geriatrics, University Hospital, Basel; berres@rheinahrcampus.de

Keywords: Orthogonal regression, principal components

This study is motivated by the administration of two memory tests, the CERAD wordlist with 10 words and the CVLT wordlist with 16 words. When diagnosing individuals the CERAD wordlist is given to more impaired individuals, because it is easier to perform. Nevertheless, a sample of individuals from the Memory Clinic of the University Hospital, Basel, is available who have done both tests. A reasonable linear relationship between both variables should a unique line to predict *Y* from *X* and *X* from *Y*, without biasing the variance of the prediction. The first principal component appears to be a good candidate. This method is also known as orthogonal regression.

We developed a variance estimator for the slope of the principal component. This variance estimator and derived confidence intervals will be compared to other estimators. Bias and coverage probabilities will be investigated in simulation studies.

The Value of Pattern-Mixture Models for Sensitivity Analysis

Mouna Akacha
Novartis Pharma AG, Switzerland; mouna.akacha@novartis.com

Keywords: Informative missingness, pattern-mixture models, sensitivity analysis

Incomplete data due to missed visits, dropouts or non-return of questionnaires are very common in the clinical trial setting. In this context, missingness usually occurs for reasons outside the control of the investigators and may be related to the primary outcome of interest, hence complicating the data analysis.

Principled approaches to deal with incomplete data such as the Multiple Imputation and the Direct Likelihood approaches, can yield valid results under a "missing at random" (MAR) missingness process. A MAR process assumes that dropout is associated with the observed data and available covariates only, but is conditionally independent of the missing outcomes. This assumption implies that patients who drop out share the same outcome distribution as patients who remain in the study (after adjusting for relevant observed data and covariates).

Although the assumption of MAR can be realistic for certain settings, in most applications it is impossible to exclude the possibility that missingness depends on unobserved data, i.e. we cannot exclude a "missing not at random" missingness process. In particular, we may be averse to the assumption that future behavior of dropouts can be modelled using future behavior of those who remain.

In this talk we discuss how Pattern-Mixture Models can be used to relax this assumption. Different assumptions about the future behavior of dropouts dependent on dropout time, reason for dropout and received treatment can be made. The stability of the conclusions across a range of different assumptions can be explored.

Effect of genotype imputation errors on the accuracy of genomic selection

Nadine Neugebauer, Robin Wellmann, Jörn Bennewitz
Universität Hohenheim, Germany; N.Neugebauer@uni-hohenheim.de

Keywords: Genomic selection, imputation

In order to reduce costs of genomic selection a combination of low density SNP chips (LD-chip) and chips with a higher density (e.g. 50K chips, HD-chip) is frequently proposed, especially for species where the economic value of a single breeding individual is limited (e.g. chicken and pigs). This study investigated for a typical sire line pig breeding population a combination of LD- and HD-chips by stochastic simulation. A mutation drift population was simulated. Male individuals from five successive generations are HD-genotyped and progeny-tested. Marker effects were estimated by G-Blup and BayesA. Young selection candidates are offspring from this reference population and are LD-genotyped with 0.3k, 0.5k,… or 10k, while this markers covered the genome with different strategies. Missing genotypes were imputed using LDMIP software that combined family information and linkage disequilibrium. Imputation errors were recorded. The accuracy of gEBV of the selection candidates (i.e. the correlation between gEBV and TBV) was analysed and compared across different imputation error rates. In particular it was investigated at which error rate it is beneficial to use only markers of the LD-CHIP, also in the reference population.

The first results show that for all imputation error rates all kinds of marker effects should be used which resulted in more moderate loss. The differences between with and without imputation are up to 20% of accuracies.

Adaptive and Sequential Designs (2)

M. Kieser

Tuesday 13. September, 9:00 – 10:30, Lecture Hall KOH-B10

Improving adaptive group sequential designs with discrete outcomes

Stefan Englert, Meinhard Kieser
University of Heidelberg, Germany; englert@imbi.uni-heidelberg.de

Keywords: Adaptive design, group sequential design, discrete outcomes

In recent years different attempts have been made to transfer group sequential designs into an adaptive setting. One main objective is to preserve the operating characteristics if no changes are performed throughout the trial.

Müller and Schäfer [1,2] presented a general method that allows such adaptive interim analyses in group sequential designs. This method does not even require specifying the intention for adaptation a priori. If no adaptations are performed, the applied design will show identical characteristics as the corresponding classical group sequential design.

In this talk it is presented how the method by Müller and Schäfer can be improved in designs with discrete test statistics where the classical group sequential design does not exhaust the nominal type I error rate. This applies, for example, to phase II designs for oncology trials where a binary endpoint is used. The presented method controls and exhausts the specified type I error rate and leads to more efficient adaptive designs as compared to a direct application of the Müller and Schäfer method. The intention for adaptation has to be specified prior to the start of the trial and cannot be made ad hoc.

The proposed method is illustrated with examples.

References

[1] H.-H. Müller and H. Schäfer (2001). Adaptive Group Sequential Designs for Clinical Trials: Combining the Advantages of Adaptive and of Classical Group Sequential Approaches, *Biometrics*, **57**:886-891

[2] H.-H. Müller and H. Schäfer (2004). A general statistical principle for changing a design any time during the course of a trial. *Statistics in Medicine*, **23**:2497-2508

Estimation of the Hazard Ratio in Adaptive Designs with Sample Size Readjustment

Sandra Ligges[1], Gernot Wassmer[2], Christine Müller[1]
[1]Fakultät Statistik, TU Dortmund, Germany; [2]ADDPLAN GmbH, Cologne, Germany;
sligges@statistik.tu-dortmund.de

Keywords: Survival endpoint, estimation of hazard ratio, adaptive designs

In adaptive designs stagewise independent data is crucial for the validity of the performed procedures. Independency is usually achieved by dealing with different patient collectives. In survival studies where patients may contribute information to subsequent stages the conventional testing procedures have to be modified in order to retain independent stagewise test statistics.

In the literature two different strategies have been pursued. Independent information can be obtained by either using increments of certain pivot statistics, e.g. logrank statistics [1], or by right-censoring and left-truncating the data at the time points of the interim analyses [2].

Estimators for the hazard ratio can be constructed based on either of these methods. In the present talk such estimators for the hazard ratio in two-armed two-stage adaptive designs with survival endpoints are presented following the general construction principle proposed by [3]. Besides, some results of a simulation study for the comparison of different estimators using inverse normal type boundaries for increments in logrank statistics [1] are shown.

References

[1] G. Wassmer (2006). Planning and Analyzing Adaptive Group Sequential Survival Trials, BMJ,**48**:714-729.
[2] A. Jahn-Eimermacher and K. Ingel (2009). Adaptive Trial Design: A General Methodology for Censored Time to Event Data, Contemporary Clinical Trials, **30**:171-177.
[3] W. Brannath, F. König and P. Bauer (2006). Estimation in Flexible Two Stage Designs, Statistics in Medicine, **25**:3366-3381.

A graphical approach for adaptive clinical trials testing multiple hypotheses

Florian Klinglmueller[1], Franz Koenig[1], Martin Posch[2]
[1]Medical University of Vienna, Austria; [2]European Medicines Agency;
florian.klinglmueller@meduniwien.ac.at

Keywords: Graphical approach, multiple endpoints, adaptive design

The graphical approach [1] provides a convenient tool for the definition of closed testing procedures based on weighted tests for each intersection hypothesis. We adopt this approach to construct adaptive tests for clinical trials with an unblinded interim analysis that reflect the complex contextual relations between multiple hypotheses in clinical trials. Adaptive designs are an attractive choice for confirmatory clinical trials as they provide type I error control while permitting certain mid-trial design modifications based on internal and external information, e.g, changing the pre-planned sample size, inserting/dropping of treatment groups and endpoints in clinical trials. The discussed approach is based on the closed testing principle combined with the conditional error principle. Starting with a closed testing procedure based on weighted Bonferroni tests we construct, for all intersection hypotheses, a second stage test at levels equal or smaller than the sum of partial conditional error levels of the initial tests [2]. In contrast to other methods [3] the knowledge of the multivariate distribution of the test statistics is not required when using marginal conditional errors making the proposed approach, suitable for, e.g. comparing treatment groups and/or multiple endpoints.

References

[1] F. Bretz et al. (2008). A graphical approach to sequentially rejective multiple testing procedures. StatMed, **28**:586-604.

[2] M. Posch, et al. (2010). Type I error rate control in adaptive designs for confirmatory clinical trials with treatment selection at interim, PharmStat.

[3] F. Koenig, et al. (2008). Adaptive Dunnett Tests for Treatment Selection, StatMed, **27**:1612-1625.

Maximum type 1 error rate inflation of conventional tests applied in trials with (inbalanced) sample size reassessment and treatment selection.

Alexandra Graf, Peter Bauer
Medical University of Vienna, Austria; alexandra.graf@meduniwien.ac.at

Keywords: Sample size reassessment, changing allocation rates, interim analysis, maximum type 1 error rate, conditional error function

Sample size reassessment in an adaptive interim analysis based on an interim estimate of the effect size can considerably increase the type 1 error rate if the conventional fixed sample size test is applied for the final analysis. We investigate the maximum type 1 error rate of the pre-planned test for comparing the means of two independent normal distributions (with common known variance) which can be yielded when sample size and allocation rate can be modified in an interim analysis. Extensions to the scenario when more than one treatment is compared to a control and treatment selection is applied in the interim analysis by carrying on only the treatment with the largest effect and the control to the second stage are investigated. Here it is assumed that a multiplicity adjusted Dunnett test is performed in the final analysis. The maximum inflation of the type 1 error rate for such types of design can be calculated by searching for the "worst case" scenarios, i.e. sample size adaptation rules in the interim analysis that lead to the largest conditional type 1 error rate. Allowing for inbalanced sample sizes, to achieve the "worst case" would require knowledge of the nuisance parameter, the unknown common treatment effect under the global null hypothesis. Although this is a rather hypothetical scenario it may be approached in practice when using a standard control treatment for which precise estimates are available from historical data.

Non-parametric Procedures in Biometry (2)

M. Neuhäuser/ F. Konietschke

Tuesday 13. September, 9:00 – 10:30, Lecture Hall KO2-F180

High-dimensional structured repeated measures under non-normality: An approximation for finite dimensions

Hans-Joachim Helms, Edgar Brunner

University Medical Center Göttingen, Germany; hans-joachim.helms@t-online.de

Abteilung Medizinische Statistik, Universität Göttingen

Keywords: High dimensional data, low-dimensional approximation, repeated measures design, non-normal distribution

The general multivariate model (Bai and Saranadasa, 1996) is applied to a group of n independent subjects which are repeatedly observed under d different conditions (time points, treatments, etc.) where d may be smaller, larger or even much larger than n. It is not assumed that the n independent observation vectors follow a multivariate normal distribution. For the statistic proposed by Chen and Qin (2010) and improved by Ellenberger and Brunner (2011), we provide an approximation which is valid for $d > n$ as well as for $d < n$ by using the Box approximation considered by Ahmad, Werner and Brunner (2008). Finally, we apply the proposed procedure to the example of the marathon runners mentioned by Ellenberger and Brunner (2011).

References

Ahmad, M. Rauf, C. Werner and E. Brunner (2008). Analysis of high dimensional repeated measures designs: The one sample case. Computational Statistics and Data Analysis, **53**:416-427.

Z. Bai und H. Saranadasa, H. (1996). Effect of highdimension: by an example of a two sample problem. Statistica Sinica **6**:311-329.

S. Chen and Y. Qin (2010). A two-sample test for high-dimensional data with applications to gene-set testing. The Annals of Statistics **38**:808-835.

D. Ellenberger and E. Brunner (2011). High-dimensional structured repeated measures under non-normality: Asymptotic distribution of quadratic forms. Biometrisches Kolloquium, Zürich.

High-dimensional structured repeated measures for two samples

Benjamin Markus Becker, Edgar Brunner
Universität Göttingen, Germany; benjamin.becker@med.uni-goettingen.de

Keywords: High dimensional data, repeated measures design, finite approximation

Similar to the one-sample statistic of Helms and Brunner, we present a two-sample statistic allowing for unequal sample sizes and unequal covariance matrices. Using a finite approximation in both the sample sizes and the number of repetitions to derive the statistics, we can circumvent any restrictions to the covariance matrix. In contrast, competing tests by [4], [1] as well as [3] rely on specific dependency structures between the repeated measures as their various asymptotics involve the number of repetitions. These asymptotics are not necessary since omitting them in our statistics does not deteriorate the performance even for the case of not normally distributed data.

References

[1] B. Zhidong and H. Saranadasa (1996). Effect of high dimension: by an example of a two sample problem. Statist. Sinica, **6**:311-329.

[2] E. Brunner (2009). Repeated Measures under Non-Sphericity. 6th St. Petersburg Workshop on Simulation, **2**:605-610.

[3] S. X. Chen and Y.L. Qin (2010). A two-sample test for high-dimensional data with applications to gene-set testing. Ann. Statist., **38**:808-835.

[4] H. Wang and M. G. Akritas (2010). Inference from heteroscedastic functional data. J. Nonparametr. Stat., **22**:149-168.

High-dimensional structured repeated measures under non-normality: Asymptotic distribution of quadratic forms

Edgar Brunner, David Ellenberger
Universität Göttingen, Germany; brunner@ams.med.uni-goettingen.de

Keywords: High dimensional data, asymptotic distribution, repeated measures design, non-normal distribution

Motivated by the example of the marathon runners, we consider a repeated measures design for one group of n independent subjects where the d > n repeated measures may have a factorial structure. We do not assume a multivariate normal distribution of the data. For testing hypotheses about the repeated measures in the general multivariate model [2], we derive a quadratic form using some results of [3]. In a one-group design, this quadratic form is asymptotically equivalent to the statistic T_n of Chen and Qin [4] which has been derived in a multivariate set-up. Here, we suggest a different variance estimator which improves the quality of the approximation to the limiting normal distribution. This estimator is L_2 ratio-consistent uniformly in the dimension d. Under the assumption of normality, this estimator was already considered in [1] and [5]. Simulations for different sample sizes, dimensions and covariance structures of the repeated measures are presented comparing the Chen-Qin statistic and the new proposed statistic.

References

[1] M. R. Ahmad, C. Werner and E. Brunner (2008). CSDA, **53**:416-427.

[2] Z. Bai, und H. Saranadasa (1996). Statistica Sinica, **6**:311-329.

[3] A. C. Bathke, S. W. Harrar, (2008). JSPI, **138**:588-610.

[4] S. Chen and Y. Qin (2010). AS, **38**:808-835.

[5] E. Brunner, B. Becker and C. Werner (2009). Technical Report, Dept. Medical Statistics, University of Göttingen.

Ranking procedures for matched pairs with missing data

Frank Konietschke
University Medical Center Göttingen, Germany; fkoniet@gwdg.de

Keywords: Rank procedures, relative effects, confidence intervals

Nonparametric methods for matched pairs with data missing completely at random are considered. It is not assumed that the observations are coming from distribution functions belonging to a certain parametric or semi-parametric family. In particular, the distributions can have different shapes under the null hypothesis. Hence, the so-called nonparametric Behrens–Fisher problem for matched pairs with missing data is considered. Moreover, a new approach for confidence intervals for nonparametric effects is presented. In particular, no restriction on the ratio of the number of complete and incomplete cases is required to derive the asymptotic results. Simulations show that for arbitrary settings of complete data and missing values, the resulting confidence intervals maintain the pre-assigned coverage probability quite accurately. Regarding the power, none of the proposed tests is uniformly superior to the other. A real data set illustrates the application.

References:

F. Konietschke, S. W. Harrar, K. Lange and E. Brunner (2011). Ranking procedures for matched pairs with missing data — Asymptotic theory and a small sample approximation. Computational Statistics and Data Analysis. In press

Health Economics

C. Schwenke and R. Nixon

Tuesday 13. September, 9:00 – 10:30, Lecture Hall KOL-F101

Estimating the costs of hospital-acquired infections: Bernoulli vs. Landmarking

Jan Beyersmann[1,2], Martin Schumacher[2]

[1]University Freiburg, Germany; [2]University Hospital Freiburg, Germany; jan@fdm.uni-freiburg.de

Keywords: Nosocomial infection, cost benefit, survival analysis, time-dependent exposure

Hospital-acquired infections are a major healthcare concern, and there is growing political pressure to improve prevention. However, infection control can easily be costly, requiring extra personnel, isolation rooms or structural measures. Cost benefit studies weigh these costs against the costs caused by infection. The latter are generally considered to arise via prolonged stay in hospital. Traditional analyses of this extra length of stay retrospectively stratified patients into infected and non-infected, leading to gross overestimation of extra stay and, hence, costs. Fueled by recent results on the nature of this so-called time-dependent bias, there is now growing consensus on the fact that the timing of infection must be adequately modelled to avoid this bias. While this fact is indisputable, we show that it is less straightforward to formalize what we *mean* by "extra length of stay". The statistical difficulty is that hazard models with time-dependent covariates do not translate into probability models without further ado. We illustrate that two common approaches to cope with this difficulty lead to different concepts of "extra length of stay". We use landmarking, which originated from oncology in the 1980s with the aim to compare cancer patients by treatment response. And we use a "partial probability" approach, which dates back as far as to Daniel Bernoulli's 1760 study on vaccination against smallpox. One implication is that the common observational studies may not suffice to fully judge effectiveness and costs of infection control measures and that randomized clinical trials may be needed.

The role of cost effectiveness analysis in NICE Technology Appraisals: current practice, challenges and future developments

Allan Joseph Wailoo
University of Sheffield, United Kingdom; a.j.wailoo@sheffield.ac.uk

Keywords: NICE, health economics, cost effectiveness

Guidance to the health service of England and Wales on the use of new and existing health technologies is provided by the National Institute for Health and Clinical Excellence (NICE). The analysis of cost effectiveness has been at the heart of appraisals at NICE since its inception in 1999. This talk will introduce the key components on cost effectiveness analysis for non specialists and highlight how these estimates are typically undertaken in the NICE context. In particular, the role of QALYs and decision analytic models will be outlined before discussing some of the challenges in incorporating the varied evidence on clinical effectiveness, health utilities and costs into this framework. The additional challenges of translating these assessments into appraisal guidance will also be outlined. The talk will then highlight areas of likely future change in NICE methods, arising from developments both in scientific methods and the decision making landscape within the National Health Service.

Biometrical Requirements for Dossiers in the Framework of Early Benefit Assessment

Ralf Bender, Yvonne-Beatrice Schüler
Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Ralf.Bender@iqwig.de

Keywords: Benefit assessment, health economics

According to the law on the reorganization of the pharmaceutical market new approved drugs have to undergo an early benefit assessment in Germany since January 2011 with the aim of a structured price regulation. The basis of the early benefit assessment is given by dossiers prepared by the manufacturer submitted to the Federal Joint Committee (G-BA), which can commission the Institute for Quality and Efficiency in Health Care (IQWiG) with an assessment of the dossier. On the G-BA website comprehensive German-language templates for the dossiers including methodological advice are available. An overview of the biometrical requirements for the dossiers is given with a special focus on indirect comparisons and surrogate endpoints.

Challenges in Agricultural Science (1)

J. Spilke

Tuesday 13. September, 9:00 – 10:30, Lecture Hall KOL-F104

On the dependence of the size of GxE interaction on the number of trials

Wieslaw Pilarczyk
Poznań University of Life Sciences, Poland; wpilar@up.poznan.pl

Keywords: GxE interaction, size of trial series

In Poland, inference on the value for cultivation and use (VCU) of new varieties is usually based on the results of series of trials performed in different environments (trial stations, locations) over a period of 3-4 years. Similar series are also performed for post-registration investigations. The size of series (number of locations) depends on economic importance of species and varies (in Poland) from a very few trials for less important species to several dozens for the most important ones. There is a permanent economical pressure on reduction of the size of series. In this paper the influence of the number of experiments on the size of genotype-environment interaction is investigated. Also the contribution of particular experimental stations (locations) to the genotype-environment interaction is assessed.

The results of four year series of 219 trials on winter wheat (years 1990-1993), three year series of 113 trials on spring barley (years 1993-1995) and five year series of 138 trials on oil-seed rape (years 2006-2010) form the basis for investigations. All trials were performed by The Research Centre For Cultivar Testing, Słupia Wielka, Poland. It has been shown, that – on average – the size of interaction expressed in terms of mean square – stabilizes when the number of (properly chosen) locations is about 15. Several experimental stations (locations) with high contribution to the interaction were indicated.

Analyzing genotypes by environment interaction by curvilinear regression

Stanislaw Mejza[1], Iwona Mejza[1], Dulce Pereira[2], Paulo Rodrigues[3], Joao Mexia[3]
[1]Poznan University of Life Sciences, Poland; [2]University of Evora, Portugal; [3]Universidade Nova de Lisboa, Portugal; smejza@up.poznan.pl

Keywords: Coincidence test; environmental indexes; genotype by environment interaction; regression analysis.

This paper deals with the analysis of the series of experiments conducted over a range of observed or potential environments in order to select genotypes that are consistently high-yielding (e.g. have high economical value). In this paper regression analysis is used to make inference concerning genotype selection. The so-called adjusted means (or some other genotypic characteristic) for genotypes usually constitute observations of the dependent variable. The problem is how to model the independent variable, i.e. the environmental indexes. In the paper we use the environmental indexes obtained by an iterative ("zig-zag") algorithm based on the joint regression approach.

The data considered here is usually used to: (i) predict and estimate the yield; and (ii) provide reliable guidance for selection of the best genotypes for planting over environments.

An inference concerning these problems is based on adapting two tests: (A) a test for parallelism of regression curves; and (B) a test of coincidence (identical regression functions). In case of failure to reject the hypothesis (A), it is worth applying test (B). The theoretical considerations are illustrated with an example of yield from a winter rye experiment.

Augmented p-rep designs

Hans-Peter Piepho[1], Emlyn Williams[2], David Whitaker[3]
[1]Universität Hohenheim, Germany; [2]Australian National University, Canberra, Australia;
[3]Ranfurly, New Zealand; piepho@uni-hohenheim.de

Keywords: Experimental design, alpha arrays, series of experiments

Early generation variety trials are very important in plant and tree breeding programs. Typically many entries are tested, often with very little resources available. Unreplicated trials using control plots are popular and it is common to repeat the trials at a number of locations. An alternative is to use p-rep designs, where a proportion of the test entries are replicated at each location; this can obviate the need for control plots. α-Designs are commonly used for replicated variety trials and we show how these can be adapted to produce efficient p-rep designs.

References

E. R. Williams, H. P. Piepho, D. Whitaker (2011): Augmented p-rep designs. Biometrical Journal **53:**19-27.

Multiple Testing (1)

L. Hothorn

Tuesday 13. September, 09:00 – 10:30, Lecture Hall KOL-F121

On the null-problem in multiple hypotheses testing

Veronika Gontscharuk, Helmut Finner
German Diabetes Center, Leibniz Center at the Heinrich-Heine-University Duesseldorf, Institute of Biometrics and Epidemiology, Germany; veronika@gontscharuk.de

Keywords: Weak dependence, asymptotic FDR control, null-problem

Suppose we are concerned with a multiple test problem with some dependence structure between test statistics or p-values. In general, depending on the underlying error rate criterion, dependency may increase or decrease the chance of false rejections. We discuss this issue and give various illustrative examples. A special type of dependence is weak dependence. In terms of p-values, weak dependence appears if the empirical cumulative distribution function of p-values under null hypotheses is asymptotically stochastically bounded by the cdf of a uniform variate on [0,1]. There is some evidence that a multiple test procedure with asymptotic control of the false discovery rate (FDR) under independence also controls the FDR under weak dependence, provided the asymptotic threshold is bounded away from 0. But if the asymptotic threshold tends to 0, the situation is completely unclear. The question arises whether the FDR is controlled in such cases. We call this issue, which is often ignored in the literature, the null-problem in multiple hypotheses testing. It will be shown that weak dependence provides no guarantee of asymptotic FDR control if the null-problem appears.

Significance and dependence occurring in gene expression analyses

Sandra Landwehr, Helmut Finner, Veronika Gontscharuk
German Diabetes Center at the Heinrich Heine University Düsseldorf, Germany;
sandra.landwehr@ddz.uni-duesseldorf.de

Keywords: Multiple testing, dependence, microarrays

Multiple test procedures are essential tools when handling data arising in biomedical research analyses. If we are interested in gene variants, expression patterns and biomarkers that are responsible for predicting certain diseases, methods that allow for a large number of hypotheses to be tested are of special importance. One example are genome-wide expression studies, in which we analyse the impact of gene expression profiles on the development of type-2-diabetes. The aim is to identify RNA transcripts which are differentially expressed. In general, the question arises whether the discrepancies between RNA transcripts of different groups which are identified as significant are actually significant or result from dependence of test statistics. Another issue that may cause problems are wrong assumptions concerning the underlying distributions. On the basis of microarray data from a large population-based study we present some approaches with which possible deviations from the model assumptions can be detected.

Expected Number of False Rejections of FDR- and FWER-Controlling Procedures Under Dependence

Marsel Scheer, Helmut Finner
Deutsches Diabetes-Zentrum an der Heinrich-Heine-Universität Düsseldorf, Leibniz-Zentrum für Diabetes-Forschung, Germany; mscheer@ddz.uni-duesseldorf.de

Keywords: Dependent test statistics, false discovery rate, familywise error rate, number of false rejections

The family wise error rate (FWER) and the false discovery rate (FDR) are very popular error measures for multiple hypothesis testing problems. A further interesting characteristic of multiple testing procedures is the expected number false rejections (ENFR). In this talk we illustrate that under dependence procedures controlling FWER or FDR may have an unpleasant inflation of ENFR.

References

[1] Y. Benjamini and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Stat., **29**:1165-1188.
[2] H. Finner, T. Dickhaus, and M. Roters (2007). Dependency and false discovery rate: Asymptotics. Ann. Stat., **35**:1432-1455.
[3] H. Finner and M. Roters (2001). Asymptotic sharpness of product-type inequalities for maxima of random variables with applications in multiple comparisons. J. Statist. Plann. Inference, **98**:127-144.
[4] H. Finner and M. Roters (2002). Multiple hypotheses testing and expected type I errors. Ann. Stat., **30**:220-238.
[5] H. Finner and M. Scheer (2011). Control of the expected number of false rejections in multiple hypotheses testing. Submitted.

Repeated significance tests for high-dimensional data

Sonja Zehetmayer[1], **Martin Posch**[2]
[1]Medical University of Vienna, Austria; [2]European Medicines Agency (EMA), London, U.K.;
sonja.zehetmayer@meduniwien.ac.at

Keywords: False discovery rate, sequential design, stopping rules

Repeated significance tests controlling the False Discovery Rate have been proposed that allow for an early stopping of the trial in studies involving large scale hypothesis testing as in microarray analysis (Posch, Zehetmayer, Bauer, 2009). To control the False Discovery Rate, multiplicity adjustment is required only for the number of hypotheses but not for the number of interim looks. This holds under suitable assumptions asymptotically, for an increasing number of hypotheses.

In this talk we investigate novel stopping rules that stop a trial early if a certain success criterion is fulfilled or a futility boundary is crossed. Using simulations we study the operating characteristics of the resulting sequential designs. Furthermore, we explore to which extend the results generalize to hypothesis tests controlling the Family Wise Error Rate.

References

M. Posch , S. Zehetmayer , and P. Bauer (2009). Hunting for Significance with the False Discovery Rate, Journal of the American Statistical Association **104**:832–840.

Bayesian Methods and Decision Theory (1)

K. Ickstadt

Tuesday 13. September, 11:00 – 12:30, Lecture Hall KOH-B10

Sensitivity analysis in Bayesian generalized linear mixed models for binary data

Malgorzata Roos, Leonhard Held
University of Zurich, Switzerland; mroos@ifspm.uzh.ch

Keywords: Bayesian Analysis, GLMMs, Hellinger distance, INLAs, sensitivity analysis

Generalized linear mixed models (GLMMs) enjoy increasing popularity because of their ability to model correlated observations. Integrated nested Laplace approximations (INLAs) provide a fast implementation of the Bayesian approach to GLMMs. However, sensitivity to prior assumptions on the random effects precision parameters is a potential problem. To quantify the sensitivity to prior assumptions, we develop a general sensitivity measure based on the Hellinger distance to assess sensitivity of the posterior distributions with respect to changes in the prior distributions for the precision parameters [1]. Although the proposed methodology holds in greater generality, we make use of the developed methods in the particular context of the well-known salamander mating data. We arrive at various new findings with respect to the sensitivity of the estimates of the model components.

References

[1] M. Roos and L. Held (2011). Sensitivity analysis in Bayesian generalized linear mixed models for binary data, Bayesian Analysis, forthcoming

Estimating the number of compartments with reversible jump MCMC

Julia C. Kärcher, Volker J. Schmid
Ludwig-Maximilians-Universität, Germany; julia.kaercher@stat.uni-muenchen.de

Keywords: Bayesian estimation, Bayesian model selection, compartment model, nonlinear regression, reversible jump MCMC

Compartmental systems describe how several so-called compartments exchange material over time. Compartmental models are for example used in biochemical experiments that aim to explore the binding behavior of proteins in living cells [2]. In such experiments, the binding state of a molecule is represented with the help of compartments: one compartment contains free (unbound) molecules and one or several compartments contain molecules bound to a binding partner. However, there may be different binding partners that bind molecules at different rates. The aim is to simultaneously estimate the parameters describing the model (binding rates and volumes) and to determine the number of compartments (and hence the number of binding partners).

To estimate the adequate model dimension from the observed concentration time curves, we treat the number of compartments as latent parameter. We choose a Bayesian framework and perform Markov chain Monte Carlo (MCMC) simulations in order to estimate the model parameters. By allowing for jumps between models of different dimensions ('reversible jumps') [1], the model dimension can be estimated simultaneously. We adapt the reversible jump MCMC algorithm such that it is suitable for the nonlinear regression problem arising in compartmental models.

The performance of the proposed procedure is evaluated for simulated data and for *in vivo* data from a biochemical experiment.

References

[1] P. J. Green (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, Biometrika **82**:711-732.
[2] B. L. Sprague and J. G. McNally (2005). Frap analysis of binding: proper and fitting, Trends in Cell Biology, **15**:84-91.

Functional uniform prior distributions for nonlinear regression

Björn Bornkamp
Novartis, Switzerland; bjoern.bornkamp@novartis.com

Keywords: Bayesian Statistics, nonlinear regression, prior distribution

In this talk I will consider the topic of finding prior distributions in nonlinear modelling situations, that is, when a major component of the statistical model depends on a non-linear function. Making use of a functional change of variables theorem, one can derive a distribution that is uniform in the space of functional shapes of the underlying nonlinear function and then back-transform to obtain a prior distribution for the original model parameters. The primary application considered in this article is non-linear regression in the context of clinical dose-finding trials. Here the so constructed priors have the advantage that they are parametrization invariant as opposed to uniform priors on parameter scale and can be calculated prior to data collection as opposed to the Jeffrey's prior. I will investigate the priors for a real data example and for calculation of Bayesian optimal designs, which require the prior distribution to be available before data collection has started (so that classical objective priors such as Jeffreys priors cannot be used).

On bayesian and generalized confidence intervals on the variance components in mixed linear models

Andrzej Michalski
Wroclaw University of Environmental and Life Sciences, Poland; apm.mich@gmail.com

Keywords: Linear models, variances components, confidence intrvals

The paper deals with construction of exact confidence intervals for the variance component s_1^2 and ratio q of variance components s_1^2 and s^2 in mixed linear models for the normal distributions $N_t(0 , s_1^2W + s^2I_t)$.

We give two classes of bayesian interval estimators depending on a prior distribution on (s_1^2 , s^2):

1) for a ratio q - by using test statistics from the decomposition of a quadratic form y'Ay for the Bayes locally best estimator of s_1^2 , [3],

2) for s_1^2 - constructed using Bayes point estimators from BIQUE class (Best Invariant Quadratic Unbiased Estimators, see [2], [3].

An idea of construction of confidence intervals using generalized p-values is also presented, [4] . Theoretical results for Bayes and for some generalized confidence intervals by simulations studies for selected experimental layouts are illustrated and compared, [1], [3].

References

[1] B. Arendackà (2005). Generalized confidence intervals on the variance component in mixed linear models with two variance components, Statistics, **39**:275-286.

[2] A. Michalski (2003). On some aspects of the optimal statistical inference of variance components in mixed linear models. Tatra Mountains Mathematical Publications, **26**:1-21.

[3] A. Michalski (2009). Bayesian and generalized confidence intervals on variance ratio and on the variance component in mixed linear models, Discussiones Mathematicae Probability and Statistics, **29**:5-29.

[4] K.W. Tsui, S. Weerahandi (1989). Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. J. Amer. Statist. Assoc., **84**:602-607.

Young Statisticians (1)

A. Berghold

Tuesday 13. September, 11:00 – 12:30, Lecture Hall KO2-F180

On generalized Hurwitz--Lerch Zeta distributions occuring in statistical inference

Dragana Jankov
J. J. Strossmayer University of Osijek, Croatia, Republic of; djankov@mathos.hr

Keywords: Riemann zeta function, Lerch zeta function, Hurwitz--Lerch Zeta function, hazard function, mean residual life function, characteristic function, Planck distribution, generalized Beta prime distribution, moment method parameter estimation

The object of the present talk is to define certain new incomplete generalized Hurwitz--Lerch Zeta functions and incomplete generalized Gamma functions. Further, we introduce two new statistical distributions named as generalized Hurwitz--Lerch Zeta Beta prime distribution and generalized Hurwitz--Lerch Zeta Gamma distribution and investigate their statistical functions, such as moments, distribution and survivor function, characteristic function, the hazard rate function and the mean residue life functions. Finally, Moment Method parameter estimators are given by means of a statistical sample of size *n*.

Estimating the risk of a Down's syndrome term pregnancy using age and serum markers. Comparison of various methods.

Kinga Sikolya, Sándor Baran, Lajos Veress
University of Debrecen, Hungary; sikolya.kinga@inf.unideb.hu

Keywords: Down's syndrome, biochemical markers, discriminant analisys, EM algorithm, logistic regression

The risk of an individual woman having a pregnancy associated with Down's syndrome is estimated given her age and her α-fetoprotein (AFP), human chorionic gonadotrophin (hCG) and gravidity specific β1-glicoprotein (SP1) levels. The classical estimation method is based on discriminant analysis under the assumption of log-normality of the marker values, but recently logistic regression is also applied for classification of the data. An improvement of the estimates of the parameters of the normal distribution is possible using EM algorithm.

In the present work we compare the performance of the two methods using a data set containing the data of almost 89.000 unaffected and 333 affected pregnancies from the eastern part of Hungary. Assuming log-normality of the marker values we also calculate the theoretical detection and false positive rates for both methods.

Hierarchical regression to adjust for multiple comparisons in a case-control study of occupational risk for lung cancer

Marine Corbin[1,2], Roel Vermeulen[3], Hans Kromhout[3], Susan Peters[3], Lorenzo Simonato[4], Lorenzo Richiardi[1], Franco Merletti[1], Neil Pearce[2,5], Milena Maule[1]

[1]Cancer Epidemiology Unit, Department of Biomedical Sciences and Human Oncology,University of Turin, Italy; [2]Centre for Public Health Research, Massey University, Wellington, New Zealand; [3]Institute for Risk Assessment Sciences, Utrecht, the Netherlands; [4]Department of Environmental Medicine and Public Health, University of Padua, Italy; [5]Department of Medical Statistics, Faculty of Epidemiology and Public Health, London School of Hygiene and Tropical Medicine, UK; marine_corbin@yahoo.fr

Keywords: Hierarchical regression, occupational study, multiple comparisons

Hierarchical regression can be used to improve on standard maximum-likelihood estimates (MLE) in the presence of multiple comparisons by adding a second-stage model and using a semi-Bayes approach to correct the estimates.

We applied hierarchical regression to a case-control study of occupational risks for lung cancer. In the first stage, we estimated the odds ratio for each occupation through logistic regression. The second-stage model was built using a Job-Exposure Matrix that classifies occupations in three categories of exposure (none, low, high) to different carcinogens (e.g. asbestos, silica, diesel). This yielded prior relative risk estimates for each occupation according to its levels of exposure to the various carcinogens. These were estimated by weighted least squares assuming a prior second-stage residual variance. Posterior (semi-Bayes) relative risk estimates were obtained by calculating a weighted average with the MLE.

We assessed the variation of posterior risk estimates with the a priori values chosen for the second-stage residual variance and compared them with standard semi-Bayes estimates (with a priori true standard deviation=0.5, implying that 95% of risks are within a 7-fold range of each other). Estimates obtained with hierarchical regression were more precise than MLE; they were also more plausible, since occupations with similar carcinogenic exposures had similar risks. Moreover, hierarchical regression shrinkage seemed to penalize less than standard semi-Bayes adjustment. This observation also held for narrower prior distributions, such as a residual standard deviation that assumes that true risks are within a 2.5-fold range of each other.

The Extended-Quasi-Likelihood-Function in Generalized Linear Models

Thorn Thaler
Nestlé Research Center, Switzerland; thorn.thaler@rdls.nestle.com

Keywords: Generalized linear model, extended-quasi-likelihood, variance function

The Generalized Linear Model (GLM) is an extension of the Linear Model. As opposed to a Linear Model, a GLM does not only allow for the usage of the normal distribution, but for all distributions from the exponential family. Estimators of the coefficients of a GLM are determined completely by the specification of the first two moments. Thus, a natural extension is to use a quasi-likelihood approach [1], where one does not specify the whole distribution, but the functional relationship between the mean μ and the variance, as expressed by the variance function var $(Y) \approx V(\mu)$. Nelder and Pregibon [2] introduce the Extended Quasi- Likelihood function (EQL function) which, inter alia, permits comparisons between different variance functions.

An R library which allows for the estimation of the parameters of a family of variance functions by means of the EQL function is now available.

References

[1] P. McCullagh (1983). Quasi-likelihood functions, The Annals of Statistics, **11**:59–67.
[2] J. A. Nelder and D. Pregibon (1987). An extended quasi-likelihood function. Biometrika, **74**:221–232.

State-of-the art solutions for class-imbalance problem. Why don't they work on high-dimensional class-imbalanced data?

Rok Blagus
Uni Ljubljana; rok.blagus@mf.uni-lj.si

Keywords: Prediction, classification

Class prediction is biased towards the majority class, when data are class-imbalanced. The bias is further increased when data are high-dimensional. In the previous years many solutions were proposed to solve the class-imbalance problem. The proposed methods are variants of simple downsizing, i.e. only a subset of samples from the majority class is chosen and included in the derivation of the classification rule, or over-sampling, i.e. minority samples are resampled in order to obtain a class-balanced distribution. The most promising methods from each approach are EasyEnsemble and SMOTE (Synthetic Minority Oversampling Technique). Both methods were shown to have a significant improvement in performance of classification algorithms when used on ordinary data. In this paper we use theoretical results as well as simulated and real high-dimensional data to show that the two methods are unable to solve the class-imbalance problem for high-dimensional data. We propose also a variant of EasyEnsemble algorithm that is useful for high-dimensional data.

Time to Event Analysis (1)

J. Beyersmann

Tuesday 13. September, 11:00 – 12:30, Lecture Hall KOL-F101

Quanitifying the correlation of paired survival times under censoring

Michael Schemper, Alexandra Kaider, Samo Wakounig
Medical University of Vienna, Austria; michael.schemper@meduniwien.ac.at

Keywords: Bivariate survival, copulas, twin studies

The analysis of correlations within pairs of survival times is of interest to many research topics in medicine, such as the correlation of survival times of twins, or of times till failure in paired organs. The dependence of such times is assumed monotonic and thus quantification by Spearman or Kendall correlation coefficients appropriate. The typical censoring of such times requires more involved methods of estimation and inference as have been developed in recent years. Here we focus on semiparametric approaches, commonly based on maximum likelihood estimation of the parameter of bivariate copula distributions, which results in estimated values of Spearman and Kendall correlation coefficients.

We have developed a novel semiparametric approach, by which the survival probabilities at death or censoring times are transformed to normal deviates. Then, normal deviates which relate to censored times are iteratively augmented, by conditional multiple imputation, until self-consistency (convergence) is obtained. For the finally augmented samples, values of Spearman and Kendall correlation coefficients are obtained.

In a Monte Carlo study of paired survival times generated under various copulas, and for censoring proportions up to 90% and underlying Spearman correlation up to 0.9, the performance of our approach compares favorably with that of the other semiparametric ones. Furthermore, we address the issue to which extent the normal copula provides reliable estimates of correlation, independent of the underlying bivariate survival distributions.

Conclusions from the Monte Carlo study are complemented by estimates of correlation of survival in the Danish Twin study, according to the approaches considered.

Copulas and frailty models for clustered survival data

Andreas Wienke
University Halle, Germany; andreas.wienke@medizin.uni-halle.de

Keywords: Survival analysis, frailty models, copulas, clustered survival data

Frailty models and copulas are useful tools to model correlated survival data. The equivalence between Archimedean copulas and shared frailty models, for example between the Clayton-Oakes copula and the shared gamma frailty model, has often been claimed. This equivalence, however, is not valid because the two approaches model the marginal survival functions in different ways, thereby leading to entirely different joint survival functions ([1], [2]). It will be shown that a similar relationship holds for the correlated gamma frailty model [3] and [4] and its related copula. We focus on the case of bivariate survival data and illustrate the similarities and differences between the two approaches based on the bivariate parametric correlated gamma frailty model applied to Danish twin mortality data. Furthermore, the question whether parameters from copulas can be interpreted as parameters from a frailty model with different baseline hazard function is considered in the parametric case.

References

[1] L. Duchateau, P. Janssen (2008). The Frailty Model. Springer, New York

[2] K. Goethals, P. Janssen, L. Duchateau (2008). Frailty models and copulas: similarities and differences. Journal of Applied Statistics, **35**:1071-1079.

[3] A. Wienke (2010). Frailty Models in Survival Analysis. Chapman & Hall/CRC, Boca Raton

[4] A. I. Yashin, J. W. Vaupel, I. A.Iachine (1995). Correlated individual frailty: An advantageous approach to survival analysis of bivariate data. Mathematical Population Studies, **5**:145-159

Software for the Analysis of shared gamma and lognormal Frailty Models in SAS and R

Katharina Hirsch, Andreas Wienke
Martin-Luther Universität Halle Wittenberg, Germany; katharina.hirsch@medizin.uni-halle.de

Keywords: Frailty models, Survival analysis, software, mixed models

In survival analysis individuals are followed over some period and the time until the transition from an initial to a final state is of particular interest. An important tool for analyzing potential risk factors of this transition is the Cox proportional hazards model. This model requires homogeneity in the study population and independence between the observations. An extension of the Cox model to deal with both, unobserved heterogeneity and clustered survival data, are frailty models. Different software tools are available for the analysis of shared frailty models such as coxph, coxme, phmm, frailtyPenal, SPGAM or SPLN3. These functions and macros are available in R and SAS, but there are also procedures in STATA and WinBugs. The user wants to know: Which software obtains the best estimation for shared frailty models? To answer this we conducted a large simulation study. We show how cluster sizes, censoring rates, and covariable distributions affect the estimation of the software. Advantages and limitations of the software are discussed in detail.

Sample size calculation in cluster randomized trials with a time to event outcome

Antje Jahn-Eimermacher, Astrid Schneider
University Medical Center Mainz, Germany; antje.jahn@unimedizin-mainz.de

Keywords: Sample size; time to event; cluster randomized

Introduction:

In cluster randomized trials treatments or interventions are randomly allocated to groups of subjects (clusters). Examples are the evaluation of health interventions which have to be implemented on a city, hospital or family level. A correlation of subjects within the same cluster can substantially decrease the power of a cluster randomized trial. Thus, sample size must be calculated by taking into account the within-cluster correlation and appropriate methods have been proposed for the comparison of means, proportions and rates [1]. We will present methods for sample size calculation if the outcome of interest is a time to event.

Methods:

We consider the Cox gamma frailty model for analyzing cluster randomized trial data. Sample size calculation is based on time to event data being exponentially distributed conditional on a frailty term shared by individuals of the same cluster. Subjects are assumed to be recruited over an accrual time and after accrual there will be an additional follow up period.

Results:

We will present the sample size formula and the resulting design effect, defined as the ratio between sample size calculated by taking and not taking into account a within-cluster correlation. We will show, how the design effect depends on trial design parameters, as the degree of within-cluster correlation, the degree of censoring, the hazard ratio and the cluster size.

References:

[1] R.J. Hayes and S. Bennett (1999). Simple sample size calculation for cluster-randomized trials, International Journal of Epidemiology, **28**:319–326.

Challenges in Agricultural Science (2)

J. Spilke

Tuesday 13. September, 11:00 – 12:30, Lecture Hall KOL-F104

Methods and Models to analyze ordered categorical data with spatial covariance - a simulation based comparison

Karen Höltl, Katrin Thamm, Norbert Mielenz, Joachim Spilke
Martin-Luther-University Halle-Wittenberg, Germany; karen.hoeltl@landw.uni-halle.de

Keywords: Ordered categorical data, spatial covariance, threshold model

The analysis of ordered categorical data plays an important role in agricultural science. Just as for continuous data, there is the question, if and how spatial covariance should be included in the analyses. Based on simulation we compared different methods to analyze categorized ordinal data and their ability to account for spatial covariance.

The evaluated methods were the:

- threshold model approach [1],
- rank method [2],
- permutation method [3],
- linear mixed model approach [4].

These methods have different abilities to include spatial covariance in the model structure. We investigated the evaluation models by the realized type one error of the t-test for a treatment contrast. The threshold model showed convergence problems. Nevertheless, if the runs had convergence, the threshold model showed the best results. The linear mixed model revealed satisfactory results as well, but solely in case of equal fixed effects (homogeneity of variances). Additionally, for practical interpretations, only the threshold model is able to provide the probabilities of the ordered categories.

References

[1] P., McCullagh (1980). Regression-Models for Ordinal Data, J. Royal Stat. Soc., Series B-Methodological, **42**:109-142.

[2] E., Brunner, and M.L. Puri (2001). Nonparametric methods in factorial designs, Stat. Papers, **42**:1-52.

[3] E., Schumacher, and M. Frisch (1995). Permutation tests to analyse ordered categorical data in simple field trial designs (in German), ZAI, **5**:107-113.

[4] C.R., Henderson (1990). Statistical methods in animal improvement: historical overview, Advances in Statistical Methods for Genetic Improvement of Livestock, Springer-Verlag, New York: 1-14.

Analysis of count data with repeated measurements – tested by simulation studies

Katrin Thamm, Karen Höltl, Norbert Mielenz, Joachim Spilke
Institute of Agricultural and Nutritional Sciences, Martin-Luther-University Halle-Wittenberg, Germany; Katrin.Thamm@landw.uni-halle.de

Keywords: Count data, generalized linear models, repeated measurements, simulation

In order to analyze count data, generalized linear models are used with increasing frequency. If there are repeated measurements per object then marginal (population-averaged) or subject-specific models are available. Our research is based on data from a trial where the thrips infestation in 20 wheat cultivars was evaluated. There were three measurements per object because a wheat ear was divided into three segments. The experimental design and the estimated model parameters were used for simulation of count data with repeated measurements. Thus, the evaluation model contains fixed cultivars and segments effects as well as random ear effects. The estimation methods were: generalized estimating equations with specification of a working correlation matrix; maximum likelihood estimation with adaptive Gauss-Hermite quadrature (ML_QUAD); maximum likelihood estimation based on Laplace-Approximation; and pseudo-likelihood estimation based on linearization of data. The models and methods were compared by convergence properties, percental Bias, mean squared error (MSE), and realizing the nominal type one error of the statistical hypothesis testing. The evaluation was carried out in SAS using the procedures GENMOD and GLIMMIX. The estimated parameters and the hypothesis testing (based on 10.000 simulations) occurred within the response scale. The estimation methods showed different convergence properties depending on the predetermined parameters. The ML_QUAD method had unbiased estimators for the fixed parameters. In comparison, the ML_Quad method showed the best convergence property, similarly MSE as well as the best realized nominal type one error.

Bayesian analysis with plant breeding data

Jens Möhring, Hans-Peter Piepho
Universität Hohenheim, Germany; moehring@uni-hohenheim.de

Keywords: Bayes, mixed model

While Bayesian analysis is widely used in animal breeding, analysis of microarrays and other fields related to agricultural sciences, it is rarely used in cultivar registration or plant breeding field trials. [1] introducted Bayesian statistics in agriculture field trials, [2] and [3] both demostrated the use of Bayesian analysis for analysing multi-environmental trials (METs). We want to investigate the use of Bayesian analysis for two special cases of field trials in plant breeding. Firstly, we want to predict heterogeneous intra-cross-variances, when the number of genotypes per cross is small. Secondly, we want to estimate genotype effects for a triannual serie, thus in a serie with limited information about year and year*genotype effects. In both cases the data themselves provide limited information for estimating precise variance components, therefore we expected the highest advantage of using Bayesian analysis. We incorporate additional information, either from other crosses, or from previous comparable series within the Bayesian analysis. Two real datasets were used to explore the ideas. Simulations were used to investigate the potential advantage in precision and computing time compared to a mixed model analysis.

References

[1] J. Besag and D. Higton (1999). Bayesian analysis of agricultural field experiments, Journal of Royal Statistical Society B **61**:691-746.

[2] C. M. Theobald, M. Talbot and F. Nabugoomu. 1992. A bayesian approach to regional and local-area prediction from crop variety trials, Journal of Agricultural, Biological, and Environmental Statistics **7**:403-419.

[3] J.W. Edwards and J.-L. Jannink (2006). Bayesian modeling of heterogeneous error and genotype by environment interaction variances, Crop Science **46**:820-833.

Sparse covariance matrices in random models for quantitative trait locus discovery in F2 populations

Norbert Reinsch, Daisy Zimmer
Leibniz-Institut für Nutztierbiologie, Germany; reinsch@fbn-dummerstorf.de

Keywords: QTL-mapping, mixed models, covariance matrices

In F2 families derived from inbred lines marker-based relationship matrices for random genetic effects (additive genetic, dominance and pairwise interactions) can be derived from simple elementary matrices, describing the covariance of those effects for known QTL genotypes. A novel kind of such elementary covariances for additive and additive-by-additive genetic effects is proposed, which reflects expected perfect correlations between genetic effects of certain genotypes and, depending on the kind of effect considered, leads to a more sparse representation of genetic covariance. It is shown theoretically and by simulated examples that these covariance matrices lead to identical restricted log-likelihood values and, hence, provide the same parameter estimates as previously reported versions, while nominal standard errors of estimated genetic effects are improved and more realistic. Computational speed is considerably enhanced, when genetic effects are estimated for each genotyped F2 individual. Moreover, when marker positions and QTL locations coincide, this kind of additive genetic covariance matrix is equivalent to models with random regression coefficients.

Multiple Testing (2)

S. Zehetmayer

Tuesday 13. September, 11:00 – 12:30, Lecture Hall KOL-F121

Calculation of Simultaneous Confidence Intervals by Constraint Propagation

Georg Gutjahr[1], Frank Bretz[2]

[1]University of Bremen, Germany; [2]Novartis Pharma AG, Switzerland; georg.gutjahr@math.uni-bremen.de

Keywords: Interval arithmetic, constraint propagation, multiple inference, computational statistics

In this talk, we describe an efficient algorithm to calculate confidence intervals of minimal length that are consistent with a given single-step test procedure. We assume that the unadjusted p-value functions of the individual tests can be extended to inclusion functions in interval-arithmetic and we briefly review the methods that are available to obtain such inclusion functions. We then express the confidence intervals as solution of a system of interval constraints and we use constraint propagation to solve this system and thereby obtain the simultaneous confidence intervals. As example, we discuss the calculation of simultaneous confidence intervals for ratios of means of normal distributions.

Multiple comparisons to both a negative and a positive control

Mario Hasler
Christian-Albrechts-Universitaet zu Kiel, Germany; hasler@email.uni-kiel.de

Keywords: simultaneous testing, three-arm trial, non-inferiority, Fieller's confidence interval, optimal allocation

This presentation addresses multiple comparisons in the presence of both a negative and a positive control. The methodology of the three-arm trial is extended to the case of many experimental treatment arms or different doses of a compound. In contrast to the classic three-arm trial, the focus is on the family-wise error type I. Normally distributed data with either homogeneous or heterogeneous group variances are considered. Explicit criteria for an optimal allocation are proposed. Depending on the pattern of heterogeneity, remarkably unbalanced designs are power-optimal. As an example, the method will be applied to a toxicological experiment.

References

[1] I. Pigeot, J. Schäfer, J. Röhmel, D. Hauschke (2003). Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. SiM, **22**:883–899.

[2] M. Hasler, R. Vonk, L.A.Hothorn (2008). Assessing non-inferiority of a new treatment in a three-arm trial in the presence of heteroscedasticity. SiM, **27**:490–503.

[3] M. Hasler, L.A. Hothorn (2008). Multiple contrast tests in the presence of heteroscedasticity. BMJ, **50**:793–800.

Test Procedures for the Assessment of the Components of Composite Endpoints

Geraldine Rauch, Meinhard Kieser
Institute of Medical Biometry and Informatics, University of Heidelberg, Germany;
rauch@imbi.uni-heidelberg.de

Keywords: Clinical trials, composite endpoints, multiple testing

Composite endpoints are increasingly used in clinical trials, particularly in the field of cardiology. Thereby, the overall impact of the therapeutic intervention is captured by including several events of interest in a single variable. In the ICH E9 Guideline [1], it is stated that 'this approach addresses the multiplicity problem without requiring adjustment to the type I error'. In fact, to demonstrate the significance of an overall clinical benefit, it is sufficient to assess the test problem formulated for the composite. However, even if a statistically significant and clinically relevant superiority is shown for the composite endpoint, there is the need to evaluate the treatment effects for the constituting components. For example, the Points to Consider on Multiplicity [2] require that "… if all cause mortality is a component, a separate analysis of all cause mortality should be provided to ensure that there is no adverse effect on this endpoint."

We propose multiple test procedures that enable decisions about the components of a composite endpoint under control of the type I error rate. The properties of these approaches in terms of required sample size and power are compared. It is shown how the issue of follow-up decisions about the components can be addressed in the planning stage. Application is illustrated by a clinical trial example.

References

[1] ICH E9. Statistical principles for clinical trials (1999)
http://www.emea.europa.eu/pdfs/human/ich/036396en.pdf (assessed 17.05.201).

[2] CPMP. Points to consider on multiplicity issues in clinical trials (2002)
http://www.emea.europa.eu/pdfs/human/ewp/090899en.pdf (assessed 17.05.2009).

Bayesian Methods and Decision Theory (2)

K. Ickstadt

Tuesday 13. September, 14:00 – 15:30, Lecture Hall KOH-B10

Incorporating utilities in Bayesian models for risk-benefit decision-making

Deborah Ashby
Imperial College London, United Kingdom; deborah.ashby@imperial.ac.uk

Keywords: Risk-benefit, decision-making, Bayes, drug regulation, utilities

There is now growing interest in the possibilities of more formal approaches to risk-benefit decision-making for drug regulation. Bayesian approaches, incorporating a decision-framework, evidence synthesis and utilities, provide a principled framework. In this talk we outline possible approaches, and demonstrate them through case-studies on regulated medicines where decision-making has been challenging. We discuss the potential for such Bayesian approaches to be used more routinely.

Benefit-Risk of Multiple Sclerosis Treatments: Lessons Learnt in Multi-Criteria Decision Analysis

Richard Nixon[1], Pedro Oliveira[2]
[1]Novartis, Switzerland; [2]RTI Health Solutions; richard.nixon@novartis.com

Keywords: Benefit-risk, decision analysis

Benefit-risk analysis comprises an approach and a set of tools for qualitative structuring and quantitative analysis of the favorable and unfavorable outcomes of a decision. Multi-Criteria Decision Analysis (MCDA) is a method used to assess benefit-risk in health care. We apply MCDA to assess the relative benefit-risk of four treatments for relapsing remitting multiple sclerosis (RRMS). Salient adverse events, clinical benefits and convenience measures are identified, and the magnitude of each of these criteria resulting from each treatment is found. The relative utilities of these magnitudes are assessed from experts and combined into an overall benefit-risk score. Sensitivity analysis is performed to identify the criteria with the greatest effects on the benefit-risk score. We reflect on lessons learnt during this exercise and the suitability of MCDA for performing benefit-risk analysis in health care. We conclude that it is a method well suited for this task, and we make suggestions for its use in practice.

Use of historical data to support the planning, analyzing and decision making of Proof of concept studies: an example in Irritable Bowel Syndrome

Rossella Belleli
Novartis AG, Switzerland; rossella.belleli@novartis.com

Keywords: Bayesian, meta-analysis, historical data, decision, proof of concept

Proof of concept study are performed in an exploratory setting where the focus is learning about the efficacy and safety of the new drug and taking a decision regarding its further development with the competitive landscape as background. It is therefore very important to take into account the available knowledge about the effect of the controls (placebo and/or competitors) for the design and the analysis of the new trial. This can be achieved by performing a Bayesian random effect meta-analysis of the relevant studies available in the public domain to predict the effect of the controls in a new trial [1]. The predicted effects and their distribution are then used as prior information in the new trial. Additionally, the prior information is used as basis to choose the decision criteria for the new trial and to optimize their operating characteristics by controlling the false positive and false negative decision rates with respect to some true scenarios. An example of using historical data for a new proof of concept trial in Irritable Bowel Syndrome is described. In this context the amount of historical control information available in the public domain is extensive and at the effect of the controls is very variable. As a consequence making the most effective use of historical data is essential and challenging at the same time.

References

[1] B. Neuenschwander (2010). Summarizing historical information on controls in clinical trials. Clinical Trials **7**:5-18.

Integration of Copy Number Variation and Gene Expression Data in Bayesian Regression Models for Prediction and Biomarker Selection

Manuela Zucknick[1], Stefan Pfister[1,2], Axel Benner[1]
[1]German Cancer Research Center, Germany; [2]University Hospital Heidelberg, Germany;
m.zucknick@dkfz-heidelberg.de

Keywords: High-dimensionality, data integration, prognostic model

Bayesian variable selection (BVS) models are an alternative to well-known sparse regularisation methods like lasso regression for prognostic modelling based on high-dimensional input spaces. A common application is the prediction of clinical endpoints like therapy response with simultaneous biomarker selection using microarray gene expression data.

High-throughput microarray technologies are also available for many other types of genomic data, and in recent years clinical researchers have begun to systematically collect genome-wide data from various sources on the DNA- and RNA-level. If data from several sources are available for the same set of biological samples, they can be analysed together in an integrative manner, with the aim of providing a more comprehensive picture of the disease biology and improving the performance of clinical prediction models.

For example, the integration of copy number variation data into gene-expression-based prognostic models promises to improve both prognostic value and interpretability of the model, because genomic aberrations are known to affect expression levels of corresponding genes. In fact, the deletion of chromosomal regions harbouring important tumour suppressor genes is a well-known cause of certain cancers.

BVS models are very flexible in their setup and are naturally well-suited to extensions allowing the integration of additional data sources. We will propose a BVS model, which combines copy number variation and gene expression data in a biologically intuitive manner. The model setup will be demonstrated, as well as aspects of the MCMC sampling algorithm and posterior inference. The model will be illustrated in an application to pediatric brain tumour data.

Clinical Trial Application

A. Berghold and J. Roehmel

Tuesday 13. September, 14:00 – 15:30, Lecture Hall KO2-F180

Multiplicity in Confirmatory Clinical Trials: New aspects and emerging principles

Norbert Benda
bfarm, Germany; Norbert.Benda@bfarm.de

Keywords: Multiplicity in confirmatory clinical trials

The CHMP points to consider on multiplicity issues in clinical trials came into operation in 2002. Since then, it has been proven to be useful for both, industry and regulators when planning and assessing confirmatory clinical trials. Meanwhile, however, methodological advances have been made in more complex multiplicity settings resulting in more powerful and flexible procedures in situations where a couple of hypotheses are to be tested, eventually with different degrees of prominence attributed to the individual questions. In line with the development of these methods an increasing complexity in the hypothesis framework emerged. On one hand, the applicant wishes to answer a couple of question within a single trial that potentially help to improve and accelerate clinical development. On the other hand it may be in the regulator's interest to ensure statistical evidence in different claims to be made with a clinical trial.

This increasing complexity which is due to different sources, as dosages, subgroups, interim looks, additional claims, and, notably, combinations of these results in new issues and poses questions on general principles that haven't been considered before, as consistency problems or the construction of simultaneous confidence intervals. Finally, the metrics used for the validity assessment of a testing procedure (the strong control of familywise type 1 error) may be questioned in specific situations, as the treatment of additional claims, safety analyses or biomarker selection.

The presentation will give a regulatory view on these questions developing and discussing principles that may help to deal with the corresponding issues.

A new approach for the simultaneous assessment of statistical significance and clinical relevance

Meinhard Kieser
Institute of Medical Biometry and Informatics, University of Heidelberg, Germany;
meinhard.kieser@imbi.uni-heidelberg.de

Keywords: Statistical significance, clinical relevance, responder analysis, effect size, relative effect, probabilistic index, sample size

In drug development it is well accepted that a successful study will not only demonstrate a statistically significant result, but also a clinically relevant effect size. Whereas standard hypothesis tests are used to demonstrate the former, it is less clear how the latter should be established. In this presentation, we first give a review of regulatory requirements with respect to the assessment of clinical relavance and present currently available methods for planning and analysis. Especially, the responder analysis approach is considered which is recommended in numerous guidelines and widely applied in practice. A new approach for the assessment of clinical relevance is proposed which is based on the so-called relative effect (or probability index). The underlying metrics is defined and motivated, and advantages of the new approach as compared to existing procedures are outlined. A method for sample size calculation for the simultaneous assessment of statistical significance and clinical relevance based on the proposed approach is given. The results are compared to those obtained when proving clinical relevance by alternative methods. Application of the new approach is illustrated by a clinical trial example.

Is there a danger of "bio-creep" with non-inferiority trials?

Primrose Beryl, Werner Vach
Clinical epidemiology Group, Department of Medical Biometry and Statistics,
Universitätsklinikum Freiburg, Germany; beryl@imbi.uni-freiburg.de

Keywords: Non-inferiority trial, pre-trial distribution, effect estimate, bio-creep

Bio-creep in Non-Inferiority(NI) trials may lead to degradation of efficacy of investigational treatment over time. Statistically speaking, bio-creep denotes the true average post-study effect being lesser than zero. This post-study effect is mainly dependent on choice of non-inferiority margin and distribution of pre-study probability which is unknown. Obligatory registration of clinical trials has resulted in availability of unbiased data on all conducted clinical trials. Our aim was to determine the pre- and post-trial distribution of true effect.

All NI trials registered in the NIH's Clinical trials register (www.ClinicalTrials.gov) carried out during 2000-2007 with non-inferiority of efficacy as primary objective were studied. Primary results from these trials were searched for in websites of NLM, IFPMA and PUBMED. Simulations and meta-analysis of the effect estimates were performed using STATA 11.

Of 114 registered NI trials, 84 met the inclusion and exclusion criteria. Of 72 completed studies, final reports were available for 69 trials which showed that the investigational treatment was superior in 7(14%), non-inferior in 55(79%) and inferior in 7(9%). The effect estimate was positive among 39(59%) of the 66 studies providing an effect estimate. We intend to derive the distribution of true effect estimates of NI trials and present this at the conference.

A high likelihood of retrieving results from registered clinical trials made it possible to calculate the pre-study distribution of true effect. The unanticipated finding of a positive average effect estimate suggests that a decline in standard treatment effect (bio-creep) is not imminent, at least on average.

QTc Analysis Employing Continuous 24 h Holter ECGs

Georg Ferber[1], Henry Holzgrefe[2]
[1]Statistik Georg Ferber GmbH, Switzerland; [2]Charles River, Reno, Nevada;
Statistik.Georg.Ferber@bluewin.ch

Keywords: Thorough QT study, time series, extreme value

The proarrhythmic potential of new drugs must be assessed with a Thorough QT study (ICH E14) where the ECG is usually measured at about 10 predefined timepoints to characterize the potential for QTc prolongation. For a negative study, the upper bound of the 95% one-sided CI for the largest time-matched QTc effect must exclude 10 ms. In practice, while digital ECGs are usually collected continuously, only short predefined segments are submitted for analysis such that <1 % of the data are evaluated. With the availability of automated ECG analysis, exclusion of 99% of the data becomes problematic. Instead of basing the analysis on a small number of selected complexes, we applied automated analysis to the continuous ECG dataset. The QTc was then expressed as the mean value for consecutive 10 min segments employing all beats that satisfied predefined quality criteria. A 10 min segment length is short enough to be commensurate with pharmacodynamic changes and long enough to allow meaningful timepoint comparisons between treatment arms.. Of note, the standard deviation for this new parameter was comparable to, or less than, that obtained with conventional analysis. The resulting longitudinal data (144 timepoints/ 24h per treatment arm) can be analysed employing a conventional intersection-union paradigm. This method is less problematic than might be anticipated given the length of the timepoint series. Alternatively, a smaller set of descriptive parameters could be derived. Importantly, bootstrapping is not considered appropriate to derive CIs for a maximum time-matched difference. Novel alternatives will be discussed.

Time to Event Analysis (2)

H. Binder

Tuesday 13. September, 14:00 – 15:30, Lecture Hall KOL-F101

Cox-based structural equation modelling with latent variables in cardiovascular trials

Lena Herich, Christine zu Eulenburg, Karl Wegscheider
Universität Hamburg, Germany; l.herich@uke.de

Keywords: Survival analysis; Cox's proportional hazards model; latent variables

In situations with complex covariable structure, structural equation models are a useful tool to display associations. Multicollinear or similar parameters can furthermore get bundled into latent variables. These extensions are also possible for survival analysis problems [1], but are uncommonly used, especially in medical research [2,3].

The present research demonstrates the application of structural equation modelling based on Cox's proportional hazards assumption for the analysis of data from different cardiological trials. It is plausible to assume the existence of latent variables measuring e.g. heart insufficiency in general, heart rate variability in particular or CAD intensity, each reflected in several signs and symptoms to be observed at baseline or during the course of a trial. While modelling observable variables only by standard Cox models, clinicians usually interpret the resulting effects as if they represent the underlying latent phenomena. Based on examples the talk will discuss the potential advantages and disadvantages of formal modelling of the latent variables as compared to a standard Cox approach.

References

[1] T. Asparouhov, K. Masyn, B. Muthen (2006). Continuous time survival in latent variable models. Proceedings of the Joint Statistical Meeting in Seattle, ASA section on Biometrics, 180-187.

[2] K. Larsen (2005). The cox proportional hazards model with a continuous latent variable measured by multiple binary indicators. Biometrics, **61**:1049-1055.

[3] T. Asparouhov, M. Boye, M. Hackshaw, A. Naegeli and B. Muthén (2009). Applications of continuous-time survival in latent variable models for the analysis of oncology randomized clinical trial data using Mplus. Technical Report.

On time-varying effects in high-dimensional survival studies

Anika Buchholz[1,2], Willi Sauerbrei[1], Harald Binder[1,2]
[1]Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Germany; [2]Freiburg Center for Data Analysis and Modeling, University of Freiburg, Germany; ab@imbi.uni-freiburg.de

Keywords: FPT, high-dimensional survival studies, time-varying effects

Microarray survival studies often focus on identifying a set of genes with significant influence on a time-to-event outcome for building a gene expression signature. Most of these signatures are assumed to have a constant effect over time, as they are usually derived using the Cox proportional hazards (PH) model. However, in microarray survival studies there might often be time-varying effects, i.e. violation of the PH assumption, for some of the genes [1]. Ignoring the presence of time-varying effects of some genes may lead to false conclusions about their influence. Hence, it is important to check for time-varying effects.

We will investigate possibilities of identifying, selecting and modelling time-varying effects in microarray survival studies. In the Cox model, the presence of time-varying effects can be explored by plotting the scaled Schoenfeld residuals against time [2]. To model such effects, a variety of approaches have been proposed in low dimensional data. Using publicly available data from cancer patients, we will illustrate how the FPT approach [3] can be used to investigate for time-varying effects and discuss potential problems such as poor power due to undersized studies.

References:

[1] D. Dunkler, M. Schemper and G. Heinze. Gene selection in microarray survival studies under possibly non-proportional hazards. Bioinformatics, **26**:784-790, 2010.
[2] T.M. Therneau and P.M. Grambsch (2000). Modeling Survival Data: Extending the Cox Model. Springer-Verlag.
[3] W. Sauerbrei, P. Royston and M. Look (2007). A new proposal for multivariable modelling of time-varying effects in survival data based on fractional polynomial time-transformation. Biometrical Journal, **49**:453-473.

Nonparametric hazard rate estimation for relative survival models

Sophie Frantal[1], **Werner Brannath**[2]
[1]Medical University of Vienna, Austria; [2]University of Bremen, Germany;
sophie.frantal@meduniwien.ac.a

Keywords: Relative survival models, hazard rate, kernel smoothing

Nonparametric estimation of hazard functions over time under random censorship is quite common, e.g., Nelson-Aalen estimates. Mueller and Wang [1] introduce methods for local data-adaptive choice of the bandwidth for Nelson-Aalen hazard estimates and proposed a new class of boundary kernels to deal with boundary effects. Boundary effects can cause misleading artefacts in the curve due to either a small number of events and/or patients at risks in a certain time interval. These effects are pronounced at boundaries where parts of ´naive' kernels are empty per definition.

As it can be difficult to interpret survival and hazard rate for a specific population on its own, relative survival gains more and more popularity. Relative survival models account for the information of the background mortality and measure the excess in the hazard of the observed data over this background mortality in a "reference" population. We apply methods of smoothing hazard rates [1] to such relative survival models [2]. The proposed relative hazard estimates use kernel methods with appropriate boundary corrections and data-dependent bandwidth allowing also left censoring. The methods are illustrated on a real data set from patients undergoing a cardiac surgery [3] and R-functions will be presented.

References

[1] Mueller and Wang (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. Biometrics, **50**:61-76.

[2] Frantal. Nonparametric hazard rate estimation for relative survival models. Diploma thesis.

[3] Lassnigg, Hiesmayr, Frantal, Brannath, Mouhieddine, Isetta, and Schmidlin. Cardiac surgical mortality, a population based study. Unpublished paper, 2011.

Modified Aalen-Johansen Estimator of the Cumulative Incidence Function for Left-Truncated Competing Risks Data

Arthur Allignol[1,2], Martin Schumacher[2], Reinhard Meister[3], Jan Beyersmann[1,2]
[1]Freiburg Center for Data Analysis and Modeling, University of Freiburg, Germany; [2]Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Germany; [3]Department of, Mathematics, University of Applied Sciences, Technische Fachhochschule Berlin, Germany; arthur.allignol@fdm.uni-freiburg.de

Keywords: Delayed entry, pregnancy, Nelson-Aalen estimator, small sample

This work is motivated by an observational study addressing the safety of exposure to statins during pregnancy, including the risk of spontaneous abortion. As women enter the cohort several weeks after conception, observations are left-truncated. Apart from spontaneous abortion, a pregnancy may end in an induced abortion or a life birth. The cumulative incidence function (CIF) is the method of choice for estimating the probability of spontaneous abortion in this setting [1]. However, as abortions happen early in pregnancy, estimation might become problematic due to very small risk sets at the beginning of the follow-up, leading to highly variable and thus potentially meaningless estimates. Estimating the CIF conditional on being event-free at some later time point is one suggested ad hoc method for obtaining reasonable estimates. Alternatively, Lai and Ying [2] proposed a modified version of the Kaplan-Meier estimator in the usual survival setting that tackle these issues and where consistency and weak convergence of the estimator on the entire range of the distribution function has been shown. We extend the Lai and Ying estimator to the competing risks setting, study finite-sample properties of the modified estimator in a simulation study and illustrate its usefulness in the statin study.

References

[1] R. Meister and C. Schaefer (2008). Statistical methods for estimating the probability of spontaneous abortion in observational studies Analyzing pregnancies exposed to coumarin derivatives, Reproductive Toxicology, **26**:31-35.
[2] Lai, TL and Ying, Z (1991). Estimating a distribution function with truncated and censored data. The Annals of Statistics, **19**:417-442.

Diagnostics 1: ROC-Analyses - Sensitivity and Specificity

C. Schwenke

Tuesday 13. September, 14:00 – 15:30, Lecture Hall KOL-F104

A smooth ROC curve estimator based on log-concave density estimates

Kaspar Rufibach
University of Zurich, Switzerland; kaspar.rufibach@ifspm.uzh.ch

Keywords: Diagnostic test, receiver operating characteristic curve, continuous-scale test, log-concave density estimation, binormal model

The receiver operating characteristic (ROC) curve is a common way of assessing the diagnostic accuracy of a diagnostic test with continuous outcome that predicts presence or absence of a binary trait, typically a disease. We introduce a new smooth estimator of the receiver operating characteristics (ROC) curve based on log-concave density estimates of the constituent distributions. We show that our estimate is asymptotically equivalent to the empirical ROC curve and that our proposed estimator exhibits an efficiency gain for finite sample sizes with respect to the standard empirical estimate in various scenarios. In addition, we illustrate that our proposed estimator is only slightly less efficient compared to the fully parametric binormal estimate in case the underlying distributions are in fact normal. These results translate to estimation of the area under the ROC curve. Computation of our proposed estimate uses the R package logcondens that implements univariate log-concave density estimation and can be done very efficiently using only one line of code. These obtained results lead us to advocate in [1] our estimate for a wide range of scenarios.

References

[1] K. Rufibach (2011). A smooth ROC curve estimator based on log-concave density estimates. Preprint, http://arxiv.org/abs/1103.1787.

Three principles for the joint evaluation of sensitivtiy and specificity in analysing a diagnostic study

Werner Vach[1], Poul Flemming Hoilund-Carlsen[2], Oke Gerke[2]
[1]Clinical Epidemiology, University Medical Center Freiburg, Germany; [2]Dept. of Nuclear Medicine, Odense University Hospital, Denmark; wv@imbi.uni-freiburg.de

Keywords: Sensitivity, specificity, sample size, joint confidence region

It is common knowledge today that sensitivity and specificity need to be evaluated together in the analysis of the accuracy of a diagnostic procedure. However, there are different ways how to perform such an evaluation.

We look at the problem as a special case of a study with two endpoints, and follow the work of Borm et al (JCE, 200/) on how we can define the success and the power of a study in the presence of multiple endpoints. We identify three different principles to define the success of a diagnostic study: A strong principle, requiring a sufficient level of both sensitivity and specificity, a weak principle, requiring only a sufficient level for one parameter, and a liberal principle, allowing some compensation between sensitivity and specificity to a clinically meaningful degree.

Although the weak principle is inadequate, it has influenced the analysis of diagnostic studies, e.g. by requiring adjustment for multiple testing. The strong principle is rather common, whereas the liberal principle is rarely used. The liberal principle can be implemented in several ways, for example by weighted averages of sensitivity and specificity or two-dimensional confidence regions for sensitivity and specificity.

We illustrate the impact of following the liberal principle on sample size calculations. Such a joint evaluation of sensitivity and specificity can lead to a substantial reduction in the necessary sample size compared to just following the strong principle.

We conclude that it is time to take the joint evaluation of sensitivity and specificity more serious in abalysing the accuracy of binary diagnostics tests.

Comparison of two diagnostic tests regarding sensitivity and specificity

Daniela Rieck, Antonia Zapf
Medical School Hannover, Germany; rieck.daniela@mh-hannover.de

Keywords: Diagnostic studies, paired samples, proportions, confidence interval

In diagnostic studies a new diagnostic test is often compared with a standard test. In phase III studies primary endpoints should be sensitivity and specificity. For the estimation of the difference between two tests regarding sensitivity and specificity the confidence intervals for the difference of two dependent rates can be used. Newcombe compared 1998 [1] some confidence intervals for improvement of the coverage probability. The favorite method was a continuity-corrected Wilson-interval. These results will be compared with other approaches; on the one hand with Agresti and Tango`s confidence intervals as improvements of the Wald interval which are well-known for not holding the type 1 error for small sample sizes. And on the other hand they are compared with non-parametric procedures [2] and the "free marginal GEE`s" (generalized estimating equations) [3]. The advantage of the "free marginal GEE's" is that repeated measures per individual can be considered.

In this presentation the different methods will be introduced and compared with an example and with simulations studies.

First results show that the "free marginal GEE's" and the nonparametric procedures perform best.

References

[1] R. G. Newcombe (1998). Improved confidence intervals for the difference between binomial proportions based on paired data, Statistics in Medicine, **17**:2635-2650.

[2] L. Lange (2011). Nichtparametrische Analyse diagnostischer Gütemaße in Clusterdaten, Dissertation, Göttingen.

[3] W. Leisenring, M. S. Pepe and G. Longton (1997). A marginal regression modeling framework for evaluating medical diagnostic tests, Statistics in Medicine, **16**:1263-1281.

Evaluation of Reader Heterogeneity in Diagnostic Trials

Katrin Haeussler, Armin Koch, Antonia Zapf
Medizinische Hochschule Hannover, Germany; katrinhaeussler@web.de

Keywords: Diagnostic trials, subjective outcome, agreement measures, type I and power simulations

In diagnostic trials with subjective outcome like in trials for imaging agents, outcome can be interpreted differently by several readers. For this reason, the Guideline on Clinical Evaluation of Diagnostic Agents from the European Medicines Agency (EMA) [1] points out that in such situations, two or more observers should make their diagnosis, if possible. An additional request in the guideline is an eligible description of intra-observer agreement, for example with the help of Cohen's kappa [2]. But it is known that the calculation of Cohen's kappa can lead to paradoxa [3].

There already exist a lot of modified kappa-coefficients and some other alternatives. To get to know which of the coefficients is really appropriate to prove poor agreement between two observers in diagnostic phase III studies in the talk, simulations for the type I error and the power will be presented. The statistical methods and the simulations are displayed in detail in the author's master thesis [4].

References

[1] Guideline on Clinical Evaluation of Diagnostic Agents, European Medicines Agency (EMA), London, 2009. Online at:

http://qibawiki.rsna.org/images/1/1a/EU.2009.08.04_Guideline_on_Clinical_evaluation_of_diagnostic_agents.pdf (last request 28/2/2011).

[2] J. Cohen (1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, **20**:37-46.

[3] A. Feinstein and D. Cicchetti (1990). High Agreement But Low Kappa: I. The Problems of Two Paradoxes. Journal of Clinical Epidemiology, **43**:543-549.

[4] K. Haeussler. Evaluation of Reader Heterogeneity in Diagnostic Trials. Master thesis in degree programme Medical Biometry/Biostatistics, University of Bremen, 2010.

Multiple Testing (3)

W. Maurer

Tuesday 13. September, 14:00 – 15:30, Lecture Hall KOL-F121

Multiple Comparisons Problems in Complex Clinical Trial Designs

H.M. James Hung[1], Sue-Jane Wang[2]
[1]FDA, United States of America; [2]FDA, United States of America;
HsienMing.Hung@fda.hhs.gov

Keywords: Active-controlled, adaptive design, trial-in-trial

Recent advances in clinical trial methodology provide a number of potentially more efficient designs to study multiple clinical hypotheses related to endpoints, doses, or patient populations in regulatory applications. Consequently statistical methods are also advanced to meet this demand, in particular, multiple comparison methods. Level of difficulty in multiplicity problems rises even more under active controlled designs and under adaptive designs. For a regulatory application, from the vantage point, multiple trials may be jointly analyzed or a single trial may be split into a number of trials to test multiple clinical hypotheses. This paper will present the challenges to the conventional frameworks of statistical inference under such designs and stipulate a number of approaches to such multiplicity problems in confirmatory clinical trials for regulatory settings.

Graphical approaches for multiple endpoint problems using weighted parametric tests

Ekkehard Glimm, Frank Bretz, Willi Maurer
Novartis Pharma, Switzerland; ekkehard.glimm@novartis.com

Keywords: Multiple testing, graphical test procedures, weighted Dunnett test

In clinical trials, the effect of a new treatment is often investigated in multiple endpoints, for example different patient characteristics (e.g. weight loss and change of HBA1C level), different doses of a drug, different time points at which the effect is measured, or combinations thereof. When the new treatment is compared to an established one or placebo, control of the familywise error rate is often required to avoid over-optimistic conclusions about the effect of the new treatment. In addition, usually some comparisons are more important than others, such that partial hierarchies of primary and secondary hypotheses arise. These challenges have triggered the development of stepwise multiple testing procedures, like the Bonferroni-Holm procedure and generalizations to gatekeeping and fallback procedures. Bretz et al. (2009, *Statistics in Medicine* 28, 586-604) have suggested a graphical approach that allows an easy, transparent description of such procedures by means of directed graphs. However, their paper restricts the investigation to Bonferroni-based procedures, i.e. methods that do not exploit knowledge about the multivariate distribution of corresponding test statistics.

The talk will first present the approach by Bretz et al. (2009) and then discuss its extension to situations where endpoints are (asymptotically) normally distributed with known correlations (which, for example, may occur when several doses of a new drug are compared with the same active control). The situation where all or some of the correlations have to be estimated from the data will also be considered.

Simultaneous confidence intervals for evaluation of multi-arm trials

Ludwig A. Hothorn
Leibniz Universität Hannover, Germany; hothorn@biostat.uni-hannover.de

Keywords: GWAS, multiple testing, interaction

Muti-arm randomized clinical trials occur in phase II and III, where commonly a placebo group and several treatment or dose groups are included. Although the reporting of multiplicity-adjusted p-values is common, this talk focuses on simultaneous confidence intervals since the ICH E9 recommends effect sizes and their confidence intervals for interpretation.

The first question arises which effect sizes are appropriate? Commonly the standardized or unstandardized difference-to-placebo is used. Alternatively the ratio-to-placebo is discussed: their advantage (interpretability in terms of k-fold change, scale-independence for multiple endpoints, larger power) and its disadvantage Non-parametric inference is rarely used, the disadvantages of the Hodges-Lehmann estimators (and their sCI) are discussed and the alternative "relative effect size" [1] is proposed. Briefly asymptotic sCI for risk differences, risk ratios and odds ratios, and hazard ratios are discussed.

The concept of multiple contrast tests is the central part of the talk and its flexibility to estimate non-stepwise sCI for many-to-one, all pairs, change-point, etc. alternatives.

Several real data examples are used to explain the particular problems, and the R packages multcomp, mratios, MCPAN and nparcomp are used for their evaluation. The statistical backbone of all approaches is the multivariate t/normal distribution with an arbitrarily-structured correlation matrix, realised in the R package mvtnorm [2]

References

[1] F. Konietschke (2009): Simultane Konfidenzintervalle fuer nichtparametrische relative Kontraste Effekte, Georg-August Universitaet Goettingen, Diss.

[2] F. Bretz, A. Genz, and L.A. Hothorn (2001). On the numerical availability of multiple comparison procedures, Biometrical Journal, **43**:645-656.

Use of modeling approaches to support dose selection at interim in adaptive designs for confirmatory clinical trials

Franz Koenig[1], Frank Bretz[2], Bjoern Bornkamp[2], Alexandra Graf[1]
[1]Medical University of Vienna, Austria; [2]Novartis Pharma AG; franz.koenig@meduniwien.ac.at

Keywords: MCP-mod, adpative designs, closed test

Due to the possibility to improve the efficiency of late phase clinical development programs[1], adaptive designs have received increased attention in the last years.

We investigate the use of modelling approaches to (i) increase the power of declaring effective dose statistically significant, (ii) support dose selection at an interim analysis. First, for a fixed sample design we apply the MCP-mod approach from [2], who suggested calculating optimal contrasts based on a-priori information about plausible dose response shapes available at the planning stage, together with the closed-testing-procedure to obtain p-values for the global trend assessment (i.e whether there is any statistical evidence for a dose-related drug effect) as well as for the pairwise comparison of individual doses against placebo.

In a second step we extend this closed MCP-Mod methodology to adaptive two-stage designs by applying an adaptive combination test to each intersection hypothesis[1,3]. Combining the data from both stages in adaptive confirmatory designs allow for flexible interim decisions based on all (interim) data of the ongoing trial while always ensuring strict type I error control. In particular, the MCP-Mod approach can be used to obtain model-based dose effect estimates at interim to guide early futility stopping and/or re-design the second stage (e.g. choice of doses, sample size) and analysis (e.g., dropping of inadequate response models).

References

[1] F. Bretz, Koenig, Brannath, Glimm and Posch (2009). Adaptive Designs for Confirmatory Clinical Trials. StatMed., **48**:574-585.

[2] F. Bretz, J. Pinheiro, M. Branson (2005). Combining multiple comparisons and modeling techniques in dose-response studies. Biometrics, **61**:738-748.

[3] P. Bauer and M. Kieser (1999). Combining different phases in the development of medical treatments within a single trial. StatMed., **18**:1833-1848.

Time to Event Analysis (3)

A. Wienke

Tuesday 13. September, 16:00 – 18:00, Lecture Hall KOL-F101

A new insight in the validation of prognostic indices based on frailty models and an application to the validation of a prognostic index in bladder cancer patients.

Catherine Legrand
ISBA, Université catholique de Louvain, Louvain-la-Neuve, Belgium)
Joint work with Paul Janssen (Hasselt University, Belgium), Luc Duchateau (Ghent University, Belgium), Richard Sylvester (EORTC, Belgium) and Vincent Ducrocq (INRA, France).

Keywords: Prognostic factors, Bayesian

Prognostic factors models investigate the relationship between patient/disease characteristics and the outcome of the patient. Factors found to be associated with the outcome are called prognostic factors and are usually combined to construct a prognostic index or score whose value can be used to classify patients in different risk groups.
A major issue when constructing a new prognostic index is its generalisibility to daily clinical practice. Besides "classical" validation technique, we recommend in this paper to investigate the heterogeneity of the prognostic index risk group hazard ratios over different centers using a frailty model. We include in this model a random center effect and a random prognostic index by center interaction; the parameter of interest being the variance components of these random effects.
We propose to base statistical inference for this model on a Bayesian approach. The variance components are thus estimated from their marginal posterior density after integrating out the fixed prognostic index effect and the random effects. As this integration cannot be performed analytically, the marginal posterior density is approximated using the Laplace integration technique. We further discuss different ways to summarize and interpret information available from this marginal posterior distribution.
We demonstrate the proposed technique using data from a pooled database of seven EORTC bladder cancer clinical trials. Using these data we investigate and interpret heterogeneity in the Allard prognostic index effects over centers.

The survAUC package: Tools to Evaluate the Prediction Accuracy of Survival Models

Matthias Schmid
University of Erlangen, Germany; matthias.schmid@imbe.med.uni-erlangen.de

Keywords: Survival analysis, prediction accuracy, risk prediction

The evaluation of prediction rules for continuous survival outcomes has become a key interest in biostatistical research. In contrast to the situation where predictions for uncensored outcomes have to be evaluated, deriving measures of prediction accuracy is not straightforward in the presence of censored observations. This is because traditional performance measures for continuous outcomes (such as the mean squared error or the R squared fraction of explained variation) are biased if applied to censored data.

The talk presents an overview of estimators to evaluate the accuracy of prediction rules for continuous survival outcomes. The focus is on estimators of discrimination indices that measure how well a prediction model separates observations having an event from those having no event (Pepe et al. 2008). A variety of these estimators is implemented in the R add-on package survAUC (Potapov et al. 2011). The package also implements R functions to estimate likelihood-based coefficients (O'Quigley et al. 2005) and measures based on scoring rules (Gerds and Schumacher 2006).

References

[1] S. Potapov, W. Adler and M. Schmid (2011). survAUC: Estimators of Prediction Accuracy for Time-to-Event Data. R package version 1.0-1. http://cran.at.r-project.org/web/packages/survAUC/index.html.

[2] M. S. Pepe, Y. Zheng, Y. Jin, Y. Huang, C. R. Parikh, W. C. Levy (2008). Evaluating the ROC performance of markers for future events. Lifetime Data Analysis **14**:86–113.

[3] T. A. Gerds, M. Schumacher (2006). Consistent estimation of the expected Brier score in general survival models with right-censored event times. Biometrical Journal **48**:1029–1040.

[4] J. O'Quigley, R. Xu and J. Stare (2005). Explained randomness in proportional hazards models. Statistics in Medicine **24**:479–489.

The trend renewal process: a useful model for medical recurrence data

Diana Pietzner, Andreas Wienke
MLU Halle-Wittenberg, Germany; diana.pietzner@medizin.uni-halle.de

Keywords: Recurrent events, colon cancer, frailty, parametric

For time to event data it is common to use the Cox model in the medical field. However, the Cox model cannot be used for recurrent events, as several observations from one subject are usually correlated. Several modifications of the Cox model exist to use for recurrence data. We suggest the use of a parametric model here, the so called trend renewal process [1]. It is composed of a trend and a renewal component that interact, and therefore provides a compromise between calendar and gap time scale. The model by Linqvist et al. [1] is extended to include a Cox type covariate term to account for observed heterogeneity. A further extension includes random effects to account for unobserved heterogeneity. We fit the suggested version of the trend renewal process to a data set of hospital readmission times of colon cancer patients [2] to illustrate the method for application to clinical data.

References

[1] B. H. Lindqvist, G. Elvebak, K. Heggland (2003). The trend-renewal process for statistical analysis of repairable systems. Technometrics, **45**:31–44.

[2] J. R. González, E. Fernandez, V. Moreno, J. Ribes, M. Peris, M. Navarro, M. Cambray and J. M. Borrs (2005). Sex differences in hospital readmission among colorectal cancer patients. Journal of Epidemiology and Community Health **59**:506–511.

Diagnostics 2: Biomarkers and Modeling

C. Schwenke

Tuesday 13. September, 16:00 – 18:00, Lecture Hall KOL-F104

Study planning for the validation of a prognostic marker

Vivian Lanius
Bayer Healthcare Pharmaceuticals, Germany; vivian.lanius@bayer.com

Keywords: Prognostic marker, time-dependent sensitivity and specificity

Diagnostic and prognostic markers play an increasingly important role in clinical practice. Ideally, a prognostic marker forecasts the likely course of a disease or a future outcome. The first step in the process of validating prognostic markers is the establishment of a correlation between the marker and the outcome of interest. The hazard ratio can be used to assess the increased relative risk to progress to a certain disease for subjects with a certain baseline marker result. However, moving from the initial association to a statistical validation entails major challenges.

In this talk we will discuss the planning of a prospective study with the aim to validate a diagnostic agent as prospective marker for a certain disease, which currently cannot be effectively treated yet. The performance of classification tools is generally measured via sensitivity and specificity, which in the context of a progostic marker are time-dependent. For a time-to-event outcome, time-dependent estimators for censored data can be used. We investigate how the expected results of such a study depend on the assumed progression rates and the hazard ratio as well as the marker prevalence. Since the prognostic impact of a marker generally depends on the time, the interpretation of the outcome of such a study requires careful consideration, also in view of the clinical utility.

Subgrouping variables for the genetic characterization of lung cancer

Christian Netzer, Jörg Rahnenführer
TU Dortmund University, Germany; netzer@statistik.tu-dortmund.de

Keywords: Survival, lung cancer, genetic data, subgroup analyses

Lung cancer is a disease with several distinct subtypes based on pathological analysis, like adenocarcinomas or squamous cell carcinomas. Many critical differences can easily be captured by genetic markers, for example based on copy number variation data. However, also within the disease types the genetic landscape of the tumors is characterized by high heterogeneity. Both clinical and genetic variables have a prognostic impact on clinical outcome, e.g. the survival of a patient or other response variables.

Common methods for analyzing such effects are the cox proportional hazard model or survival trees and survival forests. These methods assume homogeneous effects of each variable on the outcome. Then one unique parameter is used for all patients in the regression based methods and one best working split variable in the classification methods. But what if one or more variables, not necessarily having an effect on the outcome, define subgroups of patients for which an alternative parameter or splitting variable is more suitable than for the rest of the patients?

Our goal is to identify such *subgrouping variables* based on the genetic profile of the patients. In the talk we present results of subgroups of patients for which a statistical model provides a substantially better fit than for the rest of the patient cohort. The overall goal is to divide the genetic landscape of lung cancer patients in subgroups with corresponding high quality prognostic models. We discuss first results on a large lung cancer cohort with several hundred patients.

Comparison of methods for clustered data and meta-analysis in diagnostic studies

Antonia Zapf[1], Oliver Kuss[2]
[1]Medical School Hannover, Germany; [2]Martin-Luther-University Halle-Wittenberg, Germany; zapf.antonia@mh-hannover.de

Keywords: Cluster data, meta-analysis, diagnostic studies

According to the guideline for the evaluation of diagnostic tests of the European Medicine Agency (EMA) the performance of a diagnostic test should be given on a per patient basis and on a per lesion basis in case of clustered data. Clustered data here means that more than one lesion per patient is possible. Furthermore the guideline says that in a phase III diagnostic study sensitivity and specificity should be primary endpoints. A new approach to estimate them in the case of clustered data is given in the thesis of Lange [1].

Another recent research topic related to diagnostic studies are systematic reviews and meta-analyses to aggregate estimated sensitivities and specificities from several studies [2].

In both cases we have a hierarchical data structure with two levels. In the first case, lesions constitute the first level, patients the second. In the second, meta-analytic case, patients constitute the first level, studies the second. So, in principle, the same statistical models could be used for both cases where the challenge is to account for the inter-patient (first case) or inter-study (second case) correlation. An additional complexity arises because two outcomes (sensitivity and specificity) are involved.

In the talk we compare both approaches and show how they can benefit from one another.

References

[1] K. Lange (2011). Nichtparametrische Analyse diagnostischer Gütemaße in Clusterdaten, Dissertation, Göttingen.

[2] M. M. Leeflan, J. J. Deeks, C. Gatsonis, P. M. Bossuyt (2008). Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Ann Intern Med. **149**:889-97.

Nonparametric analysis of diagnostic trials regarding clustered data

Katharina Lange, Edgar Brunner
University of Göttingen, Germany; katharina.lange@medizin.uni-goettingen.de

Keywords: Diagnostic trials, ROC-Curve, AUC, clustered data, nonparametric Behrens-Fisher problem

For some years diagnostic studies have developed into an important mainstay of clinical research. Therefore the development of new statistical methods of evaluating diagnostic studies has become an influential issue in biostatistics.

Clustered data are obtained whenever several subunits (e.g. organs or vessels) of the same patient are observed, where no, several, or all subunits may be diseased or non-diseased (classified by a goldstandard). This leads to complicated correlation structures exposing a complex problem.

Konietschke and Brunner [1] develop a nonparametric method in multiple reader studies by extending an approach of Werner and Brunner [2] to small sample sizes. But restrictive assumptions regarding the sample sizes are required for both methods. As there are diagnostic studies not meeting these conditions a more general approach becomes necessary.

In this presentation a method to overcome this disadvantage is proposed. Weighed and unweighted estimators of the different measures of diagnostic accuracy as well as their asymptotic distribution and their corresponding inferencial statistics are presented.

References

[1] F. Konietschke and E. Brunner (2009). Nonparametric analysis of clustered data in diagnostic trials: Estimation problems in small sample sizes, CSDA, **53**:730–741.

[2] C. Werner and E. Brunner (2007). Rank methods for the analysis of clustered data in diagnostic trials, CSDA, **51:**5041–5054.

Miscellaneous

W. Maurer

Tuesday 13. September, 16:00 – 18:00, Lecture Hall KOL-F121

Multidimensional ordination of classification methods performance for microarray data

Małgorzata Cwiklinska-Jurkowska, Magdalena Wietlicka-Piszcz
Collegium Medicum UMK, Poland; malgosiajurkowska@o2.pl

Keywords: Performance, ordination, microarray classification,

From the reason of existing many variables selection and discriminant methods used in microarray classification problems, the arrangement of them based on resulting performance may help in further or more detailed examination. Experimental analysis of relationships between wide range of genes selection methods and discriminant methods by few exploratory methods like multidimensional metric scaling or cluster analysis based on all succeeding subsets of 100 or 30 the most important genes was performed. Thus obtained plots include joined information of performance.

The obtained information about similarity between results of genes selection methods may be beneficial in choosing the most accurate genes selection methods or in the fusion of genes selection methods in order to combine diverse or (and) effective procedures. In the same way, similarity between discrimination methods performance may indicate groups of close generalization properties. This may be useful in selection of most accurate discriminant procedure or in the construction of discriminant methods ensemble build from accurate but diverse classifiers.

The obtained results show which methods have a tendency to obtain close generalization properties. Similarities obtained by exploratory analysis methods confirm conclusions obtained by comparing learning curves and are also complementary.

The applied analysis may be regarded as the way of the condensed comparison of many methods from many points of view. Analysis of relationships between genes selection and discriminant methods based on performance assessment for microarray data classification may be useful for the decision made by explorer.

Test methods for correlated functional imaging data

Daniela Adolf, Siegfried Kropf
Otto-von-Guericke University Magdeburg, Germany; daniela.adolf@med.ovgu.de

Keywords: Correlated sample elements, block-wise permutation incl. a random shift, separated multivariate GLM

In functional magnetic resonance imaging (fMRI), data is high-dimensional and correlated in time and space. A multivariate general linear model (GLM) is considered for a fMRI session with one person

$$\mathbf{Y}=\mathbf{XB}+\mathbf{E}$$

where the noise **E** follows a multivariate normal distribution with expectation zero but covariance structure **P** in the rows and **Σ** in the columns, connected by a Kronecker product. The data matrix **Y** contains n measurements (successive fMRI scans) over p variables where $p \gg n$. The general underlying null hypothesis is a weighted contrast over several parameter vectors of **B**.

In contrast to the classical multivariate GLM, the sample vectors are correlated and **P** is supposed to be a first-order autoregressive process.

To analyze these data non-parametrically, we use a block-wise permutation method including a random shift in order to count for the temporal correlation.

But in classical permutations, it is only possible to test one single parameter vector or all. However, we want to be able to test any null hypothesis on the parameter estimates via this special permutation method. This is important because analyzing functional imaging data is particularly based on testing differences of parameter estimates.

Therefore, we use a separated multivariate linear model

$$\mathbf{Y}=\mathbf{X}_1\mathbf{B}_1+\mathbf{X}_2\mathbf{B}_2+\mathbf{E}$$

and the special null hypothesis H_0: $\mathbf{B}_2=\mathbf{0}$ that is only related to that part of the design matrix that contains the information of interest.

We will show how other hypotheses can be transformed into the separated model and can be tested via the block-wise permutation method including a random shift.

On the normality of the log odds ratio

Karl-Ernst Biebler, Bernd Jäger
Ernst-Moritz-Arndt-University Greifswald, Germany; biebler@biometrie.uni-greifswald.de

Keywords: Normality of the log odds ratio, sample size calculations

Dichotomous data are often summarized into 2x2 contingency tables. A characteristic of the tables preferentially used is the odds ratio (*OR*). The presentation deals with the asymptotic normality of the *log* (*OR*). Siegmund [2] has examined the quantitative aspects of this asymptotic quality. The rate of approach to normality of the heuristic odds ratio estimator is unknown and can be very slow near the boundary of the parameter space. Consequences are demonstrated at the example of the sample size planning of tests on the *log* (*OR*). Examples show that standard approximate methods do not always work exactly enough. Monte Carlo simulations are used for the improved sample size computations. The related programmes are developed in a SAS®-environment by Freigang [1]. They permit reliable calculation of the test sample sizes also for extreme binomial probabilities.

References

[1] F. Freigang (2011). *Stichprobenumfangsplanungen für Tests der odds ratio.* Diploma thesis. Ernst-Moritz-Arndt-University. Faculty of Natural Sciences. Greifswald

[2] D. Siegmund (1982). A sequential confidence interval for the odds ratio. *Prob.Math. Statist*., **2**:149-156.

Systematische Übersicht zur Therapie analer Inkontinenz bei Erwachsenen mit Biofeedback und Elektrostimulation

Reinhard Vonthein[1], Tankred Heimerl[2], Inke R. König[1], Christiane Wichmann[1], Claudia Hemmelmann[1], Thilo Schwandner[2], Andreas Ziegler[1]
[1]Universität zu Lübeck, Germany; [2]Universitätsklinikum Gießen, Germany;
reinhard.vonthein@imbs.uni-luebeck.de

Keywords: Systematic review, PROSPERO, minimal important difference, RCT, clinical endpoint

Ziel

Zur Behandlung der analen Inkontinenz werden häufig die konservativen Methoden Biofeedback (BF) und niederfrequente Elektrostimulation (LFS) eingesetzt. Die Evidenzlage von BF und LFS in früheren systematischen Reviews ist unklar. Ziel des Vortrags ist es, den aktuellen Stand der verfügbaren randomisierten und kontrollierten Studien unter Berücksichtigung neuer mittelfrequenter Stimulationskonzepte und der jeweils verwendeten Stromstärken zur Behandlung analer Inkontinenz zu bewerten.

Methode

Ins neue systematische Review zu BF und Elektrostimulation wurden Studien bester Evidenz eingeschlossen. Hierfür wurde die Datenbank PubMed sowie die klinischen Studienregister clinicaltrials.govund DRKS systematisch nach Originalarbeiten und systematischen Reviews durchsucht. Studien, bei denen BF oder Elektrostimulation als Adjunkt zu einer operativen Methode geprüft wurde, wurden von der Untersuchung ausgeschlossen. Suche, Selektion, Extraktion, Zusammenfassung und Darstellung folgen den Cochrane-Regeln und dem PRISMA-Statement.

Ergebnisse

Es wurden 15 randomisierte kontrollierte Studien identifiziert. Die drei ersten Studien aus den Jahren 1984 bis 1990 mussten aufgrund ungeeigneter Designs ausgeschlossen werden. Drei weitere Studien beinhalteten Prüfungen der Wirksamkeit nach Geburtstrauma mit widersprüchlichen Ergebnissen.

Unter Patienten mit durchschnittlichem Alter um 60 Jahre und gemischten Ätiologien gibt es geringe Evidenz für LFS bei teilweise verringerter Compliance. Ausschließlich für die Kombination von BF und amplitudenmodulierter Mittelfrequenzstimulation mit rund 100 mA (3T) liegen zwei Studien mit Evidenz zugunsten von 3T vor, einmal im Vergleich zu LFS, einmal im Vergleich mit BF.

Zusammenfassung

Die Evidenz von LFS und BF zur Behandlung analer Inkontinenz ist gering. Das Therapiekonzept 3T ist in einzelnen Studien sowohl LFS als auch BF überlegen.

Young Statisticians (2)

S. Mejza

Wednesday 14. September, 9:00 – 10:30, Lecture Hall KOH-B10

Gene selection procedures including correlation between genes

Joanna Grazyna Zyprych-Walczak, Idzi Siatkowski
Poznan University of Life Sciences, Poland; zjoanna@up.poznan.pl

Keywords: Microarrays, gene selection, correlation

Selection of genes is important issue in a discriminant analysis. Several existing statistical methods such as SAM (Significant Analysis of Microarrays) [1], correlation-sharing statistics [2], statistics proposed by Zuber and Strimmer in [3] as well as novel aproach for gene selection taking into consideration correlation between genes will be presented and compared. The new method allows to find the set of differentially expressed genes which have low level of correlation with each other.

References

[1] V. T. Tusher, R. Tibshirani, G. Chu (2001). Significance analysis of microarrays applied to the ionizing radiation response, Proc Natl Acad Sci, **98**:5116–5121.

[2] R. Tibshirani, L. Wasserman (2006). Correlation sharing for detection differential gene. arXiv:math/0608061v1

[3] V. Zuber, K. Strimmer (2009). Gene ranking and biomarker discovery under correlation, Bioinformatics computer Applications in The Biosciences, **25**:2700-2707.

Linear Mixed Model in Gene Selection Problem

Alicja Szabelska, Idzi Siatkowski
Poznan University of Life Sciences, Poland; aszab@up.poznan.pl

Keywords: Microarray, gene selection, linear mixed model, permutational F test

Increase of the quality and precision of the microarray technique forces application of the accurate statistical analysis of the data coming from such experiments. Following [1] standard procedures for selecting differential genes require linear model. Tests such as t test, F test or Kruskal – Wallis test are available to use for this model. In this case the independence between among all observations as well as only one source of random variation is assumed.

Although it is applicable to many microarray experiments, fixed model does not allow for multiple sources of variation nor does it account for correlation among the observations that arise as a consequence of different layers of variation. In particular, when research involves biological along with technical replicates statistical analysis could involve applying modifications of these methods. The most natural way to include several types of random variation is to use linear mixed model.

We underwent research to verify the utility of this model in the analysis of microarray data. The application of several tests introduced in [2] and [3] based on linear mixed model will be presented. The results obtained by usage of mixed and fixed models will be compared.

References

[1] G.K. Smyth (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments, Statistical Applications in Genetics and Molecular Biology, 3 (1), Article 3.

[2] J.C. Pinheiro, D.M. Bates (2000). Mixed-Effects Models in S and S-PLUS, Springer.

[3] N. M. Laird, J.H. Ware (1982). Random-Effects Models for Longitudinal Data, Biometrics, **38**:963-974.

Greenhouse gases emission in agriculture - measurements for today and prospects for the future

Elżbieta Wójcik-Gront
ewojcik.gront@gmail.com

Keywords: Emission uncertainty

Anthropogenic greenhouse gases (GHG) emission influences the global climate change. Preventing the temperature increase requires the knowledge of the GHG amount emitted. Therefore the GHG inventory has been started by the United Nation Framework Convention on Climate Change (UNFCCC) legitimated by United Nations. The mandatory emission limits were established under the Kyoto protocol. This raised a need for accurate estimation of emission values as well as their uncertainties [1, 2]. These values have to be submitted to the UNFCCC by each participating country separately for each industry sector.

Poland has reported emission using simplified approach to uncertainty estimation which is based on the assumptions that measured values are statistically independent and their probability distributions are symmetric. The aim of the presentation is to assess emission uncertainty without these assumption. The main focus in this contribution is to carefully investigate uncertainties of GHG emission in agriculture sector which has not been studied in Poland so far.

References:

[1] IPCC (1996) IPCC Guidelines for National Greenhouse Gas Inventories, vols. 1–3. Intergovernmental Panel on Climate Change, London [2] IPCC (2001) Good practice guidance and uncertainty management in national greenhouse gas inventories. In: J. Penman et al. (Eds.), IPCC National Greenhouse Gas Inventories Programme. Technical Support Unit, Hayama, Japan

Sample size calculation for the three-arm ‘gold standard’ non-inferiority design

Kathrin Stucke, Meinhard Kieser
University of Heidelberg, Germany; stucke@imbi.uni-heidelberg.de

Keywords: Non-inferiority, three-arm-trial, sample size calculation, optimal allocation, multiple comparisons

In the three-arm ‘gold standard’ non-inferiority design, an experimental treatment, an active reference and a placebo are compared. This design is becoming increasingly popular and is, whenever feasible, recommended for use by regulatory guidelines. However, comparatively few research has been done on the topic of sample size calculation for studies with this design yet.

We present a general method for sample size calculation and identification of an optimal sample size allocation for the three-arm ‘gold standard’ design. As special cases, continuous, ordered and binary outcomes are covered by this approach.

It turns out, that optimal allocation leads to a considerable decrease of the total sample size and assigns in many situations more patients to the active treatment groups than to placebo. The latter property is desirable for various reasons. The ICH E10 guideline states that ‘it is possible to make the active groups larger than the placebo group to improve precision of the active drug comparison’. Furthermore, from an ethical viewpoint ‘this may also make the trial more acceptable to patients and investigators, as there is less chance of being randomized to placebo’.

We apply our method to a recently published clinical trial in patients with major depressive disorders and show how the optimal allocation ratio can be used as a starting point to determine a sample size assignment rule that is favorable both from statistical and practical viewpoints.

Sensitivity-Preferred Strategy in Building Classifiers for High-Dimensional Data

Inoncent Agueusop[1,2], Stephan Lehr[3], Richardus Vonk[1], Katja Ickstadt[2]
[1]Bayer Pharma AG, Berlin Germany; [2]Department of Statistics, TU Dortmund, Dortmund;
[3]Baxter Healthcare, Vienna; inoncent.agueusop@bayer.com

Keywords: Sensitivity, LASSO, gene expression data.

In two-class classification problems, the correct assignment to one class may be more important than the other for specific purposes. Most classification rules, however, are designed to optimize the overall utility-function without using this weighting condition. Some new methods try to deal with the challenge, but up to now no classification rule is able to incorporate the different importances of possible misclassifications within the optimization process and hold a certain pre-specified sensitivity value at the same time. We therefore propose a new classification rule based on the optimization of the likelihood subject to the sensitivity greater than a certain threshold value, e.g., 0.9. As the classification method we choose L_1 penalized regression, which also enables us to simultaneously select sensitivity-preferred predictors. The results of this new strategy, applied on gene-expression data with binary outcomes, are compared with other sensitivity-preferred strategies.

Bayesian Methods in Translational Medicine (1)

M. Branson and L. Colin

Wednesday 14. September, 9:00 – 10:30, Lecture Hall KO2-F180

What Can we Learn from 30 Years of Bayesian Methods in Early Drug Development?

Andrew Peter Grieve
Aptiv Solutions, Germany; Andy.Grieve@aptivsolutions.com

Keywords: Bayesian, prediction, probability assessment

The genesis of the use of Bayesian methods in Pharmaceutical Research and Development is to be found in the early 1980s. In 1979 I joined Ciba-Geigy's Wissenschaftliche Rechenzentrum (WRZ) in Basel. Shortly after I joined WRZ Daniel Mandallaz and Jochen Mau developed an approach to bioequivalence assessment which they showed had a Bayesian interpretation [1]. This began a systematic investigation by WRZ colleagues of the implementation of Bayesian methods in the pre-clinical phase, particularly toxicology, and Early Development. The project was called 'Lernen aud Erfahrung' Much of what was then achievable was limited because of the lack f appropriate methodologies and computing tools. Many applications were limited to leass than 10 varaibles and that was generally achieved by analytic approximations [2]. It is of interest to trace the evolution of these early developments up to the present time. It is clear that what was advantageous in the 1980s remains advantageous today. In this talk we follow these developments pointing out what has been learnt from these early developments and what could still be learnt.

References

[1} D. Mandallaz D and J. Mau (1991). Comparison of different methods of decision makling in bioequivalence assessment. Biometrics, **37**:213-222.

[2] A. Racine, A. P. Grieve, H. Flühler and A. F. M. Smith (1986). Bayesian methods in practice: experience in the pharmaceutical industry (with discussion),. Applied Statistics, **35**:93-150.

Sample size considerations for proof of concept studies with binary outcomes

Michael Becka
Bayer Healthcare AG, Germany; michael.becka@bayer.com

Keywords: Binary outcome, proof of concept, bayes, beta-distribution

In early drug development, the step from animals and preclinical research to humans is critical. Desired effects and findings established in animal models have to be validated for humans. Preferably, this proof of concept has to take part as early as possible in the process to save costs and resources. Early studies in man are usually small in size and focused on safety, tolerability, and kinetic aspects. Beyond this primary programs, proof of concept aspects often have to be investigated separately. Using binary outcomes, bayesian techniques provide useful tools for the company's internal decision process. Besides thorough team discussions, sample size considerations form the basis for successful strategies.

A respiratory case study in early development: historical data inclusion and level-of-proof decision-making within a Bayesian framework.

Lilla Di Scala, Jouni Kerman, Beat Neuenschwander
Novartis Pharma, Switzerland; lilla.di_scala@novartis.com

Keywords: Translational medicine; Bayesian; historical evidence; level-of-proof

A respiratory proof-of-concept study, assessing the bronchodilatory effect of an experimental drug *vs*. the golden standard, provides the perfect opportunity to leverage historical information available on the active comparator, in order to complement the concurrent trial data and endeavor to keep budget and resources to a minimum. This case study suggests a framework in which the Bayesian mechanism is used to formally synthesize past information *via* a meta-analytical predictive approach and then to inform decision-making during the analysis step. The approach provides a rigorous context in which strength of evidence (level-of-proof) in favor of efficacy can be quantified and decisions on further development of the compound can be taken informatively.

Multivariable model-building with continuous variables – a comparison of flexible regression approaches

W. Sauerbrei

Wednesday 14. September, 9:00 – 10:30, Lecture Hall KOL-F101

Multivariable model-building with continuous covariates: 1. Performance measures and simulation design

Harald Binder[1,2], Willi Sauerbrei[2], Patrick Royston[3]

[1]University of Freiburg, Germany; [2]University Medical Center, Freiburg, Germany; [3]MRC Clinical Trials Unit, London, UK; binderh@imbi.uni-freiburg.de

Keywords: Continuous covariates, model selection, non-linear effects, regression models, simulation

Many approaches exist for evaluating techniques for developing multivariable regression models with potentially non-linear effects of continuous covariates. In simulation studies to evaluate new proposals and to compare model-building techniques, researchers tend to consider oversimplified settings or unrealistically complex functional forms. The true shape of functions used for data generation is rarely representative of what would be expected in biomedical applications. In addition, the mean square error of prediction is often used as the main criterion for evaluation. This is insufficient when the effect of individual variables is to be assessed, e.g., in exploratory studies in clinical epidemiology. After reviewing some of the proposals for simulation designs, we suggest a new design for normal error regression modeling that avoids oversimplification and tries to capture structure often found in biomedical settings. Specifically, we posit a non-trivial correlation structure between covariates, a challenge for many techniques of model selection. In addition to continuous covariates, binary covariates are included. As well as strong non-linear effects, some near-linear effects of covariates are considered. This allows one to judge whether a technique can distinguish between important non-linear effects and effects that might reasonably be represented by linear terms. We suggest several performance measures for capturing the potential impact of the various components of the simulation design on model structure. Some challenges of the simulation design are illustrated by diagnostic plots obtained after fitting a linear model, indicating the extent to which use of an under-specified model points towards potential non-linear effects.

Multivariable model-building with continuous covariates: 2. Comparison between splines and fractional polynomials

Willi Sauerbrei[1], Harald Binder[1], Patrick Royston[2]
[1]University Medical Center Freiburg, Germany; [2]MRC CTU London, UK; wfs@imbi.uni-freiburg.de

Keywords: Continuous covariates, fractional polynomials, model selection, splines, simulation

In observational studies, many continuous or categorical covariates may be related to a response of interest. Analyses based on splines as well as the multivariable fractional polynomial (MFP) approach can be applied to identify important variables and appropriate functional forms for continuous covariates. Whereas MFP is one well-defined procedure, many strategies based on splines have been suggested, and we chose to study two of them. The aim of an analysis often guides the level of complexity that is deemed acceptable for the final model. Spline-based strategies and MFP have tuning parameters for choosing the required level of complexity. However, it is unclear whether the strategies can equally well provide simple as well as complex models. Furthermore, a `reasonable' level of complexity may depend on the specific data situation. In a normal error regression model Therefore, we perform a comprehensive simulation study predicated on an underlying ("true") structure that realistically reflects biomedical contexts. We vary the amount of information (signal-to-noise ratio) in the data and the complexity levels for model selection. We consider prediction performance, Type I and Type II error rates, costs at the covariate level, and quantitative as well as qualitative criteria for judging selected functional forms. No one procedure performs best in all scenarios, but overall, MFP shows better performance than the multivariable spline strategies we investigated. The approaches are also applied to real data, indicating how the results from the simulation study transfer to real world applications.

Penalized splines and fractional polynomials for flexible modelling of the effects of continuous predictor variables: A comparison

Alexander M. Strasak[1], Nikolaus Umlauf[2], Ruth Pfeiffer[3], Stefan Lang[2]
[1]Hoffmann-La Roche, Switzerland; [2]University of Innsbruck, Austria; [3]National Cancer Institute, USA; alexander.strasak@roche.com

Keywords: Generalized additive models, simulation study, smoothing

P(enalized)-splines and fractional polynomials (FPs) have emerged as powerful smoothing techniques with increasing popularity in applied research. Both approaches provide considerable flexibility, but only limited comparative evaluations of the performance and properties of the two methods have been conducted to date. Extensive simulations are performed to compare FPs of degree 2 (FP2) and degree 4 (FP4) and two variants of P-splines that used generalized cross validation (GCV) and restricted maximum likelihood (REML) for smoothing parameter selection. The ability of P-splines and FPs to recover the ‘‘true’’ functional form of the association between continuous, binary and survival outcomes and exposure for linear, quadratic and more complex, non-linear functions, using different sample sizes and signal to noise ratios is evaluated. For more curved functions FP2, the current default setting in implementations for fitting FPs in R, STATA and SAS, showed considerable bias and consistently higher mean squared error (MSE) compared to spline-based estimators and FP4, that performed equally well in most simulation settings. FPs however, are prone to artefacts due to the specific choice of the origin, while P-splines based on GCV reveal sometimes wiggly estimates in particular for small sample sizes. Application to a real dataset illustrates the different features of the two approaches.

Biometrical Journal: Editor's Choice

L. Held

Wednesday 14. September, 9:00 – 10:30, Lecture Hall KOL-F104

Use of pre-transformation to cope with extreme values in important

Anne-Laure Boulesteix[1], Vincent Guillemot[2], Willi Sauerbrei[3]

[1]Uni München, Germany; [2]Ludwig-Maximilians-Universität, München, germany; [3]Uni Freiburg, Germany; boulesteix@ibe.med.uni-muenchen.de

Keywords: Biometrical Journal Editors Choice Session

Extreme values in predictors often strongly affect the results of statistical analyses in highdimensional settings. Although they frequently occur with most high-throughput techniques, the problem is often ignored in the literature. We suggest to use a very simple transformation, proposed before in a different context by Royston and Sauerbrei, as an intermediary step between array preprocessing and high-level statistical analysis. This straightforward univariate transformation identifies extreme values in continuous features and can thus be used as a diagnostic tool for outliers. The use of the transformation and its effects are demonstrated for diverse univariate and multivariate statistical analyses using nine publicly available microarray data sets. All our analyses are reproducible, following the recommendations of the Biometrical Journal.

Adaptive Dose-finding: Proof of Concept with Type I Error Control

Frank Miller
Astrazeneca, United Kingdom; Frank.Miller@astrazeneca.com

Keywords: Session Biometrical Journal Editors Choice Session

We consider an adaptive dose-finding study with two stages. The doses for the second stage will be chosen based on the first stage results. Instead of considering pair-wise comparisons with placebo, we apply one test to show an upward trend across doses. We are interested in trend tests based on a single contrast or on the maximum of multiple contrasts.

The choice of the Stage 2 doses in the adaptive design includes the possibility to add doses. If certain requirements for the interim decision rules are fulfilled, the final trend test which ignores the adaptive nature of the trial (naive test) can control the type I error. However, for the more common case that these requirements are not fulfilled, we need to take the adaptivity into account and discuss a method for type I error control. We apply the general conditional error approach to adaptive dose-finding and discuss special issues appearing in this application. We call the test based on this approach Adaptive Multiple Contrast Test.

We illustrate the theory using an example and compare the performance of several tests for the adaptive design in a simulation study.

References

[1] F. Miller (2010). Adaptive dose-finding: Proof of Concept with type I error control. Biometrical Journal, **52**:577-589.

Assessing inter-rater reliability when the raters are fixed : two

Valentin Rousson
Centre hospitalier universitaire vaudois, Switzerland; Valentin.Rousson@chuv.ch

Keywords: Biometrical Journal Editors Choice Session

Intraclass correlation (ICC) is an established tool for assessing inter-rater reliability of continuous measurements. When the raters in the sample are assumed to come from an infinite population of raters (model with random raters), ICC can be interpreted equivalently either as a correlation between repeated measurements, or as a percentage of variance due to true variability. When the raters in the samples are considered to be the whole population (model with fixed raters), however, this equivalence no longer holds such that two different concepts of ICC can be defined. In this presentation, we shall see how it is possible to estimate these two concepts and to calculate a confidence interval in a balanced design with fixed raters. While the difference between the two concepts and the two estimates is not dramatic in the case of a good reliability, it is found not negligible when the reliability is low, which may arise for example for an ICC used to measure the design effect in a sample size calculation for clustered data. Moreover, this formal distinction raises the issue how the concept of ICC is best generalized to other type of measurements, for example to categorical or ordinal measurements. What is easier/more desirable to generalize: the concept of correlation between repeated measurements, or the concept of percentage of variability not due to measurement error?

Epidemiology and Statistics

H. Ulmer and J. Wellmann

Wednesday 14. September, 9:00 – 10:30, Lecture Hall KOL-F121

The Relative Frailty Variance and Shared Frailty Models

Steffen Unkel, C. Paddy Farrington, Karim Anaya-Izquierdo
The Open University, United Kingdom; S.Unkel@open.ac.uk

Keywords: Current status data, heterogeneity, relative frailty variance, shared frailty model, time-varying frailty

The relative frailty variance among survivors provides a readily interpretable measure of how the heterogeneity of a population, as represented by a frailty model, evolves over time. In shared frailty models, the relative frailty variance is closely related to the cross-ratio function, which is estimable from bivariate survival data. Suitably rescaled, the relative frailty variance characterizes frailty distributions, and may be used to compare patterns of dependence for different models or for different data. In this talk, the possible shapes of the relative frailty variance function are investigated, to aid identification of a suitable frailty model. Flexible families of frailty distributions are introduced based on the properties of the relative frailty variance. Some ways of modelling time-varying frailties are discussed. The methods are illustrated with applications to infectious disease epidemiology using paired serological survey data.

Estimating the number of malaria infections in blood samples using high-resolution genotyping data

Amanda Ross[1,2], Cristian Köpfli[1,2], Sonja Schöpflin[1,2], Ivo Müller[3], Ingrid Felger[1,2], Tom Smith[1,2]

[1]Swiss Tropical and Public Health Institute; [2]University of Basel, Switzerland; [3]Papua New Guinea Institute of Medical Research; amanda.ross@unibas.ch

Keywords: Infectious diseases, malaria, epidemiology, genotyping, multiplicity of infection

People who live in malaria-endemic areas often harbour several infections at once. The number of infections is a useful measure which can improve the understanding of many areas of malaria epidemiology such as the dynamics of infections, parasite genetics, pathogenesis and the effects of transmission intensity. High-resolution genotyping can distinguish between infections by detecting the presence of different alleles at a polymorphic locus. However the number of infections may not be accurately counted since parasites from multiple infections may carry the same allele. We (i) propose a model to estimate the number of infections taking into account the probability of shared alleles and (ii) carry out simulations to determine the circumstances under which the number of infections is likely to be substantially underestimated due to shared alleles.

We use the observed allele frequencies to estimate the conditional probabilities of observing different numbers of genotypes given the true numbers of infections present. These probabilities are combined in a Bayesian model with the observed frequencies of genotypes and an assumed distribution for the numbers of infections. We evaluate this model using simulation and show that it can estimate the number of infections with reasonable accuracy.

Simulations indicate that the problem is not substantial for most datasets. Large disparities between the number of infections and number of observed genotypes were limited to cases with fewer than 20 alleles, fewer than 20 blood samples, a mean number of infections of more than 6 or a frequency of the most common allele of more than 20%.

Parameter estimation in a six-state model for partially observable data in chronic kidney disease

Alexander Begun[1], Andrea Icks[1,2], Ralph Brinks[1], Regina Waldeyer[1], Michael Koch[3,4], Guido Giani[1]

[1]Institute of Biometrics and Epidemiology, German Diabetes Center at the Heinrich-Heine-University, Düsseldorf, Germany; [2]Department of Public Health, Faculty of Medicine, Heinrich-Heine-University, Düsseldorf, Germany; [3]Center of Nephrology, Mettmann, Germany; [4]Clinic of Nephrology, Heinrich-Heine-University, Düsseldorf, Germany; alexander.begun@ddz.uni-duesseldorf.de

Keywords: Markov chain, parameter estimation

Markov chain models are often used to estimate the clinical course of chronic diseases. However, partially observable data are a common problem in estimating transition probabilities. This contribution concerns the estimation of parameters of a six-state Markov chain model for the course of chronic kidney disease (CKD). In the model CKD is staged by severity according to three successive ranges of glomerular filtration rate (eGFR) not requiring renal replacement therapy (RRT), dialysis, renal transplantation, and death as absorbing state. We used data of patients from a dialysis center in a German region during 2000-2011. For all patients, all visits in the center were documented with exact date and eGFR measurement. Exact data for the beginning of dialysis and renal transplantation are available, however, not for the entry in the eGFR states (partially observable data). Date of birth, sex, and diabetes status are available, too. These factors were included in the model in the form of the Cox-regression covariates. The estimates of unknown transition intensities and the coefficients of Cox regression were calculated using the maximum likelihood method. Our model-based results are in accordance with findings from clinical studies and clinical experience. E.g. we found that mortality increases significantly with reduction of GFR function and reaches its maximum in patients under dialysis. Diabetes significantly increases the mortality rate for patients under dialysis. We consider our approach to be helpful in partially observable data, e.g. in health economic models to analyze cost-effectiveness.

Modeling of continuous covariates in the mean structure of generalized estimating equations

Maren Vens, Jördis Stolpmann, Claudia Hemmelmann, Andreas Ziegler
Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Germany;
maren.vens@imbs.uni-luebeck.de

Keywords: Generalized estimating equations (GEE), multivariable fractional polynomial (MFP), backward-elimination (BE), Quasi-Likelihood Information Criterion (QIC)

Dependent outcomes are often considered in epidemiological studies. Usually these data are analyzed using standard methods neglecting the underlying dependencies. This can lead to false conclusions. Generalized estimating equations (GEE) can be used to overcome this problem. However, the modeling of continuous covariates is challenging. In our contribution, we focus on the use of multivariable fractional polynomials (MFP) for modeling the mean structure of GEE. The deviance is employed in the MFP algorithm to build the functional form, but this measure is not well defined in GEE. As an alternative, the quasi-likelihood information criterion (QIC) is used in our approach. A combination of this criterion and p-values is used in a backward elimination algorithm to get low-dimensional models. Modeling of variables and backward elimination are implemented in a SAS-Macro. The approach is illustrated using data from the Framingham Heart Study. The presented method is a promising approach for modeling continuous covariates in GEE. In future work, further methods of model selection as well as a validation using simulation studies will be conducted.

Statistics in practice (1): Double Session (for Young Statisticians) - Meta-Analysis and Experiences of Senior Statisticians

AG Nachwuchs, AG Weiterbildung, T. Mueller

Wednesday 14. September, 11:00 – 12:30, Lecture Hall KOH-B10

Introduction to Meta-Analysis (~ 45 minutes)

Andrea Berghold
Institute for Medical Informatics, Statistics and Documentation, Medical University Graz, Austria

This talk aims to prepare young statisticians for the Statistics in Practice (2) and (3) tutorial on meta-analysis. It covers the background, main ideas, methods and pitfalls. Particular topics include the history of meta-analysis, forestplots, heterogeneity and publication bias. In a second part, it gives an insight into the procedure of systematic literature reviews based on practical experience in the field of diabetes.

Practical Experiences of Senior Statisticians (~ 45 minutes)

Frank Bretz
Novartis Pharma AG, Switzerland
Joachim Gerss
Institute of Biostatistics and Clinical Research, University of Münster, Germany
Martin Kappler
Novartis Pharma S.A.S., France
Sue-Jane Wang
US Food and Drug Administration, United States of America

After university or at the end of a PhD study, the decision where to look for jobs and which steps on the way to a proper statistical career is very challenging. More information from the professional environment outside university is extremely valuable to help students in outlining their future plans.
In this session, four senior statisticians from the medical / clinical environment will give an insight into their personal career paths and talk about their current jobs. The descriptions will also include required skills and education plus an idea of the routines and tasks the jobs demand on a daily basis. Questions and lively discussions are explicitly welcome.

Bayesian Methods in Translational Medicine (2)

M. Branson and
L. Colin

Wednesday 14. September, 11:00 – 12:30, Lecture Hall KO2-F180

Comparative Bayesian escalation designs

Emmanuel Lesaffre[1,2], Paul Hamberg[3,4], Jaap Verweij[4], David Dejardin[1]
[1]Interuniversity Institute for Biostatistics and statistical Bioinformatics Katholieke Universiteit Leuven & Universiteit Hasselt, Belgium; [2]Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands; [3]Department of Internal Medicine, Sint Franciscus Gasthuis, Rotterdam, The Netherlands; [4]Department of Medical Oncology, Erasmus MC, Rotterdam, The Netherlands; Emmanuel.Lesaffre@med.kuleuven.be

Keywords: Bayesian dose escalation, adaptive design, EWOC

The primary objective of a Phase I dose escalation cancer study is to find the dose at which the drug will be tested in the subsequent phase II and III trials. In order to maximize the effectiveness of the treatment, drugs to treat cancer are combined with each other. The combination usually associates drugs that have different mechanism of action against the disease.

The vast majority of the phase I dose escalation studies for single agents and combination of agents implement a ``3+3'' dose escalation scheme to find the MTD. This phase I design has been criticized for treating too many subjects at suboptimal doses and providing a poor MTD estimate. Also, this design produces unreliable estimation of the true rate of toxicity at the optimal dose. One cause of unreliability is the nature of Phase I subjects: These subjects usually have multiple tumor types and are in advanced stages of disease. Hence, the toxicity levels observed in this population can not be generalized to the general population.

We propose a randomized Bayesian dose escalation design for combinations of drugs that takes advantage of the fact that a drug is added to a standard treatment to obtain, via Bayesian estimation, an improved estimation of the MTD and the toxicity level at the MTD.

The proposed design implements a randomization between standard treatment and the combination regimen for which we want the dose. We estimate the difference between the toxicity of the control and the combination to search for the MTD.

Combining information from healthy volunteers and patients for dose selection: case studies applying Bayesian statistics in early phase trials

Annette Sauter, Markus Niggli, Paul Jordan
Hoffmann - La Roche Ltd, Basel; annette.sauter@roche.com, markus.niggli@roche.com

Keywords: Adaptive dose selection, Bayesian decision criteria, early phase clinical trials

Dose-toxicity and dose-efficacy modeling are important in multiple ascending dose (MAD) and proof of concept (POC) studies to predict a safe and efficacious dose for future clinical trials. At the time of study conduct, only scarce data from healthy volunteers (HV) and often no data from patients (PT) are available, hence, the optimal use of all available information is important.

We present two case studies combining accumulated information applying adaptive dose selection based on Bayesian decision criteria:

1) A MAD study in HV for a drug in the central nervous disease area using a Bayesian hierarchical model accounting for severity (mild, moderate, severe) of the adverse events from the accumulated data in the SAD and the ongoing MAD.

2) A combined MAD in HV and POC study in PT for a metabolic drug whereby the selection of dose for subsequent cohorts is based on the accumulated safety, PK and efficacy data incorporated in multiple Bayesian models.

We discuss the operating characteristics of these study designs based on simulation studies. Our results show that, compared to more classical designs, the use of Bayesian statistics can lead to an increased precision of the parameter estimates and a reduction in the proportion of patients overdosed, cohort size and number of subjects per trial. However, a smaller study size may increase the study duration.

A Bayesian nonlinear mixed effects approach to analyzing data from rectal barostat experiments

Roland Fisch, Rossella Belleli
Novartis Pharma AG, Switzerland; roland.fisch@novartis.com

Keywords: Barostat experiments, mixed effects modeling, Bayesian Inference

Rectal barostat experiments are used in Gastroenterology, in order to assess visceral hypersensitivity in Irritable Bowel Syndrome (IBS). These experiments can also be used to provide evidence for the efficacy of an experimental drug intended to treat this symptomatology in IBS patients. Within each subject, the rectal volume is measured as a function of a pressure at pre-defined levels, under a variety of experimental conditions (e.g. treated and placebo). The resulting repeated measures data structure can be modeled with a Bayesian nonlinear mixed effects model. One distinct advantage of the Bayesian approach is that it allows deriving inference on any function of the parameters, without relying on asymptotics; specifically one can derive predictive distributions, with fully taking uncertainty and variability into account. These features are essential in the context of Proof of Concept (PoC) trials in early clinical development, where sample sizes are small, and PoC success criteria are often based on quantiles of predictive distributions.

Observational Studies

H.-U. Burger and A. Berghold

Wednesday 14. September, 11:00 – 12:30, Lecture Hall KOL-F101

Pharmacoepidemiological databases: Strengths, limitations, methodological challenges

Iris Pigeot, Sigrid Behr
Universität Bremen, Bremer Institut für Präventionsforschung und Sozialmedizin (BIPS), Germany; pigeot@bips.uni-bremen.de

Keywords: Pharmacoepidemiological databases

Claims databases are important sources for pharmacoepidemiological studies e.g. to assess drug safety after market approval or to conduct health outcomes research. However, since these databases are built on data collected for administrative purposes we have to cope with several limitations.

As an example we consider the German pharmacoepidemiological research database obtained from statutory health insurances (SHIs) which is used to generate safety signals, to monitor prescribed drugs and to describe drug utilization. Currently, the database covers data on drug utilization and hospitalizations of about 14 million SHI members. It allows for a population-based approach and reflects daily practice including off-label use of drugs. Independent recording of exposure and outcome data prevents reporting bias. Confounder information on medical conditions, concomitant medications and socio-demographic variables can be obtained from the database, while the assessment of confounders related to lifestyle requires supplementary data collection. Two-phase designs may overcome this limitation. For illustrative purposes, we present a nested case-control study to estimate the bleeding risk in patients treated with phenprocoumon. We will also introduce structural marginal models and instrumental variables as alternative methods to deal with limited confounder information.

Instrumental variables and propensity score methods can be applied to eliminate confounding by indication which is a crucial issue in many pharmacoepidemiological studies.

Another problem in database studies is misclassification of the outcome variable where we will show by an example that especially a high specificity in the assessment of the outcome variable should be strived for e.g. by restriction to hospital discharge diagnoses.

On observational studies for addressing safety concerns

Bharat T Thakrar, Noah J Robinson
Roche, Switzerland; bharat.thakrar@roche.com

Keywords: Observational studies, safety events

Patient safety is of paramount importance in drug development. Therefore, both the Pharmaceutical Industries and Regulatory Agencies invest considerable resources to ensure the safety of patients. Nonetheless, many safety concerns emerge only after a drug is in the market place and with many patients exposed.

Whilst novel statistical methods have helped in identifying safety concerns early in the process, due to many reasons (eg rarity and delayed onset), however, some safety concerns will continue to emerge only after approval and often considerably so.

Observational data, coupled with sophisticated statistical methodology is perhaps a unique combination in aiding the quantification and contextualization of the safety concerns after approval, and therefore in the environment of the real world use of the drug. Many recent approvals have been subject to a follow on observational studies for safety concerns. However, there are some factors limiting such investigations and, in some cases, may not even be appropriate.

In this presentation we shall discuss two examples of application of novel statistical methodology to quantify safety concerns using observational data. In the first example we shall present results from 3 observational studies investigating the association between Tamiflu and neuropsychatric events. In the second example we present the case of Epoetin beta and thromboembolic events. Here we show that despite investigating several methodologies, the available data would not provide any robust estimate which would significantly improve the precision of the current estimate of the risk.

Estimation of survival probabilities for binary time-dependent covariates

Martina Mittlboeck[1], Ulrike Pötschger[2], Harald Heinzl[1]
[1]Medical University of Vienna, Austria; [2]Childrens Cancer Research Institute, Vienna, Austria; martina.mittlboeck@meduniwien.ac.at

Keywords: Survival curves, landmark method, Kaplan-Meier estimator

The time-dependent nature of covariates in time-to-event data has to be properly addressed in the statistical analysis. In the case of a binary time-dependent covariate effect estimation and testing can be performed within the Cox regression model using standard software. However, graphical representation of survival curves is still not satisfactorily solved. Feuer et al. [1] suggested to compare survival probabilities of patients whose covariate has non-reversibly changed over time to the survival probabilities of all patients. This approach results in different estimates, if the time-dependent covariate has an effect, but does not provide representative survival curves. Simon and Makuch [2] have suggested the Landmark-method, where a predefined time interval will be skipped and the starting point for survival curves is chosen so that already in a sufficient number of patients a change in the time-dependent covariate is observed. This approach results in overestimation of survival probabilities and may be further biased by the ignorance of the initial period. We introduce a novel estimate for survival probabilities in case of an (even reversible) binary time-dependent covariate based on the Landmark-method. Properties of the new method are discussed and illustrated with the Stanford heart transplant data set.

References

[1] E.J. Feuer, B.F. Hankey, J.J. Gaynor, M.N. Wesley, S.G. Baker, J.S. Meyer (1992). Graphical representation of survival curves associated with a binary non-reversible time dependent covariate, StatMed, 11:455-474.

[2] R. Simon, R.W. Makuch (1984). A non-parametric graphical representation of the relationship between survival and the occurrence of an event: application to responder versus non-responder bias, StatMed, 3:35-44.

Genetics and Biomarker (1)

H. Binder

Wednesday 14. September, 11:00 – 12:30, Lecture Hall KOL-F104

Temporal Activation Profiles of Gene Groups

André König[1,2], Jörg Rahnenführer[1]
[1]TU Dortmund, Germany; [2]DFG: Graduate School 'Statistische Modellbildung';
akoenig@statistik.tu-dortmund.de

Keywords: Gene expression, enrichment, short time series

Time series of gene expression measurements are used to study various biological processes, e.g. immune response, cancer progression or embryonic development. Due to the costs of microarray experiments in many research projects only a few times are analyzed. Moreover, due to limited biological material or money, often none or just few replicates are considered.

We offer an approach to identify activation profiles for predefined gene sets. The use of gene sets defined by Gene Ontology (GO) or KEGG pathways is established. In distinction from other used algorithms we do neither cluster genes according to predefined expression profiles nor discriminate groups with general temporal shift.

Our bottom-up algorithm first compares the expression values at every time point with a reference distribution and identifies differentially expressed genes separately for all times. In a second step we obtain the groups whose genes are overrepresented among the differentially expressed genes per time point. This yields a characteristic and well interpretable time profile for all considered gene groups.

Some activation profiles show discontinuous activation over the observed time even after correction for multiple testing. This is for most groups implausible and hence hard to interpret. We thus developed smoothing algorithms that generate continuous activation profiles.

We present a comprehensive simulation study and an application of our methods to a mouse ovary development dataset. The simulations help in choosing optimal parameter constellations, and the results on the mouse data demonstrate the merits of our approach in terms of biologically plausible results.

SVM-based models for cancer classification and survival prediction with high-dimensional genomic predictors

Eugen Rempel, Jörg Rahnenführer
TU Dortmund, Germany; rempel@statistik.tu-dortmund.de

Keywords: Survival analysis, classification, SVM-based algorithms

The modeling of survival data with high-dimensional genomic data as covariates is difficult due to the large number of potential predictors. Especially related to cancer studies many such data sets are now becoming available. The main goals are the construction of prognostic models, classification of patient subgroups, and the identification of the crucial genes for these tasks. Especially, both survival times and binary responses are of interest as target variables for the prediction models.

Support vector machines (SVMs) have proven their usefulness in various areas of statistics and machine learning, conceptually (non-parametric approach), computationally (reducing to a convex program which can be solved efficiently), theoretically, and empirically. For our purpose, SVMs have been successfully applied in classification problems. However, for predicting survival times, the modeling challenge with high-dimensional data calls for novel tools. Survival SVMs are a promising extension that can deal with right-censored data.

We analyze and compare the potential of SVMs and survival SVMs in classification and prediction tasks for cancer cohorts. We model survival based on gene expression data, both with categories and survival times as response variables. For categorical data, evaluations are mainly based ion misclassification rates. For the survival SVMs, we evaluate the estimated models using the concordance index, the log-rank test and the integrated Brier-Score. On three large breast cancer data sets we perform a systematic comparison of the results from the SVM-based methods with results obtained with competing established methods. Software will be made available as an R package.

Prediction of progression and therapy response for cancer patients: Are high-dimensional genomic data a blessing or a curse?

Jörg Rahnenführer
TU Dortmund University, Germany; rahnenfuehrer@statistik.tu-dortmund.de

Keywords: Gene expression, classification, prediction, cancer

The collection of high-dimensional genomic data from cancer patients has become a fashionable tool for introducing new approaches in cancer research. The promise of such data, particularly gene expression data that measure gene activity and SNP data that measure variation in DNA sequence, is a better individualized therapy. However, identifying relevant single variables or combinations of variables remains challenging. Such high-dimensional data suffer from the inherent problem that any pair of individuals has too many "genomic differences" to be comparable, and many published approaches overestimate their prediction quality. In this talk, we first explain general principles how genome-wide expression data can be used to identify prognostic signatures for discriminating cancer patient subgroups. The challenge is to find a compromise between stable models with few genes that might ignore relevant information and comprehensive models with many genes that might lead to unreliable predictions. After an overview we focus on three own research aspects:

(i) How do we find the really relevant and reliable genes? A gene whose expression distribution has bimodal shape allows a natural division of patients into high and low expression group. We identify such genes and combine them to classifiers.

(ii) How do we combine the information from different patient cohorts with different characteristics? We discuss how classifiers constructed on one patient cohort can be validated on other cohorts.

(iii) How can we understand the biological differences of patient subgroups with different clinical prognosis? We develop methods for estimating differential gene regulatory networks from genome-wide data.

Analysis of small-sample gene expression and gene interactions via Bayesian hierarchical models

Mauro Gasparini[1], Anja Rockstroh[2], Christine Wells[3], Derek Kennedy[4]
[1]Politecnico di Torino, Italy; [2]Queensland University of Technology, Australia; [3]University of Queensland, Australia; [4]Griffith University, Australia; gasparini@calvino.polito.it

Keywords: qRT-PCR data, mixed models

The G3BP2 locus has been discovered and studied in [1]. In a recent set of quantitative reverse transcription polymerase chain reaction (qRT-PCR) experiments, we have been trying to study the G3BP2 transcripts and their relationships with their genomic neighborhood. The gene expression experiments we set up involved normalizing housekeeping genes, several different transcripts, several conditions of interest (different cell lines, tumor versus normal tissue), technical replicates, missing data and small sample sizes in all resulting cells. The scenario is suitable for the analysis via Bayesian hierarchical models using as response variable the continuous interpolation of cycles in the qRT-PCR, which is naturally taken to be lognormal. The construction of few normal hierarchical models will be discussed, critical points will be illustrated and a comparison with standard methods will be made.

Currently, the state of the art in analysing qRT-PCR data is based on ad hoc user-friendly software (such as REST) or on linear mixed models. We will show how the alternative analysis based on statistical Bayesian networks can be flexible enough to address the important issue of estimating gene-gene interactions.

References

[1] D. Kennedy, J. French, E. Guitard, K. Ru, B. Tocque and J. Mattick (2001). Characterization of G3BPs: tissue specific expression, chromosomal localisation and rasGAP(120) binding studies. J. Cell Biochem. **84**:173-187.

A Bayesian network approach for pathway analysis using the gene ontology database

Ronja Foraita, Janine Karl, Annika Leseberg, Frauke Günther
Bremen Institute for Prevention Research and Social Medicine (BIPS), Germany;
foraita@bips.uni-bremen.de

Keywords: Gene ontology, graphical models, pathway, principal components

Pathway analyses incorporate prior biological knowledge and focus on the association of disease with combined information of genetic variants in the same pathway. We propose a Bayesian network framework to discover pathways that are enriched with disease-associated SNPs. Bayesian networks rely on the concept of conditional independence and are able to detect direct and indirect influences, e.g. pathways that are mediated through other pathways, as well as interactions.

In our framework, we map SNPs to genes within a pathway given the external pathway knowledge from the gene ontology database (http://www.geneontology.org/). In order to reduce the large dimensionality of a pathway, we apply principal component analysis to build up a pathway score by using the first principal component for each pathway. This score is used as surrogate for the genetic variability within the pathway and is finally included in the Bayesian network analysis. Bayesian networks are learned using the tabu search algorithm with BIC as model selection criteria.

In a simulation study, we investigate the proposed method using data with realistic linkage disequilibrium structure, different effect sizes for "causal" SNPs and different pathway-disease dependency structures.

First results for the simple scenario with just direct disease-pathway associations show that only 0.6% false disease-pathway associations are found. On the contrary, true dependencies are only detected in about 50%, where small pathways have much higher detection rates than large pathways.

However, the approach is computational efficient and helps to derive parsimonious models that can be visualized in a graph.

Miscellaneous

N. Neumann

Wednesday 14. September 11:00 – 12:30, Lecture Hall KOL-F121

Using penalized splines in extended Cox-type additive hazard regression to flexibly estimate the effect of time-varying serum uric acid on risk of cancer incidence in a large population-based cohort study

Alexander M. Strasak[1], Stefan Lang[2], Thomas Kneib[3], Larry J Brant[4], Jochen Klenk[5], Wolfgang Hilbe[6], Willi Oberaigner[7], Elfriede Ruttmann[6], Lalit Kaltenbach[6], Hans Concin[8], Guenter Diem[8], Karl P Pfeiffer[6], Hanno Ulmer[6]

[1]Hoffmann-La Roche, Switzerland; [2]University of Innsbruck, Austria; [3]University of Oldenburg, Germany; [4]National Institute on Aging, USA; [5]Ulm University, Germany; [6]Innsbruck Medical University, Austria; [7]Cancer Registry of Tyrol, Austria; [8]Agency for Preventive and Social Medicine, Austria; alexander.strasak@roche.com

Keywords: Cancer incidence, Extended Cox-type additive hazard regression, Penalized Splines, Risk Factor, Serum Uric Acid

We sought to investigate the effect of serum uric acid (SUA) levels on risk of cancer incidence in men and to flexibly determine the shape of this association by using a novel analytical approach. A population-based cohort of 78,850 Austrian men who received 264,347 serial SUA measurements was prospectively followed-up for a median of 12.4 years. Data were collected between 1985 and 2003. Penalized splines (P-splines) in extended Cox-type additive hazard regression were used to flexibly model the association between SUA, as a time-dependent covariate, and risk of overall and site-specific cancer incidence and to calculate adjusted hazard ratios with their 95% confidence intervals. During follow-up 5189 incident cancers were observed. Restricted maximum-likelihood optimizing P-spline models revealed a moderately J-shaped effect of SUA on risk of overall cancer incidence, with statistically significantly increased hazard ratios in the upper third of the SUA distribution. Increased SUA (>=8.00 mg/dL) further significantly increased risk for several site-specific malignancies, with P-spline analyses providing detailed insight about the shape of the association with these outcomes. Our study is the first to demonstrate a dose-response association between SUA and cancer incidence in men, simultaneously reporting on the usefulness of a novel methodological framework in epidemiologic research.

Relation between incidence, prevalence and mortality in terms of a stochastic differential equation – formulation and application to renal replacement therapy

Ralph Brinks[1], Sandra Landwehr[1], Andrea Icks[1], Michael Koch[2], Guido Giani[1]
[1]German Diabetes Center, Duesseldorf, Germany; [2]Center of Nephrology, Mettman, Germany;
rbrinks@ddz.uni-duesseldorf.de

Keywords: Stochastic differential equation, incidence, prevalence, mortality, renal replacement therapy

Prevalence and incidence are important epidemiological characteristics of a disease. In [1] a simple model formalizing the relationship between incidence, mortality and prevalence was set up. It consists of three health states 'Normal', 'Diseased' and 'Dead'. Transitions along the paths between the three states are described as a set of ordinary differential equations (ODEs) [2]. We improved the ODE approach [3], but in diseases with low prevalence, it is advantageous to translate the deterministic ODE into a stochastic differential equation (SDE) framework [4].

Here we formulate a new two-dimensional, linear SDE relating age-specific prevalence, incidence- and mortality-density and apply it to a total population survey of patients with renal replacement therapy. Age- and sex-specific prevalence, incidence- and mortality-densities have been accounted in 2001-2010. Incidence- and mortality-densities have been used as coefficients of the SDE. The 1000 solution paths calculated by the Euler-Maruyama method are used to determine the prevalence in the age groups 40-49, 50-59, 60-69, 70-79 and 80+ years. The results are 77+/-52, 166+/-92, 332+/-134, 515+/-198, 579+/-337 (mean+/-SD, in units 1.0e-5). These values fit to the corresponding observed values 82+/-18, 184+/-31, 239+/-35, 423+/-56, 522+/-120.

References

[1] C. J. L. Murray, A. D. Lopez (1994). Quantifying disability: data, methods and results. BulletinWHO, **72**(3).

[2] J. J. Barendregt et al. (2003). A generic model for the assessment of disease epidemiology. PopHealthMetrics, 1(4).

[3] R. Brinks et al. (2010). Deriving the incidence of an incurable disease from its prevalence. Abstract, 55. GMDS annual meeting.

[4] H. Trottier (2001). Deterministic modeling of infectious diseases. InternetJInfec Dis 1(2).

Statistical modeling of flexible pooling in unbalanced experiments for gene expression data

Mihaela Pricop-Jeckstadt, Norbert Reinsch, Henrik Rudolf
Leibniz Institute for Farm Animal Biology, Germany; pricop@fbn-dummerstorf.de

Keywords: Gene expression, pooling, unbalanced design

In this talk we propose a flexible design and a linear mixed model for mRNA pooling with varying numbers of individuals and of arrays per pool in microarray experiments. mRNA pooling with varying pool sizes is favored in experimental design, for example if comparisons between subjects in a family of animals are performed where it is impossible to quantify the subgroups in advance. Enforcing balance between the pool structure of every treatment level we widen the options introduced in [1] for the design of pooled experiments and we show the unbiasedness of testing for differences in gene expression across treatments. To overcome the problem of the variance heterogeneity and to account for correlations due to repeated mixtures, replications and pooling we deploy a new random effect model and a REML approach for a more feasible testing of the chosen contrasts (see [2]). Finally, the prior results for one-colour arrays are extended to two-colour arrays.

References

[1] W. Zhang, A. Carriquiry, D. Nettleton, J. C. M. Dekkers (2007). Pooling mRNA in microarray experiments and its effect on power, Bioinformatics, **23**:1217-1224.

[2] G. K. Kenward, J. H. Roger (1997). Small sample inference for fixed effects from restricted maximum likelihood, Biometrics, **53**:983-997.

Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions

Joseph O. Ogutu, Torben Schulz-Streeck, Hans-Peter Piepho
University of Hohenheim, Germany; jogutu2007@gmail.com

Keywords: Genomic selection, lasso, elastic net, adaptive lasso, adaptive elastic net, ridge regression

Genomic selection (GS) is emerging as an efficient and cost-effective method for estimating breeding values using molecular markers distributed over the entire genome. GS involves estimating the simultaneous effects of all genes or chromosomal segments and combining the estimates to predict the total genomic breeding value (GEBV). Accurate prediction of GEBVs is a central and recurring challenge in plant and animal breeding. The existence of a bewildering array of approaches for predicting breeding values using markers underscores the importance of identifying approaches able to efficiently and accurately predict breeding values. Here, we comparatively evaluate the predictive performance of six regularized linear regression methods— ridge regression, ridge regression BLUP, lasso, adaptive lasso, elastic net and adaptive elastic net— for predicting GEBV using dense SNP markers. The four lasso penalty-based methods can simultaneously select relevant predictive markers and optimally estimate their effects.

We predicted GEBVs for a quantitative trait using a dataset on 3000 progenies of 20 sires and 200 dams and an accompanying genome consisting of five chromosomes with 9990 biallelic SNP-marker loci simulated for the QTL-MAS 2011 workshop using all the six methods. The regression models were trained with a subset of 2000 phenotyped and genotyped individuals and used to predict GEBVs for the remaining 1000 progenies without phenotypes. Predictive accuracy, the Pearson correlation between GEBVs and the simulated values using fivefold cross-validation, was higher for the elastic net, lasso and adaptive lasso than for the two ridge regression models, indicating that the lasso-type methods are promising for genomic selection.

Clinical Enpoints in Oncology (1)

H.-U. Burger and A. Berghold

Thursday 15. September, 9:00 – 10:30, Lecture Hall HAH-E-3

Biomarkers and Surrogate Endpoints In Clinical Trials

Thomas R. Fleming
University of Washington, Seattle, United States of America; fleming@seattlebiostat.com

Keywords: Surrogate, benefit-to-risk profile

One of the most important considerations in designing and evaluating clinical trials is the choice of outcome measures. Clinical endpoints are direct measures of how patients feel, function and survive. Biomarkers, including physical signs of disease, laboratory measures and radiological tests, often are considered as substitutes or "surrogates" for clinical endpoints. We discuss the definitions of clinical and surrogate endpoints, and provide examples from recent clinical trials. We provide insight into why biomarkers may fail to provide reliable evidence about the benefit-to-risk profile of interventions. We also discuss the considerations in evaluating the evidence on the validity and reliability of biomarkers prior to using them as primary endpoints in clinical trials.

Choice of endpoints in cancer clinical trials: OS, DFS, PFS?

Tomasz Burzykowski
Hasselt University, Belgium; tomasz.burzykowski@uhasselt.be

Keywords: Surrogate, endpoints

Traditionally, overall survival (OS) is considered as the endpoint of interest in cancer clinical trials. However, for many reasons, like a long duration of the follow-up or treatment cross-overs, OS can be difficult to use for the evaluation of treatment's efficacy. Thus, in practice, the use of disease-free survival (DFS) or progression-free survival (PFS) is often considered instead. In this presentation, we will briefly review the arguments for and against the use of OS, DFS, and PFS as endpoints in cancer clinical trials. The arguments will be illustrated by using real-life examples of trials. In particular, formal analyses of the validity of DFS/PFS as surrogates for OS will be reviewed.

Challenges with the Overall Survival Endpoint in Oncology Trials and Investigation of some Alternatives

Colin James Neate[1], Barbara Tong[2], Hans Ulrich Burger[3]
[1]Roche Products Ltd, United Kingdom; [2]Genentech (a member of Roche Group), USA;
[3]Hoffmann-La Roche Switzerland; colin.neate@roche.com

Keywords: Oncology, overall survival

In oncology, the two typical endpoints for assessing treatment benefit are progression-free survival (PFS) and overall survival (OS). For a considerable time there have been ongoing debates firstly as to whether PFS is an endpoint with its own clinical relevance and secondly whether PFS is a surrogate endpoint for OS in settings where OS is seen as the ultimate endpoint for demonstrating clinical benefit. Additionally, PFS is still viewed by many as an endpoint with intrinsic methodological problems and subjective elements, especially in open label trials, whereas OS seems straightforward and objectively assessable endpoint.

With improvements in the last decade in effective anticancer therapies, survival is being prolonged and many indications have started to move from life-threatening towards chronic status with multiple lines of therapies possible. This success comes not without consequences for the use of OS as a suitable endpoint. Not only are duration and size of survival studies both significantly increased due to a smaller number of events expected, an even bigger challenge comes due to long survival times and multiple uncontrolled factors coming in, such as second line therapies, which result in OS tending to start to show non-proportional hazard behaviour, leading to diminishing OS effects over time and intrinsic difficulties in defining one overall survival endpoint.

The presentation will highlight some of these difficulties in assessing OS on the basis of a number of trials. It will further introduce truncated OS endpoints as alternative endpoints to PFS and OS and will investigate the relationship between PFS, OS and truncated OS.

Validating surrogate endpoints in breast and colon cancer: A systematic literature review

Christoph Schürmann, Ralf Bender, Thomas Kaiser, Susanne Ringsdorf, Elke Vervölgyi, Volker Vervölgyi, Beate Wieseler
Institute for Quality and Efficiency in Health Care (IQWiG), Germany;
christoph.schuermann@iqwig.de

Keywords: Surrogate endpoints, systematic review, validation

Surrogate endpoints like disease free survival or progression free survival are widely used in oncology as substitutes for overall survival, but validating surrogate endpoints is difficult and not always done properly. We systematically searched the literature for studies validating surrogate endpoints in breast cancer and colon cancer. We analysed both statistical validity and reliability of the results. Statistical validity can be derived from appropriate meta-analyses, usually modelling the correlation of effects on surrogate endpoints and true endpoints given by individual studies. Reliability was rated by six quality criteria: Acceptable method of validation, robustness of the results, inclusion of all suitable studies, specificity of indication, specificity of intervention, consistent definitions of endpoints. For a surrogate endpoint to be valid a high correlation with the true endpoint in highly reliable meta-analyses is required.

We found 6 and 15 eligible studies on validation of surrogate endpoints in breast cancer and colon cancer, respectively. A high correlation was observed in some special indications only, e.g. early colon cancer. However, the reliability of the results was low for each included study: In most cases, more than two of our criteria were either not fulfilled or remained unclear.

In conclusion, it still has to be questioned, whether and/or for which specific settings surrogate endpoints like progression free survival are valid substitutes for overall survival in breast cancer or colon cancer. However, our analysis shows that adequate validation analyses could principally be conducted.

Reference

IQWiG, Report A10-05: Validity of surrogate endpoints in oncology, 2011, https://www.iqwig.de/download/A10-05_Executive_Summary_Surrogate_endpoints_in_oncology.pdf

Statistics in practice (2)

AG Nachwuchs, AG Weiterbildung, S. Roll

Thursday 15. September, 9:00 – 10:30, Lecture Hall HAH-E-11

This is a special session for “statisticians in action” which should help providing an overview of today’s environment in which statisticians are working. Both elements should help to prepare and integrate the younger generation into the biometry community, a major task to secure our future. This talk builds on the introduction to meta-analysis provided in Statistics in Practice (1) and provides further details on meta-analysis based on many examples. The topic will be further developed in the tutorial session Statistics in practice (3).

Speaker: Salanti, G

Genetics and Biomarker (2)

P. Schlattmann

Thursday 15. September, 9:00 – 10:30, Lecture Hall HAH-F-1

Strategies for integrated analysis of genome wide measurements in risk prediction models

Stefanie Hieke[1,2], Thomas Hielscher[3], Richard F. Schlenk[4], Martin Schumacher[1], Axel Benner[3], Lars Bullinger[4], Harald Binder[1,2]

[1]Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Freiburg, Germany; [2]Freiburg Center for Data Analysis and Modeling, University Freiburg, Freiburg, Germany.; [3]Division of Biostatistics, German Cancer Research Center, Heidelberg, Germany.; [4]Department of Internal Medicine III, University Hospital of Ulm, Ulm, Germany; hieke@imbi.uni-freiburg.de

Keywords: Molecular data, risk prediction models, time-to-event

In recent years, it has become feasible to obtain genome wide data at various molecular levels in parallel. Integration of these data in risk prediction models with regard to clinical endpoints could considerably improve therapy management for future patients. We systematically investigate statistical strategies to connect several molecular sources with partial overlap in the biological samples. One approach for taking biological hierarchies into account, is first to consider only one molecular source and keep the information from this source fixed in the model when incorporating the second source. Another approach is using the information from the first source for a pre-selection of the molecular entities in the second source. A further approach is to connect the information from two molecular sources in parallel, i.e. merge these sources depending on the partial overlap in the biological samples. All of these strategies can potentially provide improved outcome prediction for patients, but biological interpretation of the resulting risk prediction models will be different. We will illustrate these basic strategies in an application to survival data from acute myeloid leukemia patients. Specifically, we consider microarray-based gene expression profiling, single nucleotide polymorphisms microarrays and microRNA. Each of these can be considered as a first or second source statistically, but we will highlight specific combinations that also correspond to relevant biological questions. Thus, statistical procedures are adapted for connecting the different molecular sources in a way that is related to the underlying biology, resulting in a potential basis for molecular therapy management.

Improving SNP selection in genome-wide association studies with CAT and CAR scores

Verena Zuber, Korbinian Strimmer
University of Leipzig, Germany; vzuber@uni-leipzig.de

Keywords: Variable selection, SNP ranking, genome-wide association studies

Identification of single nucleotide polymorphisms (SNPs) connected to a trait of interest is challenging due to the high-dimensionality of the data and the correlation among SNPs. So far, standard approaches are mostly univariate and thus disregard the correlation among SNPs.

Recently, we have introduced the CAT [1] and the CAR [2] score as an efficient means to rank covariables in categorical and continuous prediction problems. By construction, CAT and CAR scores explicitly take account of correlation among predictors. In [2] we have shown that our approach often outperforms regression approaches such as elastic net and boosting.

Here, we discuss the theoretical foundations of the CAT/CAR approach and demonstrate its performance in the problem of SNP selection. Specifically, we report results from a comparison study analyzing high-dimensional SNP data provided by the GAW 17 consortium [3] and show that CAT/CAR scores lead to an enrichment of true SNPs in the ranking compared to standard analysis.

References

[1] V. Zuber and K. Strimmer (2009). Gene ranking and biomarker discovery under correlation. Bioinformatics, **25**:2700-2707
[2] V. Zuber and K. Strimmer (2010). Variable importance and model selection by decorrelation. http://arxiv.org/abs/1007.5516
[3] Genetic Analysis Workshop 17. 2011. http://www.gaworkshop.org/gaw17/

Testing and Genetic Model Selection in Genome-Wide Association Studies

Christina Loley[1,2], Inke R. König[1], Ludwig Hothorn[3], Andreas Ziegler[1]
[1]Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Germany; [2]Medizinische Klinik II, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Germany; [3]Institut für Biostatistik, Leibniz Universität Hannover, Hannover, Germany; christina.loley@imbs.uni-luebeck.de

Keywords: Delta method, generalized linear model, genetic association, genome-wide association

In genome-wide association (GWA) studies, trend test statistics based on the additive genetic model are the standard approach to test for association. But this decision can substantially reduce power if the true genetic model is a recessive or dominant one. MAX tests have been proposed to simultaneously test for additive, recessive, and dominant genetic models and to select the most likely one out of these three. P-values are estimated using permutation, although these are computationally demanding and therefore not feasible in GWA studies. To circumvent this drawback, the analytical asymptotic distribution of the MAX test statistic needs to be derived. In this contribution, we show that the asymptotic distribution of the MAX test can be approximated by an asymptotic multivariate normal distribution. Our approach is based on a generalized linear regression model with two dummy variables, and the parameters are transformed by the delta method to model specific inheritance patterns. The approach naturally allows adjustments for environmental factors or population stratification. In a simulation study, we demonstrate the validity of the method and compare its performance to existing tests. We illustrate its application by re-analyzing GWA data on Crohn's disease.

Automated investigation of genotype calling using angles and tests for unimodality

Arne Schillert, Michael Pfützenreuter, Andreas Ziegler
Universität zu Lübeck, Germany; arne.schillert@imbs.uni-luebeck.de

Keywords: Quality control, microarrays, genotype data

Incorrect genotype assignments during genotype can invalidate genome-wide association studies. Recently, we developed an algorithm to automatically detect failed genotype clusterings [1]. To this end, we plotted the contrast of the signal intensities against its sum and counted the number of points which were too close to a cluster for a different genotype. In this work, we substantially improve the algorithm by adding two features. First, we compute the angle of the first principal component because failed clusterings often result in tilted clusters. Second, we investigated tests for unimodality [2,3] to identify genotype clusters which actually consist of disjunct sub-groups as this hints at failed clusterings as well. To enable the automated analysis of all genotypes of a recent microarray experiment we transform the data massively. In particular, we convert the usually very large intensity file into many smaller chunks in DatABEL's data format [4]. The genotypes are stored as GenABEL objects. This allows fast access to different sets of SNPs and rapid evaluation. When run on a computer cluster with 50 nodes, the analysis of 3.000 individuals with 700.000 SNPs can be done over the weekend.

References

[1] Schillert et al. (2009), BMC Proc **3**:58.

[2] Larkin (1979), Behav Res Meth Instr **11**:467-468.

[3] Engelman and Hartigan (1969), J Am Stat Assoc **64**:1647-1648.

[4] Aulchenko (2007), Bioinformatics **23**:1294-1296.

Automated Allele Calling

Peter Schlattmann[1], Maryna Verba[2]
[1]University Hospital Jena, Germany; [2]Charite Universitätsmedizin Berlin;
peter.schlattmann@mti.uni-jena.de

Keywords: genome wide association study, genotype calling, finite mixture model

Visual inspection of the clusterplot for each trait-associated SNP is still the recommended strategy for ascertaining the accuracy of the genotyping (Pearson and Manolio, 2008). This requires two independent reviewers and is time consuming for genome wide association studies. Automated procedures are necessary with the advent of large-scale genotyping, which assays at least hundreds of thousands of SNPs.

Here a model based calling algorithm which can be performed unsupervised based on finite mixture models is presented. In contrast to the approach by Tao et al (2007) the number of mixture components does not need to be fixed in advance. This is achieved by applying the VEM-algorithm (Schlattmann, 2009) for bivariate normal distributions.

The validity of the method is investigated in a large scale simulation study.

References

[1] T. A. Pearson and T. A. Manolio (2008). How to interpret a genome-wide association study. Journal of the American Medical Association, **299**:1335-1344.

[2] P. Schlattmann (2009). Medical Applications of Finite Mixture Models. Berlin: Springer.

[3] Y. Y. Teo, M. Inouye, K. S. Small, R. Gwilliam, P. Deloukas, D. P. Kwiatkowski and T. G. Clark (2007). A genotype calling algorithm for the Illumina BeadArray platform. Bioinformatics, **23**:2741-2746.

Modeling and Simulation

M. Vandemeulebroecke

Thursday 15. September, 9:00 – 10:30, Lecture Hall HAH-E-10

Capture-recapture with boosting: An application to estimating the comprehensiveness of literature searches for systematic reviews

Gerta Rücker, Veronika Reiser, Edith Motschall, Harald Binder, Joerg J. Meerpohl, Gerd Antes, Martin Schumacher

University Medical Center Freiburg, Germany; ruecker@imbi.uni-freiburg.de

Keywords: Capture-recapture, Componentwise boosting, Meta-analysis, Model selection, Systematic review

Background

Capture-recapture methods were proposed to evaluate the comprehensiveness of systematic literature searches.

Objective

We investigated the statistical feasibility of capture-recapture techniques with model selection for estimating the number of missing references in literature searches.

Methods

For two systematic reviews in gastroenterology and hematology, we first compared manually selected Poisson regression models that differ with respect to included interactions. Secondly, we performed selection via componentwise boosting, which provides automatic variable selection. The proposed boosting technique is a regularized, stepwise procedure allowing to distinguish between mandatory and optional variables. Results from all models were compared based on AIC and BIC.

Results

For the first example, the best manually selected model suggested a number of 82 missing articles (95% CI [52;128]), while the boosting technique provided 127 (95% CI [86;186]) missing articles. For the second example, 140 (95% CI [116;168]) missing articles were estimated for the manually selected and 188 (95% CI [159;223] for the automatically selected model.

Conclusion

Capture-recapture analysis requires the selection of an appropriate model. Due to problems of variable selection and overfitting, manual model selection yielded large estimates, varying markedly, with broad confidence intervals. By contrast, boosting was robust against overfitting and automatically created an appropriate model for inference.

References

[1] G. Rücker, V. Reiser, E. Motschall, H. Binder, J.J. Meerpohl, G. Antes and M. Schumacher (2011). Boosting qualifies capture-recapture methods for estimating comprehensiveness of literature search for systematic reviews, Journal of Clinical Epidemiology, **in press**.

Investigations on non-inferiority: the case of the FDA draft guidance on treatments for Nosocomial Pneumonia

Joachim Roehmel[1], **Meinhard Kieser**[2]
[1]Bremen Institute for Prevention Research and Social Medicine, Germany; [2]Institut für Medizinische Biometrie und Informatik, Heidelberg, Germany; joachim.roehmel@gmx.de

Keywords: Non-inferiority margin, nosocomial infections, type I error, sample size

In a recently published FDA draft Guidance to Industry for treatments for nosocomial pneumonia [1] clinicaltudies are required with a mortality endpoint. According to historical data with effective treatment the 28 day in hospital mortality is often between 15 and 30%. There is a large difference between mortality rates of effective treatment and inadequate delayed, or inappropriate therapies [1]. This large difference makes placebo controlled trials no longer ethically defendable, and design characteristics for non-inferiority studies need to be developed. Among these characteristics, the non-inferiority margin is of particular importance. After a careful review of historical data FDA has come to the conclusion that a non-inferiority margin of 10% is often reasonable. Should however the observed mortality proportion for the active control fall below 20% the 10% margin for the difference will be replaced by a 1.67 margin for the odds ratio. The draft guidance to industry does not make any recommendation regarding (i) the analytical strategy to be used and (ii) does not make suggestions about the sample size for such studies. A third problem emerges by the guideline's proposal to base the analysis strategy on the observed mortality rate for the control treatment. This is a crucial aspect, as, for example, the proposed non-inferiority margin is not continuous (not to mention differentiable) at the transit from a difference scale to the odds ratio scale, if the words of the guideline are taken literally. We will investigate these three problems regarding the appropriateness of non-inferiority margins, test statistics, type I error control, power and sample size.

Finally the margins are quite large for a mortality endpoint. In a similar situation the EMA guideline on the development of thrombolytic agents for the treatment of acute myocardial infarction [2] recommends an absolute margin of 1% for the difference in mortality proportions in addition to a relative margin of 14% for the relative risk and the more conservative of these two margins should be used in the analysis. Sample sizes close or above 3000 per group are then necessary to achieve 90% power. This is in contrast to a 10% (1.67) margin which results in sample sizes below 700 per group. The wide margin of 10% has another aspect which is not really desirable. Assume that a sponsor is planning _to develop a new drug for nosocomial pneumonia and selects as the active comparator a drug that has in the past in clinical trials shown a mortality proportion P above 20%. Between 400 and 600 patients have to be recruited in each of the two groups in order to achieve 90% power for the demonstration of non-inferiority under the assumption that the new treatment will also have the same true mortality rate Q. The bothering aspect here, however, is that there is also a quite good chance to demonstrate non-inferiority even if the new treatment's true mortality rate is Q=P+5%, i.e. in case of an absolute 5% increase in mortality over the active control. We will try to make some suggestions on how to reduce these chances.

References

[1] Guidance for Industry: Hospital-Acquired-Bacterial-Pneumonia-and-Ventilator-Associated Bacterial-Pneumonia: Developing Drugs for Treatment. November2010

[2] CPMP/EWP/967/01, 2003

Regularization and Model Selection with Ordinal Covariates

Jan Gertheiss, Veronika Stelz, Gerhard Tutz
LMU Munich, Germany; jan.gertheiss@stat.uni-muenchen.de

Keywords: Categorical predictors, clustering, model selection, regularization

The challenge in regression problems with categorical covariates is the high number of parameters involved. Common regularization methods like the Lasso (Tibshirani, 1996), which allow for selection of predictors, are typically designed for metric predictors. If independent variables are categorical, selection strategies should be based on modified penalties. For categorical predictor variables with many categories a useful strategy is for example to search for clusters of categories with similar effects. The objective is to reduce the set of categories to a smaller number of categories which form clusters. The effect of categories within one cluster is supposed to be the same, but the (conditional) expectation of the response will differ across clusters.

In the talk, L1-penalty based methods for factor selection and clustering of categories are presented and investigated. It is distinguished between nominally and ordinally scaled covariates, with the focus on the latter ones.

Clinical Endpoints in Oncology (2)

H.-U. Burger and A. Berghold

Thursday 15. September, 11:00 – 12:30, Lecture Hall HAH-E-3

Complex clinical endpoints are present in studies in hematopoietic cell transplantation

Claudia Schmoor[1], Jan Beyersmann[2]

[1]Clinical Trials Unit, University Medical Center Freiburg, Germany; [2]Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Germany; claudia.schmoor@uniklinik-freiburg.de

Keywords: Time-to-event analysis, competing risks, multi-state, clinical trial, oncology

After hematopoietic cell transplantation various risks are present in the course of disease. Patients can experience acute graft-versus-host-disease (GvHD), chronic GvHD, need to undergo immunosuppressive therapy, a relapse can occur, patients can die after relapse, or they can die without former relapse (non-relapse mortality, NRM). Sometimes, endpoints can be reasonably combined in a composite endpoint, as e.g. relapse and NRM are combined into disease-free survival (DFS). In this case, standard survival techniques, as Kaplan-Meier estimation of the DFS probability, can be applied.

Often, interest focuses on the single components, as e.g. the effects of different GvHD prophylaxis regimens on GvHD. In this case, death without former GvHD must be considered as a competing risk for the endpoint GvHD, which must be adequately accounted for in the statistical analysis. The probability of GvHD over time is then estimated by cumulative incidence functions. The statistical comparison of patient groups is performed by modeling event-specific hazard functions using Cox regression or subdistribution hazard functions using the Fine and Gray model.

These methods are inadequate when interest focuses on time under immunosuppressive therapy. In this case, the probabilities of survival under immunosuppression and of survival free of immunosuppression (adding up to the overall survival probability) are estimated with the Aalen-Johansen estimator. Patient groups are compared by modeling the transition hazards between the states alive and free of and alive under immunosuppression using Cox regression.

The methods are illustrated and discussed in a randomized trial on the effects of Anti-T-cell globulin as GvHD prophylaxis.

Composite Cancer Endpoints versus Combined Cancer Evidence

Knut M. Wittkowski
The Rockefeller University, United States of America; kmw@rockefeller.edu

Keywords: Longitudinal data, censoring, attribution, severity

When several (repeated, censored) events with different severity and attribution need to be considered (e.g., relapse, radiation or chemotherapy, death), many currently discussed strategies have well-known deficits. Either they are not comprehensive enough or they require unrealistic assumptions to be made, with often unforeseen consequences for validity. We demonstrate how recent generalizations of Gehan's test [1] can be adopted to repeated outcomes ("endpoints").

Initially, cumulative outcomes are created to reflect, e.g., that death counts more than either radiation or chemotherapy, while the latter count more than a relapse alone. For each outcome, one determines time to first event, subsequent time to second event , and so on. Death censors allother outcomes. Attribution (death for unrelated reasons) is handled as a separate grading dimension. The univariate partial orderings are combined in a hierarchical fashion. The resulting imperfect ordering combines time-to-first-event and survival with additional relevant information without the need for proportionality assumptions to be made.

Survival alone ignores lower grade events, while time-to-first event or overall number of events may dilute rare, but severe events. Moreover, standardized "composite endpoint" ratios have the same deficiencies as standardized mortality ratios. The proposed approach fulfills the CIH E9 criteria for "a predefined algorithm", yet is flexible enough to objectively incorporates information about graded, repeated, and censored events for a wide range of complex diseases.

References

[1] J.F. Morales JF, T. Song, A.D. Auerbach, K.M. Wittkowski (2008). Phenotyping genetic diseases using an extension of mu-scores for multivariate data. Stat Appl Genet Mol Biol **2**:7-19.

Allogeneic transplants in acute myeloid leukemia (AML) – A case study on the performance of an Andersen-Gill model for a time-dependent intervening event

Manuela Zucknick[1], Richard F. Schlenk[2], Axel Benner[1]
[1]German Cancer Research Center, Germany; [2]Ulm University Hospital, Germany;
m.zucknick@dkfz-heidelberg.de

Keywords: Survival analysis, time-dependent covariable, Andersen-Gill

In acute myeloid leukemia, high-risk patients have a dismal prognosis and allogeneic hematopoietic stem-cell transplantation (allo-HSCT) is thought to be the best treatment strategy. However, allo-HSCT cannot always be performed for all intended patients, and in addition, patients get transplanted at varying times after diagnosis.

In this situation, the effect of allogeneic HSCT on survival can be estimated in an Andersen-Gill model with allo-HSCT as a time-dependent intervening event, assuming a constant treatment effect for all patients from the time of transplantation onwards and no systematic differences between patients with and without HSCT other than through factors included in the model.

We demonstrate in a systematic simulation study how the Andersen-Gill model deals with potential model misspecification. Simulations are based on data from the AMLSG HD98A study (Schlenk et al, JCO 2010), where allogeneic HSCT was intended for all high-risk patients. The following scenarios are investigated with respect to biases in the effect estimates, as well as observed power and control of the type I error rate:

Whether (and when) a patient receives HSCT,

(1) does not depend on factors related to patient survival,

(2) depends only on known factors related to survival that are included in the Andersen-Gill model, and

(3) depends on (possibly unknown) factors relevant for survival not accounted for in the model.

In addition, we also evaluate corresponding scenarios where the influence of HSCT on survival depends on clinical covariates, which is considered by an interaction term in the Andersen-Gill model.

Multiple Imputation of Missing Covariates for Multiple Survival Endpoints

Dirk Klingbiel[1,2], Shu-Fang Hsu Schmitz[1], Arnaud Roth[3], Daniel Dietrich[1], Mauro Delorenzi[4,2]

[1]SAKK Coordinating Center, Switzerland; [2]Swiss Institute of Bioinformatics, Lausanne, Switzerland; [3]Geneva University Hospital, Geneva, Switzerland; [4]University Hospital Vaud and Lausanne University, Lausanne, Switzerland; dirk.klingbiel@sakk.ch

Keywords: Missing values, multiple imputation, resampling, survival endpoints, cox models

Introduction

Multiple imputation is a common approach to tackle missing data. The consensus for survival endpoints is that the survival times and event indicators should be incorporated into the imputation models. It remains an open question how to deal with multiple survival endpoints, a common feature of clinical trials in oncology.

Methods

We took data from a randomized phase III trial in colorectal cancer, with biomarker data available for 1404 patients. Relapse free survival (RFS), overall survival (OS), and survival after relapse (SAR) were considered for analysis. Five clinical markers and two biomarkers were chosen as covariates. One of the biomarkers is associated with OS, but not with RFS. To evaluate the effects of different imputation models varying degrees of missingness were artificially imposed on the biomarkers using *missing (completely) at random* mechanisms. The results after imputation are compared with those from the complete dataset.

Results and discussion

The bias of estimates is most serious when no endpoint is included for imputation, whereas the standard error (SE) is hardly affected. For RFS and OS, including the analyzed endpoint vastly improves the results. Reassuringly, adding the other endpoint does not dilute the results. For the analysis of SAR the bias and SE are greatest with RFS alone, slightly less with OS alone, and smallest when RFS and OS are employed simultaneously.

Conclusion

We recommend including all relevant survival endpoints into imputation models. This approach is also more computationally friendly than using different imputation models for different survival endpoints in separate analyses.

Statistics in practice (3)

AG Nachwuchs, AG Weiterbildung, S. Roll

Thursday 15. September, 11:00 – 12:30, Lecture Hall HAH-E-11

This is a special session for “statisticians in action” which should help providing an overview of today’s environment in which statisticians are working. Both elements should help to prepare and integrate the younger generation into the biometry community, a major task to secure our future. This tutorial builds on the two meta-analysis sessions Statistics in Practice (1) and (2) and aims to deepen the methodology of meta-analysis with practical examples.

Speaker: Salanti, G

Genetics and Biomarker (3)

A. Ziegler

Thursday 15. September, 11:00 – 12:30, Lecture Hall HAH-F-1

Micronutrient Deficiencies and the Human Plasma Nutriproteome

Ingo Ruczinski

Department of Biostatistics, Johns Hopkins University; iruczins@jhsph.edu

Keywords: Proteins, experimental design, missing data

Micronutrient deficiencies are a major public health burden since they raise risks of morbidity, disability, growth retardation and mortality among infants, children and women of reproductive ages. Because blood plasma concentrations of proteins and other nutrients reflect biologically relevant aspects of nutritional status, a fundamental hypothesis is that a detectable "plasma nutriproteome" exists (proteins in circulation that covary with nutrient concentrations or respond to nutritional exposures) that, with foreseeable technical innovations, can provide a new platform for assessing micronutrient status of populations, and can help with targeted interventions. A large project is underway to define this nutriproteome in a young, chronically undernourished child population in South Asia by determining the strength of association between relative abundance of plasma proteins, and concentrations of multiple micronutrients, as well as related antioxidant and acute inflammation indicators. The proteome of 500 rural Nepalese children is being investigated using stable isotope masstags (iTRAQ) to identify and quantify serum proteins by high-throughput tandem mass spectrometry.

In this presentation, we discuss the issues of experimental design, data normalization within and between 8-plex iTraq runs, missing data patterns, and reproducibility. We introduce a novel and robust method to quantify peptide and protein abundance from hundreds of thousands of reporter ion spectra. We further discuss strategies to handle missing data, and contrast several statistical approaches, including mixed effects models and generalized estimation equations. To date, we have identified and quantified over 300 serum proteins, dozens of which have predictive power for micronutrient status. We show that proteomic signatures exist that explain a substantial proportion of the variability in some of the nutrients of interest.

Two-group comparisons of intensity values in omics experiments

Andreas Gleiss[1], Harald Mischak[2], Georg Heinze[1]
[1]Medical University Vienna, Austria; [2]Mosaiques Diagnostics GmbH, Hannover, Germany;
andreas.gleiss@meduniwien.ac.at

Keywords: Omics data, zero inflation, two-part tests

Many experiments conducted in molecular biology compare intensity values obtained by micro-RNA transcriptomics, proteomics or metabolomics ('Omics') procedures between two groups of independent biological samples differing in an experimental condition or in the health status of the subjects the samples were taken from. A special characteristic of such data is the frequent occurrence of zero intensity values caused by features (micro RNAs, peptides or metabolites) that cannot be detected in some samples. Zero intensities can arise either by true absence of a feature or by a signal that is below a technical limit of detection.

In the literature the distribution of observed signals is viewed either as a mixture of a binomial distribution (absence or presence of a detectable signal) and a continuous distribution (intensity if signal is present) or as a left-censored continuous distribution. Various so-called two-part tests have been proposed to compare mixture distributions between groups whereas various one-part tests have been proposed for left-censored distributions. We perform a comparative simulation study using the setting of a typical omics study, comparing several hundreds of features simultaneously. Both types of distributional assumptions as well as combinations of both are considered when comparing power and effect estimation. We discuss issues of application using an example from proteomics.

We conclude that the considered tests generally show their strengths in scenarios satisfying their respective distributional assumptions. If it is a priori unclear which distributional assumptions best fit the data at hand then a two-part Wilcoxon test can be recommended as omnibus test.

Group Effects in miRNA and related Target Gene Set Expression

Stephan Artmann[1], Klaus Jung[1], Annalen Bleckmann[1,2], Tim Beißbarth[1]
[1]Department of Medical Statistics, University Medical Center Göttingen, D-37099 Göttingen, Germany; [2]Department of Hematology and Oncology, University Medical Center Göttingen, D-37099 Göttingen, Germany; stephan.artmann@stud.uni-goettingen.de

Keywords: MicroRNA,microarray

Motivation: microRNAs (miRNAs) regulate a wide range of physiological functions and play part in many human diseases, such as cancer [1]. So far, the search for clinically relevant miRNAs has concentrated on measuring differences in miRNA expression levels between two groups, e.g. from aggressive vs. non-aggressive tumors. Additionally, however, miRNAs can directly lead to the down-regulation of a specific set of target genes, a fact which has been ignored or treated separately from miRNA expression in many miRNA studies.

Results: Expression levels of both, target genes and miRNAs, can be obtained by microarray experiments. We present a method to combine the information available on miRNAs and their target sets' expression. MiRNA-wise p-values from established component-wise testing [2] and gene set p-values from different gene set tests are combined using methodology often used in meta-analysis. In a simulation study we reveal that our combination approach is more powerful than miRNA-wise testing alone. We show that combining miRNA-wise results with 'competitive' gene set tests (as defined by [3]) maintains a pre-specified FDR.

References

[1] J. Lu, G. Getz, E.A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B.L. Ebert, R.H. Mak, A.A. Ferrando, J.R. Downing, T. Jacks, H.R. Horvitz, T.R. Golub (2005). MicroRNA expression profiles classify human cancers, Nature, 435:834-838.

[2] G. K. Smyth. Linear models, empirical bayes methods for assessing differential expression in microarray experiments, Stat Appl Genet Mol Biol, 3:Article 3.

[3] J.J. Goeman, P. Bühlmann (2007). Analyzing gene expression data in terms of gene sets: methodological issues, Bioinformatics, **23**:980-987.

Benefit Risk

Günter Heimann

Thursday 15. September, 9:00 – 10:30, Lecture Hall HAH-E-10

Benefit-Risk Assessment: A Data-Driven Approach.

Sinan B. Sarac

Keywords: Benefit-risk assessment, benefit-risk communication, data-driven method, objective scoring, clinical relevance, drug development.

Objective: The aim of the study was to develop a generally applicable and reliable data-driven benefit-risk assessment method. Regulatory agencies request transparent benefit-risk assessments of pharmaceutical products. Several general methods exist, but none are widely accepted or used.

Method: The method consists of an eight step qualitative approach with quantitative handling of data. To enable comparison of benefits and risks, they are weighted on a simple scale and to promote objectivity, scores are assigned based on data using descriptive statistics whenever possible.

Results: The results of the assessment are visualised in tornado-like diagrams, where both the importance and performance of each individual benefit and risk criterion is easily captured and communicated. Multiple trials are combined to produce an overall benefit-risk assessment.

Conclusion: The method can handle all data encountered and can be used on single or multiple trials in an individual or an overall assessment. Pharmaceutical companies and regulatory agencies are provided with a unique tool to communicate benefits and risks in a structured, transparent and objective way.

What role should formal risk-benefit decision-making play in the regulation of medicines?

Deborah Ashby
Imperial College London, United Kingdom; deborah.ashby@imperial.ac.uk

Keywords: Benefit-risk, decision-making, drug regulation

The regulation of medicine requires evidence of the efficacy and safety of medicines, and methods are well-developed to deal with the latter and to a lesser extent the former. However, until recently, assessment of risk- benefit especially in relation to alternatives has been entirely informal. There is now growing interest in the possibilities of more formal approaches to risk-benefit decision-making. In this talk, we review the basis of drug regulation, the statistical basis for decision-making under uncertainty, current initiatives in the area, and discuss possible approaches that could enhance the quality of regulatory decision-making.

Quantitative Benefit-Risk Assessment and the Role of Statistics – Is there a need for ICH E9+?

Jürgen Kübler

Novartis Pharma AG, Switzerland; juergen.kuebler@novartis.com

Keywords: Benefit-risk

Traditionally, benefits and risks of drug therapies have been analyzed separately. Medical judgment and qualitative assessments have long been considered as state of the art of the assessment of the benefit-risk balance ultimately decision making. In parallel, more advanced methods were utilized in the areas of health technology assessment (HTA) and decision analysis (DA). Over the last years, more systematic evaluation of the utility of quantitative approaches for benefit-risk assessment can be observed across the globe with contributions from regulators, academia and industry.

This presentation will provide a brief overview of activities and explores the role of statisticians in a changing environment. The quantitative assessment of benefit-risks will to be put in context with other trends and initiatives in drug development. The challenges for statisticians are manifold. The statisticians in the pharmaceutical industry long focused on experimental designs and the evaluation of efficacy, more sophisticated approaches for the assessment of safety and novel approaches for the systematic quantitative assessment of benefits and risks is required. ICH E9 can be considered as a milestone for statisticians as it provides the foundation mainly in the area of experimental design to ensure appropriate approaches for providing scientifically sound evidence of evidence of efficacy. Statisticians working in drug development now need to acquire additional skills to address the new challenges.

Application of Spatial Weather Generator for Runoff Simulation in River Catchment

Leszek Kuchar[1,2], Slawomir Iwanski[1], Leszek Jelonek[2], Wiwiana Szalinska[2]
[1]Wroclaw University of Environmental and Life Sciences, Wroclaw, Poland; [2]Institute of Meteorology and Water Resources, Wroclaw, Poland; Leszek.Kuchar@gmail.com

Keywords: Weather generator, meteorological data, climate change, river runoff

Spatial weather data are required for runoff simulation in river catchment, and used by rainfall-runoff-model to simulate daily flows for closing water-gauges. In the paper spatial weather generator SWGEN and its application to simulate runoff in river catchment for the future climate conditions according to selected scenarios is presented.

A model SWGEN generates daily values of precipitation (P), solar radiation (SR), maximum (Tmax) and minimum temperature (Tmin) parallel for k stations. The occurrence of precipitation has influence on the solar radiation and the temperature for a day by determining a status (wet or dry day) and independently generating solar radiation and temperature for a given status day. The model preserves the dependence in time and the seasonal characteristics for locations. Precipitation is generated by means of the first-order Markov chain to determine occurrence of wet/dry days, and then for the amount of rainfall multidimensional two-parameter gamma distribution. Daily values of solar radiation (SR), temperature maximum (Tmax) and minimum (Tmin) are considered as a multidimensional time series AR(1).

SWGEN model was used for runoff simulation on Kaczawa River (southwest of Poland) at discharge point Piatnica for present conditions (2010) and conditions of 2030 and 2050 according to GISS, GFDL and CCCM climate change scenarios. 300 years of generated data of potentially possible weather course, 16 stations, given time horizon and scenario were applied to hydrological model NAM.

Finally, three parameter Gamma probability distribution to fit average daily runoff for conditions of 2010, 2030 and 2050 and different climate scenarios were obtained.

Part 2: Abstracts for posters

****

The effect of regression dilution bias on the association between metabolic syndrome and brain tumours in a large population-based cohort study

Michael Edlinger[1], Håkan Jonsson[2], Tone Bjørge[3,4], Jonas Manjer[5], Pär Stattin[2], Hanno Ulmer[1]

[1]Medizinische Universität Innsbruck, Austria; [2]Umeå University, Sweden; [3]University of Bergen, Norway; [4]Norwegian Institute of Public Health, Norway; [5]Skåne University Hospital Malmö and Lund University, Sweden; Michael.Edlinger@i-med.ac.at

Keywords: Regression dilution bias, metabolic syndrome, brain tumour, cohort study

Background: Cohort studies measuring exposures solely at baseline to estimate the subsequent outcome effect are prone to bias stemming from random errors at measurement or time variations of the individual level of exposure. This phenomenon, related to the so called "regression toward the mean", could lead to underestimation of real associations. We assessed the extent of the bias in a study on metabolic factors and brain tumour risk.

Methods: 578,462 subjects from Sweden, Austria, and Norway were followed for a median of ten years (Me-Can), also utilising information from national cancer registries. Factors of the metabolic syndrome were analysed as standardised z-scores with Cox proportional hazards regression models, stratified by cohort. We used the repeated measurements available from 133,820 participants to estimate regression dilution ratios, based on linear mixed effect models, to correct for measurement error.

Results: The estimated regression dilution ratios were 0.902 for body mass index, 0.525 for systolic and 0.497 for diastolic blood pressure, 0.278 for glucose, 0.657 for cholesterol, and 0.505 for triglycerides. For body mass index and cholesterol the correction of the effect on brain tumour risk hardly made any difference. The bias was considerable for blood pressure and also for glucose with meningioma risk and for triglycerides with high-grade glioma risk. As a consequence, the precision of the effect estimates decreased.

Conclusion: Regression dilution bias in cohort studies with only one baseline measurement of exposure variables can lead to a substantial underestimation of effects, depending on the extent of the exposure measurement variation.

Does Size always Matter? A Simulation Study on the Impact of Slightly Altered True Genetic Models

Carolin Pütter[1], Karl-Heinz Jöckel[1], Heinz-Erich Wichmann[2], André Scherag[1]
[1]Institute for Medical Informatics, Epidemiology and Biometry, Germany; [2]Helmholtz Center Munich, German Research Center for Environmental Health, Institute of Epidemiology, Neuherberg, Germany; carolin.puetter@uk-essen.de

Keywords: Selective genotyping, missing heritability, GWAS, association, mixture distribution

Genome wide association studies (GWAS) revealed robust associations between single nucleotide polymorphisms (SNP) and complex traits. As the proportion of the explained phenotypic variance is still limited for most of the traits larger and larger meta-analysis are being conducted to detect additional association signals. Here we investigate the impact of the study design and more importantly the underlying assumption about the true genetic effect in a bimodal mixture situation on the power to detect additional association findings. In particular, we performed simulations of quantitative phenotypes analysed by standard linear regression (n=130,000) and artificially dichotomized case-control data sets (5,000 pairs) from the extremes of the quantitative trait analysed by standard logistic regression. Using linear regression, markers with an effect in the extremes of the traits were almost undetectable even with steadily growing sample sizes. In this situation, analysing extremes by a case-control design of a much smaller sample size had better power to detect the associated marker. Our results indicate a) that it might be worthwhile to re-analyse dichotomized versions of the available meta analysis data sets to detect new loci while it b) offers an explanation for discrepant findings pertaining to quantitative phenotypes analysed by linear regression as compared to case-control approach analysing selected samples. A real-data example of GWAS-derived SNP associations for body-mass-index in n=16,463 is provided to support the bimodal mixture assumption.

Bayesian Adaptive Markov Chain Monte Carlo Estimation of Genetic Parameters

Boby Mathew[1], Andrea Bauer[1], Petri Koistinen[2], Mikko Sillanpää[2], Jens Léon[1]
[1]University of Bonn, Germany; [2]University of Helsinki, Finland; boby.mathew@hotmail.com

Keywords: Adaptive MCMC, Gibbs sampling, Bayesian analysis

Accurate and fast estimation of genetic parameters underlying quantitative traits using mixed linear models with additive and dominance effects is of great importance in both natural and breeding populations. We propose a new fast adaptive Markov Chain Monte Carlo (MCMC) sampling algorithm for the estimation of genetic parameters in the linear mixed model with several random effects. In the learning phase of our algorithm, we use the hybrid Gibbs sampler to learn the covariance structure of the variance components. In the second phase of the algorithm, we use this covariance structure to formulate an effective proposal distribution for a Metropolis–Hastings algorithm, which uses a likelihood function where the random effects have been integrated out. Compared to the hybrid Gibbs sampler, the new algorithm showed better mixing properties and was about twice as fast to run. Our new algorithm was able to detect different modes in the posterior distribution.

Simulated versus best fitted Covariance Structures in Agricultural Field Experiments

Christel Richter[1], Bärbel Kroschewski[2]
[1]Humboldt-Universität zu Berlin, Germany; [2]Humboldt-Universität zu Berlin, Germany;
ch.richter@agrar.hu-berlin.de

Keywords: Agricultural field experiments, spatial correlation, simulation, randomization

In agricultural field experiments, correlations between neighbouring plots exist. The classical analysis model ignores these covariances because of the randomization theory based argumentation. Recently, it has been tried to clarify by increased effort whether the analysis should be changed from the model-driven to a data-driven approach that relates to the variance-covariance structure. For each concrete experiment, this changed approach requires the selection of the covariance structure from a set of covariance models. This study focuses on the relation between simulated and the best fitted model selected by AICC and on its consequences for the compliance of a given nominal Type I error of the t- and the F-test. For this study, 100 random fields with 19 different covariance structures were simulated. Randomization is a basic principle of planning of experiments and serves not only as a justification of the model without covariances. To clarify the role as well of the random field as of the randomization plan, each of the 100 fields was combined with 100 randomized plans. For the simulations, we assumed a trial with 10 dummy treatments and 4 replications laid out in a RCBD. The data was analyzed by 9 models. As expected, the model preference depends on the underlying covariance structure of the data. The best compliance of the nominal Type I error was found for the classical model; for the spatial models the results of the t-Test are better than those ones of the F-Test whereby their variation depends much stronger on the random field than on the randomized plan.

Fast Association Testing of Genotyped and Imputed SNPs as well as Gene-Environment Interactions in Case-Parent Trio Studies

Holger Schwender[1], Margaret A. Taub[2], Mary L. Marazita[3], Terri H. Beaty[2], Ingo Ruczinski[2]

[1]TU Dortmund University, Germany; [2]Johns Hopkins Bloomberg School of Public Health, Baltimore, USA; [3]University of Pittsburgh, USA; holger.schwender@udo.edu

Keywords: Association studies, single nucleotide polymorphisms, case-parent trio studies, transmission/disequilibrium test, gene-environment interaction

Case-parent trio designs considering children affected by a disease and their parents are frequently used to detect single nucleotide polymorphisms (SNPs) associated with disease. A popular procedure for identifying such genetic variations is the genotypic transmission/disequilibrium test (gTDT), which is equivalent to a Wald test based on a conditional logistic regression model. Usually, the parameters of such a model need to be estimated numerically by an iterative procedure, which can be time-consuming if this model should be fitted to hundreds of thousands of SNPs in a genome-wide association study. However, as we will show in our presentation, there exist closed-form solutions for the parameter estimates when testing a SNP with a gTDT under an additive, a dominant, or a recessive model. As exemplified by applying the gTDT to genome-wide case-parent trio data, the time required for testing all SNPs in such a study reduces from several hours to a few minutes when employing the analytic estimates instead of the conventional iterative fitting procedure. These closed-form solutions can also be used to test interactions between SNPs and binary environmental variables for association with disease. Moreover, the analytic estimates can be adapted to test fuzzy genotype calls usually determined when imputing untyped SNPs. Finally, we present a procedure that makes it feasible to compute genome-wide permutation-based p-values for the gTDTs under an additive, dominant, or recessive model, as well as for a MAX-test in which the maximum over the statistics of these three gTDTs is used as test statistic.

Bayesian Hypothesis Testing (BHT) in Single-arm Phase II Clinical Trials with Binary Endpoints

Martin Kappler[1], **Joachim Gerss**[2]
[1]Novartis, France; [2]Institute of Biostatistics and Clinical Research, University of Münster;
martin.kappler@novartis.com

Keywords: Phase II Clinical Trial, Bayesian Hypothesis Testing

When evaluating the response rate p of a new therapeutic agent using a binary endpoint in a single-arm phase II clinical trial, two competing hypotheses are established, the null hypothesis H_0: $p \leq p_0$ and the alternative hypothesis H_1: $p > p_0$. The final aim of the statistical analysis is to draw a decision either in favor of superiority of the trial agent (rejecting H_0) or futility (rejecting H_1). In order to solve the problem in a sequential approach, stopping boundaries may be defined using either frequentist methods (e.g. Simon's two-stage design [1]) or Bayesian methods. Recently, Johnson and Cook proposed an approach based on formal hypothesis testing implemented in a Bayesian framework (Bayesian hypothesis testing, BHT) [2]. A one-point prior density is used to define H_0. The prior density under H_1 assigns no mass to parameter values that are consistent with H_0 (nonlocal alternative prior density). The proposed approach has the useful property that any (e.g. over-optimistic) mis-specification of the prior density cannot increase the expected weight of evidence in favor of the trial agent. In the present work a new algorithm is derived based on the above BHT approach using a one-point prior density not only under H_0 but also under H_1. The operating characteristics of both approaches are evaluated and compared in various scenarios using simulated data. It is shown that the new algorithm is more efficient than the former BHT approach. Moreover, the new algorithm is compared to common Bayesian designs based on posterior credible intervals, as well as common frequentist designs.

References

[1] R. Simon (1989). Optimal two-stage designs for phase II clinical trials. Controll Clin Trials **10**:1-10.

[2] V. E. Johnson and J. D. Cook (2009). Bayesian design of single-arm phase II clinical trials with continuous monitoring. Clinical Trials **6**:217-226.

Comparing an alternative randomization procedure – the maximal procedure – to permuted block randomization: a simulation study

Petra Ofner-Kopeinig, Maximilian Errath, Andrea Berghold
Medical University of Graz, Austria; petra.ofner@medunigraz.at

Keywords: Randomization, selection bias, permuted block randomization, maximal procedure

Permuted block randomization is a widely used randomization procedure in clinical trials. However, there is a high chance of selection bias when using small block lengths and especially when masking is not possible. While the ICH E9 guideline suggests working with variable block lengths[1] to reduce this problem, theoretical results[2] show that this does not completely eliminate the chance of selection bias. Another approach is to choose larger block sizes. Based on this idea, Berger suggested an alternative randomization procedure: the maximal procedure[3].

The maximal procedure is a constrained randomization procedure using large block lengths (e.g. 160). All possible sequences that satisfy certain conditions are generated and each of these allocation sequences can be chosen with the same probability. The constraints are: terminal balance and a maximum tolerated imbalance. Perfect balance is not necessary. For the implementation, we used the algorithm suggested by Salama[4] for an efficient generation of constrained block allocation sequences.

In a simulation study, we explore the balance behaviour of the maximal procedure and compare it to the permuted block randomization procedure for different scenarios. The simulation tool implemented in the "Randomizer for Clinical Trials" (www.randomizer.at) is used for this purpose.

References

[1] International Conference on Harmonization (ICH) (1998). "Guidance on Statistical Principles for Clinical Trials." Federal Register. International Conference on Harmonization, E-9 Document. 63, 179, 49583-49598

[2] W. F. Rosenberger and J. M. Lachin (2002). Randomization in Clinical Trials, Theory and Practice, Wiley Series in Probability and Statistics, Section 6.5.2

[3] V. W. Berger, A. Ivanova, D. M. Knoll (2003). Minimizing predictability while retaining balance through the use of less restrictive randomization procedures. Statistics in Medicine **22:**3017-3028

[4] I. Salama, A. Ivanova, B. Qaqush (2008). Efficient generation of constrained block allocation sequenzes, Statistics in Medicine, **27**:1421–1428

Identification of differentially expressed proteins in subgroups of patients

Maike Ahrens[1], Michael Turewicz[1], Miriam Böckmann[1], Caroline May[1], Martin Eisenacher[1], Helmut E. Meyer[1], Jörg Rahnenführer[2], Christian Stephan[1]
[1]Medizinisches Proteom-Center, Germany; [2]Fakultät Statistik, Technische Universität Dortmund, Germany; maike.ahrens@tu-dortmund.de

Keywords: Minimum M statistic, subgroup detection, protein microarrays, difference in means

A common task in proteomics is the identification of differentially expressed proteins between two groups of samples, e.g. disease and control. Popular test statistics for this purpose are Student's t test, Wilcoxon's rank-sum test, and variations of both. Depending on the investigated disease, especially the disease group may contain unknown subgroups, representing different disease stages or subtypes. Whereas the two above mentioned tests assume homogeneous groups, a more appropriate test should be more sensitive to changes that are present only in subgroups.

For this purpose, the minimum M statistic was proposed. The nonparametric test is based on the probability that a given number of observations in one group have values larger than the i-th highest value in the other group. The minimal p-value across various choices of i is determined. The distribution of this minimum is not a classical extreme value distribution due to the dependency of the p-values. For multiple testing scenarios where the test statistic is then calculated separately for many proteins, the performance of the minimum M statistic regarding both types of error has to be investigated.

We therefore conduct comparative simulation studies to gain knowledge of the distribution of the minimum M statistic under the hypothesis of no group differences and under appropriate alternatives with shifts present only in subgroups, depending on the sample size. Finally, we apply the different approaches to protein expression data from 144 persons in a Parkinson's disease study with about 9000 measured proteins and exemplify critical differences on real data.

Estimation of Nonlinear Interactions in Survival Analysis

Maria Kohl[1], Meinhard Ploner[2], Max Plischke[1], Georg Heinze[1]
[1]Medical University of Vienna, Austria; [2]data-ploner.com, Brunico, Italy;
maria.kohl@meduniwien.ac.at

Keywords: Fractional polynomials, Cox regression, nonlinear effects

We consider a nephrologic study where the aim is to estimate the interaction effect of a binary risk factor, such as stage of chronic kidney disease, and a continuous risk factor, such as urine osmolality, on time to admission to dialysis or death. We assume that there could be different types of nonlinear effects across the levels of the binary factor. For continuous risk factors, the standard definition of Cox's proportional hazards model assumes a linear relationship with the log hazard. However, nonlinear effects of continuous factors can be accommodated by fractional polynomials (FP). The FP method selects so-called 'powers' to create nonlinear transformations of the original variable, which lead to the best fit in a subsequent regression analysis. Due to the optimal selection of the FP powers, reliable confidence intervals (CI) can only be obtained by the bootstrap, repeating the selection in resamples.

Here we consider less computer-intensive methods for CI estimation, namely using the model-based variance but accounting for model selection by increasing the number of degrees of freedom, and using the weighted average variance across all evaluated FP models. We performed a comparative analysis of the aforementioned medical study, and report on a simulation study investigating properties of CI computed using the two proposed variance estimators for nonlinear effects. We conclude that CI estimation for nonlinear interaction effects by the bootstrap can well be approximated by the proposed approaches. The approximation works best in the center of the range of the continuous variable.

The comparison of the performance of genes selection methods for discrimination between cirrhotic tissue and cirrhotic tissue with concomitant hepatocellular carcinoma

Magdalena Wietlicka-Piszcz, Małgorzata Ćwiklińska-Jurkowska
Collegium Medicum, Nicolaus Copernicus University, Poland; mpiszcz@cm.umk.pl

Keywords: Microarrays, classification

Microarray is a new technology allowing for simultaneous investigation of expression levels of thousands or tens of thousands of genes. Due to a very high number of genes on one microarray an important stage of microarray data elaboration is the pre-selection of features/genes for further analysis. Lee et al. (2005) mention that various methods of active gene selection applied to the same set of microarray data may give different sets of genes and consequently lead to different discrimination results.

The aim of this work was the comparison of various methods of selection of differentially expressed genes in oligonucleotide microarray dataset for differentiation between cirrhotic tissue and cirrhotic tissue with concomitant hepatocelular carcinoma. The Parametric Empirical Bayes Method, the Nonaparametric Empirical Bayes Method, the Significance Analysis of Microarrays, the Wilcoxon statistic and the permutation tests were applied to select differentially expressed genes. The genes sets generated by the examined dimensionality reduction methods were used in the construction of discrimination rules by applying discrimination methods. Each set consisted of genes ranked according to the probability of differential expression. The assessment of the constructed discrimination procedures was performed by the error estimation using 10-fold cross-validation and by the bootstrap 0.632 method.

The results suggest that the choice of a gene selection method and the choice of a discrimination method are both important. The best performance was obtained for the SVM method with radial kernel. However, the indication for a particular method of gene selection is not explicit.

Different approaches to analyze pre-post assessments of quality of life data from randomized trials

Daniel Dietrich[1,2]
[1]SAKK Coordinating Center, Bern, Switzerland; [2]Institute of Mathematical Statistics and Actuarial Science, University of Bern, Switzerland; daniel.dietrich@sakk.ch

Keywords: Pretest-posttest designs, analysis of variance, analysis of covariance

Introduction

In quality of life questionnaires there is usually a baseline assessment before treatment start and several assessments during treatment. Here we consider the simplified situation with a baseline score Y_0 and one on-treatment score Y_1, and two treatment arms A and B. The goal is to estimate the treatment effect. It can be done either with repeated measures anova or with an analysis of covariance model (Ancova).

Methods

We consider estimators of the form $(\text{mean}(Y_{A1}) - \text{mean}(Y_{B1})) - \beta\,(\text{mean}(Y_{A0}) - \text{mean}(Y_{B0}))$ with a coefficient $0 \le \beta \le 1$. The Ancova model with baseline as covariate estimates β with a value $0 < \beta < 1$, the intrapatient correlation. Ancova explicitly uses this correlation to estimate the treatment effect. Repeated measures anova also estimates the intrapatient correlation but sets either $\beta=1$ or $\beta=0$ in the formula depending on whether a treatment effect at baseline is included to the model (Anova1) or omitted (Anova0) since for randomized groups there should be only a random difference at baseline.

Results

Simulations indicate that all three models keep the alpha level under the normal and under the truncated normal distribution. The power of Anova0 is a little higher than that of Ancova and the latter has a higher power than Anova1.

Conclusions

The gain for Anova0 compared to Ancova is only little even if the intrapatient correlation is high. So Ancova seems to be the appropriate model. Nevertheless the Anova1 estimator has something logical: It incorporates individuals' profit form baseline, $Y_1 - Y_0$ which is an important measure in quality of life.

Design choices for small-scale single-arm phase II trials with non-inferiority (NI) intention for a reduced dose

Hong Sun, Shu-Fang Hsu Schmitz
SAKK, Switzerland; hong.sun@sakk.ch

Keywords: Phase II, non-inferiority, sample size

Background: In cancer patients more chemotherapy (CT) may yield better efficacy, but also more toxicity. It is plausible to investigate whether a reduced CT dose is non-inferior (NI) to the full dose. Before pursuing a phase III NI trial, a phase II trial is warranted. If resources are limited, then the scale of the phase II trials might be constrained. Design choices for small-scale phase II trials with NI intention are desired.

Methods: For an example single-arm trial with a binary endpoint, sample sizes from hypothesis testing and confidence interval (C.I.) approaches are compared under different parameter settings with the NI margin equal to the interval width of the C.I.

Results: If the success rate of full dose is 0.5, with the NI margin 0.1 for the reduced dose, type I error 5% and power 80%, the required sample size is at least 158 based on one-sided hypothesis test. Using one-sided C.I. approach with 95% confidence level and allowed width of 0.1, only 76 patients are needed. Under different settings, the sample size from the C.I. approach is always lower than that from the other approach.

Discussions: The determination of the NI margin is problematic in phase II setting as reliable previous data may not be available. The C.I. approach does not require the NI margin and does not impose a rigid go/no-go decision rule only based on the primary endpoint. This allows investigators to consider other aspects of the treatment for the final decision, hence might be preferred.

Evaluation of uni- and multivariate GCA and SCA effects of genotype on the basis of series of experiments

Zygmunt Kaczmarek[1], Elzbieta Adamska[1], Teresa Cegielska-Taras[2], Laurencja Szala[2]
[1]Institute of Plant Genetics, Poznan, Poland; [2]Plant Breeding and Acclimatization Institute, Poznan,Poland; Zkac@igr.poznan.pl

Keywords: MANOVA, series of experiments, GCA, SCA, rapeseed

A multivariate statistical approach was proposed to estimation and testing genetic and breeding parameters of winter oilseed rape genotypes. The statistical methods were used to evaluate parental forms (DH lines) on the basis of observations of progenies from line x tester mating system The aim of the study was the classification and choice the best oilseed rape genotypes with regard to seed yield and three unsaturated fatty acids: oleic (C18:1), linoleic (C18:2) and linolenic (C18:3). A series of experiments with the set of (7 x 4 =28) progenies was carried out over 3 years. The presented approach involves the use of MANOVA and other multivariate techniques for estimating parameters and testing hypotheses. These methods allowed estimation of uni- and multidimensional effects of general (GCA) and specific (SCA) combining ability of DH lines and also their interactions with environments.

The results of statistical analysis for series of experiments with line x tester progenies were also presented for individual traits and their functions. Estimates and results of testing hypotheses concerning the interactions of GCA and SCA effects of DH lines with years were given, Statistical calculations were made using in majority the computer program SERGEN [1] based on a unified methodology developed by authors of this paper.

References

[1] T. Calinski, S. Czajka, Z. Kaczmarek, P. Krajewski, I. Siatkowski (1998). SERGEN-Analysis of series of plant genetic and breeding experiments. Computer program for IBM-PC, Version 3, IGR PAN, Poznan

Effect of Genotype, Environment and their Interaction on Quality Parameters of Wheat Bread Lines

Tadeusz Adamski[1], Maria Surma[1], Zofia Banaszak[2], Zygmunt Kaczmarek[1], Anetta Kuczynska[1], Malgorzata Majcher[3], Bogumila Lugowska[2], Wiktor Obuchowski[3], Boleslaw Salmonowicz[1], Karolina Krystkowiak[1]

[1]Institute of Plant Genetics, Poznan, Poland; [2]DANKO Plant Breeding Ltd., Choryn, Poland; [3]University of Life Sciences, Poznan, Poland; Tada@igr.Poznan.pl

Keywords: GE interaction, regression analysis, stability, wheat bread

Stability of bread wheat cultivars in terms of grain quality characteristics is an important requirement for baking industries. Twenty four winter wheat genotypes of different grain hardness were assessed in multienvironment trials—with four locations and two levels of fertilization in each location. Grain samples were analysed for hardness, protein and starch content, wet gluten content, Zeleny sedimentation value, alveograph parameter and hectoliter weight. One-way analysis of variance was performed for each experiment, from which information on significance of genotype differentiation was obtained and experimental error was estimated. In the next step, the data from a series of experiments were processed using the computer program SERGEN [2] based on the methods developed by [1]. In these methods, GEI effect related to each genotype is measured by the value of the relevant *F*-statistic. The regression of GEI effects on the observed environmental means, expressed by coefficients of regression and determination and also the *F*-statistic values for testing the significance of regression and of deviations from regression, is the measure of genotype stability.

References

[1]. T. Calinski, S. Czajka, Z. Kaczmarek (1997). A multivariate approach to analysing genotype-environments interactions. Advances in Biometrical Genetics, Poznan, 3-14.

[2]. T. Calinski, S. Czajka, Z. Kaczmarek, P. Krajewski, I. Siatkowski.(1998). SERGEN - Analysis of series of plant genetic and breeding experiments. Computer program for IBM-PC, Version 3, IGR PAN Poznan

A Comparison of an Individual Patient Data (IPD) Meta-Analysis and a Literature-Based Meta-Analysis to Investigate Progression-Free Survival in Ovarian Cancer Patients

Christine zu Eulenburg[1], Sven Mahner[2], Anne Staehle[3], Alexander Reuss[4], Eric Pujade-Lauraine[5], Philipp Harter[6,9], Isabelle Ray-Coquard[7], Karl Wegscheider[1], Jacobus Pfisterer[8], Andreas du Bois[6,9]

[1]Dept of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Germany; [2]Dept of Gynecology, University Medical Center Hamburg-Eppendorf, Germany; [3]Dept of Obstetrics and Gynecology, St Marien Klinik, Karlsruhe, Germany; [4]Coordinating Center for Clinical Trials (KKS), University Marburg, Germany; [5]Dept of Oncology, Hopital Hotel-Dieu, Paris, France; [6]Dept of Gynecology and Gynecologic Oncology, Dr. Horst Schmidt Klinik, Wiesbaden, Germany; [7]Dept of Oncology, Centre Leon Berard, Lyon, France; [8]Dept of Gynecology and Obstetrics, Klinikum Solingen, Solingen, Germany; [9]Dept of Gynecologic Oncology, Kliniken Essen Mitte, Essen, Germany; C.Eulenburg@uke.de

Keywords: Individual patient data, IPD, meta-analysis, survival analysis, ovarian cancer

An individual patient data (IPD) meta-analysis is the most elaborate technique to analyze multiple survival data sets in common [1]. Yet, access to the entire dataset is an essential requirement. In contrast, a literature-based meta-analysis is less conditional, but may yield differing results not taking into account the correlations between individual patient characteristics and outcome. The objective of our study is to compare the results of both methods with respect to the prognosis of ovarian cancer patients after chemotherapy. Based on data of 3326 individuals from three independent trials we analyze the effect of time to chemotherapy on progression-free survival. To mimic a literature-based meta-analysis event rates at different time points within distinct data sets will be combined [2, 3]. The results will be compared to those from an IPD meta-analysis on individual event times. Consequences for the interpretation of literature-based meta-analyses will be discussed.

References

[1] L.A. Stewart and M.K.B. Parmar (1993). Meta-analysis of the literature or of individual patient data: is there a difference? The Lancet, **341**:418-422.

[2] K.B.G. Dear (1994). Iterative Generalized Least Squares for Meta-Analysis of Survival Data at Multiple Times, Biometrics, **50**:989-1002.

[3] L. Duchateau et al. (2001). Individual Patient- versus Literature-Based Meta-analysis of Survival Data: Time to Event and Event Rate at a Particular Time Can Make a Difference, an Example Based on Head and Neck Cancer. Controlled Clinical Trials **22**:538-547.

Spatial demography influences the spreading of infectious diseases and optimal policy making

Markus Schwehm
ExploSYS GmbH, Germany; markus.schwehm@explosys.de

Keywords: Spatial demography, vaccination coverage, spatial modelling, policy making

How much intervention is necessary to contain an epidemic? Classical modelling approaches derive the necessary intervention level from the estimated basic reproduction number R_0. For example, for Measles with a very high basic reproduction number R_0~16 a necessary vaccination coverage of about 95% is derived. However, this approach implicitly assumes a homogeneously mixed population where everybody can infect anybody with equal probability. The influenza pandemic 2009 has revealed, that infection transmission models with homogeneous mixing can not explain the observed infection dynamics. More successful are age-structured models that use a POLYMOD contact matrix to connect the age groups: children have much more and more intensive contacts with other children than children with adults or adults with adults. As a consequence, the fraction of children in a population has a large impact on the spread of a disease. This fraction decides whether interventions like school closures are successful and what vaccination coverage is necessary to prevent an outbreak. However, the children are not uniformly distributed in a population. For example, Switzerland has about 21% children (individuals below age 20). Looking at Swiss cantons, we have a range from 17 to 29% children and a range from 16 to 32% children in the districts. Considering the non-uniform spatial distribution of children in a population, this leads to opportunities to improve policy making by regionally adequate recommendations for vaccination coverage and other interventions.

Impact of matching and analysis approaches on estimates of association based on case-control data

Regina Riedl[1], Ruth Pfeiffer[2], Andrea Berghold[1]
[1]Institute for Medical Informatics, Statistics and Documentation, Medical University Graz, Austria; [2]Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA; andrea.berghold@medunigraz.at

Keywords: Case-control study, propensity score, simulation

In case-control studies the presence of confounders, defined as variables associated with both, exposures and disease outcome, may induce a bias in the estimates of association. We conduct a simulation study to investigate the performance of several matching methods in combination with regression adjustments for the analysis of case-control data. We analyse the data based on the entire population and based on matched case-control data sets using two different matching designs. The first design is based on matching on the propensity score, defined as the conditional probability of receiving exposure given a set of observed baseline covariates. Using this approach, matched sets of cases and controls with similar values of the propensity score are formed. In the second design matching is performed based on the measured baseline covariates. Given the matched case-control data, the following models are used to estimate the association of the main exposure with outcome: a logistic regression model that adjusts for all measured covariates, a logistic regression model that adjusts for the estimated propensity score, a conditional logistic regression based on the matched sets and stratification into quintiles of the propensity score and calculation of the summary odds ratio using the Mantel-Haenszel method. We explore the performance of the different designs and outcome models under various settings.

When the exposure is not associated with outcome, all methods correctly estimate this null association. However, when there is association, there was substantial bias in the estimates from the propensity score adjusted model in analyses based on the entire population. This bias was less severe for rare diseases. In analyses based on the matched data there was little bias (<10%) if the matching resulted in balance for all covariates.

Copy number estimation using nonparametric monotonic Bayesian regression

Helena Janzen[1], **Katja Ickstadt**[1], **Björn Bornkamp**[2]
[1]TU Dortmund, Germany; [2]Novartis Pharma AG, Switzerland; janzen@statistik.tu-dortmund.de

Keywords: Copy number, nonparametric monotonic Bayesian regression

Genotyping and copy number estimation is fundamental for the analysis of data from genome wide association studies (GWAS). Initially, intensity values are derived by scanning fluorescence intensities of hybridized DNA on genotyping arrays. Based on these intensity values reliable estimates for the assignment of the underlying genetic structure at given DNA loci, such as the genotype or the number of copies, are derived. Carvalho et al. [1] provide a widely-used algorithm to genotype single nucleotide polymorphisms (SNPs), which applies a corrected robust linear mixture model (CRLMM). This model is robust with regard to batch effects, e.g. due to multiple laboratories. Whereas CRLMM is utilized to estimate the genotypes, the allele-specific copy number estimates conditional on the CRLMM genotypes are obtained by using a linear multilevel model. However, an explicit deviation from linearity for large copy numbers is observed [2].

To model this nonlinear relationship we adapt the nonparametric monotonic regression approach by Bornkamp and Ickstadt [3] for copy number estimation. Our approach addresses saturation in the fluorescence intensity occurring with increasing copy number.

References

[1] B. Carvalho, H. Bengtsson, T.P. Speed and R.A. Irizarry (2007). Exploration, Normalization, and Genotype Calls of High-Density Oligonucleotide SNP Array Data. Biostatistics, **8**:485–499.

[2] R. Scharpf, I. Ruczinski, B. Carvalho, B. Doan, A. Chakravarti and R. Irizarry (2011).

A multilevel model to address batch effects in copy number estimation using SNP arrays, Biostatistics, **12**:30– 50.

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

Increasing the Resolution of GWAS Results based on u-Statistics for Multivariate Data (µGWAS) Through the Use of HapMap Data and Information Content

Knut M. Wittkowski
The Rockefeller University, United States of America; kmw@rockefeller.edu

Keywords: GWAS, Diplotype, giSNP, grid computing, GPU

Almost a decade after completion of the Human Genome Project, the advances hoped for have not yet been realized. After early successes with diseases where one locus confers all or most risk, the common statistical approaches based on (sets of) p-values computed one SNP at a time, have often produced ambiguous results when applied to common, complex diseases. Neither increasing the sample size to (tens of) thousands of subjects nor using multivariate methods based on the linear model necessarily increase resolution to detect epistasis when several genetically identically SNPs are present within an LD block..

Nonparametric methods are, in principle, better suited, yet they were never broadly developed for multivariate data due to their substantial computational demands. The recent technological advances in cluster and grid computing have spurred the extension of u-scores for multivariate data (µScores) to reflect hierarchical structures, in general, and, recently, the sequence of SNPs on a chromosome to provide a robust and valid statistical approach for multi-locus and multi-allelic data, termed µGWAS [1].

To improve the ability of µGWAS to reliably detect intragenic regions, we propose to first increase sensitivity by excluding genetic intervals spanning recombination hotspot from the diplotypes being formed, based on information from HapMap. Second, to use information content of the µScores (proportion of non-ambiguous pairwise orderings) to guard against false positives.

We present several applications based on recent analyses of several data sets of different size.

References

[1] K. M. Wittkowski and T. Song (2010). Nonparametric methods for molecular biology. Methods Mol Biol, **620**:105-53.

Evaluation of main and interaction effects with user-defined contrasts in the cell means model

Andreas Kitsche
Leibniz Universität Hannover, Germany; kitsche@biostat.uni-hannover.de

Keywords: Interaction contrasts, Simultaneous Confidence Intervals, Cell Means Model, linear model

In horticultural and agricultural research often experiments with two or more treatment factors are set up. Using the analysis of variance and the corresponding F-tests for the main and interaction effects offers only global inference. Particularly when the research interest is to shed some light on the source of interaction a method that gives further indication about the nature of the significant interaction effects is required.

The presented approach provides simultaneous inference for all tetrad interaction contrasts to detect the source of interaction [1]. Since many experiments involve a complex treatment structure of at least one factor (i.e. a control group is added to a treatment factor or the levels are ordered) the choice of all tetrad contrasts seems to be unsuitable. Therefore the contrasts should be chosen on the nature of the structure of the experimental design. Taking the treatment structure into account leads to user specified interaction contrasts. In addition to adjusted p-values we recommend the use of simultaneous confidence intervals to present the direction and magnitude and, possibly, the relevance of the comparison of interest [2]. The proposed approach is applied to an illustrative example from a horticultural trial [3].

References

[1] K. R. Gabriel, J. Putter, and Y. Wax (1973). Simultaneous Confidence Intervals for Product-Type Interaction Contrasts, *Journal of the Royal Statistical Society. Series B (Methodological)* **35,:**234–244.

[2] T. Hothorn, F. Bretz and P. Westfall (2008). Simultaneous inference in general parametric models, *BIOMETRICAL J,* **50:**346–363.

[3] R. G. Petersen (1985). *Design and analysis of experiments,* Dekker, New York, NY.

List of Contributors

ISBN 978-3-033-03083-1