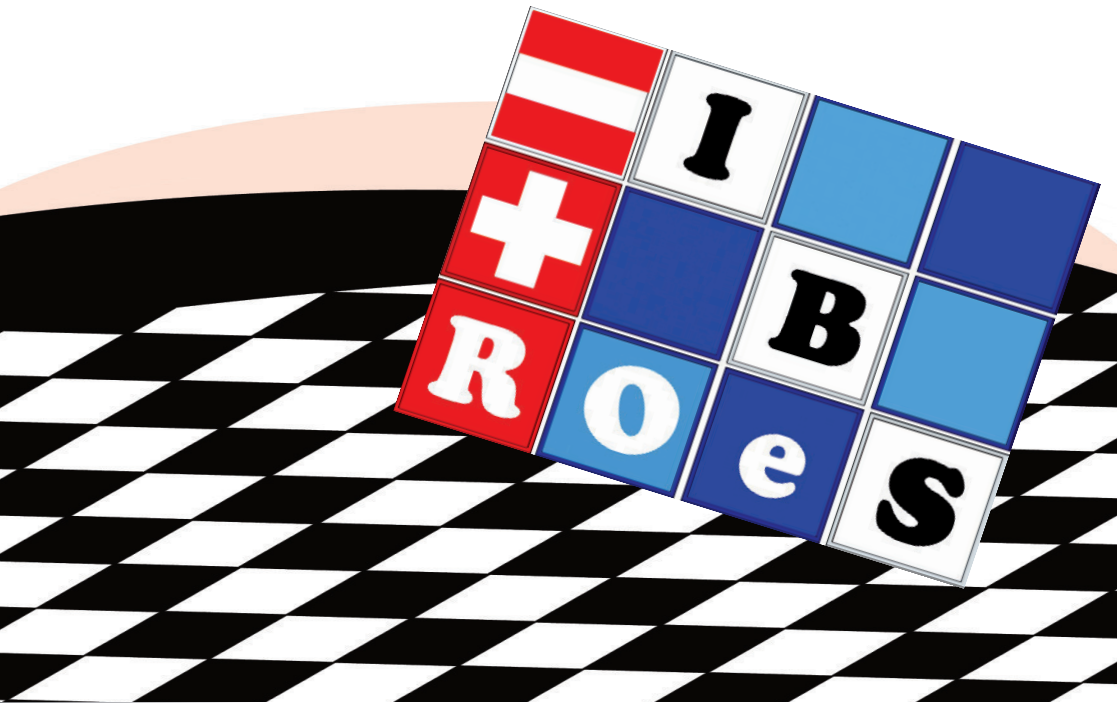
Department of Medical Statistics,
Informatics and Health Economics
Innsbruck Medical University

Sabrina B. Neururer
Hanno Ulmer
(editors)

ROeS 2013

9th - 12th September 2013. Dornbirn, Austria

Conference Program
Conference Proceedings



Sabrina Barbara Neururer
Hanno Ulmer
(editors)

ROeS 2013

9^{th} – 12^{th} September 2013. Dornbirn, Austria

Conference Program

Conference Proceedings

organized by

Department of Medical Statistics,
Informatics and Health Economics

Innsbruck Medical University

Dear Participant,

It is our great pleasure to welcome you to Dornbirn for this conference organized by the Department of Medical Statistics, Informatics and Health Economics of Innsbruck Medical University. The conference has brought together leading scientists from academia and industry across the disciplines of biostatistics, mathematical statistics, epidemiology, as well as clinical trials and promises to be a highly interactive event.

This conference features an exceptional program that includes the latest developments. It offers excellent networking and collaboration opportunities for scientists from a variety of research fields.

The conference venue, Dornbirn, the largest town in the Austrian state of Vorarlberg, benefits from its favorable location in a diverse cultural and natural setting, close to Liechtenstein, Switzerland, and Germany. It is a friendly, lively small city which has been able to keep its cozy traditional rustic character. This location comprises a perfect combination of a picturesque town in an impressive landscape.

On behalf of all of who contributed to the organization of this conference we would like to thank all our speakers, financial supporters, reviewers, and attendees, and extend a warm welcome to you in Dornbirn.

We hope you will enjoy your stay.

With best wishes,

Sabrina B. Neururer
Conference Director

Hanno Ulmer
ROeS President

SCIENTIFIC PROGRAM COMMITTEE

Hanno Ulmer (chair)
Andrea Berghold
Michael Branson
Frank Bretz
Hans Ulrich Burger
Michael Edlinger
Ekkehard Glimm
Georg Göbel
Georg Heinze
Leonhard Held
Claudia Lamina
Willi Maurer
Martina Mittlböck
Franz König
Barbara Kollerits
Karl Peter Pfeiffer
Martin Posch
Regina Riedl
Valentin Rousson
Shu-Fang Hsu Schmitz
Helga Wagner

LOCAL ORGANIZING COMMITTEE

Sabrina Barbara Neururer (chair)
Josef Fritz
Lalit Kaltenbach
Joachim Masser

REVIEWERS

We thank the following reviewers for their contribution and support of the ROeS 2013

Andrea Berghold
Michael Branson
Frank Bretz
Hans Ulrich Burger
Michael Edlinger
Josef Fritz
Ekkehard Glimm
Georg Göbel
Georg Heinze
Leonhard Held
Shu-Fang Hsu Schmitz
Lalit Kaltenbach
Barbara Kollerits
Franz König
Claudia Lamina
Joachim Masser
Willi Maurer
Martina Mittlböck
Sabrina Barbara Neururer
Martin Posch
Regina Riedl
Valentin Rousson
Hanno Ulmer
Helga Wagner

GENERAL INFORMATION

Venue

The conference venue is at University of Applied Sciences Vorarlberg (Fachhochschule Vorarlberg), 1 Hochschulstrasse, Dornbirn 6850, Austria.

Cloakroom

There is limited space for storing coats and bags at the venue. Items are left at own risk.

Security

The delegate badge must be worn at all times when inside the conference venue.

Smoking

Please refrain from smoking anywhere on the premises. For smoking please visit the venue forecourt.

Refreshments

There will be refreshment breaks each day. Refreshments will be provided in the foyer. Please refer to the program for the break times. For lunch you are free to visit the "Mensa" of the conference venue at your own expense.

Social Program

Please also note the Get Together, the Poster Session including finger foods and the Conference Dinner & Excursion (see Social Program). Do not forget to register for these events for there is only limited space. In case, you have not registered yet, you can check at the conference desk, whether there are still places available.

Wireless network

A wireless network is available. You can find the access information in your conference folder.

Posters

The conference posters will be displayed throughout the whole conference in the foyer. We kindly ask all poster presenters to be available for questions while refreshment breaks. For details about the poster session, please refer to the program.

Certificates of Attendance

Upon having attended the ROeS 2013, delegates will receive a certificate of attendance. I will be handed over to them at the conference desk.

TABLE OF CONTENTS

PROGRAM

1. SCIENTIFIC PROGRAM

Monday, 9th September 2013

Time	Name	Room
09:00 – 10:30	Comparing Clinical Measurements I (Short Course)	W2 11/12
09:00 – 10:30	Statistical Methods in Dose Finding I (Short Course)	W2 06
10:30 – 11:00	Coffee Break	Foyer
11:00 – 12:30	Comparing Clinical Measurements II (Short Course)	W2 11/12
11:00 – 12:30	Statistical Methods in Dose Finding II (Short Course)	W2 06
12:30 – 13:30	Lunch Break	-
13:30 – 14:00	Opening Event	W2 11/12
14:00 – 15:30	Regulatory Topics and Guidelines I	W2 11/12
14:00 – 15:30	Sub-Group Analysis	W2 06
15:30 – 16:00	Coffee Break	Foyer
16:00 – 17:30	Regulatory Topics and Guidelines II	W2 11/12
16:00 – 17:30	Multiple Comparisons	W2 06

09:00 – 10:30 o'clock **Room W2 11/12**

Comparing Clinical Measurements I (Short Course)

Short Course (Part 1): Bendix Carstensen

* Statistical Analysis of Method Comparison Studies (B. Carstensen)

09:00 – 10:30 o'clock **Room W2 06**

Statistical Methods in Dose Finding I

Short Course (Part 1): Byron Jones, Björn Bornkamp

* Statistical Methods in Dose Finding (B. Jones, B. Bornkamp)

11:00 – 12:30 o'clock **Room W2 11/12**

Comparing Clinical Measurements II

Short Course (Part 2): Bendix Carstensen

* Statistical Analysis of Method Comparison Studies (B. Carstensen)

11:00 – 12:30 o'clock **Room W2 06**

Statistical Methods in Dose Finding II

Short Course (Part 2): Byron Jones, Björn Bornkamp

* Statistical Methods in Dose Finding (B. Jones, B. Bornkamp)

13:30 – 14:00 o'clock **Room W2 11/12**

Opening Event

- Welcome Address: Hanno Ulmer (ROeS President)
- Welcome Address: Andrea Kaufmann (Mayoress Dornbirn)
- Welcome Address: John Hinde (IBS President)
- Official Conference Opening: Hanno Ulmer (ROeS President)

14:00 – 15:30 o'clock **Room W2 11/12**

Regulatory Topics and Guidelines I

Chair(s): Franz König | Willi Maurer

- EMA Guidance Documents: Overview of Recent Developments (T. Lang)
- Diversity of Regulatory Guidelines in the Absence of Harmonization (G. Rosenkranz)
- Statistical Challenges in Active Controlled Trial Designs and Multiple Comparisons in Regulatory Applications (H.J. Hung)

14:00 – 15:30 o'clock **Room W2 06**

Sub-Group Analysis

Chair(s): Thomas Jaki

- Enrichment designs for the development of personalized medicines (A. Graf, F. König, M. Posch)
- A framework to assess the added value of subgroup analyses when the overall treatment effect (TE) was significant (H. Sun, W. Vach)
- Subgroup Selection in Clinical Trials (S. Strohmaier, T. Jaki)
- Performance Characteristics of Biomarker-Based Patient Selection Rules in Adaptive Enrichment Designs (J. Krisam, M. Kieser)

16:00 – 17:30 o'clock **Room W2 11/12**

Regulatory Topics and Guidelines II

Chair(s): Franz König | Willi Maurer

- Regulatory Guidelines as a Starting Point for Methodological Research – The Case of Bioequivalence (M. Kieser, S. Schnaidt)
- The Changing Environment for Drug Development and Drug Licensing in Europe after ICH (J. Röhmel)

16:00 – 17:30 o'clock **Room W2 06**

Multiple Comparisons

Chair(s): Thomas Jaki

- Statistical Criteria for Components of Primary Composite Endpoint to Support Benefit-Risk Assessment at an Interim Analysis Using Blinded Data (J. Brumm)
- A Parametric Rüger Test and Diagnostic Trials (R. Ristl, C. Gartner, T. Lang, M. Posch)
- Test Procedures and Efficacy Claims for Composite Endpoints and its Components (S. Schüler, G. Rauch, M. Kieser)
- Consistency-Adjusted Alpha Allocation Methods for Composite Endpoints (G. Rauch, M. Kieser)

Tuesday, 10th September 2013

Time	Name	Room
09:00 – 10:30	Recent Advances in Time-to-Event Models	W2 11/12
09:00 – 10:30	Health Technology Assessment (HTA)	W2 06
10:30 – 11:00	Coffee Break	Foyer
11:00 – 12:30	Multiregional Trials	W2 11/12
11:00 – 12:30	Multistate Models – Cure Models	W2 06
12:30 – 14:00	Lunch Break	-
14:00 – 15:30	Flexible and Adaptive Designs	W2 11/12
14:00 – 15:30	Analysis of Registries and Large Databases	W2 06
15:30 – 16:00	Coffee Break	Foyer
16:00 – 17:30	Growth Curves	W2 11/12
17:30 – 18:00	Statistics in Drug Development	W2 11/12
18:00 – 20:00	Poster Session (with Finger foods)	Foyer

09:00 – 10:30 o'clock **Room W2 11/12**

Recent Advances in Time-to-Event Models

Chair(s): Georg Göbel

- Understanding Individual Heterogeneity: What is the Potential of the Frailty Approach (O.O. Aalen)
- Practical Implications of Competing Risk Models – How to Interpret the Results to Clinicians (M. Smastuen)
- Understanding Treatment Mechanisms in Clinical Survival Trials – A Causal Approach (S. Strohmaier, T. Pedersen, T. Lange, Ø. Borgan, O.O. Aalen)

09:00 – 10:30 o'clock **Room W2 06**

Health Technology Assessment (HTA)

Chair(s): Hans Ulrich Burger

- Biometrical Issues in Health Technology Assessment (R. Bender)
- Designing Drug Development to Optimize HTA Submissions (K. Facey)
- Statistics in Health Technology Assessment: Issues and Challenges (Y. Castro)
- Measuring an Effect Size from Dichotomized Data: Contrasted Results whether Using a Correlation or an Odds-Ratio (V. Rousson)

11:00 – 12:30 o'clock **Room W2 11/12**

Multiregional Trials

Chair(s): Michael Branson | Hans Ulrich Burger

- Multiregional Clinical Trials: Introduction to Basic Features and Discussion of Special Issues (B. Jones, P. Gallo)
- Heterogeneity in Multiregional Studies (J. Röhmel)
- Recent Statistical Development in Global Clinical Trial Strategy (H.J. Hung)

11:00 – 12:30 o'clock **Room W2 06**

Multistate Models – Cure Models

Chair(s): Georg Göbel | Martina Mittlböck

- Cure Models Based on Univariate and Bivariate Random Effects (A. Wienke)
- A New Approach Based on Pseudo-Values for Assessing the Effect of a Binary Time-Dependent Covariate on Patient Survival (U. Pötschger, H. Heinzl, A. Attarbaschi, M. Minkov, M. Mittlböck)
- Multistate Models with Multiple Time Scales (B. Carstensen, S. Iacobelli)

14:00 – 15:30 o'clock **Room W2 11/12**

Flexible and Adaptive Designs

Chair(s): Franz König

- Treatment Selection in Multi-Arm, Multi-Stage Clinical Trials (T. Jaki, D. Magirr, N. Stallard)
- Case Studies for Adaptive Treatment Arm Selection and Population Enrichment Designs (G. Wassmer)
- Type I Error Rate Control in Adaptive Clinical Trials with blinded interim Analysis (M. Posch, M. Malina)
- Population Enrichment Designs for Molecularly Targeted Therapies in Oncology (C. Mehta)

14:00 – 15:30 o'clock **Room W2 06**

Analysis of Registries and Large Databases

Chair(s): Andrea Berghold

- Analysis of Large Registry-Based Study of Prescription Drugs and Road Traffic Crashes (M. Avalos, L. Orriols)
- Mammography Screening and Cancer Registration in Tyrol/Austria (W. Oberaigner)
- Observational Studies, Matching and Propensity Scores: An Application to Colorectal Cancer Data (J. Hinde, C. Dooley)

16:00 – 17:30 o'clock **Room W2 11/12**

Growth Curves: GAMLSS, SITAR and Austrian Kids

Chair(s): Georg Heinze

- The Use of GAMLSS in Growth Curve Estimation (D. Stasinopoulos)
- Development of New Austrian Height and Weight References (A. Gleiss, G. Häusler, M. Schemper)
- SITAR Growth Curve Analysis to Evaluate Interventions and Life Course Outcomes (T. Cole)

17:30 – 18:00 o'clock **Room W2 11/12**

Statistics in Drug Development

Chair(s): Hanno Ulmer

- Developing Drugs for Reimbursement – Current and Future Challenges for Statisticians (M. Scott)
- Company presentation of Numerus Limited

18:00 – 20:00 o'clock **Foyer**

Poster Session (with Finger foods)

Chair(s): Hanno Ulmer | Sabrina Neururer

- Camylobacter in Human and Chicken (W. Wei, L. Held, G. Schuepbach)
- "Randomizer for Clinical Trials" – A Web-Based Randomization Service is On-Line for 10 Years (P. Ofner-Kopeinig, M. Errath, A. Berghold)

- Simultaneous Confidence Intervals for Adaptive Graph-Based Multiple Test Procedures (F. Klinglmueller)
- Non-Parametric Control Chart Approach and Visualisation for the Surveillance of Notifiable Infectious Diseases in Baden-Wuerttemberg (I. Zoellner, F. Koehler)
- Bayesian Adaptive Randomization in Early Development Oncology (P. Vlachos)
- The Applicability of Site-Specific Proportion Cured Models in the Small Cancer Registry of Tyrol (M. Edlinger, H. Ulmer, M. Smastuen, W. Oberaigner)
- The Use of Antitranspirants with Aluminium Salts and its Relation to Breast Cancer (C. Linhart, N. Concin, S. Taucher, E. Morandi, J. Kowalski, H. Ulmer)
- Statistical Methods in the Metabolic Syndrome and Cancer Project (Me-Can) (H. Ulmer, E. Edlinger, S. Strohmaier, H. Jonsson, T. Stocks, T. Bjorge, J. Manjer, G. Nagel, P. Stattin)
- Prognostic Scores to Support Decision Making for Women with Breast Cancer: Results of a Retrospective Data Analysis (M. Arvandi, B. Jahn, H. Fiegl, G. Göbel, U. Siebert)
- Modeling Repeated Observations in Longitudinal Studies with Mixed-Effects Models and Survival Analysis (J. Fritz, H. Ulmer)

Wednesday, 11th September 2013

Time	Name	Room
09:00 – 10:30	Risk Prediction I	W2 11/12
09:00 – 10:30	Mixed Data Analysis	W2 06
10:30 – 11:00	Coffee Break	Foyer
11:00 – 12:30	Risk Prediction II	W2 11/12
11:00 – 12:30	Young Statisticians	W2 06
12:30 – 13:15	Arthur Linder Award	W2 11/12
13:15 – 13:45	Lunch Break	Foyer
13:45 – 14:45	ROeS General Assembly	W2 11/12

09:00 – 10:30 o'clock **Room W2 11/12**

Risk Prediction I

Chair(s): Michael Edlinger

- Shrinkage of Regression Coefficients: Old Concepts and New Ideas (D. Dunkler, G. Heinze)
- On Criteria for Evaluating Risk Prediction Models for Public Health Applications (R. Pfeiffer)
- Assessing the Incremental Predictive Value of Markers: Towards Modern Reclassification Measures and Graphical Displays (E. Steyerberg)

09:00 – 10:30 o'clock **Room W2 06**

Mixed Data Analysis

Chair(s): Helga Wagner

- Joint Analysis of Mixed Outcome Data: Issues and Challenges (K. Chough)
- Hierarchical Modeling of Endpoints of Different Types with Generalized Linear Mixed Models (C. Faes)
- Updating the Probability of Success of Clinical Development Programs Over Time (M. Abt, P. Jordan)

11:00 – 12:30 o'clock **Room W2 11/12**

Risk Prediction II

Chair(s): Michael Edlinger

* Changes of Body Mass Index in Relation to Mortality: Results of a Cohort of 42,099 Adults (G. Nagel, K. Rapp, H. Ulmer, H. Concin, J. Klenk)
* Evaluating Disease Predictors with Repeated Measurements: A Longitudinal Study of Alzheimer's Disease (L. Brant)
* Some Applications of Risk Modeling in Breast Cancer Prevention (M. Gail)

11:00 – 12:30 o'clock **Room W2 06**

Young Statisticians

Chair(s): Andrea Berghold

* Efficient Data-Augmented MCMC Methods for Binomial Logit Models (A. Fussl, S. Frühwirth-Schnatter, R. Frühwirth)
* Problems of INLA in Generalized Linear Mixed Models for Binary Responses (R. Sauter, L. Held)
* Analysis of Association on Non-Product Spaces (A. Klimova)
* Statistical Methods for Improvement of Motif Finding Algorithms (Ž Stepančič)
* Overrunning Data Methods: Comparisons Based on Real Data Trials (N. Soriani, I. Baldi, D. Gregori)

12:30 – 13:15 o'clock **Room W2 11/12**

Arthur Linder Award

Chair(s): Hanno Ulmer

* Blood Transfusions and the Subsequent Risk of Cancers in the United States Elderly (R. Riedl, E.A. Engels, J.L. Warren, A. Berghold, W. Ricker, R. Pfeiffer)
* Adaptive Graph-Based Multiple Testing Procedures (F. Klinglmueller, F. König, M. Posch)

13:45 – 14:45 o'clock **Room W2 11/12**

ROeS General Assembly

- Welcome Address (Hanno Ulmer, ROeS President)
- Report on bi-annual Activities (Hanno Ulmer)
- Report from IBS (Andrea Berghold)
- Reports from ROeS Sections
    - Wiener Biometrische Sektion (Gerhard Svolba)
    - Basler Biometrische Sektion (Michael Branson)
    - Sektion Kärnten-Steiermark (Andrea Berghold)
- Financial Report (Shu-Fang Hsu Schmitz)
- Report from the Audit Committee (Franz König, Hans Ulrich Burger)
- Elections
- Matters related to ROeS
    - Vereinsgründung Österreich
    - MCP
- Next ROeS/CEN meetings (designated ROeS president)

Thursday, 12th September 2013

Time	Name	Room
09:00 – 10:30	Statistical Methods in Genetic Epidemiology	W2 11/12
09:00 – 10:30	Statistical Methods in Infectious Disease Epidemiology	W2 06
10:30 – 11:00	Coffee Break	Foyer
11:00 – 12:30	Statistical Evaluation of Biomarkers and Clinical Tools	W2 11/12
11:00 – 12:30	Biometrics – From Bench to Trials	W2 06
12:30 – 13:15	Closing Event	W2 11/12
13:30 – 15:00	Group-Sequential and Adaptive Design for Confirmatory Trials using East (Workshop)	W2 11/12

09:00 – 10:30 o'clock **Room W2 11/12**

Statistical Methods in Genetic Epidemiology

Chair(s): Claudia Lamina

* Optimal Strategies to Detect Stratum-Sensitive Genetic Effects (I. Heid)
* Max-test to Evaluate Genetic Association Studies for Continuous and Time-to-Event Traits (L. Hothorn, E. Herberich)
* Using Weighted Genetic Predisposition Scores to Evaluate Gene-Environment Interaction Effects on Lipid Levels (C. Lamina, L. Forer, S. Schönherr, B. Kollerits, J. Ried, C. Gieger, A. Peters, H.E. Wichmann, F. Kronenberg)
* Using the LASSO and Ridge Regression in Case-Control Studies with Zero-Inflated Predictors (M. Kohl, G. Heinze)

09:00 – 10:30 o'clock **Room W2 06**

Statistical Methods in Infectious Disease Epidemiology

Chair(s): Leonhard Held

* Estimating MMR Vaccine Coverage from Australian Sero-Surveillance Data (N. Hens, J. Wood, N. Goeyvaerts)
* Recent Development in Identifying Transmission Routes of Healthcare Associated Infections using Whole Genome Sequence Data (T. Kypraios)
* Sample Size Calculations and the Temporal Dependency between an Infection and its Sequela in Randomised Controlled Trials (S. Herzog, A. Berghold)
* Modelling Power-Law Spread in Infectious Diseases (S. Meyer, L. Held)

11:00 – 12:30 o'clock **Room W2 11/12**

Statistical Evaluation of Biomarkers and Clinical Tools

Chair(s): Barbara Kollerits

* Development in Prostate Cancer Risk Predition Tools in Response to Changes in Clinical Landscape (D. Ankerst)
* Comparison and Evaluation of Cardiac Biomarkers in Patients with Intermittent Claudication: Results from the CAVASIC Study (B. Kollerits, G. Sturm, C. Lamina, A. Hammerer-Lercher, B. Rantner, M. Stadler, T. Ziera, J. Struck, P. Klein-Weigel, G. Fraedrich, F. Kronenberg)
* Stopping Rules for Sequential Trials in High-Dimensional Data (S. Zehetmayer, M. Posch, A. Graf)

11:00 – 12:30 o'clock **Room W2 06**

Biometrics – From Bench to Trials

Chair(s): Shu-Fang Hsu Schmitz

* A Novel Method to Estimate the Minimum Effective Dose for Monotone and Non-Monotone Dose-Response Relationships (M.J. Wolfsegger, G. Gutjahr, W. Engl, T. Jaki)
* Nonparametric Multivariate Density Estimation Using Mixtures (X. Wang, Y. Wang)
* The Use of Propensity Scores in Observational Trials (J. Alsop)
* Double Criteria Design for a Minimaly Important Difference in Single-Arm Single-Stage Phase II Trials with Binary Primary Endpoint (S.-F. Hsu Schmitz)

12:30 – 13:15 o'clock **Room W2 11/12**

Closing Event

13:30 – 15:00 o'clock **Room W2 11/12**

Group-Sequential and Adaptive Design for Confirmatory Trials using East

2. SOCIAL PROGRAM

Monday, 9th September 2013

Start Time	Name	Room
17:30	Get Together	Foyer

Wednesday, 11th September 2013

Start Time	Name	Info
16:00	Excursion (Walking-Tour to "Karren")	hikers
17:40	Meeting at Cable Car "Karren"	cable car users
18:00	Dinner	for all delegates

To attend our conference dinner and excursion you have to be registered in advance. Only limited seats are available.

There are two different possibilities to reach the "Panoramarestaurant Karren" high above Dornbirn, which provides a breathtaking view over Lake Constance and the surrounding area and where our dinner will take place.

1. Walking-Tour
   You can reach the "Panoramarestaurant Karren" by a 90-minutes walking-tour. Therefore you need sturdy shoes and be in quite a good shape for you have to pass approximately 530 meters in height.
   Meeting point: Foyer of the conference venue
   Meeting time: 16:00 o'clock

2. Cable Car
   The "Karren Seilbahn" (Karren cable car) will take you up to the "Panoramarestaurant Karren" in approximately 10 minutes.
   Meeting point: Lower Terminus of "Karren Seilbahn"
   Meeting time: 17:40 o'clock

For further information about the restaurant and the cable car see www.karren.at.

PROCEEDINGS

(sorted by first name of first author)

1. INVITED SPEAKER ABSTRACTS

DEVELOPMENT OF NEW AUSTRIAN HEIGHT AND WEIGHT REFERENCES

Andreas Gleiss, Medical University of Vienna, Vienna, Austria
Gabriele Häusler, Medical University of Vienna, Vienna, Austria
Michael Schemper, Medical University of Vienna, Vienna, Austria

In the clinical evaluation of growth disorders and growth monitoring, sex and age dependent height reference curves derived from a relatively small and special sample of Swiss children between 1954 and 1970 have been in use in Austria so far [1]. Due to the secular trend and potential national differences new reference curves needed to be developed on a broad national basis.

For this purpose a sample of nearly 14500 children and adolescents between 4 and 19 years of age was drawn via schooling institutions [2]. Sampling was stratified by provinces taking into account the age and sex specific population proportions in each province. An existing R implementation for Generalized Additive Models for Location, Scale and Shape (GAMLSS) was employed for estimating percentile curves [3]. The flexible Box-Cox Power Exponential distribution was used to describe the distributions of height and weight at each age, while the dependence of the four distributional parameters on age was modelled by cubic spline functions. The degrees of freedom for these splines were selected by an optimization procedure using information criteria. The functional dependence was further adapted, if necessary, according to paediatric or goodness-of-fit considerations.

The various steps in the basic estimation process of reference curves will be presented in detail. The quantification of the uncertainty in extremely low percentiles (such as 0.62%, crucial in diagnostics) using bootstrap methods will be demonstrated. The experiences gained from this case study of growth curve estimation may provide guidance for planning and analysis of future studies.

[1] Prader, A., Largo, R. H., Molinari, L. & Issler, C. 1989. Physical growth of Swiss children from birth to 20 years of age. First Zurich longitudinal study of growth and development. Helvetica paediatrica acta. Supplementum, 52, 1-125.
[2] Gleiss, A., Lassi, M., Blümel, P., Borkenstein, M., Kapelari, K., Mayer, M., Schemper, M., Häusler, G. 2013. Austrian height and body proportion references for children aged 4 to under 19 years. Annals of Human Biology, in press.
[3] Stasinopoulos, D.M., & Rigby, R.A. 2007. Generalized additive models for location scale and shape (GAMLSS) in R. Journal of Statistical Software, 23, 1-46.

CURE MODELS BASED ON UNIVARIATE AND BIVARIATE RANDOM EFFECTS

Andreas Wienke, University of Halle, Halle, Germany

Traditional analysis of time-to-event data is based on the assumption that all individuals are susceptible to the event of interest and will eventually experience this event if the follow-up is sufficiently long. To overcome this strong assumption cure models were introduced into the field of survival analysis. In the first part of the talk univariate cure models based on random effects models (frailty models) will be considered. Here a random latent variable contains the information whether individuals are at risk or never-at-risk, respectively. In the simplest case a binary variable (two-point frailty) is used. The terminology to describe the never-at-risk group varies from field to field. It includes "long-term survivors" or "cured" in epidemiology, "nonsusceptibles" in toxicology, "stayers" in finite Markov transition models of occupational mobility, the "nonfecundable" in fertility models, and "nonrecidivists" among convicted criminals just to mention a few possible applications.

More complex is an approach applying a variable with compound Poisson distribution introduced by Aalen [1]. This distribution has probability mass at zero (cured fraction) and a continuous part which accounts for heterogeneity in the susceptible fraction.

In the second part of the talk a bivariate extension of the compound Poisson model using a correlated frailty approach [2, 3] is presented and illustrated by application to breast cancer data of 5857 Swedish female twin pairs to estimate the size of the susceptible fraction.

[1] Aalen OO (1992) Modelling heterogeneity in survival analysis by the compound Poisson distribution. Annals of Applied Probability 4, 951–972
[2] Wienke A, Lichtenstein P, Yashin A (2003) A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. Biometrics 59, 1178 – 1183
[3] Wienke A, Ripatti S, Palmgren J, Yashin A (2010) A bivariate survival model with compound Poisson frailty. Statistics in Medicine 29, 275–83

STATISTICAL ANALYSIS OF METHOD COMPARISON STUDIES

Bendix Carstensen, Steno Diabetes Center, Gentofte, Denmark

You should attend this course if you ever have come across a clinician that asked you which of two methods are the better for measuring blood-pressure (say). Or a researcher trying to convince you that a correlation of 0.93 is compelling evidence that two methods are in agreement.

The comparison of two methods of measurement using the so-called "Bland-Altman" procedure of plotting the difference against the mean for each pair of observations has become the de facto standard for analysis of method comparison studies without replicates [1, 2]. But still some papers appear that rely on correlations in some form or another. The Bland-Altman procedure produces limits of agreement, that is, prediction limits for the difference between a measurement by one method and a measurement by another. The Bland-Altman approach has been expanded with practical methods for the analysis of situations with replicate measurements and for multiple methods [3].

This half-day course will give an introduction to the standard procedures, and put these in a proper modeling framework. More elaborate designs of method comparison studies, in particular studies with replicate measurements by each method will also be covered [4].

I will illustrate the methods by examples from the literature, and demonstrate the use of the MethComp package for R. There will be opportunities to try out the MethComp package on your own computer, either using the examples or your own data.

---

[1] JM Bland and DG Altman. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, i:307–310, 1986.
[2] JM Bland and DG Altman. Measuring agreement in method comparison studies. Statistical Methods in Medical Research, 8:136–160, 1999.
[3] B Carstensen, J Simpson, and LC Gurrin. Statistical models for assessing agreement in method comparison studies with replicate measurements. International Journal of Biostatistics, 4(1):Article 16, 2008.
[4] B. Carstensen. Comparing Clinical Measurement Methods: A practical guide. Wiley, 2010.

STATISTICAL METHODS IN DOSE FINDING

Björn Bornkamp, Novartis Pharma AG, Basel, Switzerland
Byron Jones, Novartis Pharma AG, Basel, Switzerland

The selection of a dose for confirmatory Phase III trials and potential market authorization is one of the most difficult decisions to make in the whole drug development process. Indeed, the poor selection of dose(s) for confirmatory trials has been a persistent problem in the last decade. One of the main reasons for this is an inadequate knowledge of the dose response relationship (both for efficacy and safety).

In this training we will introduce and discuss methods for dose-response testing and estimation that provide a solid basis for all subsequent dose selection strategies and decisions. We will focus on dose-response modelling techniques (nonlinear regression) and the MCPMod methodology, which is a strategy for designing and analyzing dose-finding studies under uncertainty about the underlying dose-response model.

Case studies, based on real clinical trials, will be used to illustrate the use of the methodology in practice.

We will use the R package DoseFinding, which is publicly available from CRAN, to implement all the methods presented in this training.

Outline of the course:

Session 1: Introduction to dose-finding studies
Session 2: Dose-response modelling
Session 3: MCPMod (Multiple Comparisons and Modelling) methodology
Session 4: Optimal Designs / Adaptive Designs

MULTIREGIONAL CLINICAL TRIALS: INTRODUCTION TO BASIC FEATURES AND DISCUSSION OF SPECIAL ISSUES

Byron Jones, Novartis Pharma AG, Basel, Switzerland
Paul Gallo, Novartis Pharma AG, Basel, Switzerland

We describe the main features of multiregional clinical trials (MRCTs) from statistical and regulatory points of view. It is not uncommon for trials to be run across multiple parts of the globe, for various reasons, including to obtain a desired total number of patients, and to satisfy regulators desiring representation within their areas of responsibility. This raises potential challenges for interpretation, because a treatment may have different effects in different regions, perhaps due to factors of genetics, culture, or local medical practice but false signals of heterogeneity may often arise because of the smaller numbers of subjects within individual regions. Although the analysis and interpretation of an MRCT has some similarities to investigation of other potentially important subgroups, a unique aspect of MRCTs is that regulators very naturally tend to focus on the results that are most relevant to their own regions. The suggestion of a meaningful signal of a differential treatment effect across regions can therefore raise serious concerns. Two MRCTs in which such signals arose, PLATO and MERIT, will be used to illustrate the difficulties of interpreting and explaining trial results in the presence of suggested treatment by region interaction.

HIERARCHICAL MODELING OF ENDPOINTS OF DIFFERENT TYPES WITH GENERALIZED LINEAR MIXED MODELS

Christel Faes, Hasselt University, Diepenbeek, Belgium

In many health surveys a large collection of measurements are recorded, often containing measurements of different types. The most common non-commensurate situation is that of a continuous, often assumed normally distributed outcome, and a binary or ordinal outcome. In addition, it is not uncommon that these measurements are recorded repeatedly in time or that sampling units are clustered. In this presentation, general interest is in the analysis of multivariate hierarchical data, also-called multivariate multi-level data. The strength of association between the different endpoints as well as the association of measurements within the same sampling unit might be important questions of interest. Several approaches to construct a joint model exist. The use of generalized linear mixed models (GLMM) via the specification of shared or correlated random effects and correlated residuals is a flexible methodology in this setting, and will be illustrated using several examples. Extension to more than two outcomes in the GLMM context is straightforward, but computational problems arise as the number of outcomes increases. To avoid maximization of the full likelihood expression, inference can be based the pseudo-likelihood methodology, also called composite likelihood. The idea is to replace the full likelihood by a function that is easier to evaluate to reduce the computational costs. A convenient pseudo-likelihood function in this setting is one that replaces the joint density by the product of all pairwise densities, the so-called pairwise pseudo-likelihood.

POPULATION ENRICHMENT DESIGNS FOR MOLECULARLY TARGETED THERAPIES IN ONCOLOGY

Cyrus Mehta, Cytel Inc., Cambridge, MA, United States of America

The development of molecularly targeted therapies for certain types of cancers (e.g., Vemurafenib for advanced melanoma with mutant BRAF; Cetuximab for metastatic colorectal cancer with KRAS wild type) has led to the consideration of population enrichment designs that explicitly factor-in the possibility that the experimental compound might differentially benefit different biomarker subgroups. In such designs, enrollment would initially be open to a broad patient population with the option to restrict future enrollment, following an interim analysis, to only those biomarker subgroups that appeared to be benefiting from the experimental therapy. While this strategy could greatly improve the chances of success for the trial, it poses several statistical and logistical design challenges. Since late-stage oncology trials are typically event driven, one faces a complex trade-off between power, sample size, number of events and study duration. This trade-off is further compounded by the importance of maintaining statistical independence of the data before and after the interim analysis and of optimizing the timing of the interim analysis. This talk will highlight the crucial role of simulation-guided design for resolving these difficulties while nevertheless maintaining strong control of the type-1 error.

THE USE OF GAMLSS IN GROWTH CURVE ESTIMATION

Dimitrios Stasinopoulos, London Metropolitan University, London, United Kingdom of Great Britain and Northern Ireland

Since the introduction of GAMLSS, Rigby and Stasinopoulos [1], the models have been used in a variety of applied fields one of each is the construction of reference charts for child growth curves, Borghi et al. [2], Cole et al. [3] In this area GAMLSS complements the standard LMS method of Cole and Green [4] used widely for the construction of growth curves, see Rigby and Stasinopoulos [5, 6].

In this talk, after a general overview of the GAMLSS models, the authors will describe his experience with modelling growth curves and in particular with the following problems related to growth curve fitting:

i) the automatic selection of the degrees of freedom of the smoothers
ii) the different diagnostic tools for checking the adequacy of the fitted model
iii) the comparison of the parametric LMS methods to non-parametric quantile regression techniques
iv) the construction of growth curves for censored/truncated data.

[1] Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape, (with discussion). Appl. Statist., 54: 507–554.
[2] Borghi, E. de Onis, M. Garza, C. Van den Broeck, J.E. Frongillo, E. A. Grummer-Strawn, L. Van Buuren, S. Pan, H. Molinari, L. Martorell, R. Onyango, A. W. and Martines, J. C. (2006). Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. Statistics in Medicine, 25: 247–265.
[3] Cole, T., Stanojevic, S., Stocks, J., Coates, A., Hankinson, J. and Wade, A. (2009). Age-and size-related reference ranges: A case study of spirometry through childhood and adulthood. Statistics in Medicine, 28: 880–898.
[4] Cole, T. J. and Green, P. J. (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. Statistics in Medicine., 11: 1305–1319.
[5] Rigby, R. A. and Stasinopoulos, D. M. (2004). Smooth centile curves for skew and kurtotic data modelled using the Box-Cox Power Exponential distribution. Statistics in Medicine, 23: 3053–3076.
[6] Rigby, R. A. and Stasinopoulos, D. M. (2006). Using the Box-Cox t distribution in GAMLSS to model skewness and kurtosis. Statistical Modelling, 6: 209–229.

DEVELOPMENTS IN PROSTATE CANCER RISK PREDICTION TOOLS IN RESPONSE TO CHANGES IN CLINICAL LANDSCAPE

Donna Ankerst, Technical University Munich, Garching b. Munich, Germany

The clinical landscape surrounding early detection for prostate cancer has experienced changes over the past several years, necessitating corresponding responses for prostate cancer prediction tools. The primary change has been recognition of the significant amount of over-detection (detection of cancers that would not have led to mortality in the patient's lifetime and whose treatment induces harm), necessitating the distinction of clinically-significant from non-significant prostate cancer as outcomes. The prostate biopsy technique has moved from a standard of 6-cores to a 12-core sampling scheme, increasing the likelihood of detection of both prostate cancer and high-grade prostate cancer. Two of the leading prostate cancer prediction tools are built on primarily 6-core biopsy cohorts and may need recalibration. A large fraction of unexplained variability among international cohorts in prevalence of prostate cancer and operating characteristics of risk prediction tools has been characterized that needs to be reckoned with by future models. Finally, new markers that have successfully passed the development pipeline into the market require incorporation into modern risk prediction tools. In this talk, I address recent and ongoing updates by the prostate cancer risk community in response to these challenges.

ASSESSING THE INCREMENTAL PREDICTIVE VALUE OF MARKERS: TOWARDS MODERN RECLASSIFICATION MEASURES AND GRAPHICAL DISPLAYS

Ewout Steyerberg, Erasmus MC, Rotterdam, Netherlands

New markers may improve prediction of diagnostic and prognostic outcomes. We aimed to review various measures an graphical displays to assess the incremental predictive value of markers over standard, readily available characteristics.

Widely used traditional measures include the area under the receiver operating characteristic (ROC) curve (AUC). We evaluated the incremental value of adding HDL to a Cox regression model in 3264 subjects from the Framingham study to predict 10-year risk of coronary heart disease (n=183 events, 5.6%, [1]). The AUC difference (ΔAUC) between a model with and without HDL was small in numerical value (0.012 for adding HDL with continuous risk; 0.029 with 20% threshold).

A new but already quite popular measure for incremental value is the Net Reclassification Improvement (NRI). Using a 20% threshold to classifiy subjects as high risk, the 2 NRI components are the net percentages of correctly reclassified patients with events (11/183=6.0%) and without events (-5/3081=-0.2%). Their sum is the NRI (0.058, which is higher in numerical value than ΔAUC, 0.029).

Another recent measure is the net benefit (NB), which is based on decision analysis. The NB is the net fraction of true-positive (TP) classifications penalized for false-positive (FP) classifications: NB= (TP – w FP)/N, with w defined by the odds of harms:benefits, or odds (classification threshold). With a threshold of 20%, w = 0.2 / (1-0.2)=0.25. ΔNB can hence be calculated as (11 – 0.25*5) / 3264=0.30%.

For better understanding of the NRI and NB, we propose a “Net Reclassification Risk” graph. This simple graph allows us to focus on the number of patients and event rates of the 2 reclassified groups: those reclassified from high to low risk (H/L, n=29, 10% event rate) and those reclassified from low to high risk (L/H, n=45, 31% event rate). We note that at most 45*0.31 – 29*0.10 = 11 events can be extra identified (ΔTP=11, NRIevents 11/183, 6%). The negative side is that 45*(1-0.31) – 29*(1-0.10) = 5 extra overtreatments are expected (ΔFP=-5, NRInonevents -5/3081, -0.2%). The burden of overtreatment is explicitly weighted by 0.25 in the NB calculation, leading to the 0.30% estimate ((11-1.25)/3264=0.30%).

We conclude that various traditional and more modern measures can be used to quantify the incremental predictive value of a marker. Important insights may be obtained by a simple graph for the Net Reclassification Risk (‘NRR graph’).

[1] Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157-72

Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011;30:11-21
Steyerberg EW, Pencina MJ, Lingsma HF, Kattan MW, Vickers AJ, Van Calster B. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. Eur J Clin Invest 2012;42:216-28
Van Calster B, Vickers AJ, Pencina MJ, Baker SG, Timmerman D, Steyerberg EW. Evaluation of Markers and Risk Prediction Models: Overview of Relationships between NRI and Decision-Analytic Measures. Med Decis Making 2013;33:490-501
Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565-74

DIVERSITY OF REGULATORY GUIDELINES IN THE ABSENCE OF HARMONIZATION

Gerd Rosenkranz, Novartis Pharma AG, Basel, Switzerland

The ICH process may have been lengthy and tedious, but when it came to fruition it delivered a series of guidelines that were applicable in what were then the three major areas of pharmaceutical drug development. Today this process seems to have come to a standstill and many local or regional health authorities are back to setting their own priorities for the creation of new guidance documents. Recent examples from a statistical perspective comprise the guidelines on adaptive designs or the handling of missing data.

To give an impression of potential consequences of an independent guideline development model, we present two examples. One is concerned with an area that sounds very much standard, bioequivalence, where the variations between guidelines of different regions are in the details but numerous. The other example reminds of the situation that studies may have to be designed and analyzed according to different primary endpoints to satisfy regulatory needs of different regions.

CASE STUDIES FOR ADAPTIVE TREATMENT ARM SELECTION AND POPULATION ENRICHMENT DESIGNS

Gernot Wassmer, Aptiv Solutions, Germany

In confirmatory adaptive designs, at interim stages it is possible to select and de-select treatment arms or pre-defined subpopulations where the latter may be applied in population enrichment designs. Control of familywise Type I error rate in the strong sense is possible through the application of the combination testing principle together with the closed testing principle (e.g., [1]). Using these principles, the way of how to perform the treatment arm or subset selection and the sample size recalculation needs not to be pre-specified. 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. We provide case studies where through extensive simulations the operating characteristics of the designs were assessed.

[1] Bretz, F., König, F., Brannath, W., Glimm, E., Posch, M.: Tutorial in Biostatistics: Adaptive designs for confirmatory clinical trials. Statistics in Medicine 28: 1181-1217, 2009.

STATISTICAL CHALLENGES IN ACTIVE CONTROLLED TRIAL DESIGNS AND MULTIPLE COMPARISONS IN REGULATORY APPLICATIONS

H.M. James Hung, US Food and Drug Administration, Silver Spring, United States of America

Active controlled trial designs and multiple comparisons are often encountered in regulatory applications. The draft guidance documents developed by US FDA on non-inferiority trials and multiple endpoints present many challenging issues and unsettled evolving subjects. A fundamental problem with an active controlled trial design without a placebo arm is one of whether a non-inferiority margin can ever be determined (related to assay sensitivity and constancy of the effect of the selected active control). In many practices, the feasibility of a clinical trial often supersedes the consideration of testing the doses of a test treatment with multiple endpoints under the principle of controlling studywise type I error rate. On the other hand, there are scenarios where controlling studywise type I error alone may not be sufficient. In this talk, I shall revisit some essential features laid out in the FDA draft guidance documents regarding the two topics, highlight the cores of the aforementioned challenges and suggests possible approaches to meet these challenges.

RECENT STATISTICAL DEVELOPMENT IN GLOBAL CLINICAL TRIAL STRATEGY

H.M. James Hung, US Food and Drug Administration, Silver Spring, United States of America

More and more clinical trials in regulatory applications are multi-regional or global trials. Literature has provided abundant research work devoted to the topic of multi-regional trial in the last few years. For this topic, interestingly, statistical development is almost always a forefront in many aspects, such as developing statistical models for quantifying the concept of consistency or inconsistency. On the other hand, the problem of regional consistency is regarded as a subgroup problem. The matter of the fact is that the consistency-inconsistency analysis is most often a post hoc exploration. In this talk, I shall discuss about the recent advances in statistical modeling for consistency versus inconsistency and their impacts. A number of future research tasks will be stipulated.

OPTIMAL STRATEGIES TO DETECT STRATUM-SENSITIVE GENETIC EFFECTS

Iris Heid, University of Regensburg, Regensburg, Germany

While genome-wide association meta-analyses have clearly been a milestone in the understanding of the genetics of complex diseases, current efforts tackle more complex questions like whether or not there are sex-differences in genetic effects. Further questions are whether restricting to allegedly more homogeneous subgroups would evolve subgroup-specific effects. These questions can be generalized to the question of detecting stratum-sensitive genetic effects. This work presents current and novel approaches to screen for stratum-sensitive effects and evaluates them with regard to type 1 error and power distinguishing between one-stage and two-stage approaches. Several approaches including the screening for the minimum of the stratified p-values, the p-value of between-strata difference, a pseudo-joint test p-value and the naive approach of screening the p-value overall strata followed by a test of between-strata difference were evaluated. Surprisingly, the naive approach of screening for the p-value overall strata followed by a test of between-strata difference turned out to be the optimal strategy to detect stratum-sensitive effects that do not point into opposite direction. The test for between-strata difference is the optimal strategy to detect between-strata difference of opposite effect direction. The fact that none of the opposite direction effects have been detected might reflect the lack of systematic screens, the lack of use of the optimal strategy, or the lack of such effects in general. Future studies applying the optimal strategies will further our knowledge upon the full extent of stratum-sensitive genetic effects as well as upon the existence of opposite effect direction signals.

THE CHANGING ENVIRONMENT FOR DRUG DEVELOPMENT AND DRUG LICENSING IN EUROPE AFTER ICH

Joachim Röhmel, Bremen, Germany.

The International Conference on Harmonisation in the last decade of the 20th century was a strong initiative by regulators and pharmaceutical drug developers from Europe, Japan, and the USA. This initiative aimed at the harmonisation of drug development, a process starting with discovery and covering all steps up to the successful license for marketing access. The rules developed in this process had to be satisfactory to regulatory requirements in all three regions. The greatest success claimed by this initiative was the planning reliability for drug developers in all regions. The rules were very clear about what roads had to be taken for a successful drug development. The ICH E9 biostatistical guideline was developed rather late in this process. However, rather unexpectedly, it became soon necessary to interpret important items from this guideline, and to provide more details for fostering wider application of biostatistical principles. As these interpretations were done separately and independently in the regions, one is attempted to say that this already was the beginning of a drifting apart soon after establishing harmonized rules. However, situation for drug developer became even worse with the emergence of national health technology assessment (HTA) agencies. The problem is that national HTA agencies are bound to respect the national health insurance systems, and in particular have to care that these systems are not overloaded by excessive expenditures. In the ICH process it was possible to overcome differences in the divergent national regulatory procedures, in part because in Europe this process was strongly supported by the simultaneously emerging desire for growing together in the European Union. At present there is, however, no similar force in sight that could motivate national HTA agencies also to look for common rules. Quite interesting different approaches related to biostatistics have been developed by HTA agencies, for example with respect to use of subgroup analyses, evaluation of clinical relevance, imputation of missing data but have not yet found general acceptance or rejection. This, however, would be necessary to bring lost planning reliability back to drug development.

HETEROGENEITY IN MULTIREGIONAL STUDIES

Joachim Röhmel, Bremen, Germany.

As the name indicates, multiregional studies are conducted in many regions. Regions can have very different culture, standards, norms and traditions. This impacts many parts of clinical studies' conduct. Main areas were strong differences are sometimes visible are patient information, informed consent, handling of inclusion and exclusion criteria. More importantly these inter regional differences can also influence clinical monitoring and the assessment of primary and secondary endpoints. Suspicions about the usefulness of clinical trial data generated in a foreign region vary between national health services, but can be strong and can lead to compromising a complete study program. Some recent examples will be discussed.

OBSERVATIONAL STUDIES, MATCHING AND PROPENSITY SCORES: AN APPLICATION TO COLORECTAL CANCER DATA.

John Hinde, NUI Galway, Galway, Ireland
Cara Dooley, TCD, Dublin, Ireland

In observational studies various techniques have been proposed to account for the lack of explicit design and randomisation. Popular approaches include matching and inverse probability weighting, often based on the propensity score. But how do these techniques work with datasets where the number of potential controls far exceeds the number of cases? We will consider this problem and explore some possible approaches. This will be illustrated in a survival analysis setting using a study on the effect of inflammatory bowel disease on survival in colorectal cancer patients using a dataset from the Irish Cancer Registry.

DESIGNING DRUG DEVELOPMENT TO OPTIMIZE HTA SUBMISSIONS

Karen Facey, University of Glasgow, United Kingdom of Great Britain and Northern Ireland

All healthcare providers seek to balance the competing factors of increasing healthcare costs and limited financial resources, alongside greater consumer expectations for the highest quality care and greater patient choice. To manage in this paradigm, it is essential to have transparent, systematic methods for determining which treatments will be made available to healthcare consumers and at what price. One solution is the use of Health Technology Assessment (HTA). HTA is a multidisciplinary field of policy analysis that systematically assesses the medical, social, ethical and economic implications of the use of health technologies (including medicines). HTA Agencies exist worldwide and although their methods of working may differ, their underlying principles are common to all: to inform evidence-based healthcare policy decisions. HTA takes evidence from a wide variety of sources, including the drug development programme, published literature and from health service sources. It often uses complex modelling to determine the clinical and cost effectiveness of a product compared to a current standard of care and that helps payers make decisions about reimbursement and/or pricing. This requires a paradigm shift in drug development, to consider not only the safety, efficacy and quality requirements of the regulator, but also the "fourth hurdle" of HTA.

The implications of HTA and its future development in Europe will be discussed in relation to drug development programmes and the statistician's role in demonstrating the (HTA) value of a product.

JOINT ANALYSIS OF MIXED OUTCOME DATA: ISSUES AND CHALLENGES

Keumhee Chough, University of Alberta, Edmonton, Canada

Multivariate data of mixed types occur frequently in many fields of science research. Conventional tools generally rely on the assumption that the data (or some suitable transformations) follow a normal distribution, although such an assumption does not directly apply. Analyses of such data have created new challenges that have made it necessary to develop new statistical techniques and methodologies. Some remarkable advances have been made over the past two decades. In this talk, we review major advances concerning the analysis of mixed data, highlighting significant recent methodological developments, such as the copula. The focus will be on important and lasting contributions to mixed-data methodology, to facilitate and promote further research, including some new directions for research.

de Leon and Carriere Chough. (2013) Analysis of Mixed Data. CRC Press.
de Leon and Carriere Chough. (2007). General mixed-data model: extension of general locatin and grouped continuous models. CJS. 533-548.
de Leon and Carriere Chough. (2010). Mixed-outcome data. In: Encyclopedia of Biopharmaceutical Statistics. 3rd ed. 817-822. Informa Healthcare.

EVALUATING DISEASE PREDICTORS WITH REPEATED MEASUREMENTS: A LONGITUDINAL STUDY OF ALZHEIMER'S DISEASE

Larry Brant, National Institute on Aging, Fallston, United States of America

Hierarchical or mixed-effects models have been used to model longitudinally collected data from individuals or clusters from many diverse areas of agricultural, biological and medical sciences. Recently, mixed models along with Bayes' theorem have provided a method for predicting future events based on the corresponding distribution of longitudinal trends from variables measured over time, and the calculated posterior probabilities of the occurrence of a given event. The method is illustrated by examining the ability of numerous behavioral, physiological, and biochemical measures collected over the entire adult lifespan of individuals participating in the Baltimore Longitudinal Study of Aging (BLSA) to predict diagnosis of incident AD. From 1576 participants aged 20-97 years and initially free of AD from the BLSA, we applied our longitudinal prediction method to generate posterior probabilities for AD for each of 15 candidate measures separately by sex and APOE ε4 carrier status.

During mean follow-up of 14.4 years, 6.7% of participants were diagnosed with AD. All candidate measures predicted AD, but with varying degrees of accuracy (represented by area under the receiver operating characteristic curve, AUC) and mean lead times (MLT). For women ε4 non-carriers, LDL was the most accurate predictor (AUC = 0.80, MLT = 10.7 years), followed by total cholesterol, whereas for women ε4 carriers the most accurate predictors were depressive symptoms (AUC = 0.78, MLT = 9.1 years), followed by systolic blood pressure (SBP). For men ε4 non-carriers and carriers, mean arterial pressure was the most accurate predictor (AUC = 0.71, MLT = 8.4 years and AUC = 0.76, MLT = 8.2 years). In general, lipids were better predictors in APOE ε4 non-carriers and BP measures in APOE ε4 carriers. Age modified the effects of depressive symptoms, SBP, and LDL on AD probability. In conclusion, longitudinal changes in blood pressure, lipid levels, and depressive symptoms are variably powerful predictors of AD diagnosis, depending on age, sex, APOE status. Widely available clinical measures collected longitudinally may be useful for early prediction of AD.

ANALYSIS OF A LARGE REGISTRY-BASED STUDY OF PRESCRIPTION DRUGS AND ROAD TRAFFIC CRASHES

Marta Avalos, Univ. Bordeaux Segalen - INSERM U897, Bordeaux, France
Ludivine Orriols, Univ. Bordeaux Segalen - INSERM U897, Bordeaux, France

Drugs (alcohol, illicit, prescription or over-the-counter) have a potential effect on the skills needed for driving, a task that involves a wide range of cognitive, perceptual and psychomotor activities.

In what concerns medicinal drugs, disentangling their impact on road traffic crashes is a complex issue for several reasons: 1/ the large variety of pharmaceutical classes, with various prevalence of use in the general population; 2/ the confounding underlying health conditions; 3/ the potential medicinal benefits of drugs that may lead to improved rather than impaired driving ability; 4/ the adaptive behaviors: whilst on medication, people may pay more attention to compensate for changes in perceived risk; 5/ the dose, cumulative dose and duration of drug consumption (prevalent, intermittent, incident); 6/ the co-consumption and interaction of drugs,… [1]. As a result, there is a relatively small epidemiological literature examining the associations between medicinal drugs and impaired driving.

A major approach to this complex subject relies on the use of population-registries data such as those conducted in UK [2, 3], Norway [4], Finland [5] or France [6]. An advantage of registry-based research is the potential to study associations between rare exposures and outcomes in a population large enough to provide sufficient precision. On the other hand, these studies can lead to high-dimensional and/or large-scale datasets, whose analysis needs to be addressed using appropriate statistical and computational techniques.

In this work, we report the use of shrinkage techniques [7] in exploratory analysis of the large study of prescription drugs and road traffic crashes described in Orriols and colleagues [6]; Avalos and colleagues, 2012, 2013 [8-12]. We insist on practical aspects, by detailing model selection and stability problems, implementation issues and discussing the obstacles for the use of these techniques in epidemiological studies.

---

[1] Elvik R. Risk of road accident associated with the use of drugs: A systematic review and meta-analysis of evidence from epidemiological studies. Accid Anal Prev. 2012.
[2] Barbone F, McMahon A, Davey P, Morris A, Reid I, McDevitt D, MacDonald T. Association of road-traffic accidents with benzodiazepine use. Lancet 1998; 352:1331–1336.
[3] Gibson J, Hubbard R, Smith C, Tata L, Britton J, AWFogarty. Use of self-controlled analytical techniques to assess the association between use of prescription medications and the risk of motor vehicle crashes. Am J Epidemiol 2009; 169:761–768.
[4] Engeland A, Skurtveit S, Mørland J. Risk of road traffic accidents associated with the prescription of drugs: a registry-based cohort study. Ann Epidemiol. 2007 Aug;17(8):597-602.

[5] Karjalainen K, Blencowe T, Lillsunde P. Substance use and social, health and safety-related factors among fatally injured drivers. Accid Anal Prev. 2012 Mar;45:731-6. doi: 10.1016/j.aap.2011.09.044.
[6] Orriols L, Delorme B, Gadegbeku B, Tricotel A, Contrand B, Laumon B, Salmi L, Lagarde E. Prescription medicines and the risk of road traffic crashes: A french registry-based study. PLoS Med 2010; 7(11): e1000366. doi:10.1371/journal.pmed.1000366.
[7] Tibshirani R. Regression shrinkage and selection via the lasso. J. Royal. Statist. Soc B. 1996; 58:267–288.
[8] Avalos, M, Duran-Adroher, N, Thiessard, F, Grandvalet, Y, Orriols, L, and Lagarde, E. Prescription-drug-related risk in driving comparing conventional and lasso shrinkage logistic regressions. Epidemiology 2012a, 23:706–12.
[9] Avalos, M, Grandvalet, Y, Duran-Adroher, N, Orriols, L, and Lagarde, E. Analysis of multiple exposures in the case-crossover design via sparse conditional likelihood. Stat. Med. 2012b, 31:2290–302.
[10] Avalos, M, Grandvalet, Y, Pouyes, H, Orriols, L, and Lagarde, E. clogitLasso: an R package for high–dimensional analysis of matched case–control and case–crossover data. Proceedings of the Tenth International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics (CIBB 2013), Nice, France. 2013a.
[11] Avalos, M., Orriols, L., Pouyes, H., Grandvalet, Y., Thiessard, F., and Lagarde, E. Variable selection on large case-crossover data: application to a registry–based study of prescription drugs and road–traffic crashes. Technical report, 2013b, Bordeaux School of Public Health, University Bordeaux Segalen.
[12] Avalos, M., Pouyes, H., Grandvalet, Y., Orriols, L., and Lagarde, E. High–dimensional variable selection in individually matched case-control studies. Technical report, 2013c, Bordeaux School of Public Health, University Bordeaux Segalen.

A NOVEL METHOD TO ESTIMATE THE MINIMUM EFFECTIVE DOSE FOR MONOTONE AND NON-MONOTONE DOSE-RESPONSE RELATIONSHIPS

Martin J Wolfsegger, Baxter Innovations GmbH, Vienna, Austria
Georg Gutjahr, Competence Center for Clinical Trials, University of Bremen, Bremen, Germany
Werner Engl, Baxter Innovations GmbH, Vienna, Austria
Thomas Jaki, Lancaster University, United Kingdom of Great Britain and Northern Ireland

Estimation of the minimum effective dose (MED) is an objective of many clinical and non-clinical dose-finding trials. It can be addressed using a multiple comparison procedure or modeling approach. Here, we present a new multiple comparison method for estimation of the MED as a combination of Dunnett's procedure and the step-down application of two-sample t-tests. The new approach controls the type-I error rate of underestimating the true MED and performs often better than the available alternatives.

TYPE I ERROR RATE CONTROL IN ADAPTIVE CLINICAL TRIALS WITH BLINDED INTERIM ANALYSIS

Martin Posch, Medical University of Vienna, Austria
Magdalena Malina, Medical University of Vienna, Austria

It is well known that adaptations of an ongoing trial based on unblinded interim data may result in biased estimates of treatment effects and increased type I error rates unless the adaptations are accounted for in the statistical analysis.

In this work we quantify the potential inflation of type I error rates resulting from adaptations based on blinded interim data. Here we use the fact that if a treatment shows an effect on secondary or safety endpoints also a formally blinded data set contains information on the treatment allocation. If, at the same time, the treatment has no effect on the primary endpoint, adaptations based on this partially unblinded data may result in an inflation of the type I error rate. For the setting of a parallel group comparison of normally distributed endpoints we quantify the maximal type I error inflation that may result from sample size reassessment based on blinded data from secondary endpoints. We determine the sample size reassessment rules leading to the maximum type I error inflation and investigate the impact of the effect sizes in the secondary endpoints on the error rate.

REGULATORY GUIDELINES AS A STARTING POINT FOR METHODOLOGICAL RESEARCH – THE CASE OF BIOEQUIVALENCE

Meinhard Kieser, University of Heidelberg, Heidelberg, Germany
Sven Schnaidt, University of Heidelberg, Heidelberg, Germany

There exist a huge number of regulatory guidelines in various fields of clinical research, many of them addressing biostatistical topics. It is widely recognized that these documents may be extremely helpful in making recommendations towards achieving greater harmonization thus facilitating registration of new drugs and devices. There are considerable differences between the styles of these documents, ranging from presenting principles only (e.g. ICH E9 Guideline [1]) to discussing (or even prescribing) specific methods (e.g. EMA Guideline on Missing Data in Confirmatory Clinical Trials [2]). In many cases, it is not evident how to adequately and efficiently address these principles or how recommended methods perform in a specific clinical trial situation at hand. For biostatisticians, these situations are most welcome opportunities for methodological research. In our presentation, we would like to demonstrate for the case of bioequivalence trials how the ambiguity which is frequently an intrinsic feature of regulatory documents (see, e.g., the EMA Guideline on the Investigation of Bioequivalence [3]) may stimulate methodological research. Examples are given where this approach may lead to new procedures or a deeper insight in the characteristics of known methods, respectively.

---

[1] International Conference on Harmonization (ICH) Topic E9. Note for Guidance on Statistical Principles for Clinical Trials. (CPMP/ICH/363/96), September 1998. Available from: http://www.emea.europa.eu/docs/en_GB/document_library /Scientific_guideline/2009/09/WC500002928.pdf.
[2] Committee for Medicinal Products for Human Use (CHMP). Guideline on Missing Data in Confirmatory Clinical Trials. (EMA/CPMP/EWP/1776/99 Rev. 1), 2 July 2010. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf.
[3] Committee for Medicinal Products for Human Use (CHMP). Guideline on the Investigation of Bioequivalence. (CPMP/EWP/QWP/1401/98 Rev. 1 / Corr **), 20 January 2010. Available from: http://www.ema.europa.eu/pdfs/human/qwp/140198enrev1fin.pdf.

PRACTICAL IMPLICATIONS OF COMPETING RISK MODELS – HOW TO INTERPRET THE RESULTS TO CLINICIANS

Milada Smastuen, Cancer Registry of Norway, Oslo, Norway

Competing risk models are a recognised methodology in survival analysis however they are still rarely used and they are hard to interpret by clinicians.

The Kaplan Meier approach is still often the method of choice even though it might result in wrong and misleading results. In this talk some methodological issues will be presented and illustrated with a ”real life example”.

Disease specific survival in colon cancer has been studied for decades however often using a wrong methodology. With increasing cancer survival an increasing proportion of patients die of other causes. Furthermore colon cancer survival varies depending on the stage of the disease and the length of follow up varies as well. Thus competing risk should be incorporated when analysing cancer specific mortality and morbidity. In this context some methodological issues will be adressed when trying to depict and compare cumulative disease specific survival in relation to stage.

The method of Fine and Grey will be discussed and some suggestions will be presented how competing risk regression results could be presented in a way that is understandable and meaningful to clinicians.

SOME APPLICATIONS OF RISK MODELING IN BREAST CANCER PREVENTION

Mitchell Gail, National Cancer Institute, Bethesda, Maryland, United States of America

We use the term “risk” to denote absolute risk, namely the probability that a particular event will occur in a defined time interval. Absolute risk depends on the age at the beginning of the time interval, on other covariates that may be present, on the duration of the interval, and on competing risks, that can diminish the absolute risk of the event of interest. Models of the absolute risk of developing breast cancer have applications in disease prevention. Such models can be used to assist in designing prevention trials, in assessing the potential population reductions in absolute risk from a prevention program, in implementing a prevention strategy focused on high-risk subjects, and in allocating preventive resources to those at highest risk. Absolute risk may also be useful in counseling individuals on prevention strategies, and, more formally, to assist in weighing the risks and benefits of a preventive intervention. We review some of these applications and some methods used to assess the usefulness of risk models.

Gail MH. Personalized estimates of breast cancer risk in clinical practice and public health. Stat Med. 2011 May 10;30(10):1090-104

ESTIMATING MMR VACCINE COVERAGE FROM AUSTRALIAN SERO-SURVEILLANCE DATA

Niel Hens, Hasselt University, Diepenbeek, Belgium
James Wood, University of New South Wales, Sydney, Australia
Nele Goeyvaerts, Hasselt University, Diepenbeek, Belgium

The effectiveness of childhood immunization programs depends on the achieved vaccination coverage, vaccine efficacy against infection and the persistence of vaccine-induced immunity. Whereas estimates for the latter two are often available, good coverage data is not always obtainable and comparability between studies is often compromised because of different data collection methods. In 2000, Gay developed a method to estimate trivalent vaccination coverage from readily available trivariate serological data on measles, mumps and rubella. His method was indirectly published in a paper by Altmann and Altmann [1], who derived exact solutions for Gay's modelling equations. Recently Goeyvaerts et al. [2] proposed a general likelihood-based marginal model framework to extend Gay's model explicitly accounting for possible associations between disease-specific exposure probabilities.

Applying their method to data from Belgium, vaccination coverage estimates in line with estimates from routinely collected coverage data were obtained. These results were based on one cross-sectional serological sample only and consequently ignored waning of antibodies and exposure to the wild-type virus. Using data collected at multiple timepoints, we propose a novel cohort model accounting for waning and boosting mechanisms. Using simulations we show that our method outperformsthe model without waning (or Gay's model) and using data from Australia anno 1998, 1999 and 2007 we obtain vaccination coverage, seroconversion and waning estimates in agreement with data from other sources.

[1] Altmann D, Altmann K. Estimating vaccine coverage by using computer algebra. IMA Journal of Mathematics Applied in Medicine and Biology 2000; 17:137–146. DOI: 10.1093/imammb/17.2.137.
[2] Goeyvaerts, N., Hens, N., Theeten, H., Aerts, M., Van Damme, P. and Beutels, P. Estimating vaccination coverage for the trivalent measles-mumps-rubella vaccine from trivariate serological data. Statistics in Medicine 2012; 31:1432-1449. DOI: 10.1002/sim.4481

UNDERSTANDING INDIVIDUAL HETEROGENEITY: WHAT IS THE POTENTIAL OF THE FRAILTY APPROACH?

Odd O. Aalen, University of Oslo, Oslo, Norway

Unobserved individual heterogeneity, also called frailty, is a major concern in application of survival analysis. Hazard rates do not give direct information on the change over time in the individual risk, but are in addition strongly influenced by selection effects operating in the population. The individuals surviving up to a certain time will on the average be less frail than the original population. Models are reviewed that account for this phenomenon.

Application to cancer incidence data will be given, with examples from testicular cancer, osteosarcoma and Hodgkin's lymphoma. We shall also review how frailty models can be used to assess and understand familial association of cancer. Frailty analysis shows that apparently modest familial association implies a much greater variation in risk than one should intuitively think.

BIOMETRICAL ISSUES IN HEALTH TECHNOLOGY ASSESSMENT

Ralf Bender, Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany

Health technology assessment (HTA) in Germany is currently influenced by the law on the reorganization of the pharmaceutical market. According to this law, new approved drugs have to undergo an early benefit assessment 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, which can commission the Institute for Quality and Efficiency in Health Care (IQWiG) or other HTA agencies with an assessment of the dossier. In the framework of early benefit assessment biometrical issues such as identification of relevant subgroups, application of indirect comparisons, validation of surrogate endpoints, and classification according to the extent of added benefit are of interest. An overview of the processes within early benefit assessment is given and biometrical issues are discussed.

ON CRITERIA FOR EVALUATING RISK PREDICTION MODELS FOR PUBLIC HEALTH APPLICATIONS

Ruth Pfeiffer, National Cancer Institute, National Institutes of Health, Rockville, United States of America

I propose and study novel criteria to assess the usefulness of models that predict risk of disease incidence for screening and prevention, or the usefulness of prognostic models for management following disease diagnosis. The proportion of cases followed, PCF(p), is the proportion of individuals who will develop disease who are included in the proportion p of individuals in the population at highest risk. The proportion needed to follow-up, PNF(q), is the proportion of the general population at highest risk that one needs to follow in order that a proportion q of those destined to become cases will be followed. I also propose the integrated PCF, iPCF, and iPNF, the integrated PNF, obtained by integrating PCF and PNF over a range of values of q or p. Under the assumption that the risk model is well calibrated PCF, PNF, iPCF and iPNF can be estimated based on observed risks in a population alone. When the risk models are not well calibrated, PCF, PNF, iPCF and iPNF can be estimated consistently from case control data when the outcome prevalence in the population is known, and from cohort data, with baseline covariates and observed health outcomes. I illustrate the criteria with novel models that predict incidence of endometrial, ovarian and breast cancer.

Pfeiffer RM. Extensions of criteria for evaluating risk prediction models for public health applications, Biostatistics, 2013
Pfeiffer RM, Gail MH. Two criteria for evaluating risk prediction models. Biometrics, 2011

TREATMENT SELECTION IN MULTI-ARM, MULTI-STAGE CLINICAL TRIALS

Thomas Jaki, Lancaster University, United Kingdom of Great Britain and Northern Ireland
Dominik Magirr, Lancaster University, United Kingdom of Great Britain and Northern Ireland
Nigel Stallard, Warwick Medical School, United Kingdom of Great Britain and Northern Ireland

Adaptive designs that are based on group-sequential approaches have the benefit of being efficient as stopping boundaries can be found that lead to good operating characteristics with test decisions based solely on sufficient statistics. The drawback of these so called "pre-planned adaptive" designs is that unexpected design changes are not possible without impacting the error rates. "Flexible adaptive designs", and in particular designs based on p-value combination, on the other hand can cope with a large number of contingencies at the cost of reduced efficiency.

In this presentation we focus on so called multi-arm multi-stage trials which compare several active treatments against control at a series of interim analysis. We will focus on the methods by Stallard and Todd [1] and Magirr et al. [2], two different approaches which are based on group-sequential ideas, and discuss how these "pre-planned adaptive designs" can be modified to allow for flexibility. We then show how the added flexibility can be used for treatment selection and evaluate the impact on power in a simulation study. The results show that a combination of a well chosen pre-planned design and an application of the conditional error principle to allow flexible treatment selection results in an impressive overall procedure.

[1] Stallard, N, & Todd, S. 2003. Sequential designs for phase III clinical trials incorporating treatment selection. Statistics in Medicine, 22, 689-703.
[2] Magirr, D, Jaki, T, & Whitehead, J. 2012. A generalised Dunnett test for multi-arm, multi-stage clinical studies with treatment selection. Biometrika, 99, 494-501.

EMA GUIDANCE DOCUMENTS: OVERVIEW OF RECENT DEVELOPMENTS

Thomas Lang, AGES Medizinmarktaufsicht, Austria

The mandate of EMA's Biostatistics Working Party (BSWP) includes – among other responsibilities -the task to revisit existing methodological/statistical EMA guidance documents in order to evaluate whether the recommendations contained therein are still applicable and up to date in context of evolving new methodologies in the broad field of biostatistics. In addition, the mandate describes the task to evaluate the need for new guidance documents in order to react to /reflect recent discussions related to the planning and conduct of statistical analyses in clinical trials.

The 'Points to Consider on Multiplicity Issues in Clinical Trials' was adopted 10 years ago, and approaches the problem of multiplicity in clinical trials in a rather pragmatic (and user-friendly) fashion, describing frequently encountered multiplicity issues in clinical trials and corresponding necessities/options to account for these issues. The progress in the development of methods and new concepts in this area since adoption of this guideline triggered the question of whether the Guidance document needs an update. In November 2012, a workshop was held at EMA on the topic. The talk will give a brief overview of the current discussions.

A further topic of the talk will be the emerging methodological issues concerning subgroup analyses in clinical trials and their impact on regulatory decision making. In 2010, the concept paper 'Need for a guideline on the use of subgroup analyses in randomised controlled trials' was issued, followed by a public consultation phase. A draft of a new EMA guidance on this topic is expected to be published in summer this year. The talk will highlight the most important aspects from regulatory perspective, and will address issues like pre-specification of subgroups, replication of evidence and 'biological plausibility' in the assessment of internal consistency.

Other BSWP initiatives in relation to guideline improvement/development will be mentioned, e.g. the update of the 'Points to Consider on Adjustment for Baseline Covariates', or the development of a 'Reflection paper on Statistical Methodology for the comparative assessment of Quality attributes in drug development'.

SITAR GROWTH CURVE ANALYSIS TO EVALUATE INTERVENTIONS AND LIFE COURSE OUTCOMES

Tim Cole, University College London, London, United Kingdom of Great Britain and Northern Ireland

An important element of life course epidemiology is the investigation of possible relationships between the pattern of growth during childhood and later health outcomes. SITAR (SuperImposition by Translation And Rotation) is a form of growth curve analysis that summarises the growth patterns of a group of children over time (e.g. height or weight) as a single summary growth curve, and at the same time defines each individual's growth curve in terms of three random effects that transform their curve to closely match the summary curve. The three parameters are both biologically and geometrically meaningful: size is an up-down shift of each individual curve along the measurement axis (defining relative mean size), tempo is a left-right shift along the age axis (defining the relative timing, or tempo, of growth events such as puberty), and velocity is a shrinking-stretching of the age scale (which defines relative velocity). The mean growth curve is estimated as a cubic regression spline.

Using a series of examples it is shown that the SITAR model fits growth data extremely well, explaining up to 99% of the age-specific variance, depending on the measurement and when in childhood it is measured. So a strength of SITAR is that once the model has been fitted the original data can effectively be ignored, and the focus shifts to the fitted mean growth curve and the triplets of size-tempo-velocity SITAR random effects that define each subject's growth pattern. The model can also include fixed effects to adjust for factors such as grouping, sex or (in infancy) gestation.

The talk will describe the development of and rationale for the SITAR model, including examples both of growth as the outcome of a previous intervention and also where growth is the exposure relating to a later outcome. The strengths and weaknesses of the model will be discussed.

Cole TJ, Donaldson MD, Ben-Shlomo Y. SITAR--a useful instrument for growth curve analysis. Int J Epidemiol 2010;39:1558-66.
Gault EJ, Perry RJ, Cole TJ, et al. Effect of oxandrolone and timing of pubertal induction on final height in Turner's syndrome: randomised, double blind, placebo controlled trial. BMJ 2011;342:d1980.
Johnson L, Llewellyn CH, van Jaarsveld CHM, et al. Genetic and environmental influences on infant growth: prospective analysis of the Gemini twin birth cohort. PLoS ONE 2011;6:e19918.
Prentice A, Dibba B, Sawo Y, et al. The effect of prepubertal calcium carbonate supplementation on the age of peak height velocity in Gambian adolescents. Am J Clin Nutr 2012;96:1042-50.

A NEW APPROACH BASED ON PSEUDO-VALUES FOR ASSESSING THE EFFECT OF A BINARY TIME-DEPENDENT COVARIATE ON PATIENT SURVIVAL

Ulrike Pötschger, Children's Cancer Research Institute, Vienna, Austria
Harald Heinzl, Medical University Vienna, Vienna, Austria
Andishe Attarbaschi, St. Anna Children's Hospital, Vienna, Austria
Milen Minkov, Children's Cancer Research Institute, Vienna, Austria
Martina Mittlböck, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria

Statistical challenges in the comparison of allogeneic stem-cell transplantation (SCT) with conventional chemotherapy for the treatment of paediatric leukaemia are the motivation for this work. Here, the anticipated non-proportional hazards, the waiting time to transplant, and the primary interest in long-term outcome need to be considered simultaneously. Available statistical methods show several limitations that may lead to wrong conclusions. Thus, novel approaches of statistical analyses are needed that adjust for waiting time without relying on the proportional hazard assumption. The pseudo-value regression technique provides an attractive alternative for modelling time-fixed covariates in a non-proportional hazards situation. We generalize the original pseudo-value approach to allow for a binary non-reversible time-dependent covariate. Like in the original approach, survival rates at pre-specified time-points are evaluated using generalized linear models. Survival expectations of patients with transition to a non-terminal state (e.g. SCT) are estimated conditional on the observed waiting time distribution and compared to baseline survival rates (e.g. without SCT). The new approach adjusts for both, immortal time bias and negative selection bias. The practical value of this approach is illustrated with real data from childhood leukaemia. A simulation study that mimics published examples was performed to investigate the statistical properties of the proposed model. Regardless of whether the proportional hazard assumption holds or not, the estimated parameters are unbiased. A further advantage of the proposed approach is the possibility to straightforwardly study the impact of waiting time on survival, a common secondary aim in many studies. In paediatric oncology the main interest usually is long-term outcome. The proposed approach provides a reasonable and previously not available statistical tool to directly address this goal in a clinically common but methodologically difficult situation.

STATISTICS IN HEALTH TECHNOLOGY ASSESSMENT: ISSUES AND CHALLENGES

Yovanna Castro, F. Hoffmann-La Roche, Basel, Switzerland

This presentation aims to show some of the challenges the pharmaceutical industry faces in the Health Technology Assessment setting. Once a product is approved by regulatory agencies, another important step is to get reimbursed by payers where each country has its own regulations and requirements.

One of the first questions we face is what is the treatment effect of a new compound compared to treatments currently available on the market? The issue is that in a clinical trial it is often not possible to compare all existing drugs against the product of interest. We have to rely on indirect comparisons and specifically network meta-analysis. A brief overview of the method will be presented.

Another issue arises when patients cross-over from the control arm to receive the new treatment before the end of the trial for various reasons. We should then address the question, what is the "real" effect of the new compound given there were some control patients who switched to the new treatment? Several methods have been proposed in the literature, and some of them with their assumptions and drawbacks will be discussed. This review will by no means be exhaustive.

The last point is the use of real world data. Clearly, a randomized clinical trial does not necessarily reflect everything that can occur in "real life" situations. Information regarding real world data can be incorporated into statistical models, but there are several issues regarding the use of this data that will be presented.

These are some of the challenging questions we need to address in order to support the reimbursement process. Payers are interested in knowing how new treatments perform versus all established standards of care in routine clinical setting and we should attempt to address that question as accurately as possible.

Jansen, J. P., Fleurence, R., et al.. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices, part 1. Value in Health, 14(4), 417-428. "Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices, part 1." Value in Health 14.4 (2011): 417-428.
Dias, S., Welton, N.J., Sutton, A., Aedes, A.E. NICE DSU Technical Support Document 1: Introduction to evidence synthesis for decision making. (2011) Available from http://www.nicedsu.org.uk.
Robins, J. M., and Tsiatis, A. A. "Correcting for non-compliance in randomized trials using rank preserving structural failure time models." Communications in Statistics-Theory and Methods 20.8 (1991): 2609-2631.
Branson, M., and Whitehead, J. "Estimating a treatment effect in survival studies in which patients switch treatment." Statistics in medicine 21.17 (2002): 2449-2463.

2. CONTRIBUTED SPEAKER ABSTRACTS

ENRICHMENT DESIGNS FOR THE DEVELOPMENT OF PERSONALIZED MEDICINES

Alexandra Graf, Medical University of Vienna, Vienna, Austria
Franz König, Medical University of Vienna, Vienna, Austria
Martin Posch, Medical University of Vienna, Vienna, Austria

If the response to treatment depends on genetic biomarkers, it is important to identify (sub-)populations where the treatment has a positive benefit risk balance. One approach to identify relevant subpopulations are subgroup analyses where the treatment effect is estimated in biomarker positive and biomarker negative groups. Subgroup analysis are challenging because different types of risks are associated with inference on subgroups: On the one hand, ignoring a relevant subpopulation one could miss a treatment option due to a dilution of the treatment effect in the full population. Even, if the diluted treatment effect can be demonstrated in an Overall population, it is not ethical to treat patients that do not benefit from the treatmemt, if they can be identified in advance. On the other hand selecting a spurious sub-population is not without risk either: it might increase the risk to approve a inefficient treatment (inflating the type 1 error rate), or may wrongly lead to restricting an efficient treatment to a too narrow fraction of a potential benefiting population. The latter can not only lead to a reduced revenue from the drug, but is also unfavourable from a public health perspective. We investigate these risks for non-adaptive study designs that allow for inference on subgroups using multiple testing procedures as well as adaptive designs, where subgroups may be selected in an interim analysis. Quantifying the risks with utility functions the characteristics of such adaptive and non-adaptive designs are compared for a range of scenarios.

COMPARISON AND EVALUATION OF CARDIAC BIOMARKERS IN PATIENTS WITH INTERMITTENT CLAUDICATION: RESULTS FROM THE CAVASIC STUDY

Barbara Kollerits, Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria
Gisela Sturm, Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria
Claudia Lamina, Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria
Angelika Hammerer-Lercher, Central Institute of Medical and Chemical Laboratory Diagnostics, University Hospital Innsbruck, Innsbruck, Austria
Barbara Rantner, Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria
Marietta Stadler, 3rd Medical Department of Metabolic Diseases and Nephrology, Hietzing Hospital, Vienna, Austria
Tim Ziera, Research Department, B.R.A.H.M.S GmbH/Thermo Fisher Scientific, Hennigsdorf, Germany
Joachim Struck, AdrenoMed AG, Hennigsdorf, Germany
Peter Klein-Weigel, Clinic for Angiology, HELIOS Klinik Berlin-Buch, Berlin, Germany
Gustav Fraedrich, Department of Vascular Surgery, Innsbruck Medical University, Innsbruck, Austria
Florian Kronenberg, Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria

Various methods exist for evaluating the performance of prediction models and their pros and cons are discussed [1-3]. Routinely used measures such as the test of deviances on nested models, the C-index and new upcoming tools like integrated discrimination improvement (IDI) and net reclassification improvement (NRI) were applied to estimate the comparative performance of four cardiac markers (MR-proADM, MR-proANP, CT-proET-1, NT-proBNP) in peripheral arterial disease (PAD). This was done in the CAVASIC Study, a male cohort of 238 patients diagnosed with symptomatic PAD and 245 age- and diabetes-matched controls.

Models including the four markers separately showed significantly improved model fits compared to the model including only age and classical cardiac risk factors (=reference model). These distinct models showed that the C- index for all four markers was not significantly different from the C-index of the reference model. A positive and significant IDI ($p<0.05$) was observed for all four markers when compared to the reference model. However, the IDI increase was about twice as high for log-NT-proBNP (absolute IDI 0.053; relative IDI 0.15, $p<0.0001$) compared to MR-proADM, MR-proANP and CT-proET-1, respectively. Additionally, MR-proADM (absolute IDI 0.02; relative ID 0.17, $p=0.002$) and CT-proET-1 (absolute IDI 0.02 relative IDI 0.17, $p=0.002$) provided even improved reclassification as compared

to a model containing age and the established heart failure parameter log-NT-proBNP. The improved reclassification shown by IDI was confirmed by applying NRI. This study in a male cohort of patients with intermittent claudication and age- and diabetes-matched controls indicated that NT-proBNP is the strongest marker for PAD risk determination. Nevertheless, MR-proADM and CT-proET-1 provided significant additive information beyond classical cardiac risk factors and also in direct comparison to NT-proBNP [4]. IDI and NRI were more sensitive than the C-index in determining the discriminative power of the applied prediction models containing the four investigated markers.

[1] Sundström J, Byberg L, Gedeborg R, et.al. Useful tests of usefulness of new risk factors: tools for assessing reclassification and discrimination. Scand J Public Health. 39(4):439-41, 2011.
[2] Pencina MJ, D'Agostino RB Sr, Demler OV. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. Stat Med. 31(2):101-13, 2012.
[3] Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med. 30(1):11-21, 2011.
[4] Kollerits B, Sturm G, Lamina C, Hammerer-Lercher A, Rantner B, Stadler M, Ziera T, Struck J, Klein-Weigel P, Fraedrich G, Kronenberg F. Comparison and Evaluation of Cardiac Biomarkers in Patients with Intermittent Claudication: Results from the CAVASIC Study. Clin Chem. 59:692-702, 2013.

MULTISTATE MODELS WITH MULTIPLE TIME SCALES

Bendix Carstensen, Steno Diabetes Center, Gentofte, Denmark
Simona Iacobelli, Centro Interdipartimentale di Biostatistica e Bioinformatica Universita Tor Vergata, Roma, Italy

The normal approach to multistate modeling is to assume a common underlying time-scale (Markov property) for all transitions between states. This greatly simplifies calculation of transition probabilities, because they all reduce to multiplication of infinitesimal transition probability matrices, which are easily derived from fitted models for the transition intensities. This approach is implemented in e.g. R-packages mstate and etm.

However, in studies with progression through disease states it is untenable from a clinical point of view to assume that transitions rates do not (additionally) depend on either time of entry into a state or time spent in a state or even both. In this paper we will argue that whether this is the case or not, choice of timescale(s) is an empirical question, not something to be decided a priori. Thus it should be routinely checked in multistate modeling which timescale(s) provide the best description of the transition rates.

We will use an example from bone marrow transplant in leukaemia treatment to illustrate how this can be checked using simple parametric (spline) models for the rates, and how these models allow standard likelihood-ratio tests for the relevant hypotheses.

Finally we will give example of the use of Lexis objects from the Epi package[1, 2] for R, and specifically show how simulation tools can be used to estimate state occupancy probabilities from complex multistate models with multiple time scales.

[1] Bendix Carstensen and Martyn Plummer. Using Lexis objects for multi-state models in R. Journal of Statistical Software, 38(6):1{18, 1 2011.
[2] Martyn Plummer and Bendix Carstensen. Lexis: An R class for epidemiological studies with long-term follow-up. Journal of Statistical Software, 38(5):1{12, 1 2011.

USING WEIGHTED GENETIC PREDISPOSITION SCORES TO EVALUATE GENE-ENVIRONMENT INTERACTION EFFECTS ON LIPID LEVELS

Claudia Lamina, Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria
Lukas Forer, Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria
Sebastian Schönherr, Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria
Barbara Kollerits, Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria
Janina Ried, Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany
Christian Gieger, Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany
Annette Peters, Institute of Epidemiology II, Helmholtz Zentrum München, Neuherberg, Germany
H.-Erich Wichmann, Institute of Epidemiology I, Helmholtz Zentrum München, Neuherberg, Germany
Florian Kronenberg, Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria

In recent years, a large number of genetic association studies with complex phenotypes have been published, mainly due to the availability of genome-wide gene-chips and worldwide collaborations which bring these data together. For lipids, 95 genetic loci have been identified, explaining roughly about 25-30% of their genetic variance [1]. Yet, the majority of variance is still unexplained. Unidentified gene-environment interaction effects (GxE) might partly explain this missing heritability. Genome-wide GxE studies on specific Single-Nucleotide Polymorphisms (SNPs) could not satisfy this expectation due to lack of power.

To overcome this limitation, we used weighted genetic predisposition scores to evaluate gene-obesity interaction effects on HDL (HDL-C), LDL (LDL-C) and total cholesterol (TC). From the population-based studies KORA F3 (n=1406) and KORA F4 (n=1515), we derived imputed genotypes for 104 lipid-associated SNPs. Specific unbiased weights for each SNP were inferred from linear regression estimates in KORA F4. Weighted genotypic predisposition SNP-scores were then calculated for each lipid trait in KORA F3. Interaction terms of SNP-scores with each of the obesity parameters (BMI, waist-hip-ratio, waist circumference) were included in age-and sex-adjusted linear regression models on HDL-C, LDL-C and TC.

Our investigation revealed a significant interaction effect between obesity parameters and a SNP-score on HDL-C: for increasing levels of obesity, the combined effect of

HDL-C-increasing alleles on HDL-C is attenuated [2]. Accounting for gene-obesity interaction substantially increased the proportion of HDL-C variance explained: By inclusion of SNP-obesity interaction terms, additional 3.0-3.5% of HDL-C variance can be explained, while the SNPs alone explain about 9%. Still, the majority of phenotypic variance remains unexplained. No significant interaction could be found for LDL-C or TC. We also propose a graphical method for the presentation and interpretation of such interaction effects [3], specifically showing the effect of the SNP-score on lipids for varying values of the obesity parameters.

[1] Teslovich TM et al. Biological, Clinical, and Population Relevance of 95 Loci Mapped for Serum Lipid Concentrations. Nature. 2010 Aug 5;466(7307):707-13.
[2] Lamina C, Forer L, Schönherr S, Kollerits B, Ried JS, Gieger C, Peters A, Wichmann HE, Kronenberg F. Evaluation of gene-obesity interaction effects on cholesterol levels: A genetic predisposition score on HDL-cholesterol is modified by obesity. Atherosclerosis. 2012 Dec;225(2):363-9.
[3] Lamina C, Sturm G, Kollerits B, Kronenberg F. Visualizing interaction effects: A proposal for presentation and interpretation. J Clin Epidemiol. 2012 Aug;65(8):855-62.

SHRINKAGE OF REGRESSION COEFFICIENTS: OLD CONCEPTS AND NEW IDEAS

Daniela Dunkler, Medical University of Vienna, Vienna, Austria
Georg Heinze, Medical University of Vienna, Vienna, Austria

Shrinkage of regression coefficients in prognostic models is required to correct the optimism of such models caused by variable selection. Estimated shrinkage factors are multiplied with the regression coefficients in order to reduce the prediction error. While global shrinkage modifies all coefficients by the same factor, Sauerbrei's [1] parameterwise (PW) shrinkage factors differ between coefficients, their values depending on each variable's prognostic relevance.

We extend the shrinkage methodology by proposing joint shrinkage factors (JSFs). A JSF may be useful to describe the common prognostic relevance for a set of highly correlated or intrinsically tied design variables, e.g., when modeling nonlinear effects or interactions, and to shrink the corresponding coefficients in order to improve predictive accuracy. Application of and differences between PW and JSFs are demonstrated in an example study involving fractional polynomials (FP). In this study, we observe implausibly low PW shrinkage factors for the two components of an FP of degree 2. This typically occurs because the independent prognostic relevance of each of the FP's two correlated design variables, corresponding to the same explanatory variable, is low. Their JSF, however, assumes a meaningful value, quantifying the shrinkage needed because of selection of that particular explanatory variable.

In a short excursus, we describe how DFBETA residuals can be used to approximate leave-one-out (LOO)-resampling shrinkage estimates with substantial savings in computational demand. LOO-type and DFBETA-type shrinkage factors are practically undistinguishable with large sample sizes (N), and still similar with moderate N. The proposed approximation may be particularly desired when resampling is applied or in simulation studies.

Joint shrinkage factors are a useful extension to existing methodology for intrinsically tied design variables, both in quantifying their prognostic relevance and in correcting prognostic models for overfit.

The R package 'shrink' implements the discussed approaches to shrinkage.

---

[1] Sauerbrei W. The Use of Resampling Methods to Simplify Regression Models in Medical Statistics. Journal of the Royal Statistical Society: Series C (Applied Statistics) 1999; 48(3):313-329.

CHANGES OF BODY MASS INDEX IN RELATION TO MORTALITY: RESULTS OF A COHORT OF 42,099 ADULTS

Gabriele Nagel, Institute of Epidemiology and Medical Biometry, Ulm University, Germany
Kilian Rapp, Germany
Hanno Ulmer, Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Innsbruck, Austria
Hans Concin, Austria
Jochen Klenk, Germany

The VHM&PP (Vorarlberg Health Monitoring & Prevention Program) is a population-based risk factor surveillance program in Vorarlberg, the westernmost province of Austria. The screening examinations were performed by local physicians according to a standard protocol. Anthropometric measures were carried out with participants wearing light indoor clothes and no shoes. Within the VHM&PP population-based prospective cohort of 42,099 Austrian men and women with at least three BMI measurements we investigated the relationship of BMI at baseline and two subsequent BMI change intervals of five years each with all-cause mortality using Cox proportional Hazard models.

During a median follow-up of 12 years 4,119 deaths were identified. The lowest mortality rates were found in persons with normal weight or overweight at baseline and stable BMI over 10 years. Weight gain ($\geq$ 0.10 kg/m²/year) during the first five years was associated with increased mortality in overweight and obese people. For weight gain during both time intervals mortality risk remained significantly increased only in overweight (1.39 (95% confidence interval: 1.01; 1.92)) and obese women (1.85 (95% confidence interval: 1.18; 2.89)). Weight loss ($<$ -0.10 kg/m²/year) increased all-cause mortality in men and women consistently.

BMI change over time by accepted WHO BMI categories showed no increased mortality risk for people who remained in the normal or overweight category for all three measurements. In contrast, HRRs for stable obese men and women were 1.57 (95% CI: 1.31; 1.87) and 1.46 (95% CI: 1.25; 1.71), respectively.

Our findings highlight the importance of weight stability and obesity avoidance in prevention strategies.

CONSISTENCY-ADJUSTED ALPHA ALLOCATION METHODS FOR COMPOSITE ENDPOINTS

Geraldine Rauch, University of Heidelberg, Heidelberg, Germany
Meinhard Kieser, University of Heidelberg, Heidelberg, Germany

Composite endpoints are often used as primary efficacy endpoints, particularly in the field of oncology and cardiology. These endpoints combine several time-to-event variables of interest within a single time-to-first-event outcome. Thereby, it is intended to enlarge the expected effect size and thus to increase the power of the clinical trial. However, the interpretation of composite endpoints can be difficult, if the effects of the single components are of different magnitude or even point in adverse directions. Therefore, it might not be adequate to judge the efficacy of the new intervention exclusively on the composite effect. A possible solution would be to include at least the most relevant in a multiple confirmatory test strategy. However, a superiority test problem for an individual component is usually not realistic in these settings as the expected individual effects are small.

Alosh and Huque [1, 2] and Li et al. [3] recently proposed consistency-adjusted alpha allocation methods which can be used and extended to address this particular problem. We discuss several alpha allocation methods for composite endpoints, compare their power properties and apply the methods to several clinical trial examples. Moreover, we face the general problem of correlation-adjusted local significance levels focusing on the special correlation structure between composite endpoints and their components.

[1] Alosh M, Huque MF (2009): A flexible strategy for testing subgroups and overall population. Stat Med 28:3-23.
[2] Huque MF, Alosh M (2012): A consistency-adjusted strategy for accomodating and underpowered primary endpoint. J Biopharm Stat 22: 160-179.
[3] Li H, Sankoh AJ, D'Agostino Sr RB (2012): Extension of adaptive alpha allocation methods for strong control of the family-wise error rate. Stat Med, Epub ahead of print.

A FRAMEWORK TO ASSESS THE ADDED VALUE OF SUBGROUP ANALYSES WHEN THE OVERALL TREATMENT EFFECT(TE) WAS SIGNIFICANT

Hong Sun, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Freiburg, Germany
Werner Vach, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Freiburg, Germany

Although statisticians have already noticed and discussed the problem, it is still not clear whether we should perform and how to perform such subgroup analyses when the overall TE is significant in both single clinical trials and systematic reviews. So a framework is needed to assess and compute the long term effect of different strategies to perform subgroup analysis.

We propose two performance measures to evaluate the average post-study TE for patients in all studies (E) and fraction of patients with a negative TE in the positive studies (P). Eight decision rules are applied to different assumptions of subgroup specific and individual TE. Optimistic, moderate and pessimistic scenarios are assumed for true TE.

A hierarchical structured simulation study is set up with the fixed power of 90% and an assumed constant TE for each study in particular case of binary outcomes with TE presented as risk difference, a constant sample size 280 is required. Patients were allocated in 1:1 ratio to two hypothetical treatment subgroups. 10000 simulation runs applied to each scenario. Moderate scenario is most common and realistic scenario among three for true TE.

We demonstrate that there are decision rules for subgroup analysis which decrease P and increase E simultaneously comparing to the situation of no subgroup analysis. These rules are much more liberal than the usual significance testing, since there is a high risk to decrease E using the latter.

Hasford J, Bramlage P, Koch G, Lehmacher W, Einhaeupl K, Rothwell P.M. Inconsistent trial assessments by the National Institute for Health and Clinical Excellence and IQWiG: standards for the performance and interpretation of subgroup analyses are needed. J Clin Epidemiology 2010;63:1298-1304.
Bender R, Koch A, Skipka G, Kaiser Th, Lange S. No inconsistent trial assessments by NICE and IQWiG: different assessment goals lead to different assessment results regarding subgroup analysess. J Clin Epidemiol 2010;63:1305-1307.
Ikeda K and Bretz F. Sample size and proportion of Japanese patients in multi-regional trials. PHARMACEUTICAL STATISTICS 2010;9:207-216.

STATISTICAL CRITERIA FOR COMPONENTS OF A PRIMARY COMPOSITE ENDPOINT TO SUPPORT BENEFIT-RISK ASSESSMENT AT AN INTERIM ANALYSIS USING BLINDED DATA

Jochen Brumm, F. Hoffmann - La Roche, Basel, Switzerland

Stopping a confirmatory clinical trial for cardiovascular outcomes using a composite endpoint at an interim analysis typically requires - besides the criterion on the primary endpoint - supportive evidence of efficacy in the components of the composite endpoint. These analyses are typically not adequately powered to ensure a prospectively specified p-value, so the criteria for the benefit risk assessments are often given in a qualitative nature, using words like “results should not be in contradiction” or “directionally consistent effects”.

Here we provide quantitative guidance for decision makers by modeling the distribution function of the observed hazard ratio for the components of the primary endpoint, given the observed hazard ratio for the primary endpoint. This is accomplished using the asymptotic multivariate normal distribution of the log hazard ratios, estimating the correlation by sampling from the blinded data available.

This distribution is used to indicate what hazard ratios on the secondary endpoint can be expected under various treatment effect assumptions, allowing to distinguish analyses that should be considered for decision making from hopelessly underpowered analyses which yield no information on the efficacy of the drug.

PERFORMANCE CHARACTERISTICS OF BIOMARKER-BASED PATIENT SELECTION RULES IN ADAPTIVE ENRICHMENT DESIGNS

Johannes Krisam, University of Heidelberg, Heidelberg, Germany
Meinhard Kieser, University of Heidelberg, Heidelberg, Germany

In the planning stage of a clinical trial investigating a potentially targeted therapy there is commonly a high degree of uncertainty whether the treatment is more efficient (or efficient only) in a subgroup as compared to the whole population. Recently developed adaptive designs enable to plan an efficacy assessment both for the whole population and a subgroup and to select the target population mid-course based on interim results (see, e.g., [1-4]). Frequently, predictive biomarkers are used in these trials for identifying patients more likely to benefit from a drug. The performance of the applied subset selection rule is crucial for the overall characteristics of the design.

We investigate the features of biomarker-based subgroup selection rules to be applied in adaptive two-stage designs: Methods are developed that allow an evaluation of the operational characteristics of rules for selecting the target population thus enabling to choose an appropriate strategy. Optimal decision rules are derived for the situation of uncertain assumptions about treatment effects by modeling the uncertainty about parameters by prior distributions. We evaluate the performance of these optimal decision rules in terms of Type I error rate and power in the setting of an adaptive enrichment design. Both the situation of perfect and imperfect (e.g. sensitivity and/or specificity smaller than 100%) biomarkers are considered.

[1] Wang, S. J., O'Neill, R. T., Hung, H. M. J. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. Pharmaceutical Statistics 2007; 6: 227-244.
[2] Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, Racine-Poon A. Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. Statistics in Medicine 2009; 28:1445–1463.
[3] Jenkins M, Stone A, Jennison C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. Pharmaceutical Statistics 2011; 10: 347–356.
[4] Friede T, Parsons N, Stallard N. A conditional error function approach for subgroup selection in adaptive clinical trials. Statistics in Medicine 2012; 31: 4309–4320.

THE USE OF PROPENSITY SCORES IN OBSERVATIONAL TRIALS

Jonathan Alsop, Numerus Limited, Kirchentellinsfurt, Germany

Propensity score analysis has been around for 30 years but has really only filtered into the consciousness of clinical trial statisticians during the past five or so years. Analyses using propensity scores aim to reduce bias due to confounding in observational trials. Here, the focus is on the propensity to be treated, whereby some patients are more likely to receive certain treatments compared with other patients. Perhaps the most problematic aspect of propensity score analysis is the matching of propensity scores, where the aim is to matche patients with a similar propensity to receive a treatment. The talk will highlight the various pitfalls associate with this method, and will include real-life examples of the use of propensity scores (including a observational cohort of Pulmonary Arterial Hypertension patients).

MAX-TEST TO EVALUATE GENETIC ASSOCIATION STUDIES FOR CONTINOUS AND TIME-TO-EVEN TRAITS

Ludwig Hothorn, Leibniz University Hannover, Germany
Esther Herberich, LMU, Germany

The association between a di-allelic marker and a quantitative trait (in population-based designs) can be analysed by i) ANOVA-F-test, ii) linear regression model or iii) maximum test for i) any heterogeneity, ii) an additive mode or iii) one of pre-selected mode of inheritances.

Most studies use case-control designs and therefore 2-by-k table data, some quantitative traits, such as BMI and time-to-event data. The max-test for 2-by-k table was described by Freidlin et al. [1]. We demonstrate the analysis of normal distributed data with substantial variance heterogeneity, scores data (by means of a non-parametric approach according to Konietschke et al. [2]) and for censored time-to-event data. Related real data example will be evaluated by means of the R packages multcomp and nparcomp.

[1] Freidlin B, Zheng G, Li Z, Gastwirth J: Trend tests for case-control studies of genetic markers: Power, sample size and robustness. Human Heredity 2002, 53(3):146-152.
[2] Konietschke, F;L.A. Hothorn, Brunner, E. Rank-based multiple test procedures and simultaneous confidence intervals Electronic Journal of Statistics Vol. 6 (2012) 737–758, DOI: 10.1214/12-EJS691

USING THE LASSO AND RIDGE REGRESSION IN CASE-CONTROL STUDIES WITH ZERO-INFLATED PREDICTORS

Maria Kohl, Medical University of Vienna, Austria
Georg Heinze, University of Vienna, Austria

Many experiments conducted in molecular biology compare intensity values between two groups of independent biological samples which are obtained by microRNA transcriptomics, proteomics or metabolomics ('Omics') procedures. The groups differ in experimental condition or health status of the subjects from which the samples were taken from. Zero intensity values are frequently encountered and caused by features (e.g., microRNAs, peptides or metabolites) which are below detection levels or are completely absent. Thus, each feature is characterized by two components, a binary variable D discriminating zero from non-zero intensity, and a continuous variable X with log2 of intensity values.

By means of several plasma-proteomic case-control studies, we evaluate several strategies to build a multivariable predictor discriminating cases from controls. Features with excess frequencies of zeros are first filtered out, and the elastic net (combination of LASSO and ridge regression) is used for further feature selection and model estimation: model 1 only considers D for each feature, model 2 only considers X for each feature, model 3 considers D and X as non-related variables, and model 4 selects features by performing the LASSO on X only and refits the model with ridge regression on the selected X and corresponding D. For a zero intensity measurement, models 2, 3 and 4 need assumptions about the corresponding value of X. Thus, hyperparameters of model development are a) the minimum required proportion of non-zeros for each feature, b) the positioning of the zeros in X, and c) the lambda parameters for the elastic net, all of which can be tuned by leave-one-out cross-validation. Model performance is assessed by means of cross-validated misclassification rate, discrimination index and explained variation.

Overall, the four strategies select between 6 and 46 features, where fewer features are selected by model 1. The performance of models 2-4 is similarly good, while model 1 is slightly inferior. However, the differences in the performance measures between the modeling strategies are small in absolute terms.

UPDATING THE PROBABILITY OF SUCCESS OF CLINICAL DEVELOPMENT PROGRAMS OVER TIME

Markus Abt, F. Hoffmann-La Roche AG, Basel, Switzerland
Paul Jordan, F.Hoffmann-La Roche Ltd., Switzerland

Clinical development programs for new medicines can take several years. Already based on results from early studies, quantification of the probability of success (PoS) of a Phase 3 program is of interest before it is initiated to make informed portfolio decisions. Later, when the first Phase 3 pivotal trial(s) are under way, information external (e.g., completion of parallel studies) or internal (e.g., passing a futility analysis) to the Phase 3 trial(s) may become available. Emerging information at such milestones during the lifecycle of a molecule can trigger requests to re-evaluate the PoS of the Phase 3 program. The presentation will discuss a Bayesian approach to update PoS as new information becomes available. We will define PoS and illustrate how the calculation differs depending on whether external or internal information is obtained. The technical implementation is illustrated with an example from the cardiovascular area. Prior to results at milestones becoming available, teams should discuss how PoS would change depending on hypothetical scenarios. Such considerations by the clinical team are essential to ensure that also the clinical expectations are met when the statistical PoS update has been performed.

DEVELOPING DRUGS FOR REIMBURSEMENT - CURRENT AND FUTURE CHALLENGES FOR STATISTICIANS

Martin Scott, Numerus, Tübingen, Germany

Drug research and development has reached a crucial turning point. No longer is the demonstration of efficacy, safety and quality sufficient for successful launch of a new product. A fourth, and perhaps, far more difficult hurdle now has to be proven: effectiveness. Regardless of whether “comparative” or “cost” effectiveness is implied, the goal is essentially one and the same, namely to bring meaningful benefit to patients at a cost that can be justified.

This talk will look at the current challenges of developing drugs for reimbursement and will try to encourage statisticians to think of what can be done if the demands of regulatory agencies and payors can be satisfied, both now, and in the future. In particular, it will explore the possibilities of how clinical development may be reformed so that the interests of all parties concerned are met, including those of sponsors.

A PARAMETRIC RÜGER TEST AND DIAGNOSTIC TRIALS

Robin Ristl, Medical University Vienna, Vienna, Austria
Christian Gartner, Austrian Medicines and Medical Devices Agency, Vienna, Austria
Thomas Lang, AGES Medizinmarktaufsicht, Austria
Martin Posch, Medical University of Vienna, Austria

Rüger [1] suggested a procedure to test the intersection hypothesis $H = H1 \cap \ldots \cap Hn$ based on conservative tests of n individual hypotheses Hi and to reject H if at least k of the individual hypotheses can be rejected at level $\alpha k/n$. For pre-specified k, the type I error rate of this test is bounded by α, regardless of the joint distribution of the test statistics.

For the case of multivariate normally distributed test statistics and $n = 3$, $k = 2$ we investigate a parametric variant of the Rüger test where the unadjusted level ? is applied instead of $\alpha k/n$ to test the individual hypotheses. For equicorrelated test statistics we show analytically that this test controls the familywise type I error rate at level α. For general covariance matrices numerical investigations suggest that the maximum type I error is α, too. We conclude that for trivariate normally distributed test statistics the test procedure does not require adjustment of the critical value α applied to the individual tests.

An application of the Rüger test in the regulatory setting are clinical trials for diagnostics, where three readers independently assess the same diagnostic images. For example, binomial tests can be performed for each reader to test the null hypotheses that the sensitivity is at or below a certain threshold. If this hypothesis can be rejected for two out of three readers, the trial is considered successful. Because the test statistics are asymptotically multivariate normal, this procedure corresponds to a parametric Rüger test for the null hypotheses that for each of the three clinical experts the sensitivity is at or below the threshold. The analytical and numerical results on the parametric Rüger test show that the proposed testing strategy is conservative and can be applied without type I error adjustment in this situation.

---

[1] Rüger, B. (1978), Das maximale Signifikanzniveau des Tests: Lehne H0 ab, wenn k unter n gegebenen Tests zur Ablehnung führen." Metrika, 25, 171-178.

Committee for medicinal products for human use, European Medicines Agency (2009), “Appendix 1 to the guideline on clinical evaluation of diagnostic agents (CPMP/EWP/1119/98 REV. 1) on imaging agents,” Doc. Ref. EMEA/CHMP/EWP/321180/2008.

MODELLING POWER-LAW SPREAD OF INFECTIOUS DISEASES

Sebastian Meyer, University of Zurich, Zurich, Switzerland
Leonhard Held, Division of Biostatistics, Institute for Social and Preventive Medicine, University of Zurich, Switzerland

Short-time human travel behaviour may be described by a power law with respect to distance x, i.e. f(x) ∝ x^{-d} with positive decay parameter d [1]. We incorporate this information in space-time models for infectious disease surveillance data to better capture the dynamics of disease spread.

Two previously established model classes are extended, which both decompose disease risk additively into endemic and epidemic components: a multivariate time series model for aggregated surveillance counts proposed by Held and Paul [2], and a space-time point process model for individual point-referenced data proposed by Meyer et al. [3]. In both frameworks, the power-law spread is embedded into the epidemic component and its decay parameter d is estimated simultaneously with all other unknown parameters using (penalised) maximum-likelihood inference.

The performance of the new approach is investigated by a re-analysis of count data on influenza in 140 administrative districts of Southern Germany (2001-2008), and individual cases of invasive meningococcal disease in Germany (2002-2008). In both applications, the power-law formulation substantially improves model fit and predictions. Implementation in the R package surveillance allows to apply the approach in other settings.

[1] Brockmann, D., Hufnagel, L., and Geisel, T. (2006). The scaling laws of human travel. Nature, 439(7075):462–465.
[2] Held, L. and Paul, M. (2012). Modeling seasonality in space-time infectious disease surveillance data. Biometrical Journal, 54(6):824–843.
[3] Meyer, S., Elias, J., and Höhle, M. (2012). A space-time conditional intensity model for invasive meningococcal disease occurrence. Biometrics, 68(2):607–616.

SAMPLE SIZE CALCULATIONS AND THE TEMPORAL DEPENDENCY BETWEEN AN INFECTION AND ITS SEQUELA IN RANDOMISED CONTROLLED TRIALS

Sereina Herzog, Medical University of Graz, Graz, Austria
Andrea Berghold, Medical University of Graz, Graz, Austria

The success of an intervention (e.g. screening) that aims to reduce infection sequela by reducing the exposure is influenced by the temporal dependency between infection and sequela. Hence, sample size calculations are also affected. Pelvic inflammatory disease (PID) is a sequela of the sexually transmitted Chlamydia trachomatis infection for which the timing of development is unknown. A randomised controlled trial (RCT) investigating the effect of a single round screening for chlamydia on PID incidence over a 12 months period of follow-up found no significant reduction in PID incidence. The authors argued that the trial was underpowered. We re-examined their assumptions about the PID incidence and the relative risk (RR) and investigated how different temporal dependency assumptions can influence the sample size calculation.

We used a compartmental model to examine three hypothetical processes for PID development: at the start, at the end, or throughout the duration of infection.2 For all processes we assumed that, of all women infected, the same fraction will develop PID in the absence of an intervention. We determined the fraction such that the PID incidence in the control group equals the incidence used in the sample size calculation for the RCT. The incidences in the intervention group derived by the model are used for the sample size calculations. The results are compared to the two sample size calculations done for the RCT: 2% PID incidence and 0.48 RR; 3% PID incidence and 0.44 RR.

The first sample size calculation conducted in the trial using log-rank test was compatible with the assumption that PID develops through infection period. In the second calculation, the relationship was compatible to the process where PID develops at the end of the infection.

The incidence and RR determine the underlying assumption about the temporal dependency between an infection and its sequela and the assumption can vary considerably even for small changes in incidence and RR.

Oakeshott P, Kerry S, Aghaizu A, Atherton H, Hay S, Taylor-Robinson D et al. Randomised controlled trial of screening for Chlamydia trachomatis to prevent pelvic inflammatory disease: the POPI (prevention of pelvic infection) trial. BMJ 2010; 340:c1642.
Herzog SA, Althaus CL, Heijne JCM, Oakeshott P, Kerry S, Hay P, Low N. Timing of progression from Chlamydia trachomatis infection to pelvic inflammatory disease: a mathematical model. BMC Infect Dis 2012; 12(1):187.

DOUBLE CRITERIA DESIGN FOR A MINIMALLY IMPORTANT DIFFERENCE IN SINGLE-ARM SINGLE-STAGE PHASE II TRIALS WITH A BINARY PRIMARY ENDPOINT

Shu-Fang Hsu Schmitz, Novartis Pharma AG, Basel, Switzerland

The incorporation of a minimally important difference (MID) for sample size calculation has been proposed many times in the literature [1-4]. The purpose is to ensure a positive decision based on both statistical significance and clinical relevance. For a given measure of the primary endpoint (D), a positive result will be declared using the following double criteria: (i) the confidence interval (CI) does not cover the null hypothesis value (d0) and (ii) the point estimate is not worse than the MID, which is a clinically equivocal value. The corresponding frequentist approach to sample size calculation for such double criteria designs sets power at 50% at the MID for a given type I error (alpha) and d0. The Bayesian approach requires that the posterior risk of D<d0 is at most alpha and the posterior risk of DMID is at least 50%. Most publications present applications of such designs for between-arm comparisons based on normal distributions or approximation. Here we investigate whether the double criteria can always be satisfied for single-arm single-stage phase II designs with a binary primary endpoint. Following the Fleming's design [5, 6] in the frequentist approach, a trial would be declared as positive if the observed number of successes is greater than or equal to the given critical value. In this situation the point estimate is always greater than the MID, hence satisfies condition (ii). Concerning the one-sided CI, if it is calculated using more exact methods like Clopper-Pearson, Wilson [7], Agresti-Coull [8], or Jeffreys [9], then condition (i) is always satisfied. This is not always the case if simple normal approximation is used to calculate CI. For Bayesian designs using different beta prior distributions under practical settings, condition (i) for the credible interval and condition (ii) for the point estimate are always satisfied.

[1] Harris RJ, Quade D (1992). The minimally important difference significant criterion for sample size. Journal of Educational Statistics 17(1):27-49

[2] Nicewander WA, Price JM (1997). A consonance criterion for choosing sample size. The American Statistician 51(4):311-317

[3]. Kieser M, Hauschke D (2005). Assessment of clinical relevance by considering point estimates and associated confidence intervals. Pharmaceutical Statistics 4:101-107. DOI: 10.1002/pst.161

[4] Neuenschwander B, Rouyrre N, Hollaender N, Zuber E, Branson M (2011). A proof of concept phase II non-inferiority criterion. Statistics in Medicine 30(13):1618-27. DOI: 10.1002/sim.3997

[5] Fleming TR (1982). One-sample multiple testing procedure for Phase II clinical trials. Biometrics 38:143-151

[6] A'Hern RPA (2001). Sample size tables for exact single-stage phase II designs. Statistics in Medicine 20:859-866

[7] Wilson EB (1927). Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association 22: 209–212

[8] Agresti A, Coull BA (1998). Approximate is better than 'exact' for interval estimation of binomial proportions. The American Statistician 52: 119–126

[9]. Brown LD, Cai TT, DasGupta A (2001). Interval estimation for a binominal proportion. Statistical Science 16(2):101-133

STOPPING RULES FOR SEQUENTIAL TRIALS IN HIGH-DIMENSIONAL DATA

Sonja Zehetmayer, Medical University of Vienna, Vienna, Austria
Martin Posch, Medical University of Vienna, Vienna, Austria
Alexandra Graf, Medical University of Vienna, Vienna, Austria

Sequential trials have been proposed that allow for an early stopping of the trial in studies involving large scale hypothesis testing as in microarray experiments. To control the False Discovery Rate, multiplicity adjustment is required only for the number of hypotheses but not for the number of interim looks (under suitable assumptions asymptotically for an increasing number of hypotheses).

In this talk we introduce novel stopping rules that stop a trial early if a certain success criterion is fulfilled based on the proportion of rejected hypotheses. Using simulations studies we investigate the actual size of the error rate and 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.

UNDERSTANDING TREATMENT MECHANISMS IN CLINICAL SURVIVAL TRIALS - A CAUSAL APPROACH

Susanne Strohmaier, University of Oslo, Norway
Theis Lange, University of Copenhagen, Denmark
Ørnulf Borgan, University of Oslo, Norway
Odd O. Aalen, University of Oslo, Norway

Regulatory restrictions for clinical survival trials usually require the application of survival methods that solely utilize baseline covariates and the intention-to-treat principle, even though important covariates are measured repeatedly over time, patients withdraw and underlying study plans allow patients to switch treatments or change dosage. Thereby a lot of useful information is ignored that could help to explain how treatments work. We aim to illustrate how dynamic path analysis can shed more light on underlying treatment dynamics.

Fosen and co-authors proposed a model for dynamic path analysis, which makes it feasible to distinguish between direct, indirect and total effects in the presence of a survival outcome and time-dependent covariates, a concept particularly useful for understanding mechanisms of treatment effects. We suggest how this method can be utilized and extended to incorporate a larger amount of routinely collected information with a focus on a causal interpretation.

The methods are applied to a large-scale secondary prevention trial comparing different statin treatment strategies. Previously performed post-hoc analyses aimed to assess the amount of treatment effect explained by various blood lipid measures and to adjust for adherence in a rather heuristically way. We show to which extend these results can be confirmed by a mediation analysis and which additional insights can be gained by exploiting information on the development of lipid levels over time as well as modelling the processes of treatment and dosage switching in a more sophisticated way.

Analysis guidelines developed by regulatory authorities aiming to answer the main treatment question typically only request utilizing baseline covariates, while important information on internal covariates is often ignored. We show how the method of dynamic path analysis can be used to provide important insights when it comes to the secondary objective of understanding treatment mechanisms.

SUBGROUP SELECTION IN CLINICAL TRIALS

Susanne Strohmaier, University of Oslo, Norway
Thomas Jaki, Lancaster University, United Kingdom of Great Britain and Northern Ireland

A randomized controlled clinical trial is considered the gold standard in providing evidence that a (new) treatment is clinically effective in treating a targeted disease. With the growing interest in personalised medicine and individualised therapies extensions from the conventional single population design to more flexible design approaches, which allow more informative evaluation in patients responding differently to medication, have become increasingly important.

More specifically, we considered two stage clinical trials, where at an interim analysis after the first stage the patient subpopulation showing the most promising treatment effect is selected for further recruitment in the second stage. The statistical considerations are similar to those proposed for the related treatment selection problem and mainly fall into one of two families of approaches: On the one hand, there is the group sequential approach, where test statistics obtained from accumulated data after each stage are compared to appropriate boundary values. On the other hand, there is the p-value combination approach, where p-values obtained from stage-wise data are combined using a pre-specified combination function and compared to respective critical values at the final analysis. We compared both approaches in the context of the subpopulation selection problem in terms of type I error and power in a simulation study.

TEST PROCEDURES AND EFFICACY CLAIMS FOR COMPOSITE ENDPOINTS AND ITS COMPONENTS

Svenja Schüler, University of Heidelberg, Heidelberg, Germany
Geraldine Rauch, University of Heidelberg, Heidelberg, Germany
Meinhard Kieser, University of Heidelberg, Heidelberg, Germany

Composite endpoints combine several events of interest within a single time-to-first-event variable. They are increasingly used in clinical trials, particularly in the field of cardiology and oncology. In the ICH E9 Guideline, it is stated that "this approach addresses the multiplicity problem without requiring adjustment to the type I error" [1]. In fact, to demonstrate the significance of an overall clinical benefit, it is sufficient to assess the test problem formulated for the composite. However, the effect observed for the composite does not necessarily reflect the effects for the components. Therefore, it would be desirable that the sample size for clinical trials using composite endpoints provides enough power not only to detect a clinically relevant superiority for the composite but also to address the components in an adequate way. Including the most relevant component effects in an efficacy claim assessed by a confirmatory test strategy could overcome this problem, however imposes the problem of multiplicity [2].

We discuss several alternative efficacy claims for clinical trials with composite endpoints based on multiple test procedures and give recommendations for different situations most common met in clinical practice. For example, we consider the case that the most relevant component (death) contributes the lowest number of events to the composite or the alternative case that there exist two (composite) endpoints of possible interest but it is uncertain which one will correspond to the larger effect in the planning stage. The properties of these multiple test approaches in terms of required sample size and power are compared. Applications are illustrated by clinical trial examples.

[1] ICH E9. Statistical Principles for Clinical Trials. 1999.
http://www.emea.europa.eu/pdfs/human/ich/036396en.pdf (assessed 17.05.2013).
[2] Rauch G, Beyersmann J. Planning and evaluating clinical trials with composite time-to-first-event endpoints in a competing risk framework. Statistic in Medicine (2013), doi: 10.1002/sim.5798 (Epub ahead of print).

.

RECENT DEVELOPMENTS IN IDENTIFYING TRANSMISSION ROUTES OF HEALTHCARE ASSOCIATED INFECTIONS USING WHOLE GENOME SEQUENCE DATA

Theodore Kypraios, University of Nottingham, United Kingdom of Great Britain and Northern Ireland

Healthcare-associated infections (HCAIs) remain a problem worldwide, and can cause severe illness and death. It is estimated that 5-10% of acute-care patients are affected by nosocomial infections in developed countries, with higher levels in developing countries.

Statistical modelling has played a significant role in increasing understanding of HCAI transmission dynamics. For instance, many studies have investigated the dynamics of MRSA transmission in hospitals, estimating transmission rates and the effectiveness of various infection control measures. However, uncertainty about the true routes of transmission remains and that is reflected on the uncertainty of parameters governing transmission.

Until recently, the collection of whole genome sequence (WGS) data for bacterial organisms has been prohibitively complex and expensive. However, technological advances and falling costs mean that DNA sequencing is becoming feasible on a larger scale.

In this talk we first describe how to construct statistical models which incorporate WGS data with regular HCAIs surveillance data (admission/discharge dates etc) to describe the pathogen's transmission dynamics in a hospital ward. Then, we show how one can fit such models to data within a Bayesian framework accounting for unobserved colonisation times and imperfect screening sensitivity using efficient Markov Chain Monte Carlo algorithms. Finally, we illustrate the proposed methodology using MRSA surveillance data collected from a hospital in North-East Thailand.

MEASURING AN EFFECT SIZE FROM DICHOTOMIZED DATA: CONTRASTED RESULTS WHETHER USING A CORRELATION OR AN ODDS-RATIO

Valentin Rousson, Lausanne University Hospital, Lausanne, Switzerland

It is well known that dichotomizing continuous data has the effect to decrease statistical power when the goal is to test for a statistical association between two variables [1, 2]. Modern researchers however are not only focusing on statistical significance, but also on an estimation of the "effect size" (i.e. the strength of association between the variables) to judge whether a significant association is also clinically relevant [3]. In this presentation, we are interested in the consequences of dichotomizing continuous data on the value of an effect size in some classical settings. To be able to compare the strength of association before and after dichotomization, we need to use a concept which applies both to continuous and binary variables. Two possible choices are Pearson's correlation, which is directly related to the statistical power, also when dealing with binary variables [1], and for this reason would be a natural choice, or Agresti's generalized odds-ratio [4], which corresponds to the usual odds-ratio in the case of binary variables and which might be easier to interpret. It turns out that the consequences of dichotomization on the strength of association will not be the same whether using a correlation or an odds-ratio: whereas the value of a correlation is typically decreased by a factor sqrt(pi/2) after each dichotomization, the value of an odds-ratio is at the same time raised to the power sqrt(2). From a descriptive statistical point of view, it is thus not clear whether dichotomizing continuous data leads to a decrease or to an increase of the effect size, as illustrated using a data set to investigate the relationship between motor and intellectual functions in children and adolescents.

[1] Cohen J (1983). The cost of dichotomization. Applied Psychological Measurement, 7, 249-253.
[2] MacCallum RC, Zhang S, Preacher KJ and Rucker DD (2002). On the practice of dichotomization of quantitative variables. Psychological Methods, 7, 19-40.
[3] DeCoster J, Iselin AMR and Gallucci M (2009). A conceptual and empirical examination of justications for dichotomization. Psychological Methods, 14, 349-366.
[4] Agresti A (1980). Generalized odds-ratio for ordinal data. Biometrics, 36, 59-67.

MAMMOGRAPHY SCREENING AND CANCER REGISTRATION IN TYROL/AUSTRIA

Willi Oberaigner, TILAK/UMIT, Innsbruck, Austria

Managing large datasets in epidemiology poses a number of problems. Therefore, it was our aim to analyse problems and demonstrate solutions for working with large datasets in epidemiology based on two projects, namely evaluating the mammography screening program in Tyrol and the population-based cancer registry of Tyrol.

Starting from our experiences with these two projects, we will discuss solutions for data flow, data privacy, aspects of data quality like completeness, comparability, validity and timeliness and we will demonstrate results of statistical analysis.

Solutions for data privacy questions are very demanding and partly limiting, for example, problems with inviting women to mammography screening depending on her last screening visit. Pseudonymised data seems to be a good solution, however poses problems with data checks. Legal obligations for cancer registration are no sufficient means of reaching the goal of completeness in an epidemiological cancer registry. Reaching a high level of data quality needs strict definition and rules, but also appropriate training for data suppliers. Whether time-consuming projects for checking validity such as reabstraction can be realised has to be decided. In our experience, a balance has to be found between strict obligation of data plausibility rules and accepting datasets with acceptable errors.

Completeness of cancer registration in Tyrol has been shown to reach 97.5% at two years after end of year of diagnosis. For all cancer cases except non-melanoma skin cancer, survival analysis resulted in age-adjusted relative five-year survival estimates of 64.1 and 65.6 for woman and men respectively, which is as good as estimates from the SEER registries in the USA.

In the mammography screening program in Tyrol, two year attendance rate was estimated at 59% which is comparable to results from neighbouring countries. Recall to further assessment was fairly small at 1.4% while interval cancer rate is in the desired range according the EU guidelines.

Working with large datasets needs sound methodological background but also a well-established network between epidemiologists and data suppliers and experiences in order to reach complete and valid datasets.

NONPARAMETRIC MULTIVARIATE DENSITY ESTIMATION USING MIXTURES

Xuxu Wang, The University of Auckland, Auckland, New Zealand
Yong Wang, The University of Auckland, Auckland, New Zealand

A new method is proposed for nonparametric multivariate density estimation, which extends a general framework that has been recently developed in the univariate case based on nonparametric and semipara-metric mixture of normal distributions. The major challenge to a multivariate extension is the dilemma that one can not maximize directly the likelihood function with respect to the whole component covariance matrix, since the likelihood is unbounded for a singular covariance matrix. To select the covariance matrix does not solve the problem of unboundedness and moreover is this would be at least computationally demanding if not infeasible. We consider using a volume parameter h to enforce a minimal restriction on the covariance matrix so that, with h fixed, the likelihood function is bounded and its maximization can be successfully carried out with respect to all the remaining parameters. The role played here by the scalar h is just the same as by the bandwidth in the univariate case and its value can be determined by a model selection criterion, such as the Akaike information criterion. New efficient algorithms are also described for finding the maximum likelihood estimates of these mixtures under various restrictions on the covariance matrix. Empirical studies using simulated and real-world data show that the new multivariate mixture-based density estimator performs remarkably better than two state-of-the-art kernel-based density estimators.

Scott, D. W. (1992). Multivariate Density Estimation: Theory, Practice, and Visualization. New York: Wiley.
Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. London: Chapman and Hall.
Wang, Y. (2007). On fast computation of the non-parametric maximum likelihood estimate of a mixing distribution. Journal of the Royal Statistical Society, Ser. B, 69, 185–198.
Wang, Y. (2010). Maximum likelihood computation for fit- ting semiparametric mixture models. Statistics and Comput- ing, 20, 75–86.
Wang, Y. and C.-S. Chee (2012). Density estimation using non- parametric and semiparametric mixtures. Statistical Mod- elling, 12, 67–92.

3. YOUNG STATISTICIANS ABSTRACTS

EFFICIENT DATA-AUGMENTED MCMC METHODS FOR BINOMIAL LOGIT MODELS

Agnes Fussl, Johannes Kepler University Linz, Linz, Austria
Sylvia Frühwirth-Schnatter, Vienna University of Economics and Business, Vienna, Austria
Rudolf Frühwirth, Austrian Academy of Sciences, Vienna, Austria

Our work considers efficient Bayesian methods for data originating from experiments where binary outcome variables are aggregated in terms of binomial outcome variables and the data are modeled by a binomial regression model. Such models arise, for instance, when data take the form of two-way or three-way contingency tables or when repeated measurements are available for identical covariate patterns.

To perform MCMC sampling, we rewrite the binomial logit model as an augmented model which involves some latent variables called random utilities. It is straightforward, but inefficient, to use the individual random utility model (RUM) representation based on the binary observations reconstructed from each binomial observation. Alternatively, we suggest a new aggregated difference random utility model (dRUM) representation of the binomial logit model. The parameters are estimated by using three different MCMC algorithms: a data-augmented Metropolis-Hastings sampler, an auxiliary mixture sampler and a novel hybrid auxiliary mixture sampler. A comparison of their performances within a comparative case study on various data sets shows that the modifications lead to a considerable reduction of computing time and a noticeable gain in efficiency. The corresponding source codes of the different samplers are included in the R package binomlogit.

Fussl, A. (2012). binomlogit: Efficient MCMC for Binomial Logit Models. R package version 1.0. http://CRAN.R-project.org/package=binomlogit
Fussl, A., S. Frühwirth-Schnatter and R. Frühwirth (2013). Efficient MCMC for Binomial Logit Models. ACM Transactions on Modeling and Computer Simulation, Volume 23, Number 1, Article 3, 21 Pages, 2013.

ANALYSIS OF ASSOCIATION ON NON-PRODUCT SPACES

Anna Klimova, Institute of Science and Technology (I.S.T.) Austria

In some problems that appear in, for example, feature selection and multistage capture-recapture procedures, the sample space is not the Cartesian product of the individual sample spaces and patterns of association cannot be described using partitions in a contingency table. In such cases association can be analysed within the relational model framework proposed by Klimova, Rudas, & Dobra [1]. A relational model is generated by a class of subsets of cells in a contingency table, and, under the model, the cell probabilities are products of the effects associated with the generating subsets. The stochastic properties of relational models with the overall effect are similar to those of log-linear models, but the properties of relational models without the overall effect are surprisingly different. A pattern of association that may lead to a relational model without the overall effect is illustrated in the talk using a model of independence between congenital malformations. An extension of the iterative proportional fitting procedure can be used to compute the MLE under relational models with or without the overall effect. The algorithm is implemented in the gIPFrm R package and is demonstrated in the talk. The material presented here is joint work with Tamas Rudas.

[1] A. Klimova, T. Rudas, A. Dobra (2012). Relational models for contingency tables. J. Multivariate Anal., 104, 159--173.

_____

A. Klimova (2012). Coordinate-Free Exponential Families on Contingency Tables. PhD thesis. Advisers: Tamas Rudas and Thomas Richardson.

A. Klimova, T. Rudas (2013). gIPFrm: Generalized Iterative Proportional Fitting for Relational Models. R package version 1.1.

K.B.M. Källén, E.E. Castilla, M.D.G. Dutra, P. Mastroiacovo, E. Robert, B.A.J. Källén (1999). A modified method for the epidemiological analysis of registry data on infants with multiple malformations. International Journal of Epidemiology, 28, 701-710.

OVERRUNNING DATA METHODS: COMPARISONS BASED ON REAL DATA TRIAL

Nicola Soriani, University of Padua, Padova, Italy
Ileana Baldi, University of Padua, Padova, Italy
Dario Gregori, University of Padua, Padova, Italy

Sequential trial designs foresee one or more interim analyses (IA) before the full sample size has been reached. Such IA have the primary purpose to terminate the trial when futility or superiority of one of the interventions becomes clear, according to prespecified stopping rules.

Overrunning occurs when data continue to be collected also if a stopping criterion has been reached. Overrunning data collected according to the trial protocol are considered valid and should be included in the analyses but they could influence the results and change the conclusions.

Over the years many proposals to deal with overrunning were proposed. Deletion method includes overrunning data, ignoring the interim analysis that has led to the stopping of the trial. The methods of combining p-values rely on the idea to make two different analyses, one on the sequential portion of the data and one on the overrunning part, and to combine them by weighting their p-values. The repeated confidence interval approach is a further alternative to adopt for the overrunning problem.

A multicenter phase III trial, for non-inferiority of a Test drug compared to a Reference drug, is the motivating example. The trial was designed assuming response rates for Test and Reference drug respectively of 50% and 45%, a non-inferiority margin of 15%, a power of 80% and a 2.5% one-sided significance level. The non-inferiority of the Test drug emerged at the first of three IA. However, the majority of the pre-planned patients had already been recruited so the trial sponsor decided to analyze the full sample to confirm the non-inferiority hypothesis.

This motivating example is used as the basis for a simulation study aimed at comparing different methods for overrunning under a variety of data generating mechanisms. Preliminary results show similar behaviors for deletion and combining p-values methods while repeated confidence interval approach seems to be the most conservative.

Fleming TR, Harrington DP, O'Brien PC. Designs for group sequential tests. Controlled Clinical Trials. 1984;5(4):348-361.
Jennison C, Turnbull BW. Interim analyses: the repeated confidence interval approach. Journal of the Royal Statistical Society. Series B (Methodological). 1989:305-361.
Hall W, Ding K. Sequential tests and estimates after overrunning based on p-value combination. Pushing the

Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh. 2008;3:33-45.
Whitehead J. Overrunning and underrunning in sequential clinical trials. Controlled Clinical Trials. 1992;13(2):106-121.

PROBLEMS OF INLA IN GENERALIZED LINEAR MIXED MODELS FOR BINARY RESPONSES

Rafael Sauter, University of Zurich, Zurich, Switzerland
Leonhard Held, Division of Biostatistics, Institute for Social and Preventive Medicine, University of Zurich, Zurich, Switzerland

Clustered observations such as longitudinal data are preferably analyzed with mixed models. Non-normal responses belonging to the exponential family are modeled in analogy with generalized linear models by generalized linear mixed models (GLMM). Maximum likelihood (ML) estimation of a GLMM is usually based on the marginal likelihood and requires numerical integration with e.g. Gauss Hermite quadrature approximation, which may lead to unstable parameter estimates depending on the number of quadrature points. Bayesian or approximate Bayesian estimation such as the integrated nested Laplace approximation (INLA) [1] of GLMM's is an alternative approach, which takes the hierarchical structure of GLMM's into account but requires prior assumption about the hyperparameters. When comparing INLA to MCMC inaccuracies by INLA are revealed in this paper. The deficiencies of INLA are further investigated by comparing ML estimation, MCMC and INLA for fixed hyperparameters at values corresponding to the ones obtained by ML estimation. Applications to clinical trial data on toenail infections illustrate that the ML estimates are in fact closer to MCMC posterior distributions than the results by INLA.

[1] Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). Journal of the Royal Statistical Society - Series B 71 319–392.

STATISTICAL METHODS FOR IMPROVEMENT OF MOTIF FINDING ALGORITHMS

Živa Stepančič, Arctur d.o.o., Nova Gorica, Slovenia

With expansion of genome data bases, due to rapid development of sequencing technology, the need for accurate and faster motif finding algorithms is still a large challenge in biostatistics. A number of motif finding algorithms were already developed in recent years, leading with Gibbs motif sampler and its variations.

Although some improvements towards time complexity in these variations were made, the accuracy of their results is still not optimal. Furthermore, recent research implied that the correlation between expressions of genes in correlated species needs to be addressed as well in the algorithms.

Thus the implementation of reliable statistical methods for improvement of motif finding algorithms is called for in solving this particular challenge. Our aim is to evaluate some of the possible solutions concerning computational complexity of motif finding algorithms and application of statistical tools needed for better and more accurate interpretation of their results.

Thompson, A.W. et al., The Gibbs Centroid Sampler, Nucleic Acids Research, 2007, Vol. 35, W232-W237
Das, M.K. & Dai, H., A survey of DNA motif finding algorithms, BMC Bioinformatics, 8 (2007), (suppl 7):S21
Lawrence, C.E. et al., Detecting subtle sequence signals: A Gibbs sampling stategy for multiple alignment, Science, 8 (1993),Vol. 262, 208-214
Tompa, M. et al., Assessing computational tools for the discovery of transcription factor binding sites, Nature Biotechnology, 23 (2005), 137-144

.

4. POSTER ABSTRACTS

THE USE OF ANTIPERSPIRANTS WITH ALUMINIUM SALTS AND ITS RELATION TO BREAST CANCER

Caroline Linhart, Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Innsbruck, Austria
Hanno Ulmer, Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Innsbruck, Austria
Nicole Concin, Department of Obstetrics and Gynecology, Innsbruck Medical University, Innsbruck, Austria
Susanne Taucher, Department of Obstetrics and Gynecology, Innsbruck Medical University, Innsbruck, Austria
Evi Morandi, Department of Plastic, Reconstructive and Aesthetic Surgery, Innsbruck Medical University, Innsbruck, Austria.
Chris Exley, Birchall Centre, Lennard-Jones Laboratories, Keele University, Staffordshire ST5 5BG, UK
Herbert Lindner, Biocenter, Division of Clinical Biochemistry and Protein Micro-Analysis Facility, Innsbruck Medical University, Innsbruck, Austria
Heribert Talasz, Biocenter, Division of Clinical Biochemistry and Protein Micro-Analysis Facility, Innsbruck Medical University, Innsbruck, Austria

Studies of antiperspirants containing aluminum salts and their effect on breast cancer have shown conflicting results. We designed a study consisting of two parts. Case-control study: History of antiperspirant use will be compared between a group of 262 female breast cancer patients aged 20-85 years (n=131 cases) and age-matched controls (n=131) without breast cancer. A personal interview regarding individual hygiene, life-style, and aluminum exposure will be performed. The study questionnaire is partly based on the MARIE study of the German Cancer Research Centre. Study part II: Aluminum measurement of breast tissue. Case group: 40 Patients with breast cancer that need breast surgery because of malign diagnosis. Control group: 40 Patients (without any history of breast cancer) that will undergo a breast reduction surgery. Breast tissue will be collected for determination of aluminium amount with atomic absorption spectroscopy (AAS). If possible, one breast specimen will be sampled in the breast quadrant close to the axilla and one specimen in a quadrant away from the axilla. These patients will also be interviewed with the same study questionnaire.

SIMULTANEOUS CONFIDENCE INTERVALS FOR ADAPTIVE GRAPH-BASED MULTIPLE TEST PROCEDURES

Florian Klinglmueller, Medical University of Vienna, Vienna, Austria

Simultaneous confidence intervals are presented that are compatible with adaptive graph-based multiple test procedures. That is, the intersection of the resulting confidence region with an elementary null hypothesis is empty, if and only if the underlying test procedure rejects that null hypothesis. The approach extends related approaches for adaptive closed test procedures based on combination tests and the conditional error rate principle [1], to adaptive weighted Bonferroni-type procedures based on partial conditional error rates. Consequently, knowledge of the joint distribution of test statistics is not required. Furthermore, if no interim adaptations are performed, the preplanned test and corresponding simultaneous confidence intervals also control the unconditional FWER.

_____

[1] Magirr D., Jaki T., Posch M., Klinglmueller F. – Simultaneous confidence intervals that are compatible with closed testing in adaptive designs. - Biometrika - to appear

STATISTICAL METHODS IN THE METABOLIC SYNDROME AND CANCER PROJECT (ME-CAN)

Hanno Ulmer, Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Innsbruck, Austria
Michael Edlinger, Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Innsbruck, Austria
Susanne Strohmaier, University of Oslo, Norway
Håkan Jonsson, Epartment of Radiation Sciences, Oncology, UmeåUniversity, Sweden
Tanja Stocks, Department of Surgical and Perioperative Sciences, Urology and Andrology, Umeå University, Sweden
Tone Bjorge, Department of Public Health and Primary Health Care, University of Bergen, Bergen, Norway
Jonas Manjer, Department of Surgery, Malmö University Hospital, Lund University, Malmö, Sweden
Gabriele Nagel, Institute of Epidemiology, University of Ulm, Germany
Pär Stattin, Department of Surgical and Perioperative Sciences, Urology and Andrology, Umeå University, Sweden

In the Metabolic syndrome and Cancer (Me-Can) project health examination data from seven cohorts in Norway, Sweden, and Austria were pooled to a total of 578,700 participants. The research aim of Me-Can was to longitudinally investigate possible associations between single metabolic risk factors, the metabolic syndrome as entity, and subsequent risk of cancer. Metabolic risk factors include body mass index, blood pressure, total cholesterol, triglycerides and glucose. Outcomes were defined as cancer incidence and cancer mortality outcome information came from national and local cancer registries and were linked to the health examination data. This poster describes the statistical methods used in the Me-Can project. These include standard survival analysis techniques, such as the Cox proportional hazards regression, correction for measurement error and regression dilution bias, lag-time analyses, and the use of splines in order to model non-linear associations. As a proxy for the metabolic syndrome as an entity a score of the sum of the standardized z-scores of the single metabolic factors was constructed.

NON-PARAMETRIC CONTROL CHART APPROACH AND VISUALISATION FOR THE SURVEILLANCE OF NOTIFIABLE INFECTIOUS DISEASES IN BADEN-WUERTTEMBERG

Iris Zoellner, Baden-Wuerttemberg State Health Office, Stuttgart, Germany
Fee Koehler, Stuttgart University, Stuttgart, Germany

In Baden-Wuerttemberg, the surveillance of infectious diseases is one of the tasks of the State Health Office in Stuttgart. The usual statistical reporting of notified infectious disease cases in weekly tables seemed not very appropriate for the detection of outbreaks and case clusters in time due to different seasonal patterns for different diseases and notifiable categories like norovirus infections. For this reason we use a method for the visualisation of infectious disease data which is based on a non-parametric modification of control charts specificly for the surveillance of infectious diseases. These infectious disease control charts allow a visual comparison of current data with the data observed in the last five years and a timely detection of unusual clustering in time (outbreaks).

The method should be applicable to all notifiable diseases and was therefore developed without parametric assumptions about the distribution of case numbers and seasonal or other trends in these numbers. For different infectious diseases we observed different dis-tributions in time (seasonal variations). The visualisation in each control chart for the differ-ent notifiable categories includes the presentation of the variation in the number of notified cases of this disease observed in the last five years for the weeks 1 to 52 and the currently notified number of cases per week.

This visualisation technique will be shown for different notifiable infectious diseases and illustrated by different outbreak examples as the Legionellosis outbreak in Ulm 2009/10, the epidemics caused by infections with EHEC in 2011 and an Adenovirus outbreak in 2012 in the Southwest of Baden-Wuerttemberg. Weekly reports for Baden-Wuerttemberg are published under http://www.gesundheitsamt-bw.de/oegd/Fachservice/IfSG-Wochenberichte/Seiten/wochenberichte.aspx [1].

_____

[1] Landesgesundheitsamt Baden-Württemberg(2013): Infektionsberichte 2013 http://www.gesundheitsamt-bw.de/oegd/Fachservice/IfSG-Wochenberichte/Seiten/wochenberichte.aspx

MODELING REPEATED OBSERVATIONS IN LONGITUDINAL STUDIES WITH MIXED-EFFECTS MODELS AND SURVIVAL ANALYSIS

Josef Fritz, Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Innsbruck, Austria
Hanno Ulmer, Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Innsbruck, Austria

Mixed-effects models applied on data with repeated observations allow a correction of measurement error. They can also be used to model time courses and to detect age-dependent patterns, for example in vital signs or laboratory parameters. Furthermore, mixed-effects models can yield subject-specific random effects to estimate the underlying time-trend for each subject. These random effects can be used as input in other statistical models, like survival analysis via Cox regression. In doing so, the Cox regression model can be extended in that way that not only the impact of a predictor variable at a single time-point, but also the trend of this predictor over time is accounted for in the analysis of survival. An application of this approach to data of the Vorarlberg Health Monitoring and Promotion Programme (VHM&PP) will be presented.

PROGNOSTIC SCORES TO SUPPORT DECISION MAKING FOR WOMEN WITH BREAST CANCER: RESULTS OF A RETROSPECTIVE DATA ANALYSIS

Marjan Arvandi, UMIT - University for Health Sciences, Medical Informatics and Technology, Hall i.T., Austria
Beate Jahn, UMIT - University for Health Sciences, Medical Informatics and Technology, Hall i.T., Austria
Heidelinde Fiegl, Department of Gynecology and Obstetrics, Innsbruck Medical University, Innsbruck, Austria
Georg Göbel, Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Innsbruck, Austria
Uwe Siebert, UMIT - University for Health Sciences, Medical Informatics and Technology, Hall i.T., Austria

The role of CHAC1-mRNA expression(CHAC1) in breast cancer patients was investigated recently by Goebel et al. The study showed that high CHAC1 besides other risk factors such as socioeconomic factors, type of cancer, etc. could be an independent indicator for elevated risk of breast cancer recurrence.

The goal of our study was to develop a multivariate prognostic score for combining multiple significant prognostic factors that is easy to apply in clinical practice.

Original patient-level data were taken from Goebel et al study containing information on 106 patients with primary breast cancer. Potential predictors included age, tumour size and grade, histological type, therapies (endocrine, radiation and chemo), menopausal, lymphnode status, hormone receptor(HR), HER-2/neu status and CHAC1. Risk factors were analysed using a time-independent Cox proportional hazard model for the two main outcomes relapse-free survival (RFS) and overall survival (OS). A multivariate prognostic score was calculated based on regression coefficients of statistically significant prognostic factors. We performed sensitivity analyses for different cutoff points. Survival distributions were visualized via Kaplan-Meier curves and tested by log-rank tests. Significance level was set at $p<0.05$.

For OS, age, tumour size, HR and CHAC1 were significant predictors. Low risk group (53%) includes patients with one or two risk factors high risk group (47%) includes patients with >2 risk factors. For RFS, age, CHAC1 and radiation therapy were significant predictors. Low risk group (38%) includes patients with one risk factor high risk group (62%) includes patients with >1 risk factors.

The prognostic scores derived by our analysis can be easily used by clinicians and decision makers besides current decision devices. Next steps involve validation of these scores with independent data sets to assess validity and generalizability.

The Competence Center Oncotyrol is funded within the scope of the COMET – Competence Centers for Excellent Technologies through BMVIT, BMWFJ, through the province of Salzburg and the Tiroler Zukunftsstiftung/Standortagentur Tirol.

The COMET Program is conducted by the Austrian Research Promotion Agency (FFG).

THE APPLICABILITY OF SITE-SPECIFIC PROPORTION CURED MODELS IN THE SMALL CANCER REGISTRY OF TYROL

Michael Edlinger, Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Innsbruck, Austria
Hanno Ulmer, Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Innsbruck, Austria
Milada Smastuen, Cancer Registry of Norway, Oslo, Norway
Willi Oberaigner, TILAK/UMIT, Innsbruck, Austria

On data of the relatively small cancer registry of Tyrol, proportion cured models were applied to evaluate their applicability with limited numbers. The main focus was on the assessment of the up-to-date survival level of 25 major cancer types. For this purpose, mixture cure model estimates were calculated with the period approach to attain proportions cured and, of the so called uncured, median survival times. For some cancer sites the model could not be applied, because it did not converge and therefore no estimates were obtained. Among the sites for which estimates were available, the proportion cured levels ranged from 4.6% (95% CI 0.2% to 9.0%) for male pancreas cancer to 74.0% (95% CI 64.4% to 83.6%) for cervix cancer. For the fatal cases, the lowest median survival amounted to 0.3 years (95% CI 0.1 to 0.6) with male acute myeloblastic leukaemia (AML) and the highest to 2.7 years (95% CI 1.2 to 6.0) with male larynx cancer. Altogether, estimates were achieved for 14 sites among females and 15 among males. Of these, the results seem reliable among women for stomach, colon, rectum, pancreas, lung, cervix, ovary, central nervous system/brain and AML cancer and among men for head/neck, oesophagus, stomach, colon, rectum, liver, pancreas, lung, kidney, central nervous system/brain and AML cancer. In conclusion, it is shown that even data from a cancer registry covering a rather small region can be utilised to derive up-to-date survival estimates of various cancer types. With some restrictions these models seem to be usable in circumstances where there are limited numbers of cancer cases, such as in a population-based cancer survival study or when a relatively small cancer group of a national registry is investigated. Because the estimates are up to date, they could also be advantageously applied for monitoring temporal site-specific survival trends.

BAYESIAN ADAPTIVE RANDOMIZATION IN EARLY DEVELOPMENT ONCOLOGY

Pantelis Vlachos, Cytel Inc, Switzerland

Evaluate through simulations the operating characteristics of Bayesian Adaptive Randomization (BAR) in proof-of-concept dose selection studies in oncology.

A proof-of-concept study (placebo, low and high dose of experimental treatment) was used to evaluate the BAR characteristics. The statistical design links treatment assignment probabilities to performance of respective arms. Smooth adaptation of assignment probabilities is guaranteed by a tuning parameter controlling how the randomization is influenced by data. A standard block randomization with equal assignment probabilities is used in the first batch of subjects. A Bayesian model summarizing prior information has been implemented with priors assuming no activity in order not to unbalance the randomization. Higher prior variance was given to the experimental arms reflecting uncertainty on the drug activity. The posterior probability that placebo has better performance than experimental treatment arms will be used to reject a null hypothesis of no drug activity. A screening design was used to calculate the maximum sample size and to compare operating characteristics of the two methods.

Simulations evaluated different scenarios of activity, priors, and rejection regions, showing consistent allocation of subjects with the performance observed in the study. Final Bayesian analyses were more powerful while controlling the same alpha level.

"RANDOMIZER FOR CLINICAL TRIALS" – A WEB-BASED RANDOMIZATION SERVICE IS ON-LINE FOR 10 YEARS

Petra Ofner-Kopeinig, Medical University of Graz, Graz, Austria
Maximilian Errath, Medical University of Graz, Graz, Austria
Andrea Berghold, Medical University of Graz, Graz, Austria

The Institute for Medical Informatics, Statistics and Documentation of the Medical University of Graz provides an online randomization service for multi-center clinical trials since 2003 – "Randomizer for Clinical Trials" (http://www.randomizer.at/). Six randomization methods are implemented including, completely at random, biased coin, big stick, minimization, urn model and permuted blocks randomization. The service provides an immediate access to the randomization database and to the trial's audit trail, provides email notifications, reduces transcription errors, communication delays and expenses. It also has a powerful simulation tool and was GCP-certified in 2005.

Users of this application do not need to have any special computing skills. Investigators randomize patients with only a few clicks by simply completing an on-screen form with patient details and are then immediately notified of the treatment allocation. But since it is a self-serve randomization service knowledge about setting up clinical trial designs and randomization procedures is necessary.

Meanwhile the service has been used for more than 200 studies by about 1600 registered users. Permuted block randomization is the most often used randomization procedure, followed by the minimization method and the urn model. The allocation method employed seems to depend on the field and / or region people are working. Surprisingly few support requests did arise. However, there were challenges in double-blind clinical trials due to complex trial logistics.

CAMPYLOBACTER IN HUMAN AND CHICKEN

Wei Wei, Institute for Social und Preventive Medizine, University Zurich, Zurich, Switzerland
Leonhard Held, Institute for Social and Preventive Medicine, University Zurich, Zurich, Switzerland
Gertraud Schuepbach, Veterinary Public Health-Institut Vetsuisse-Fakultät, Universität Bern, Bern, Switzerland

Campylobacteriosis is the most common food-associated infectious disease in Switzerland since 1995 [1]. Contact to or ingestion of raw/undercooked poultry are considered as dominant risk factors for infection [2]. Currently, the impact of this main transmission source is rarely analyzed. In this study, we aim to investigate the relationship between the disease incidence among humans and prevalence of chicken carrying bacteria in Switzerland from 2008 to 2009. An analysis of the temporal trends of Campylobacter infection in humans and chicken could be helpful for understanding disease dynamics. Two possible applications are: an evaluation of the effectiveness of future intervention measures in chicken on reducing human disease incidence, and an early detection of future outbreaks among humans.

A model-based approach [3] is used to describe the pattern of the disease by incorporating a time trend, seasonal effect and auto-correlation. A retrospective time series analysis of counts on disease cases is of main interest. Prevalence of the bacteria-infected chicken is included either as an additive or a multiplicative effect in the model. As a primary result, chicken prevalence with a lag of five weeks impacts the disease in human most. This result will be further validated on additional data.

[1] Lutz, A., Buttner, S. and Schupbach, G. (2010).Vergleich der jahrlichen Campylobacter Pravalenz von Mastpoulets und Menschen. Tech. rep., Bundesamt fur Veterinarwesen, Veterinary Public Health Institute.
[2] Held, L., Graziano, G., Frank, C. and Rue, H. (2006). Joint spatial analysis of gastrointestinal infectious diseases. Statistical Methods in Medical Research 15 465{480.
[3] Paul, M., Held, L. and Toschke, A. (2008). Multivariate modelling of infectious disease surveillance data. Statistics in Medicine 27 6250-6267

SUPPORTERS

We gratefully acknowledge the following organizations for their generous support of the ROeS conference 2013 in Dornbirn.

Platinum Supporters

numerus

Discover Statistical Power

"We consider that statistics is a crucial strategic component in any clinical development program, which can deliver substantial savings and increase chances of regulatory & reimbursement success. As such, we strive to be a long-term strategic partner for all our clients. Numerus provides statistical expertise on a pro-active basis, with fast turn-around times, and to the highest of quality standards."

www.numerus-ldt.com

Cytel

"The clinical development of drugs and devices is crucial for human welfare. Our mission is to improve success rates in this endeavor. We do this by improving the design and implementation of clinical trials through the innovative application of statistical science, operations research, and information technology. We are leaders in the design and implementation of adaptive clinical trials."

www.cytel.com

Silver Supporter

ORION PHARMA *www.orionpharma.de*

Bronze Supporter

MED EL *www.medel.com/at*

MSD Be well *www.msd.at*

Other Supporters

The **City of Dornbirn** and the **Federal State of Vorarlberg** support ROeS 2013 social events.

Organized by …

MEDICAL STATISTICS INFORMATICS HEALTH ECONOMICS

Department of Medical Statistics, Informatics and Health Economics

Innsbruck Medical University

MEDIZINISCHE UNIVERSITÄT INNSBRUCK

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES