www.iscb2014.info

International Society for Clinical Biostatistics 35th Annual Conference

FINAL PROGRAMME

24 - 28 August 2014
Vienna, Austria

ISCB35 Vienna 2014

Welcome to the ISCB 2014 Conference

It is the first time that an ISCB conference takes place in Vienna, Austria. The venue is Vienna's main University building with arcades full of busts of famous scholars, who had taught here in past centuries. They remind us of Vienna's long tradition, not only in Music and Arts, but also in Science, and Medicine. The current building of 1865 has a nice green yard, ideal for scientific exchanges in relaxed atmosphere. Shady parks (Rathauspark) and rose gardens (Volksgarten) are close. Furthermore, many of Vienna's numerous sights are within walking distance from the venue. Explore, why Vienna has been such an inspiring place for centuries, and still is ranked top in the annual Mercer Quality of Living Surveys of 223 cities on this globe!

The traditional excursions on Tuesday, as well as pre- and post-conference trips permit to get a deeper knowledge and understanding of Vienna, of Austria, and also of our close neighbours; you find various suggestions in this booklet and on our website www.iscb2014.info. The optional conference dinner takes place in one of Vienna's finest historic 19th century palaces, the Palais Ferstel, with live Viennese ballroom-music and dancing.

Like previous annual conferences of ISCB the 2014 conference provides a scientific forum for international exchange of theory, methods and applications of biostatistics in medical research and practice among clinicians, statisticians and members of other disciplines, such as epidemiologists, clinical chemists and clinical pharmacologists, working or interested in the field of clinical biostatistics. The following pages give details on the scientific topics of the conference, the invited sessions, the conference courses, and on the mini-symposia on Thursday morning.

We are convinced that ISCB35 in 2014 will be as attractive scientifically as the unique location certainly is.

We cordially welcome you to Vienna and to ISCB35,

Michael Schemper
(Chair, Local Organizing Committee)

Georg Heinze
(Chair, Scientific Programme Committee)

Table of contents

Welcome Letter 2
Commitees 4
Programme Overview 6
Detailed Programme (Monday) 10
(Tuesday) 20
(Wednesday) 26
(Thursday) 36
Posters 38
Information for Presenters 52
ISCB Awards 53
Conference Sponsors 54
General Information 55
Social Programme 56
Conference Tours 58
Overview of the Conference Venue 60
Imprint 63

ISCB35 Vienna 2014

International Society for Clinical Biostatistics (ISCB)

The International Society for Clinical Biostatistics (ISCB) was founded in 1978 to stimulate research into the principles and methodology used in the design and analysis of clinical research and to increase the relevance of statistical theory to the real world of clinical medicine.

Executive Committee 2014

President Koos Zwinderman, Netherlands
Vice-President David W. Warne, Switzerland
Treasurer KyungMann Kim, USA
Secretary Vana Sypsa, Greece

Michal Abrahamowicz, Canada
Keith Abrams, UK
Tomasz Burzykowski, Belgium
Zsolt Lang, Hungary
Saskia Le Cessie, Netherlands
Chris Metcalfe, UK
Jeremy Taylor, USA
Zdenek Valenta, Czech Republic

Scientific Programme Committee

Georg Heinze (Medical University of Vienna, Austria; Chair)
Michal Abrahamowicz (McGill University, Montreal, Canada)
Axel Benner (German Cancer Research Center, Heidelberg, Germany)
Hans Ulrich Burger (Hoffmann La Roche, Basel, Switzerland)
Ronald Geskus (Academic Medical Center, Amsterdam, Netherlands)
Els Goetghebeur (University of Gent, Belgium)
Miguel Hernán (Harvard School of Public Health, USA)
Karen Leffondré (University of Bordeaux Segalen, France)
Stefan Michiels (Institut Gustave Roussy, Villejuif, France)
Alessandra Nardi (University of Rome Tor Vergata, Italy)
Martin Posch (Medical University of Vienna, Austria)
Jenö Reiczigel (Szent István University Budapest, Hungary)
Michael Schemper (Medical University of Vienna, Austria; Chair of LOC)
Ian White (MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK)

Commitees

Additional reviewers

Zsolt Lang (Szent István University, Budapest, Hungary)
Cecile Proust-Lima (University of Bordeaux Segalen, France)
Thomas Waldhör (Medical University of Vienna, Austria)
Manuela Zucknick (German Cancer Research Center, Heidelberg, Germany)

Local Organizing Committee

Michael Schemper (Chair)
Martin Posch (Treasurer, Deputy Chair)
Andreas Gleiss (Public Relations, 2nd Deputy Chair)
Daniela Dunkler (Sponsoring)
Andreas Futschik
Alexandra Graf (Public Relations)
Maria Kohl
Franz König (Sponsoring)
Julia Singer
Ursula Taborsky / Viktoria Frysak (Secretary)

Congress Secretariat

AIM Group International
Mariannengasse 32
1090 Vienna, Austria

Sunday 24 August

	Hörsaal 26	Hörsaal 34	Hörsaal 16	Hörsaal 24
09:00 10:30	**Course 1:** Handling missing outcome data in clinical trials	**Course 2:** Data analysis with competing risks and multiple states	**Course 3:** Extension of frailty models for recurrent or clustered survival data with prediction	**Course 4:** Statistical methods in Systems Medicine
10:30 11:00	Coffee break			
11:00 12:30	continued	continued	continued	continued
12:30 14:00	Lunch			
14:00 15:30	continued	continued	**Course 6**: Interaction analysis	**Course 5**: Data and Safety Monitoring Board workshop
15:30 16:00	Coffee break			
16:00 17:30	continued	continued	continued	continued

Monday 25 August

	Auditorium Max	Hörsaal 7	Hörsaal 21
09:00 10:30	**I1**: On trial: integrated care pathways, EBM and EHR	**C01**: Randomized clinical trials	**C02**: Missing data
10:30 11:00	Coffee break		
11:00 12:30	Welcome address **IP**: President's invited speaker: Tom Louis		
12:30 14:00	Lunch		
14:00 15:30	**I2**: Beyond R packages	**C07**: Clinical trials	**C08**: Prediction models: case studies
15:30 16:00		Coffee break & **P1** Poster session	
16:00 17:30	**I3**: Inverse probability weighting techniques	**C13**: Personalized and stratified medicine I	**C14**: Meta-analysis
19:00		Welcome Reception (City Hall)	

Programme Overview

	Hörsaal 41	Hörsaal 34	Hörsaal 16	Hörsaal 31
	C03: Regression modelling in epidemiology	**C04**: Dose finding studies	**C05**: High-dimensional data analyses I	**C06**: The biostatistician's toolbox I
	Coffee break			
	Lunch			
	C09: Latent variable methods	**C10**: Genome-wide association studies	**C11**: Functional data analysis	**C12**: Vaccine studies and infectious diseases
	Coffee break & **P1** Poster session			
	C15: Longitudinal data analysis I	**C16**: High-dimensional data analysis II	**C17**: Adaptive designs I	**C18**: Binary and count data analysis
	Welcome Reception (City Hall)			

Tuesday 26 August

	Auditorium Max	Hörsaal 7	Hörsaal 21
09:00 10:30	**I4**: New methods to control for unmeasured confounding	**C19**: Development of prediction models	**C20**: Individual patient data meta-analysis
10:30 11:00	Coffee break & **P2** Poster session		
11:00 12:30	**I5**: Prediction to support clinical decision making	**C25**: Personalized and stratified medicine II	**C26**: Network meta-analysis
12:30 14:00	Lunch		
14:00	Conference Tours		
19:00	Young Researchers Lounge ("Heuriger" in Nußdorf)		

Wednesday 27 August

	Auditorium Max	Hörsaal 7	Hörsaal 21
09:00 11:00	**S1**: STRATOS initiative: first results & future steps	**C31**: Variable selection in high-dimensional models	**C32**: Longitudinal data analysis II
11:00 11:30	Coffee break & **P3** Poster session		
11:30 12:30	Annual General Meeting		
12:30 14:00	Lunch		
14:00 15:30	**S2**: The power of data sharing (Panel discussion)	**C37**: Causal inference from observational studies I	**C38**: Patient-centered outcomes
15:30 16:00	Coffee break & **P4** Poster session		
16:00 17:30	**I6**: Statistical methods for poly-omics studies	**C43**: Causal inference from observational data II	**C44**: Validation of prediction models
19:00	Conference Dinner (Palais Ferstel)		

Thursday 28 August

	Hörsaal 7	Hörsaal 21
09:00 10:30	**Mini-Symposium 1**: Statistical challenges in the epidemiology of aging	**Mini-Symposium 2**: Genomics-based personalized medicine (joint with the IGES)
10:30 11:00	Coffee break	
11:00 12:30	MS1 continued	MS2 continued
	IGES Conference (Imperial Riding School Renaissance Hotel)	

	Hörsaal 41	Hörsaal 34	Hörsaal 16	Hörsaal 31
	C21: Survival analysis and competing risks	**C22**: Surrogate and composite endpoints	**C23**: Design and analysis of clustered studies	**C24**: Group-sequential designs
	Coffee break & **P2** Poster session			
	C27: Survival analysis I	**C28**: Marginal structural models	**C29**: Clinical trial designs	**C30**: Adaptive designs II
	Lunch			
	Conference Tours			
	Young Researchers Lounge ("Heuriger" in Nußdorf)			

	Hörsaal 41	Hörsaal 34	Hörsaal 16	Hörsaal 31
	C33: Relative and net survival	**C34**: Methodology	**C35**: The biostatistician's toolbox II	**C36**: Issues in multiple testing
	Coffee break & **P3** Poster session			
	Lunch		12:35 - 13:15 STRATOS public meeting (Elise-Richter-Saal)	
	C39: Multistate models and competing risks I	**C40**: Model performance evaluation	**C41**: Survival analysis II	**C42**: Poly-omics studies & Systems Biology
	Coffee break & **P4** Poster session			
	C45: Multistate models and competing risks II	**C46**: Multiple imputation	**C47**: Special types of censored data	**C48**: Drug development
	Conference Dinner (Palais Ferstel)			

	Hörsaal 24
	Course 7: Designing adaptive clinical trials
	Coffee break
	continued
	IGES Conference (Imperial Riding School Renaissance Hotel)

Detailed Programme

Monday, 25th August 2014

Session **I1 On trial: integrated care pathways, evidence based medicine and EHRs**

Organizer Els Goetghebeur

Room Auditorium Max 9:00 - 10:30

I1.1		Michael Campbell	Integrated care pathways, evidenced based medicine and electronic health records: an overview
I1.2		Erica Moodie	From idealized to realized: learning about dynamic treatment regimens using electronic medical records?
I1.3		Richard Emsley	A causal inference approach to optimising care pathways in type 2 diabetes using electronic health records

Session **C01 Randomized clinical trials**

Chair Franz König

Room Hörsaal 7 9:00 - 10:30

C01.1	9:00 - 9:18	Gerd Rosenkranz	Analysis of clinical trials requiring rescue medication
C01.2	9:18 - 9:36	Miriam Tamm	Assessment of chronological bias in randomized clinical trials
C01.3	9:36 - 9:54	Stephen Walter	An analytic framework for randomised trial designs that take patient preferences into account
C01.4	9:54 - 10:12	Douglas Thompson	Covariate adjustment has similar benefits in small and large randomised controlled trials
C01.5	10:12 - 10:30	Mohamed-Amine Bayar	Long-term evaluation of different designs of a series of phase III clinical trial in rare cancers: a simulation study

Session **C02 Missing data**

Chair Ian White

Room Hörsaal 21 9:00 - 10:30

C02.1	9:00 - 9:18	Karla Diaz Ordaz	A comparison of multiple imputation methods for hierarchical data when there is whole cluster non-response
C02.2	9:18 - 9:36	Guido Thoemmes	Pattern mixture models applied to clinical trials for chronic pain
C02.3	9:36 - 9:54	Anneke Grobler	CD4+ counts in a 3-arm longitudinal clinical trial with substantial missing data: a sensitivity analysis
C02.4	9:54 - 10:12	Jacques-Emmanuel Galimard	Proposition of a Multiple Imputation approach for MNAR mechanism using Heckman´s model
C02.5	10:12 - 10:30	Finbarr Leacy	Allowing for nonignorable missingness in HIV status using multiple imputation with delta-adjustment: applications to causal mediation analysis and prevalence estimation

Session **C03 Regression modelling in epidemiology**

Chair Thomas Waldhör

Room Hörsaal 41 9:00 - 10:30

C03.1	9:00 - 9:18	Werner Brannath	Interpretation of linear regression coefficients under model miss-specifications
C03.2	9:18 - 9:36	Carolin Jenkner	Simulation study to assess and compare strategies for modelling two continuous covariates with a spike at zero
C03.3	9:36 - 9:54	Martin Scharpenberg	A new measure of association based on non-linear regression
C03.4	9:54 - 10:12	Svenja Burger	Weighted mean impact analysis
C03.5	10:12 - 10:30	Yonas Ghebremichael-Weldeselassie	Non-parametric Self controlled Case Series Method

Session **C04 Dose finding studies**

Chair Thomas Jaki

Room Hörsaal 34 9:00 - 10:30

C04.1	9:00 - 9:18	Xavier Paoletti	Dose finding methods based on longitudinal ordinal data: Realistic prior hypotheses identified from 49 phase I studies
C04.2	9:18 - 9:36	Haige Shen	A Bayesian approach to oncology combination dose-finding
C04.3	9:36 - 9:54	Markus Lange	Bayesian optimal clinical trial design for monoclonal antibodies

C04.4	9:54 - 10:12	Amy Cotterill	Dose-escalation strategies which utilise subgroup information
C04.5	10:12 - 10:30	Gareth James	Continual reassessment method for dose escalation clinical trials in oncology: A comparison of prior approaches using AZD3514 data

Session	**C05 High-dimensional data analyses I**		
Chair	Andreas Gleiss		
Room	Hörsaal 16	9:00 - 10:30	
C05.1	9:00 - 9:18	Ian James	Assessment of positivity in ELISPOT assays based on FDR-type and mixture procedures
C05.2	9:18 - 9:36	Peter Ström	Estimating individual peptide effects from pooled ELISPOT data
C05.3	9:36 - 9:54	Ying Xu	Estimation of Antibody Concentration from Multiple Dilutions Data
C05.4	9:54 - 10:12	Katherine Lee	Mixed models for the analysis of brain magnetic resonance imaging data
C05.5	10:12 - 10:30	Julia Forman	Comparison of Analysis Approaches for Multi-Level Vascular Imaging Data

Session	**C06 The biostatistician's toolbox I**		
Chair	Jeno Reiczigel		
Room	Hörsaal 31	9:00 - 10:30	
C06.1	9:00 - 9:18	John Newell	Translational Statistics and Dynamic Nomograms
C06.2	9:18 - 9:36	Lara Lusa	Dynamic graph generation and data analysis of complex data: a web-application based on R and shiny
C06.3	9:36 - 9:54	Reinhard Vonthein	The grammar of parametric boxplots
C06.4	9:54 - 10:12	Roger Marshall	Structure and interpretation of classification and regression trees
C06.5	10:12 - 10:30	Ralitza Gueorguieva	Classification and regression trees for moderator effects in clinical trials

Session	**IP President's invited speaker**		
Organizer	Koos Zwinderman		
Room	Auditorium Max	11:30 - 12:30	
IP.1		Thomas Louis	Bayes, why bother?

Session	**I2 Beyond R packages: getting our methods into standard software**		
Organizer	Georg Heinze		
Room	Auditorium Max	14:00 - 15:30	
I2.1		Ian White	Writing and developing statistical software: the statistical methodologist´s view
I2.2		Yannis Jemiai	Beyond wild horses: developing innovation at Cytel
I2.3		Robert Rodriguez	An inside perspective on the development of SAS statistical software

Session	**C07 Clinical trials**		
Chair	Lisa Hampson		
Room	Hörsaal 7	14:00 - 15:30	
C07.1	14:00 - 14:18	Lehana Thabane	Dealing with Challenges in Design and Analysis of Clinical Trials in Long-term Care
C07.2	14:18 - 14:36	Ann-Kristin Leuchs	Estimating efficacy and effectiveness using data retrieved after treatment non-compliance
C07.3	14:36 - 14:54	Joost van Rosmalen	Including historical data in the analysis of clinical trials using the modified power prior: practical and theoretical issues
C07.4	14:54 - 15:12	Karin Meiser	A robust Bayesian meta-analytic-predictive approach to borrow strength from historical information in thorough QT studies
C07.5	15:12 - 15:30	Chris Roberts	Power and Sample Size of Trials with a Partially Nesting Design for Binary Outcomes

Session	**C08 Prediction models: case studies**		
Chair	Robin Henderson		
Room	Hörsaal 21	14:00 - 15:30	
C08.1	14:00 - 14:18	Richard Holubkov	Multidimensional Assessment of the Predictive Ability of a Trichotomous-Outcome Model in the NICHD Collaborative Pediatric Critical Care Research Network [CPCCRN]
C08.2	14:18 - 14:36	Khanh Lam Phung	Developing dynamic prediction models for acute diseases

C08.3	14:36 - 14:54	Abdel Douiri	A patient-specific predicting tool for functional recovery after stroke
C08.4	14:54 - 15:12	Mousumi Banerjee	A Tree Based Model for Thyroid Cancer Prognostication
C08.5	15:12 - 15:30	Veronika Weyer	Stratified weighted regression for subgroup signatures from prognostic models with molecular data

Session **C09 Latent variable methods**

Chair Jeremy MG Taylor

Room Hörsaal 41 14:00 - 15:30

C09.1	14:00 - 14:18	Annette Kifley	Treatment effect estimation in latent variable models with structural misspecification
C09.2	14:18 - 14:36	Esra Kurum	Joint modeling of longitudinal binary and continuous responses
C09.3	14:36 - 14:54	Mathieu Bastard	Joint modelling of longitudinal and time-to-event data: a comparison between shared random-effect and latent class model
C09.4	14:54 - 15:12	Luohua Jiang	Longitudinal patterns of stages of change and lifestyle intervention outcomes - a latent class analysis with distal outcomes
C09.5	15:12 - 15:30	Anaïs Rouanet	Joint latent class model for longitudinal data and competing interval-censored events : Application to the study of Alzheimer's disease

Session **C10 Genome-wide association studies**

Chair Yudi Pawitan

Room Hörsaal 34 14:00 - 15:30

C10.1	14:00 - 14:18	Paola Ferrario	Entropy-based statistics to detect gene-gene interactions
C10.2	14:18 - 14:36	Anand Vidyashankar	Network Approach to Identify Gene-by-Secondary Phenotype Interactions in GWAS
C10.3	14:36 - 14:54	Taesung Park	Gene-gene interaction analysis of correlated phenotypes
C10.4	14:54 - 15:12	Donghwan Lee	Estimating the rediscovery rate for assessing the validation of genome-wide association studies

C10.5	15:12 - 15:30	Niki Dimou	A multivariate method for meta-analysis of multiple outcomes in genetic association studies

Session **C11 Functional data analysis**

Chair Tim Ramsay

Room Hörsaal 16 14:00 - 15:30

C11.1	14:18 - 14:36	Kathrine Frøslie	Functional data analysis of temporal glucose curves compared with gold standard measurements of insulin sensitivity and beta-cell function
C11.2	14:18 - 14:36	Nicholas Tarabelloni	Unsupervised classification of functional data based on covariance structures
C11.3	14:36 - 14:54	Mohamed Amine Benadjaoud	Functional data analysis in radiobiology and radiation epidemiology
C11.4	14:54 - 15:12	Nicolas Sauvageot	Use of finite mixture models for dietary patterns analysis
C11.5	15:12 - 15:30	Linda Abrahamsson	A statistical model of breast cancer tumour growth with estimation of screening sensitivity as a function of mammographic density

Session **C12 Vaccine studies and infectious diseases**

Chair Fabrice Bailleux

Room Hörsaal 31 14:00 - 15:30

C12.1	14:00 - 14:18	W. Dewé	Taking into account strains heterogeneity in the estimation of vaccine efficacy against seasonal influenza
C12.2	14:18 - 14:36	Ruth Keogh	Estimating the effects of time-since-exposure using case-control data, motivated by a study of vaccine efficacy over time
C12.3	14:36 - 14:54	Rodolphe Thiébaut	Integrative analysis of high-dimensional data in clinical trials: an example in HIV vaccine development
C12.4	14:54 - 15:12	Lauren Rodgers	Assessing vaccine effectiveness using observational data in the presence of hidden confounders
C12.5	15:12 - 15:30	Angela Noufaily	Modeling reporting delays for outbreak detection of infectious diseases

Session	**I3 Inverse probability weighting techniques**		
Organizers	Ronald Geskus and Karen Leffondré		
Room	Auditorium Max	16:00 - 17:30	
I3.1		Ronald Geskus	Iterative inverse probability weighting
I3.2		Thomas Gerds	Bias-variance trade off in IPCW: Is it possible to hear the curse of dimensionality in a random forest?
I3.3		Tianxi Cai	Inverse probability weighting methods for biomarker evaluation with case cohort studies
Session	**C13 Personalized and stratified medicine I**		
Chair	KyungMann Kim		
Room	Hörsaal 7	16:00 - 17:30	
C13.1	16:00 - 16:18	Willi Sauerbrei	Interaction of treatment with a continuous variable: simulation study of significance level and power for several methods of analysis
C13.2	16:18 - 16:36	Hong Sun	Comparing a marker based stratified treatment strategy with the standard treatment in a randomized clinical trial
C13.3	16:36 - 16:54	Maren Kechel	A framework for comparing methods for marker-based selection of treatment change
C13.4	16:54 - 17:12	Harriet Sommer	Analyzing treatment-by-subgroup interactions in time-to-event data – comparison of two multivariate approaches
C13.5	17:12-17:30	Yudi Pawitan	Bounds for causal interaction
Session	**C14 Meta-analysis**		
Chair	Zsolt Lang		
Room	Hörsaal 21	16:00 - 17:30	
C14.1	16:00 - 16:18	Dimitris Mavridis	Addressing continuous missing outcomes in pair-wise and network meta-analysis
C14.2	16:18 - 16:36	Dean Langan	The impact of choice of heterogeneity estimator in meta-analysis
C14.3	16:36 - 16:54	Evan Kontopantelis	A re-analysis of the Cochrane Library data: the dangers of unobserved heterogeneity in meta-analyses
C14.4	16:54 - 17:12	Andrea Benedetti	Estimating heterogeneity when pooling proportions in a meta analysis
C14.5	17:12 - 17:30	Martin Schumacher	Meta-analysis and the surgeon general's report on smoking and health

Session	**C15 Longitudinal data analysis I**		
Chair	Erica Moodie		
Room	Hörsaal 41	16:00 - 17:30	
C15.1	16:36 - 16:54	Inês Sousa	Longitudinal models with outcome dependent follow-up times
C15.2	16:36 - 16:54	Michael Crowther	Joint modelling of longitudinal and survival data incorporating delayed entry: application to longitudinal mammographic breast density and breast cancer survival
C15.3	16:36 - 16:54	Diklah Geva	Dynamic time process model for the association among two longitudinal markers in the presence of survival: application to healthABC cohort
C15.4	16:54 - 17:12	Eleni-Rosalina Andrinopoulou	Combined dynamic predictions using joint models of multiple longitudinal outcomes and competing risk data
C15.5	17:12 - 17:30	Cécile Proust-Lima	Development and validation of individualized dynamic predictions based on repeated biomarker data according to scenarios of new treatments
Session	**C16 High-dimensional data analysis II**		
Chair	Stefan Michiels		
Room	Hörsaal 34	16:00 - 17:30	
C16.1	16:00 - 16:18	Selen Yilmaz Isikhan	A comparison of prediction models for gene expression data by resampling techniques
C16.2	16:18 - 16:36	Putri Novianti	An application of sequential meta-analysis to gene expression studies
C16.3	16:36 - 16:54	Rok Blagus	Ensemble classifiers in the high-dimensional setting with class-imbalanced data
C16.4	16:54 - 17:12	Isabell Hoffmann	Combining techniques for screening and evaluating interaction terms on high-dimensional time-to-event data
C16.5	17:12 - 17:30	Simone Wahl	Comparing models of location and scale for genome-wide DNA methylation data

Session **C17 Adaptive designs I**

Chair Tim Friede

Room	Hörsaal 16	16:00 - 17:30	
C17.1	16:00 - 16:18	Giuseppe Palermo	Dose-escalation using safety and biomarker data: A Bayesian adaptive approach
C17.2	16:18 - 16:36	Wai Yin Yeung	Bayesian adaptive dose-escalation procedures utilizing a gain function with binary and continuous responses
C17.3	16:36 - 16:54	Simon Bond	Adaptive dose-finding designs to identify multiple doses that achieve multiple response targets
C17.4	16:54 - 17:12	James Wason	Bayesian adaptive designs for biomarker trials with biomarker discovery
C17.5	17:12 - 17:30	Lisa Law	Design of telehealth trials – introducing adaptive approaches

Session **C18 Binary and count data analysis**

Chair Saskia le Cessie

Room	Hörsaal 31	16:00 - 17:30	
C18.1	16:00 - 16:18	Tong-Yu Lu	Multiple comparisons of treatments with highly skewed ordinal responses
C18.2	16:18 - 16:36	Ralf Bender	Calculating confidence intervals for risk differences by means of MOVER-R
C18.3	16:36 - 16:54	Randi Grøn	Misspecified Poisson regression models for large-scale registry data: Problems with "large n and small p"
C18.4	16:54 - 17:12	Linda Sharples	Bayesian analysis of zero-inflated beta regression models with application to quality of life and functional outcomes
C18.5	17:12 - 17:30	Yalcin Yavuz	Count data analysis in nutrition clinical trials

Tuesday, 26th August 2014

Session **I4 New methods to control for unmeasured confounding**

Organizer Michal Abrahamowicz

Room Auditorium Max 9:00 - 10:30

I4.1	Rolf Groenwold	Bias sensitivity analysis of unmeasured confounding
I4.2	Lawrence McCandless	A Bayesian perspective on unmeasured confounding in large administrative databases
I4.3	M. Abrahamowicz, R. Burne	New statistical methods for using validation subsamples to adjust for unmeasured confounders in survival analysis

Session **C19 Development of prediction models**

Chair Thomas Gerds

Room Hörsaal 7 9:00 - 10:30

C19.1	9:00 - 9:18	Gary Collins	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement
C19.2	9:18 - 9:36	Paul Blanche	The multi-split testing approach for choosing between 2 prediction strategies
C19.3	9:36 - 9:54	Emmanuel Ogundimu	The impact of events per variable on the predictive performance of the Cox model
C19.4	9:54 - 10:12	Laure Wynants	The number of events per variable needed to build logistic prediction models in clustered data: a simulation study
C19.5	10:12 - 10:30	Menelaos Pavlou	Review and evaluation of penalised likelihood methods for risk prediction in data with few events

Session **C20 Individual participant data meta-analysis**

Chair Andrea Berghold

Room Hörsaal 21 9:00 - 10:30

C20.1	9:00 - 9:18	Thomas Debray	How to appraise Individual Participant Data (IPD) meta-analysis in diagnostic and prognostic risk prediction research
C20.2	9:18 - 9:36	Neta Zach	Being PRO ACTive- What can a clinical trials database reveal about ALS?
C20.3	9:36 - 9:54	Matteo Quartagno	Missing data in individual patient data meta-analysis
C20.4	9:54 - 10:12	Thomas Debray	Multiple imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE
C20.5	10:12 - 10:30	Lorenzo Tanadini	Analysis of repeated ordinal measurements and trial planning in a rare neurological disorder

Session **C21 Survival analysis and competing risks**

Chair Nadine Binder

Room Hörsaal 41 9:00 - 10:30

C21.1	9:00 - 9:18	Harald Binder	Stagewise pseudo-value regression for time-dependent effects on the cumulative incidence
C21.2	9:18 - 9:36	Matthieu Resche-Rigon	Imputing missing covariate values in presence of competing risks
C21.3	9:36 - 9:54	Laetitia Teixeira	Evaluation of a peritoneal dialysis program using semiparametric multi-state models in the presence of competing risks
C21.4	9:54 - 10:12	Giuliana Cortese	Predicting optimal cumulative doses for breast cancer chemotherapy via competing risks regression models
C21.5	10:12 - 10:30	Luise Cederkvist	The liability-threshold model for case-control family studies applied to censored time to event data

Session **C22 Surrogate and composite endpoints**

Chair Martina Mittlböck

Room Hörsaal 34 9:00 - 10:30

C22.1	9:00 - 9:18	Jasmin Link	A new audit strategy to detect possible bias in the evaluation of progression free survival
C22.2	9:18 - 9:36	Theis Lange	A causal inference/mediation analysis based approach for assessing pseudo end-points applied to to ovarian cancer trials
C22.3	9:36 - 9:54	Christoph Schürmann	Bias assessment of surrogate threshold effects in simplified correlation based validation approaches

C22.4	9:54 - 10:12	Steffen Witte	Extension of win-ratio: analyzing a composite endpoint considering the clinical importance order among components
C22.5	10:12 - 10:30	Marcel Wolbers	Weighted comparisons of composite endpoints

Session **C23 Design and analysis of clustered studies**
Chair Stephen Walter

Room Hörsaal 16 9:00 - 10:30

C23.1	9:00 - 9:18	Shaun Seaman	Methods for observed-cluster inference when cluster size is informative
C23.2	9:18 - 9:36	Lisa Yelland	Generalised estimating equation methods for analysing continuous outcomes when cluster size is informative
C23.3	9:36 - 9:54	Neil Wright	Choosing covariates and the effects of covariate adjustment in the analysis of CRTs
C23.4	9:54 - 10:12	Andrew Forbes	Sample size and analysis considerations for cluster randomised crossover trials with unbalanced cluster sizes and binary data
C23.5	10:12 - 10:30	Esther de Hoop	A multidisciplinary approach to benefits and drawbacks of the stepped wedge cluster randomized design

Session **C24 Group-sequential designs**
Chair Kit Roes

Room Hörsaal 31 9:00-10:30

C24.1	9:00 - 9:18	Hsiao Yin Liu	Group sequential monitoring of response-adaptive randomised clinical trials with censored survival data
C24.2	9:18 - 9:36	Michael Grayling	Group-sequential designs for crossover trials
C24.3	9:36 - 9:54	Ingeborg van der Tweel	Optimal sequential clinical trials in small populations
C24.4	9:54 - 10:12	L Hampson	Group sequential designs for verifying whether effective drug concentrations are similar in adults and children
C24.5	10:12 - 10:30	Toshimitsu Hamasaki	Group-sequential strategies when considering multiple outcomes as co-primary in clinical trials

Session **I5 Prediction to support clinical decision making**
Organizer Alessandra Nardi

Room Auditorium Max 11:00 - 12:30

I5.1		Nils Lid Hjort	What price Cox regression? Ranking predictions from semiparametric and parametric hazard regression models via focused information criteria
I5.2		Jeremy Taylor	Individualized predictions of event times using joint longitudinal-survival models
I5.3		Robin Henderson	Model selection and ensemble predictive performance

Session **C25 Personalized and stratified medicine II**
Chair Ulrich Mansmann

Room Hörsaal 7 11:00 - 12:30

C25.1	11:00 - 11:18	Stephen Senn	Mastering variation: Variance components and personalised medicine
C25.2	11:18 - 11:36	Julien Tanniou	Subgroup analyses: time to be specific about their goals
C25.3	11:36 - 11:54	Nicole Heßler	On the evaluation of predictive biomarkers with dichotomous endpoints: a comparison of the linear and the logistic probability models
C25.4	11:54 - 12:12	Lucinda Billingham	Design dilemmas in the multi-drug, genetic-marker-directed, non-comparative, multi-centre, multi-arm phase II National Lung Matrix Trial
C25.5	12:12 - 12:30	Shogo Nomura	A new framework using G-estimation for placebo-controlled randomized phase 3 trials with extensive crossovers for biomarker-driven molecularly targeted oncology agents

Session **C26 Network meta-analysis**
Chair Ralf Bender

Room Hörsaal 21 11:00 - 12:30

C26.1	11:00 - 11:18	Orestis Efthimiou	A multi-state Markov model for network meta-analysis of studies with missing data

C26.2	11:18 - 11:36	Jochem König	Investigating consistency of mixed treatment comparisons by approximating sub-networks
C26.3	11:36 - 11:54	Gerta Rücker	Visualisation of networks in meta-analysis
C26.4	11:54 - 12:12	Lorenz Uhlmann	Bayesian network meta-analysis for cluster-randomized trials
C26.5	12:12 - 12:30	Adriani Nikolakopoulou	Precision of the estimates from a network meta-analysis model and their role in planning future studies

Session **C27 Survival analysis I**

Chair Maria Grazia Valsecchi

Room Hörsaal 41 11:00 - 12:30

C27.1	11:00 - 11:18	Laura Antolini	Survival probability with non-reversible time varying treatment indicator: theoretical quantities and nonparametric estimators
C27.2	11:18 - 11:36	Ulrike Pötschger	Using pseudo-values for comparing long-term survival after stem-cell transplantation (SCT) with long-term survival after chemotherapy
C27.3	11:36 - 11:54	Irene Schmidtmann	Assessing effects of treatment change on survival when the measurement pattern of covariates and events are dependent
C27.4	11:54 - 12:12	Catherine Schramm	Clustering for treatment effect on recurrent events
C27.5	12:12 - 12:30	Antje Jahn-Eimermacher	A total time approach for the simulation of recurrent event data when planning a clinical trial

Session **C28 Marginal structural models**

Chair Rolf Groenwold

Room Hörsaal 34 11:00 - 12:30

C28.1	11:00 - 11:18	Ryan Kyle	Addressing measurement error in time-varying covariates through the use of SIMEX-adjusted marginal structural models
C28.2	11:18 - 11:36	Genevieve Lefebvre	A caution on the use of stabilized weights in marginal structural models
C28.3	11:36 - 11:54	Jessica Kasza	Dialysis, catheter use and mortality: challenges in applying marginal structural models to data from a clinical registry
C28.4	11:54 - 12:12	Aksel Jensen	Non-specific effects of vaccines on child morbidity examined with a marginal structural model for recurrent events.
C28.5	12:12 - 12:30	Vanessa Didelez	Robustness and efficiency in instrumental variable models with covariates

Session **C29 Clinical trial designs**

Chair Chris Metcalfe

Room Hörsaal 16 11:00 - 12:30

C29.1	11:00 - 11:18	Constanze Schulz	Patient-oriented randomization - a new clinical design
C29.2	11:18 - 11:36	Irene Rebollo-Mesa	A novel modified standard-gamble task to measure patients' preferences for biomarker-led care
C29.3	11:36 - 11:54	Wenle Zhao	Response adaptive randomization in large phase III confirmative clinical trials with binary outcomes – Benefits are unlikely
C29.4	11:54 - 12:12	Thomas Jaki	Incorporating feasibility assessment in the design of clinical studies
C29.5	12:12 - 12:30	Richard Hooper	Some novel alternatives to parallel group designs for pragmatic clinical trials

Session **C30 Adaptive designs II**

Chair Yannis Jemiai

Room Hörsaal 31 11:00 - 12:30

C30.1	11:00 - 11:18	Florian Klinglmueller	Estimation after blinded sample size reassessment
C30.2	11:18 - 11:36	Stavros Nikolakopoulos	A two-stage adaptive design for small clinical trials
C30.3	11:36 - 11:54	Dominic Magirr	Adaptive designs for time-to-event trials
C30.4	11:54-12:12	Geraldine Rauch	Adaptive designs for two candidate primary time-to-event endpoints
C30.5	12:12-12:30	Sheetal Solanki	Backward image confidence intervals for adaptive group sequential trials

Wednesday, 27th August 2014

Session **S1 STRATOS (Strengthening Thinking about Analyses of Observational Studies) initiative: first results & future steps**

Organizers Willi Sauerbrei and Harbajan Chadha-Boreham

Room	Auditorium Max	9:00 - 11:00	
S1.1		Marianne Huebner	Setting the stage with initial data analyses
S1.2		Ewout Steyerberg	Evaluation of incremental value of a marker: a historic perspective on the Net Reclassification Improvement
S1.3		Michal Abrahamowicz	Review of methods used in recent observational epidemiological studies to select variables and their functional forms [STRATOS Task Group 2]
S1.4		Els Goetghebeur	Causal questions and principled answers: a guide through the landscape for practicing statisticians

Session **C31 Variable selection in high-dimensional models**

Chair Rainer Spang

Room	Hörsaal 7	9:00 - 10:48	
C31.1	9:00 - 9:18	Nils Ternès	An extension of the lasso penalization to reduce false positive selection in high-dimensional Cox models
C31.2	9:18 - 9:36	Benjamin Hofner	Biomarker discovery: controlling false discoveries in high dimensional situations
C31.3	9:36 - 9:54	Philippe Bastien	Deviance residuals based sparse PLS and sparse kernel PLS regression for censored data
C31.4	9:54 - 10:12	Paul Newcombe	Weibull regression with Bayesian variable selection to identify prognostic biomarkers of breast cancer survival
C31.5	10:12 - 10:30	Leonhard Held	Approximate Bayesian model selection with the deviance statistic
C31.6	10:30 - 10:48	Magdalena Malina	A novel variable selection method for Monte Carlo Logic Regression

Session **C32 Longitudinal data analysis II**

Chair Cecile Proust-Lima

Room	Hörsaal 21	9:00 - 10:48	
C32.1	9:00 - 9:18	Delphine Courvoisier	Mixed-effects location scale model for time to event data
C32.2	9:18 - 9:36	Kazem Nasserinejad	Bayesian growth mixture models to distinguish hemoglobin value trajectories in blood donors
C32.3	9:36 - 9:54	Nina Breinegaard	Pairwise residuals and diagnostic tests for misspecified dependence structures in models for binary longitudinal data.
C32.4	9:54 - 10:12	Marie Vigan	Evaluation of LRT in joint modelling of repeated time-to-event and longitudinal data using nonlinear mixed effects models.
C32.5	10:12 - 10:30	Rachael Hughes	Estimation of the linear mixed integrated Ornstein-Uhlenbeck stochastic model

Session **C33 Relative and net survival**

Chair Catherine Quantin

Room	Hörsaal 41	9:00 - 10:48	
C33.1	9:00 - 9:18	Amel Mahboubi	Flexible modeling of continuous covariates in Net Survival: additive vs multiplicative model
C33.2	9:18 - 9:36	Célia Touraine	An excess hazard model adjusting for lack of additional life table variables
C33.3	9:36 - 9:54	Nathalie Grafféo	Generalization of a log-rank type test to compare net survival distributions
C33.4	9:54 - 10:12	Florence Gillaizeau	Additive relative survival multistate semi-Markov model
C33.5	10:12 - 10:30	Coraline Danieli	Diagnostic tools for model building in net survival: use and comparison of two methods to test the proportional hazards assumption
C33.6	10:30 - 10:48	Roch Giorgi	Oblique decision trees for spatial clusters detection of net cancer survival rates

Session	C34 Methodology		
Chair	Andreas Futschik		
Room	Hörsaal 34	9:00 - 10:48	
C34.1	9:00 - 9:18	Yingjie Zhang	Constrained ordination analysis with an increased number of bell-shaped response functions with applications in metagenomics
C34.2	9:18 - 9:36	Gaj Vidmar	Multivariate statistical process control for mixed-type data: an overview and a simulation study
C34.3	9:36 - 9:54	Leigh Blizzard	Case-wise diagnostics for the multinomial log-link regression model
C34.4	9:54 - 10:12	Janet Peacock	Interpreting small differences in mean z-scores in sick populations: does dichotomisation help?
C34.5	10:12 - 10:30	Corine Baayen	Confidence bounds for monotone dose-response relationships
C34.6	10:30 - 10:48	Daniel Nevo	Simpler is better: a comparison of methods for construction of fetal reference charts

Session	C35 The biostatistician's toolbox II		
Chair	Daniela Dunkler		
Room	Hörsaal 16	9:00 - 10:48	
C35.1	9:00 - 9:18	Emilie Peron	Pharmacodependence: new graphical representations.
C35.2	9:18 - 9:36	Ying Zhang	Constructing robust confidence intervals for drug utilization time series data
C35.3	9:36 - 9:54	Howard Thom	Using constraints to compare state structures in cost-effectiveness decision models
C35.4	9:54 - 10:12	Md. Abu Manju	Optimal and maximin sample sizes for multicentre cost-effectiveness trials
C35.5	10:12 - 10:30	David Robertson	Correcting for bias in the detection and validation of informative diagnostic tests
C35.6	10:30 - 10:48	Dankmar Böhning	Modelling and choice of cutoff in meta-analysis of diagnostic studies with varying cut-off value

Session	C36 Issues in multiple testing		
Chair	Hans-Ulrich Burger		
Room	Hörsaal 31	9:00 - 10:48	
C36.1	9:00 - 9:18	Sylvia Schmidt	An informative modification of the fallback procedure
C36.2	9:18 - 9:36	Franz König	Confirmatory testing for a beneficial treatment effect in dose-response studies using MCP-Mod
C36.3	9:36 - 9:54	Toshifumi Sugitani	Graph based multiple testing strategies for confirmatory adaptive enrichment designs
C36.4	9:54 - 10:12	Georg Gutjahr	Likelihood ratio tests for multiple nonlinear models
C36.5	10:12 - 10:30	Victoria Vickerstaff	Are multiple outcomes analysed appropriately in randomised controlled trials? A systematic review.
C36.6	10:30 - 10:48	Robin Ristl	A multiple testing procedure for three primary endpoints

Session	S2 The power of data sharing: advancing research for everyone's benefit? (Panel discussion)		
Organizers	Martin Posch and Franz König		
Room	Auditorium Max	14:00 - 15:30	
S2		Panel discussion	The power of data sharing: advancing research for everyone´s benefit?

Session	C37 Causal inference from observational studies I		
Chair	Lawrence McCandless		
Room	Hörsaal 7	14:00 - 15:30	
C37.1	14:00 - 14:18	Florent Le Borgne	Adjusted survival curves by using inverse probability of treatment weighting: the comparison of three adapted log-rank tests.
C37.2	14:18 - 14:36	Marie Reilly	Inverse probability weighting of overmatched nested case-control data to enable estimation of main effects and interactions
C37.3	14:36 - 14:54	Menglan Pang	Performance of Targeted Maximum Likelihood Estimation in Point-exposure Studies Using High-dimensional Covariate Data

C37.4	14:54 - 15:12	Achmad Efendi	The impact of pCR after neoadjuvant chemotherapy in patients with large operable breast cancer on survival outcomes: A causation analysis
C37.5	15:12 - 15:30	Susanne Strohmaier	Causal mediation analysis in a clinical survival trials - Can statistics help to understand treatment mechanisms?

Session **C38 Patient-centered outcomes**

Chair Mike Campbell

Room Hörsaal 21 14:00 - 15:30

C38.1	14:00 - 14:18	Werner Vach	The design of diagnostic studies - another case for STRATOS?
C38.2	14:18 - 14:36	Olivier Collignon	Methodological issues in developing scores and cut-offs of rheumatoid arthritis activity
C38.3	14:36 - 14:54	Thuva Vanniyasingam	Determining optimal fractional factorial designs of discrete choice experiments using d-efficiency: Application in addiction services
C38.4	14:54 - 15:12	Phillip Gichuru	Developing robust scoring methods for use in child assessment tools.
C38.5	15:12 - 15:30	Antoine Barbieri	Random effect models for quality of life analysis in oncology

Session **C39 Multistate models and competing risks I**

Chair Martin Schumacher

Room Hörsaal 41 14:00 - 15:30

C39.1	14:00 - 14:18	Menggang Yu	Illness death models and their applications in cancer research
C39.2	14:18 - 14:36	Yuko Palesch	Multi-state model for analysis of modified Rankin Scale in acute stroke trials: a new approach with a twist.
C39.3	14:36 - 14:54	Micha Mandel	The illness death model under left truncated and right censored data
C39.4	14:54 - 15:12	Liesbeth de Wreede	Multi-state models for treatment success after stem cell transplantation
C39.5	15:12 - 15:30	Nadine Binder	Comparing multistate approaches for reducing the bias of relative risk estimates from cohort data with missing information due to death

Session **C40 Model performance evaluation**

Chair Geir Edil Eide

Room Hörsaal 34 14:00 - 15:30

C40.1	14:00- 14:18	Babak Choodari-Oskooei	A new measure of predictive ability in a survival model: the total gain statistic
C40.2	14:18 - 14:36	Jerome Lambert	A note on the time-profile of time-depndent area under the ROC curve for survival data
C40.3	14:36 - 14:54	Morten Fagerland	A unified approach for testing goodness of fit in binary, multinomial, and ordinal logistic regression models
C40.4	14:54 - 15:12	Juan Carlos Pardo-Fernandez	Nonparametric estimation of covariate-specific summary indices of ROC curves through regression models
C40.5	15:12 - 15:30	Janez Stare	On bias of measures of explained variation for survival data

Session **C41 Survival analysis II**

Chair Zdenek Valenta

Room Hörsaal 16 14:00 - 15:30

C41.1	14:00 - 14:18	Jiri Zelinka	Kernel estimation of hazard function for orthopedic data
C41.2	14:18 - 14:36	Michael Mayer	Quantile regression and prediction intervals for survival data
C41.3	14:36 - 14:54	Aris Perperoglou	A special case of the reduced rank model for modelling time varying effects in survival analysis
C41.4	14:54 - 15:12	Andrea Callegaro	Estimating probability of non-response to treatment with survival data.
C41.5	15:12 - 15:30	Kalaivani Mani	A application of Frailty modeling for family level clustering of Infant Mortality in Empowered Action Group States in India

Session	**C42 Poly-omics studies & Systems Biology**		
Chair	Wessel van Wieringen		
Room	Hörsaal 31	14:00 - 15:30	
C42.1	14:00 - 14:18	Johanna Mazur	A stratified boosting approach for combining gene expression measurements from different platforms to identify prognostic markers
C42.2	14:18 - 14:36	Aeilko Zwinderman	Weighted penalized canonical correlation analysis to integrate multiple omics-data
C42.3	14:36 - 14:54	Aslihan Gerhold-Ay	Prediction performance as a measure for optimal mapping of methylation and RNA-seq data
C42.4	14:54 - 15:12	Chen Suo	Integration of somatic mutation, gene expression and functional data in predicting human breast cancer survival
C42.5	15:12 - 15:30	Yessica Fermin	Nonparametric mixture modelling of dynamic bayesian networks derives the structure of protein-networks in adhesion sites

Session	**I6 Statistical methods for poly-omics studies**		
Organizers	Axel Benner and Manuela Zucknick		
Room	Auditorium Max	16:00 - 17:30	
I6.1		Rainer Spang	From associations to mechanical understanding - data integration and causal inference in genomics
I6.2		Marina Vannucci	Bayesian models for integrative genomics
I6.3		Wessel van Wieringen	Do we gain by jointly analyzing multiple types of genomics data?

Session	**C43 Causal inference from observational data II**		
Chair	Marie Reilly		
Room	Hörsaal 7	16:00 - 17:30	
C43.1	16:00 - 16:18	Regina Riedl	Using different propensity score matching methods to construct comparable control groups for disease management program evaluation
C43.2	16:18 - 16:36	Peter Austin	Double propensity-score adjustment: a solution to incomplete matching
C43.3	16:36 - 16:54	Marie-Quitterie Picat	A structural equation modelling approach to explore the role of interferon-a on chronic immune activation in successfully treated HIV-infected patients
C43.4	16:54 - 17:12	Kjetil Røysland	Independent censoring in survival analysis: A causal approach.

Session	**C44 Validation of prediction models**		
Chair	Ewout Steyerberg		
Room	Hörsaal 21	16:00-17:30	
C44.1	17:12 - 17:30	Werner Vach	The need for a third dimension in the external validation of clinical prediction rules
C44.2	16:00 - 16:18	Daan Nieboer	Multiple validation of prediction models: a framework for summarizing and interpreting results
C44.3	16:18 - 16:36	Kym Snell	Summarising the performance of prognostic models developed and validated using multiple studies
C44.4	16:36 - 16:54	Leandro García Barrado	Incorporating retrospective information to reduce the sample size of prospective diagnostic-biomarker-validation designs.

Session	**C45 Multistate models and competing risks II**		
Chair	Hein Putter		
Room	Hörsaal 41	16:00-17:30	
C45.1	16:00 - 16:18	Mia Grand	Regression models for expected length of stay
C45.2	16:18 - 16:36	Susanne Weber	A multistate model to assess the impact of menstrual status in premenopausal breast cancer patients
C45.3	16:36 - 16:54	Michael Lauseker	Variable selection in the illness-death model
C45.4	16:54 - 17:12	Francesca Ieva	Statistical models for improving prognosis of chronic cardiovascular diseases: hazard reconstruction and clustering of patients affected by heart failure
C45.5	17:12 - 17:30	Anh Nguyen Duc	Smooth non-parametric estimation of the cumulative incidence functions for arbitrarily censored data

Session	C46 Multiple imputation		
Chair	Shaun Seaman		
Room	Hörsaal 34	16:00 - 17:30	
C46.1	16:00 - 16:18	Kyoji Furukawa	A multi-stage multiple imputation in a large-scale cohort study
C46.2	16:18 - 16:36	Nicole Erler	Sequential imputation for large epidemiological data sets
C46.3	16:36 - 16:54	Laura Rodwell	Comparison of methods for imputing limited-range variables: a simulation study
C46.4	16:54 - 17:12	Jammbe Musoro	Validation of prediction models based on lasso regression with multiply imputed data
C46.5	17:12 - 17:30	Siobhan Crichton	Impact of incomplete follow-up when exploring associations between baseline characteristics and outcome in a longitudinal study

Session	C47 Special types of censored data		
Chair	Janez Stare		
Room	Hörsaal 16	16:00 - 17:30	
C47.1	16:00 - 16:18	Maja Pohar Perme	Analysing disease recurrence with missing at risk information
C47.2	16:18 - 16:36	Kaspar Rufibach	Weibull regression for a right-censored endpoint with one censored and an arbitrary number of non-censored covariates
C47.3	16:36 - 16:54	Sylvie Scolas	Accelerated failure time model with interval censored data and cure
C47.4	16:54 - 17:12	Aysun Cetinyurek Yavuz	Semiparametric Bayesian frailty model for clustered interval-censored data

Session	C48 Drug development		
Chair	Julia Singer		
Room	Hörsaal 31	16:00 - 17:30	
C48.1	16:00 - 16:18	Christopher Weir	Bayesian response-adaptive design development: practical experiences from the DexFEM trial
C48.2	16:18 - 16:36	Marietta Kirchner	Sample size optimization for phase II/III drug development programs
C48.3	16:36 - 16:54	Sylwia Bujkiewicz	Bayesian meta-analytical methods to incorporate multiple surrogate endpoints in drug development process
C48.4	16:54 - 17:12	Daniel Saure	Sequential meta-analyses of safety data
C48.5	17:12 - 17:30	Toshiro Tango	On the three-arm non-inferiority design including a placebo

Thursday, 28th August 2014

Session **M1 Statistical challenges in the epidemiology of aging**

Organizers Carole Dufouil and Karen Leffondré

Room	Hörsaal 7	09:00 - 12:30	
M1.1		Scott Hofer	Methodological challenges in the epidemiology of aging from a reproducible research perspective
M1.2		Niels Keiding	Survival analysis aspects of the epidemiology of ageing
M1.3		Cécile Proust-Lima	Modelling issues in the longitudinal study of cognitive aging
M1.4		Marcus Koeller	Medical and conceptual challenges in conducting studies of the elderly

Session **M2 Genomics-based Personalized Medicine**

Organizers Andreas Ziegler and Georg Heinze

Room	Hörsaal 21	09:00 - 12:30	
M2.1		Martin Filipits	Application of genomic tests in breast cancer management
M2.2		Joan Bailey-Wilson	Risk prediction models using family and genomic data
M2.3		Bertram Müller-Myhsok	The importance of appropriate quality control in -omics studies as required for personalized and stratified medicine
M2.4		Andreas Ziegler	Study designs for predictive biomarkers

Posters

Poster session P1

Monday, 25th August 2014. 15:30 - 16:00

P1.1 Bayesian methods in biostatistics

P1.1.1	Soheila Aghlmandi	Coverage properties of Bayesian 95% probability intervals for odds ratio and relative risk
P1.1.17	Fabrice Paul Bailleux	Comparison of two methods for futility analysis in vaccine efficacy trials
P1.1.20	Sophie Bastide	Choosing a gold standard: support of Bayesian inference methods for diagnostic accuracy of new biomarkers in pediatric urinary tract infection
P1.1.21	Sophie Bastide	Methodological review of Bayesian inference methods used in clinical decision rules and diagnosis studies
P1.1.32	Sylvie Chevret	A Bayesian hybrid adaptive design for phase III survival trials
P1.1.36	Emílio Augusto Coelho-Barros	Copula functions in the presence of cure fraction
P1.1.43	Daniela Dunkler	Reliable confidence intervals for fractional polynomials: a simulation study
P1.1.107	Leacky Kamau Muchene	Determination of the minimum effective dose for correlated dose-response data using Bayesian variable selection (BVS) models
P1.1.110	Stavros Nikolakopoulos	On adapting the sample size in a Bayesian clinical trial in small populations
P1.1.118	Rhiannon K Owen	Bayesian analysis of parametric frailty models for repeated event data: estimating unreported event times using interval data
P1.1.124	David Pau	Phase II study to assess the safety of bevacizumab with neoadjuvant chemotherapy in ovarian cancer using a Bayesian approach
P1.1.130	Alfonso Ramos	Some inferential results in branching processes in random environments
P1.1.164	Chitra Tirodkar	Bayesian methods in adaptive dose finding

P1.2 Design and analysis of clinical trials

P1.2.5	Félix Almendra-Arao	Introducing continuity correction for the Laster-Johnson-Kotler non-inferiority asymptotic test
P1.2.7	Alberto Alvarez-Iglesias	An online calculator for futility interim monitoring rules in randomised clinical trials
P1.2.18	Elleny Balder	Best-after-breast design: challenges of nutrition intervention studies in infants
P1.2.28	Per Broberg	When does an interim analysis not jeopardise the type I error rate ?
P1.2.33	Chieh Chiang	Use of an adaptive approach to design and evaluation of multi-regional clinical trials
P1.2.35	Cristian Ciria	Appropriate or inappropriate use of REMARK guidelines
P1.2.47	Monia Ezzalfani	Using longitudinal toxicity score to detect time trend in dose-finding trials. Application to 19 phase I studies
P1.2.48	Mohammad Fayaz	Standardizing safety analyses for clinical trials: a story of success in the making
P1.2.57	Aenne Glass	Trial-situations where stratification in randomization is advantageous - results of a simulation study
P1.2.59	Moisés Gómez-Mateu	CompARE: web platform to choose the primary endpoint of a randomized clinical trial
P1.2.63	Stefanie Hayoz	Effect of one-patient clusters on power in cluster-randomized trials
P1.2.64	Stefanie Hayoz	Comparison of design options for phase IB clinical trials in oncology: simulation results
P1.2.67	Gerald Hlavin	Adapted levels of evidence for small populations
P1.2.68	Julia Hocke	Retrospective evaluation of the futility analysis in LUME-Lung 2, a phase III trial of nintedanib plus pemetrexed for NSCLC patients
P1.2.86	Sophie Knahl	Comparison of two-stage versus two separate single-stage settings for bioequivalence studies with crossover design
P1.2.88	Franz König	Adaptive two-stage bioequivalence trials with early stopping and sample size re-estimation
P1.2.89	Sergii Krasnozhon	Adaptive designs for confirmatory model based decisions using MCP-Mod
P1.2.93	Martin Law	Residual plots for censored data: a new approach
P1.2.97	Kun Liu	Longitudinal cluster analysis with application to identify mortality associated SOFAtrends in critical care medicine
P1.2.100	James T Martin	Using simulation to examine the effects of varying cluster size on the precision of stepped wedge cluster randomised trials

P1.2.106	Katy E Morgan	Analysis of cluster randomised cross-over trials with binary outcomes
P1.2.109	Thu Thuy Nguyen	Optimal sampling times for pharmacokinetic modelling of a cocktail of phenotyping drugs
P1.2.112	Petra Ofner-Kopeinig	Generalization of the big stick randomization rule to more than two treatment groups and unequal allocation rates
P1.2.114	Neil O'Leary	The implications of differential clustering for the analysis of binary outcomes in cluster randomised trials
P1.2.121	Sameer Parpia	Treatment crossovers in time-to-event non-inferiority randomized trials of radiotherapy in subjects with breast cancer
P1.2.122	Ozge Pasin	An application of non-parametric factorial MANOVA ın health research
P1.2.142	Silvana Romio	Comparison of different methods for controlling false positives in adverse event reports analysis
P1.2.146	Sarah Simpson	Adaptive crossover designs for phase II dose-finding trials
P1.2.166	Juan V Torres-Martin	Statistical methods for centralised risk-based monitoring in clinical trials
P1.2.167	Juan V Torres-Martin	Re-sampling methods for internal model validation in diagnostic and prognostic studies: review of methods and current practice
P1.2.168	Juan V Torres-Martín	Adaptive increase sample size with count endpoints: the path from statistical simulation to the development of an explicit formula
P1.2.172	Diane Uschner	Comparison of different allocation procedures in clinical trials in small population groups with respect to accidental and selection bias
P1.2.178	Anand N Vidyashankar	Factors associated with fever control and identification of subgroups in sepsis trials: a regularization based approach
P1.2.195	Wenle Zhao	Benefit and cost of clinical trial operation quality monitoring - experiences and lessons from the ProTECT III trial

P1.3 Pharmacoepidemiology and drug development

P1.3.51	Brooke L Fridley	Pharmacogenetic study of delayed hyperbilirubinemia in a cohort of 4,000 infants
P1.3.123	David Pau	Choice-based conjoint (CBC) analysis to evaluate patient´s perception regarding their erythropoiesis stimulating agent (ESA) treatment in chronic kidney disease (CKD)
P1.3.136	Mafalda Ribeirinho	Fast track assessment of generics in Portugal

Poster session P2

Tuesday, 26th August 2014, 10:30 - 11:00

P2.1 Longitudinal data analysis

P2.1.54	Susana García	Six- and 12-month follow-up of an interdisciplinary treatment of patients with fibromyalgia: results of a randomised trial
P2.1.60	Stefan Hantel	Assessment of the bias introduced by excluding patients from the analysis set due to missing post-randomization
P2.1.74	Karel Hrach	Progress of bilateral monitoring - case study
P2.1.77	Susan C.E. Ifeagwu	Assessing the ´General Health Questionnaire´ and ´Center of Epidemiological Studies Depression Scale´ for depression screening: stroke and cancer patients
P2.1.92	Anna M. Laszlo	Investigating trend in the rate of suicide using regression methods in Hungary between 1963 and 2011
P2.1.119	Agnieszka Pac	Health-related mortality predictors among Krakow older citizens. 25-year follow-up study
P2.1.129	Cécile Proust-Lima	An IRT longitudinal model for graded repeated responses: IADL and ADL hierarchy and functional dependency trajectories in the elderly
P2.1.147	Rajvir Singh	Factors associated with under five child mortality in mothers´ employed in agriculture, India
P2.1.161	Sophie Tezenas du Montcel	Disease evolution of spinocerebellar ataxia type 2 patients: interruption of follow-up considerations

P2.2 Methodological issues and case studies in epidemiology

P2.2.16	Shaghayegh Bagher	Use of propensity score in plastic and reconstructive research; rare complication events in elective surgeries
P2.2.29	Vilya A. Bulgakova	Pharmacoepidemiological characteristics of influenza-like illness in hospitalized children in Russia: a case-control study

P2.2.41	Maria A.J. de Ridder	Sensitivity analysis for possible bias due to event-dependent observation periods in self-control case series analysis
P2.2.52	Ikuko Funatogawa	Smoking statistics in the mid-1920s birth cohorts
P2.2.53	Ikuko Funatogawa	Non-monotonic trends in smoking statistics
P2.2.65	Harald Heinzl	Exemplifying the usefulness of combining difference and equivalence tests in spatial maps
P2.2.79	Bo-Hyoung Jang	Medical use of allergic rhinitis under two healthcare system in South Korea
P2.2.85	Catherine Klersy	Multidimensional outcome. Does the interpretation change with the analysis method?
P2.2.105	Makiko N Mieno	Sensitivity analysis for the misclassification of competing outcomes in a cohort study in Japan
P2.2.115	Raymond Omollo	Timing for definitive cure in clinical trials for visceral leishmaniasis
P2.2.127	Janne Pitkäniemi	Etiological age-components of cervical cancer in Finland in 1953-2011
P2.2.132	Jeno Reiczigel	Location-scale tests for non-negative data with skewed distribution, with focus on parasitology research
P2.2.155	Jennifer A Summers	The challenges of conducting a multidisciplinary trilogy of studies: diagnostic accuracy & use of the SeHCAT test in England
P2.2.156	Jennifer A Summers	The value of using modern epidemiological approaches in studying past influenza pandemics: combining history, war and statistical methods
P2.2.163	Bhaskar Thakur	Assessment of neighbourhood effect on neonatal mortality: translation of area level variance in odds ratio scale in multilevel logistic regression
P2.2.182	Katalin Virág	Regression models for rare events – stroke mortality rates over the last 30 years in Hungary
P2.2.183	Thomas Waldhoer	Dependence of the effect of altitude on infant as well as maternal related variables on birth weight
P2.2.185	Wei Wei	Time series analysis of Campylobacter incidence in Switzerland

P2.3 Methods for handling missing data

P2.3.2	Urko Aguirre	Multiple imputation is not necessary for performing analyses in pre-post studies
P2.3.44	Iris Eekhout	A new method for significance testing of categorical covariates after multiple imputation
P2.3.137	Marianne Riksheim	Missing categorical data: the influence of imputation technique on regression analysis in an opioid maintenance treatment setting
P2.3.140	Laura Rodwell	Imputation of an ordinal exposure derived from a semi-continuous variable with missing data: a simulation study
P2.3.169	Chi-Hong Tseng	Regularized approach for missing data problem
P2.3.188	Steffen Witte	What if my doctor would be as receptive to innovations in therapies as to innovations in statistical methods?
P2.3.196	Vera Daniela Zietemann	Anemia is a risk factor for poor cognitive outcome after ischemic stroke

P2.4 Penalized methods in high- and in low-dimensional regression analyses

P2.4.23	Mohamed Amine Benadjaoud	Dimensional reduction in the flexible B-spline Cox model using functional principal components analysis
P2.4.56	Angelika Geroldinger	Firth's bias reduction method revisited: software implementation boosts application

P2.5 Statistical methods for systems biology and genetics

P2.5.40	Lizzy De Lobel	A two-stage approach to test for gene-gene interactions in family data based on within-family and between-family information
P2.5.78	Ivana Ihnatova	Topology-based pathway analysis of microarray and RNA-Seq data: an evaluation of existing methods
P2.5.116	Wouter Ouwerkerk	Contribution of alternative splicing variants to gene expression variation
P2.5.120	Mira Park	Multi-purpose SNP selection method in genetic association study
P2.5.138	Ralph C.A. Rippe	Measurement Error in GWAS: what have we missed?
P2.5.144	Fabian Schroeder	Classification in high-dimensional feature spaces

P2.5.194	Fatemeh Zamanzad Ghavidel	A non-homogeneous hidden Markov-model for gene mapping based on whole-genome sequencing data

P2.6 Software aspects of efficient statistical analyses

P2.6.26	Mariann Borsos	Easy-to-use R-application to evaluate bio-equivalence studies
P2.6.30	Yuan-Chin I. Chang	Boosting diagnosis performance of biomarkers with nonparametric logistic type classification functions
P2.6.73	Karel Hrach	Repeated observations design analysed with ANOVA tools in MS-Excel

Poster session P3 Wednesday, 27th August 2014, 11:00 - 11:30

P3.1 Survival analysis, multistate models and competing risks

P3.1.19	Mathieu Bastard	Competing-risk regression model to explore risk factors associated with lost to follow-up prior to antiretroviral therapy: a multicentric observational cohort
P3.1.22	Alexander Begun	Description of disease progression and relevant predictors in diabetic foot ulcer patients using a Markov chain model
P3.1.27	Rebecca H Boucher	Reconstructing individual patient level data: a simulation approach
P3.1.69	M HP Hof	Joint modelling of multiple longitudinal markers and recurrent events of multiple types
P3.1.75	Marianne Huebner	Analyzing clinical pathways in observational studies: pitfalls and approaches
P3.1.111	Masako Nishikawa	Remedy for 'IntCox' in partly interval-censored survival data
P3.1.125	Markus Pfirrmann	Comparison of survival between allogeneic haematopoietic stem cell transplantation and continued drug treatment when differentiating between risk groups at diagnosis
P3.1.148	Milada Cvancarova Småstuen	Modeling cause-specific survival in cancer patients compared to the general population, a large population based study
P3.1.149	Ana Sousa	Cox model with multiple events: an application to mammography screening intervals in the Portuguese primary health care system
P3.1.150	Jacqueline Stephen	Statistical modelling of biomarkers incorporating non-proportional effects for survival data

P3.1.151	Jacqueline Stephen	Impact of length of follow-up on the evaluation of prognostic scores with an example using two breast cancer studies
P3.1.157	Prafulla Kumar Swain	On the estimation of survival of HIV/AIDS patients on anti-retroviral therapy: an application to interval censored data
P3.1.160	Laetitia Teixeira	The relevance of joint modelling of longitudinal and competing risks data in the analysis of a peritoneal dialysis program
P3.1.162	Ammarin Thakkinstian	An illness-death model of chronic kidney disease progression
P3.1.170	Umaporn Udomsubpayakul	An illness-death model of HIV infection
P3.1.189	Willy Wynant	New insights on therapy choices in non-small cell lung cancer using a flexible extension of the standard Cox's model
P3.1.191	Sean Yiu	Heterogeneous M/M/1 type queuing models

P3.2 Diagnostic studies

P3.2.91	Zsolt Lang	Sample size calculations for confidence limits of prevalence of disease adjusted for estimated sensitivity and specificity
P3.2.95	Ji Eun Choi	Systematic review and meta-analysis of diagnostic accuracy of FDG-PET in dementia and Alzheimer's disease
P3.2.101	Pablo Martínez-Camblor	Smooth time-dependent ROC curve estimators
P3.2.103	Elizabeth J McKinnon	Seasonality in testing for systemic lupus erythematosis
P3.2.117	Aleksander Jerzy Owczarek	Clinical factors affecting bias between different eGFR measurements based on the weighted Deming regression
P3.2.153	Ana Subtil	When do latent class models outperform an imperfect gold standard? A problem revisited
P3.2.154	Ana Subtil	Bayesian latent class models for the evaluation of diagnostic tests in multiple populations

P3.3 Analysis of electronic health records

P3.3.49	Fanny Feuillet	A validation algorithm for detecting dose increase from longitudinal data of psychotropic drug users, using Monte Carlo simulation
P3.3.96	Ruoran Li	How much of socioeconomic differences in breast cancer patient survival can be explained by stage at diagnosis and treatment?
P3.3.152	Adam J Streeter	Adjustment for hidden confounding in the analysis of pneumococcal vaccination effectiveness using electronic health records

P3.4 Comparative effectiveness and outcomes research

P3.4.25a	Laura J Bonnett	Short- versus long-term outcomes after treatment for Tuberculosis
P3.4.34	Jimin Kim	The association of inhaled bronchodilators with the risk of acute myocardial infarction
P3.4.83	Jimin Kim	Economic evaluation of cervical cancer screening strategy
P3.4.113	Mercy Ofuya	Dichotomising highly skewed outcome data using a distributional method: a simulation study
P3.4.143	Kinga Salapa	Comparison of classification models for sex determination of Polish skulls
P3.4.180	Estela Vilhena	Quality of life in Portuguese cancer patients. A structural equation modeling application
P3.4.181	Estela Vilhena	A structural equation modeling application to test mediation of optimism between stigma and quality of life in Portuguese obese patients

P3.5 Development and validation of clinical prediction models

P3.5.3	Urko Aguirre	Logistic regression and linear discrimant analysis for assessing factors related to genetic anemia: a comparison of both approaches
P3.5.8	Gareth Ambler	An investigation of performance measures developed to validate risk models for survival data
P3.5.10	Inmaculada Arostegui	PREVEXEPOC: a computer tool for risk stratification of patients with exacerbated COPD based on a predictive severity scoring system
P3.5.11	Junichi Asano	Assessing the prediction accuracy of cure in the Cox proportional hazards cure model

P3.5.25	Laura J Bonnett	External validation of a prognostic model
P3.5.37	Olivier Collignon	Partial least square discriminant analysis of neurological outcome after cardiac arrest using bispectral index
P3.5.45	Joie Ensor	Internal-external cross-validation (IECV) for prognostic model research using data from multiple studies: potential & pitfalls
P3.5.55	Susana Garcia-Gutierrez	Creation and validation of a predictive model to assess poor outcomes in acute decompensated heart failure
P3.5.62	Andrea Harnos	Model building using learning methods to identify SNPs related to the pharmacokinetics of high-dose methotrexate in pediatric acute lymphoblastic leukemia
P3.5.82	Vineet Kumar Kamal	Prediction of outcome after severe and moderate head injury by classification and regression tree technique
P3.5.145	Beverley M Shields	A clinical diagnostic model using biomarkers and clinical characteristics for the identification of MODY patients
P3.5.173	Kirsten Van Hoorde	Comparing different methods to develop prediction models for polytomous outcomes
P3.5.174	David van Klaveren	Assessing the influence of case-mix heterogeneity on the discriminative ability of a risk model: the model-based concordance-index
P3.5.175	David van Klaveren	A censoring-robust concordance measure for proportional hazards regression models in external validation data: the calibrated Gönen and Heller estimator

Poster session P4

Wednesday, 27th August 2014, 15:30-16:00

P4.1 Meta-analysis and network meta-analysis

P4.1.4	Ikhlaaq Ahmed	Developing and validating risk prediction models in an individual participant data meta-analysis
P4.1.38	Christophe Combescure	Meta-analysis of single-arm survival studies: a distribution-free approach for estimating summary survival curves with random effects
P4.1.42	Thomas P A Debray	Publication bias tests for survival data: a simulation study
P4.1.80	Yannan Jiang	Meta-analysis of mobile phone-based interventions for smoking cessation trials in different countries

P4.1.87	Min Jung Ko	Diagnostic accuracy of Pap testing and human papillomavirus DNA testing in cervical cancer screening in Korea
P4.1.108	Dittapol Muntham	Mean platelet volume and coronary artery disease: a systematic review and meta-analysis
P4.1.131	Sasivimol Rattanasiri	The association between the IRF6 genes and non-syndromic cleft lip with or without cleft palate: Systematic review and meta-analysis
P4.1.134	Kirsty M Rhodes	Bayesian meta-analysis without MCMC
P4.1.135	Laureen Ribassin-Majed	Individual patient data network meta-analysis with a time-to-event endpoint in head and neck cancer
P4.1.177	Pablo E Verde	Bayesian evidence synthesis and combining randomized and nonrandomized results: a case study in diabetes
P4.1.186	Christopher J Weir	Meta-analysis of continuous outcomes: systematic review of methods available for dealing with missing mean and standard deviation values
P4.1.190	Kazue Yamaoka	Effect of lifestyle and metformin for delaying or preventing type 2 diabetes: a network meta-analysis

P4.2 Modeling infectious diseases

P4.2.24	Stéphanie Blaizot	A predictive model for HIV spreading in hyperendemic settings
P4.2.58	Patrícia Filipe	A dynamic regression analysis of pulmonary tuberculosis incidence
P4.2.66	Sereina A Herzog	Modelling infectious disease parameters using serological data
P4.2.102	Amy Matser	Determining the risk of Neisseria gonorrhoeae infection by meeting location among men who have sex with men in Amsterdam

P4.3 Novel designs and methods for simulations

P4.3.13	Nada Assi	Bi-cross-validation for the choice of optimal number of non-zero loadings in sparse PCA methods
P4.3.70	Richard Hooper	Calculating sample size for cluster-randomised trials with mid-point sample size re-assessment

P4.4 Observational studies and causal inference methods

P4.4.14	Safia Awan	Prevalence of internet use and Internet addiction disorder among medical students: a case from low income country
P4.4.15	Salma Ayis	Quality of life (QoL) after stroke in five European populations assessed by health survey form (SF-12) and EuroQoL (EQ-5D)
P4.4.46	Christine Eulenburg	Discrepancies in exercise therapy prescriptions after hip replacement: a multicenter survey of surgeons, rehabilitation physicians and physiotherapists
P4.4.94	Jaebong Lee	Application of propensity-score matching: comparison of CT and ultrasonography in the effect on negative appendectomy
P4.4.99	Mohammad Reza Maracy	The mediating role of mental health in the relations between dietary behaviors and general health
P4.4.104	Kirsten Mehlig	The distribution of apolipoprotein E genotype in relation to age and origin of birth
P4.4.126	Ronnie Pingel	Estimating the variance of a propensity score matching estimator: another look at right heart catheterization data
P4.4.171	Yukari Uemura	Sensitivity analysis of dissociative principal strata effect: application to a bone fracture prevention trial
P4.4.192	Ho Ming Yuen	Does a predisposition to kidney disease originate during prenatal development? A cohort study from the Born in Bradford Project

P4.5 Other

P4.5.6	Henar Alonso	Meta-analysis of genome-wide gene-environment interactions on colorectal cancer
P4.5.9	Caroline Ameling	Dutch national cohort for environmental health issues. The DUELS Study
P4.5.12	Shuichiro Asano	Correlation between the degree of grade of astrocytomas and the current value of motor-evoked potential during brain surgery
P4.5.61	Markus Harden	Calculation of target range values on a continuous scale in immunosuppression following solid organ transplantation
P4.5.71	Sayed Mohsen Hosseini	Population-based metabolic syndrome risk score and its determinants: the Isfahan Healthy Heart Program

P4.5.72	Sayed Mohsen Hosseini	Comparison of reference curves of anthropometric indices in two national studies conducted among Iranian children in 2003-2004 and 2009-2010
P4.5.76	Sabir Hussain	IL-1A gene promoter region polymorphism and the risk of familial CAD in a Pakistani population
P4.5.81	Victor Lih Jong	Selecting a classification function for class prediction in gene expression data
P4.5.84	Jimin Kim	Risk of pneumonia in patients with respiratory disease according to type of inharler devices
P4.5.90	Roman Kreuzhuber	Finding variant X - a framework for identification of causative variants in WGS-data
P4.5.98	Jay Mandrekar	Use of factor analysis in assessments of clinical teaching evaluations
P4.5.128	Antoine Poncet	How to choose a two-sample test for continuous variables: a new solution to an old problem
P4.5.133	Zsofia Reiczigel	Properties of ANOVA and its alternatives under violation of their assumptions: a simulation study
P4.5.139	María Xosé Rodríguez-Álvarez	Estimation of the ROC curve with a time-dependent disease variable in the presence of covariates
P4.5.141	Jo Røislien	“Numberstories”: Teaching statistics to kids - and clinicians
P4.5.158	Andrzej Szymański	A model of population dynamics based on fuzzy cellular automata theory
P4.5.159	Jérôme Tanguy	Studying latent hydration level through multivariate longitudinal biological indicators: application of a latent process mixed model
P4.5.165	Leman Tomak	Item analysis and evaluation of the examinations in Faculty of Medicine
P4.5.176	Thuva Vanniyasingam	An empirical comparison of methods for a conjoint analysis survey of knowledge translation in women's addiction agencies
P4.5.179	Juan Vilar	Outlier detection in functional time series: an application to mortality rates
P4.5.184	Yanzhong Wang	Split and merge techniques for Gaussian mixture learning
P4.5.187	Tore Wentzel-Larsen	Preanalytical variation: computing variance estimates from systematic differences and models for clinical practice

Information for Presenters

Instructions for Poster Presentations

The Poster Exhibition is scheduled from Monday to Wednesday (25 – 27 August 2014).

Set-up and take down time
Set-up from: Sunday, 24 August 2014, morning

Please install your poster at the poster board marked with your poster number. Adhesive double-sided tape will be provided on-site.

Take down from: Wednesday, 27 August 2014, 16:00

Posters which are not removed by Thursday 11:00 will be disposed of.

Poster exhibition venue
The Poster Exhibition will take place in the Arkadenhof (arcades of the inner courtyard) in the mezzanine of the conference venue.

Each poster is assigned to one of four Poster Sessions which take place during one of the coffee breaks from Monday to Wednesday. Please stay close to your poster during your assigned Poster Session.

Instructions for Oral Presentations

You can find the session in which your presentation is scheduled on the website of the conference and in the conference program. Please make your presence known to the session chair at least 10 minutes before your session starts and be present during the entire session in which your presentation is scheduled.

Length of presentation
The time allocated for contributed presentations is **15 minutes**, plus 3 minutes for discussion. The length of invited presentations is determined by the respective session organizer. We kindly ask you to stay within the time limit in order to give the other speakers enough time for their talks, and to give the audience the opportunity to ask questions.

Transfer of presentation
All invited and contributed oral presentations will be using computers provided by the conference organization. No personal laptop or notebook computers will be allowed for invited and contributed oral presentations.

The presentations can be uploaded to the computer of the respective lecture hall in any break prior to your session. A 'Helper' will assist and assign your file to the correct session folder. At that instance quickly check your presentation.

We recommend that you have a backup copy of your presentation on your own memory stick available during your session.

Equipment in the presentation room
All conference rooms are equipped with the following:
- One stand-alone computer
- One projector
- One wireless microphone
- One laser pointer

ISCB Awards

Conference Awards for Scientists (CAS)

C41.1 Kernel estimation of hazard function for orthopedic data
Jiri Zelinka, Czech Republic

C02.3 CD4+ counts in a 3-arm longitudinal clinical trial with substantial missing data: a sensitivity analysis
Anneke Grobler, South Africa

Student Conference Awards (SCA)

C15.4 Combined dynamic predictions using joint models of multiple longitudinal outcomes and competing risk data
Eleni-Rosalina Andrinopoulou, Netherlands

C24.1 Group sequential monitoring of response-adaptive randomised clinical trials with censored survival data
Hsiao Yin Liu, United Kingdom

C35.5 Correcting for bias in the detection and validation of informative diagnostic tests
David Robertson, United Kingdom

C31.1 An extension of the lasso penalization to reduce false positive selection in high-dimensional Cox models
Nils Ternès, France

Special Issue of Statistics in Medicine

As in previous years a special issue of Statistics in Medicine will be published with contributions from this ISCB conference. Please refer to the Author Guidelines of Statistics in Medicine for instructions and use the ScholarOne Manuscripts system at http://mc.manuscriptcentral.com/sim for submissions. During the submission process please select ‚Special Issue Paper' as article type, and indicate in the cover letter that the submission is for the ISCB35 special issue.
The deadline for submission is 1 November 2014.

Georg Heinze
Chair, Scientific Programme Committee

Conference Sponsors

Gold Partners

Baxter

Cytel

F. Hoffmann-La Roche

SAS - The Power to Know

Stadt Wien

Vienna Convention Bureau

Silver Partner

BioMed Central
(Sponsor of a scientific award)

CRC Press

Medical University of Vienna

Springer publishing

Wiley

Bronze Partner

Austrian Airlines

MPS Research Unit

Österreichische Statistische Gesellschaft

General Information

Opening Hours of Registration Desk

Sunday,	24 August 2014:	08:00 – 18:00
Monday,	25 August 2014:	08:00 – 18:00
Tuesday,	26 August 2014:	08:00 – 14:00
Wednesday,	27 Augsut 2014:	08:00 – 18:00
Thursday,	28 August 2014:	08:00 – 13:00

Lunches and Refreshments

Refreshments during the breaks and lunch (Monday to Wednesday, 12:30 – 14:00) are included in the conference fees. Lunch on Sunday is included only if a full-day or two half-day courses are booked.

Public Transportation

The Vienna Transport Authority (Wiener Linien) offers a modern and efficient network. Operating hours are from 5:00 to 0:30, night buses from 0:30 to 5:00. On weekends the underground lines are operated 24 hours.

Tickets for tram, bus or underground lines can be purchased in tobacco shops, at ticket machines (directly located in the underground stations) or at ticket machines in the tram. For more information see www.wienerlinien.at. **Free tickets** are offered to participants of the conference at the registration desk.

City Bike Wien is a free city bike in Vienna - more information on www.citybikewien.at.

Smoking

Smoking is not allowed at the conference venue.

Telephone

The international access code is +43 for Austria and +431 for Vienna. Emergency Telephone numbers are 112 (European emergency call), 122 (Fire department), 133 (Police), 144 (Ambulance & emergency services), 141 (Emergency doctor), 512 20 78 (Night and weekend dental services), 1550 (Pharmacies open during nights and Sunday).

Business Hours and Shopping

Shopping hours are Monday to Friday from 09:00 to 18:00 and Saturday from 09:00 to 12:00, supermarkets have longer opening hours. Most shops are closed on Sundays. Visitors from non-EU countries should ask for VAT refunds when purchasing goods.

Banks and Exchange

Banks are generally open from Monday to Friday, from 08:00 to 12:30 and from 13:30 to 15:00 (Thursday from 08:00 to 12:30 and from 13:30 to 17:00). Most Austrian Banks offer currency exchange facilities during opening hours and can also change Traveller's cheques. Automated cash machines (ATMs) are located outside most banks, where cash can be withdrawn 24 hours a day.

Social Programme

Welcome Reception

Monday, 25 August 2014, 19:00 - 21:30
Festsaal of the Vienna City Hall
Friedrich-Schmidt-Platz 1
(entrance: Felderstraße 1A)

Meet your international colleagues and enjoy a pleasant evening with snacks and drinks in the great ballroom (Festsaal). The City Hall is close to the conference venue and can be reached by a short walk along the Ringstrasse. For conference participants the Welcome Reception is included in the conference fee. For each accompanying guest € 20,- will be charged.

The City Hall is one of the most splendid amongst the numerous monumental buildings along Vienna's Ringstrasse. Designed by Friedrich Schmidt, it was erected between 1872 and 1883. The architecture of the Ringstrasse is dominated by historicism. In Historicism various stylistic elements of the past were combined into a style in its own right. Friedrich Schmidt however orientated himself just on one particular epoch. The City Hall was built in gothical style, with a tower similar to gothic cathedrals. Today the City Hall is the head office of Vienna's municipal administration."

Conference Dinner

Wednesday, 27 August 2014, 19:00 - 24:00
Großer Ferstelsaal of the Palais Ferstel
Freyung 2

Equipped with magnificent chandeliers, parquet flooring and elaborate ceiling murals, the Großer Ferstelsaal, designed in the style of romantic historicism, is one of the most magnificent and beautiful ballrooms in Vienna. This unique room decorated in warm beige tones provides a perfectly harmonious setting for the conference dinner.

Great food and drinks will be offered for this dinner. Afterwards, the music band Divertimento Viennese will play traditional Viennese and other classic pieces for dancing.
Enjoy the last evening of the conference with your international colleagues in the fantastic ambience of Palais Ferstel!

Young Researchers Lounge

Tuesday, 26 August 2014, 19:00
Heuriger Schübel-Auer in Nußdorf

This is the perfect chance to meet other young researchers in statistics. Get to know what is going on in other groups, or just have a glass of wine together with fellow young researchers and share your research experience with others.

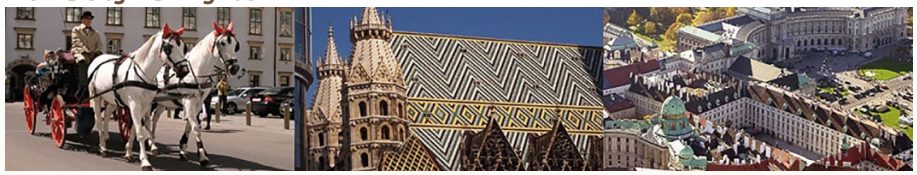
The Young Researchers Lounge will take place at a 'Heurigen', a rustic and authentic style wine tavern in Nußdorf, one of Vienna's charming and romantic wine growing villages surrounded by soft hills and vineyards. Nußdorf can be reached directly from the conference venue by tram 'D' (exit at terminal stop).

The free Conference Tour 'Hiking in the Vienna Woods' will end at the same 'Heurigen'.

Conference Tours

On Tuesday, 26th of August, a wide variety of tours in and around Vienna will be offered. All tours will start from the conference venue, Vienna's main university building, at 14:00. The Vienna City Walking Tour will also be offered on Sunday, 24th of August, at 15:00.

Vienna City Walking Tour



This classical city walk will visit well known historic buildings and monuments in Vienna's city centre. The tour will conclude with a visit to the town's landmark, the gothic cathedral St. Stephen and its catacombs.
Price: € 25, includes English speaking guide, entrance fee to catacombs
Duration: 3,5 hrs. Min. number of participants: 15

Vienna City Music Walking Tour

More famous composers have lived in Vienna than in any other city. The tour will visit places where composers lived and worked, and concludes with a visit to the House of Music.
Price: € 29, includes English speaking guide, entrance fees to Pasqualati House & House of Music. Duration: 4 hrs. Min. number of participants: 20

Schönbrunn Palace and Park

The tour to Schönbrunn Palace, the former summer residence of the Habsburg dynasty, includes a visit to the state rooms and private apartments of the emperor Franz Joseph and his wife. There is ample time for walking through the impressive huge park up to the cafe in Gloriette.
Price: € 38, includes bus shuttle, English speaking guide, entrance fee.
Duration: 4 hrs. Min. number of participants: 20

Art Nouveau

Vienna was one of the birthplaces of modernity at the beginning of the last century. Art Nouveau has a well-established place in Viennese architecture. At the turn of the 20th century, prominent architects such as Otto Wagner and Josef Hoffmann designed world-renowned buildings; several of these will be visited on the tour.
Price: € 42, includes English speaking guide, bus, and entrance fees to Secession & Church of Steinhof. Duration: 4 hrs. bus tour. Min. number of participants: 20

Schloss Hof Palace and Gardens

Prince Eugene's former summer palace Schloss Hof, built in the 18th century, is one of the most beautiful settings of baroque design. You will enjoy the unique baroque gardens, the magnificent rooms of the chateau, and the idyllic manor farm.
Price: € 57, includes bus shuttle, English speaking guide, entrance fee.
Duration: 4.5 hrs. Min. number of participants: 20

Boat Tour in the Donau-Auen National Park

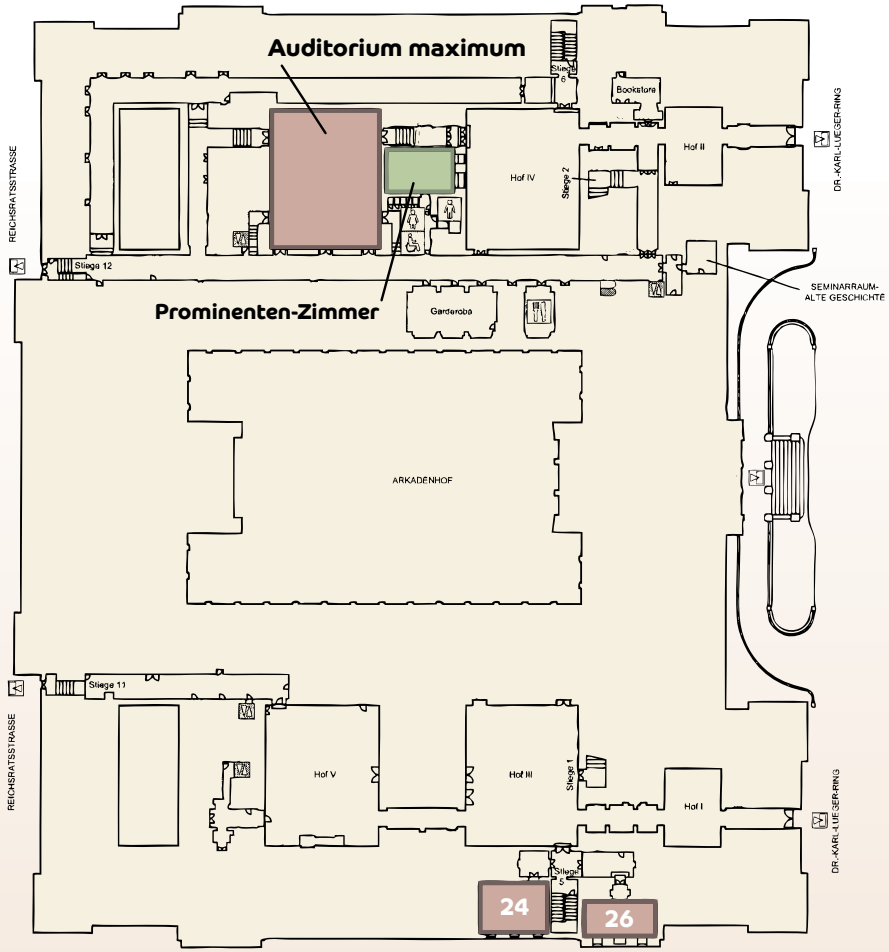
With National Park rangers as your guides, experience one of the most beautiful former branches of the Danube in a canoe - the Stopfenreuth Wetlands. Observe the lush flora and be on the lookout for rare fauna while paddling across calm waters.
Price: € 55, includes bus shuttle, English speaking guide, boat tour
Duration: 5 hrs. Min. number of participants: 18

Hiking in the Vienna Woods

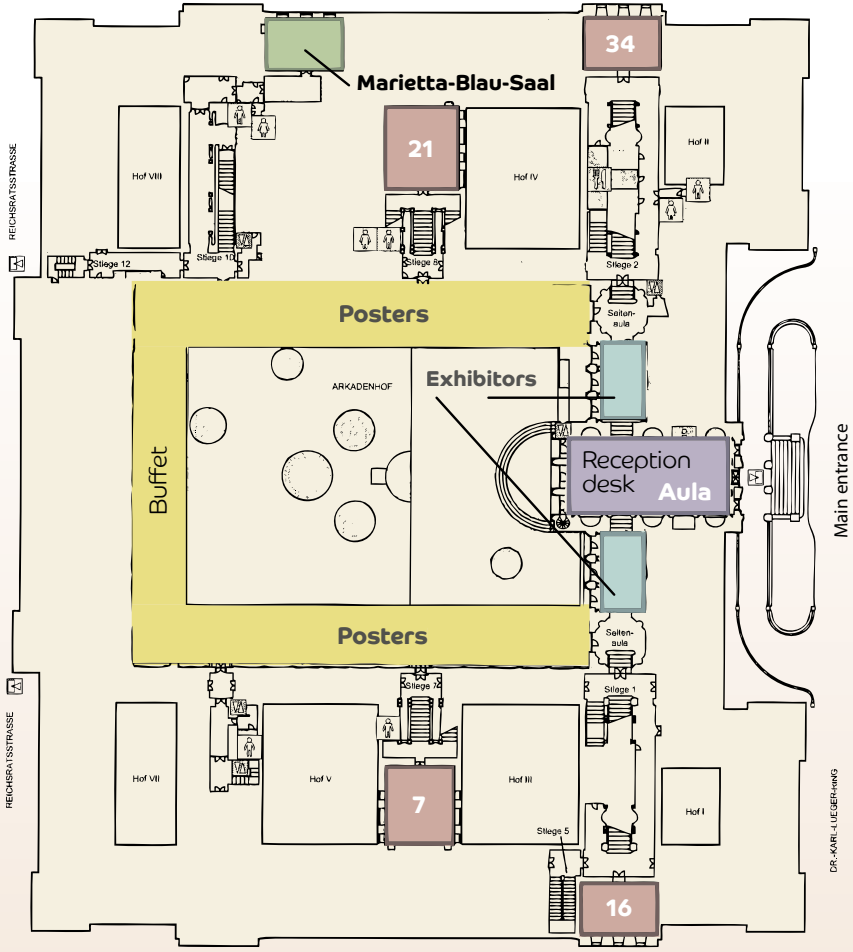
Explore the favourite recreation resort of the Viennese on a hike through the Vienna Woods. Walk from the Schwarzenbergpark to the Stefaniewarte on the Kahlenberg from where you have a breathtaking view over the city.
Price: free of charge, includes English speaking guide, tram ticket & entrance ticket to Stephaniewarte. Duration: approx. 4 - 5 hrs walking tour (ca 15 km)

Overview of the Conference Venue

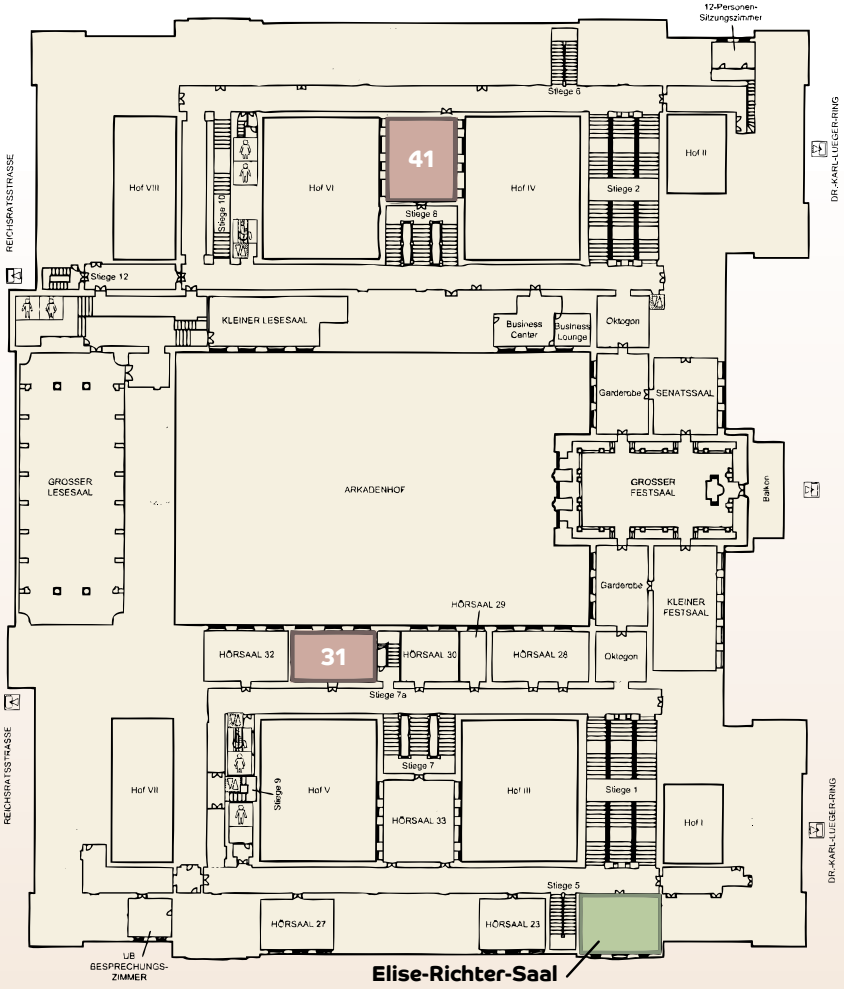
Basement (Tiefparterre)



Mezzanine (Hochparterre)



Level 1 (1. Stock)



Imprint

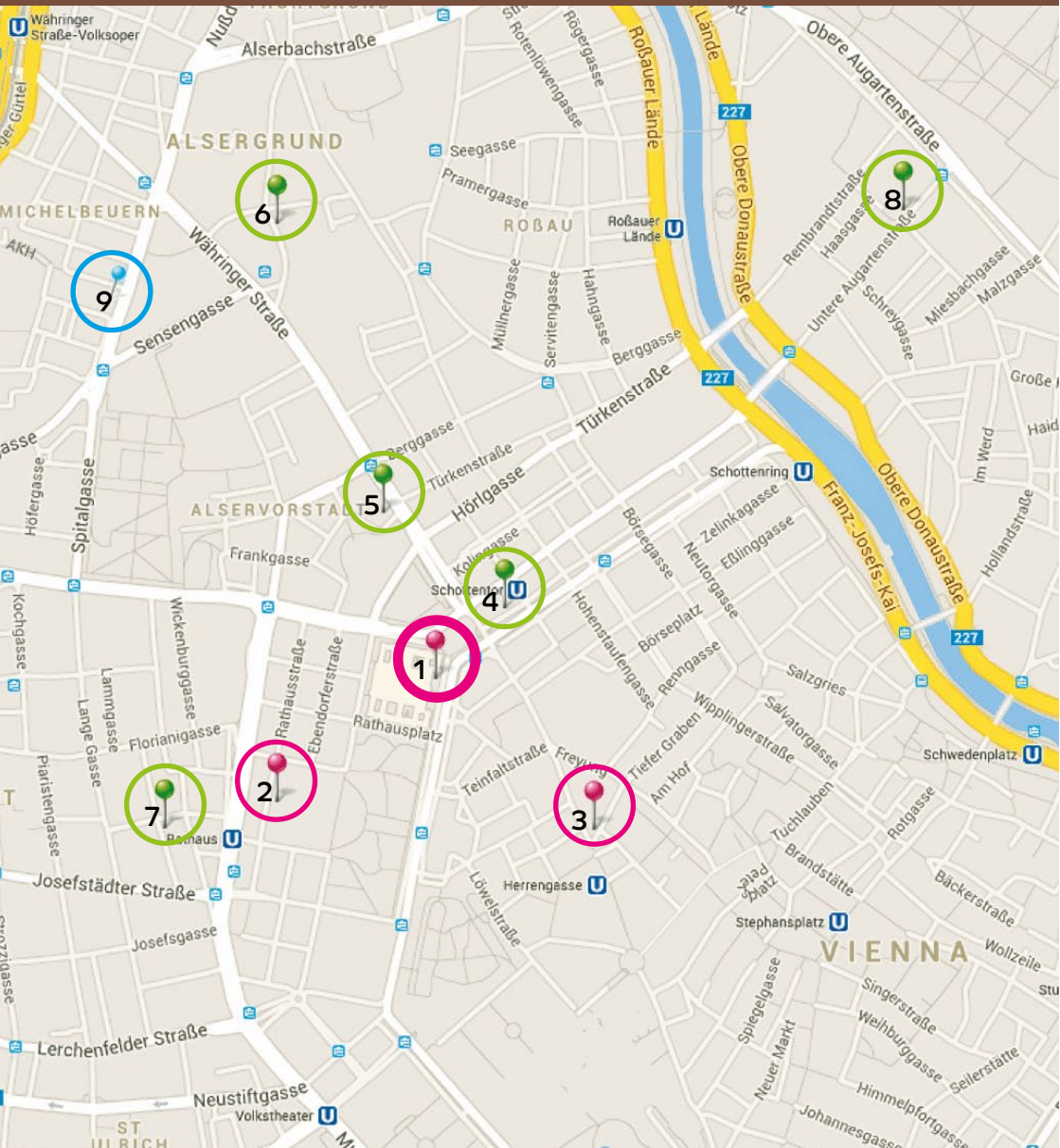
Prof. Michael Schemper
Section for Clinical Biometrics
Center for Medical Statistics, Informatics, and Intelligent Systems
Medical University of Vienna
Spitalgasse 23
1090 Vienna
Austria

michael.schemper@meduniwien.ac.at

Editorial deadline: 23 July 2014

Map & Locations

1. University of Vienna (Conference Venue)
2. Vienna City Hall (Welcome Reception)
3. Palais Ferstel (Conference Dinner)
4. Hotel de France
5. Hotel Regina
6. Hotel Boltzmann
7. Hotel Alpha
8. ÖJAB Haus Niederösterreich
9. Medical University of Vienna



www.iscb2014.info

International Society for Clinical Biostatistics 35th Annual Conference

ABSTRACT BOOK

24 - 28 August 2014
Vienna, Austria

ISCB35
Vienna 2014

Abstract Book

ISCB 2014 • 35th Annual Conference of the
International Society for Clinical Biostatistics
24 - 28 August 2014 • Vienna, Austria
www.iscb2014.info

ISCB35
TABLE OF CONTENTS

Programme Overview 6

Oral Sessions

Sunday, 24th August 2014

Pre-conference Courses:

Conference courses - Full-day	Course 1 - Course 2	9
Conference courses - Half-day – Morning	Course 3 - Course 4	9
Conference courses - Half-day – Afternoon	Course 5 - Course 6	10

Monday, 25th August 2014

9:00-10:30

Invited session:

I1 On trial: integrated care pathways, evidence based medicine and EHRs	I1.1 - I1.3	11

Contributed sessions:

C01 Randomized clinical trials	C01.1 - C01.5	12
C02 Missing data	C02.1 - C02.5	13
C03 Regression modelling in epidemiology	C03.1 - C03.5	14
C04 Dose finding studies	C04.1 - C04.5	15
C05 High-dimensional data analyses I	C05.1 - C05.5	17
C06 The biostatistician's toolbox I	C06.1 - C06.5	18

11:30-12:30

IP President's invited speaker	IP.1	20

14:00-15:30

Invited session:

I2 Beyond R packages: getting our methods into standard software	I2.1 - I2.3	21

Contributed sessions:

C07 Clinical trials	C07.1 - C07.5	22
C08 Prediction models: case studies	C08.1 - C08.5	23
C09 Latent variable methods	C09.1 - C09.5	24
C10 Genome-wide association studies	C10.1 - C10.5	25
C11 Functional data analysis	C11.1 - C11.5	27
C12 Vaccine studies and infectious diseases	C12.1 - C12.5	28

16:00-17:30

Invited session:

I3 Inverse probability weighting techniques	I3.1 - I3.3	30

Contributed sessions:

C13 Personalized and stratified medicine I	C13.1 - C13.5	31
C14 Meta-analysis	C14.1 - C14.5	32
C15 Longitudinal data analysis I	C15.1 - C15.5	33
C16 High-dimensional data analysis II	C16.1 - C16.5	34
C17 Adaptive designs I	C17.1 - C17.5	36
C18 Binary and count data analysis	C18.1 - C18.5	37

Tuesday 26th August 2014

9:00-10:30

Invited session:

I4 New methods to control for unmeasured confounding	I4.1 - I4.3	39

Contributed sessions:

C19 Development of prediction models	C19.1 - C19.5	40
C20 Individual participant data meta-analysis	C20.1 - C20.5	41
C21 Survival analysis and competing risks	C21.1 - C21.5	42
C22 Surrogate and composite endpoints	C22.1 - C22.5	44
C23 Design and analysis of clustered studies	C23.1 - C23.5	45
C24 Group-sequential designs	C24.1 - C24.5	46

11:00-12:30

Invited session

I5 Prediction to support clinical decision making	I5.1 - I5.3	48

Contributed sessions:

C25 Personalized and stratified medicine II	C25.1 - C25.5	49
C26 Network meta-analysis	C26.1 - C26.5	50
C27 Survival analysis I	C27.1 - C27.5	51
C28 Marginal structural models	C28.1 - C28.5	53
C29 Clinical trial designs	C29.1 - C29.5	54
C30 Adaptive designs II	C30.1 - C30.5	55

Wednesday, 27th August 2014

9:00-11:00

Invited session

S1 STRATOS (Strengthening Thinking about Analyses of Observational Studies) initiative: first results & future steps	S1.1 - S1.4	57

9:00-10:48

Contributed sessions:

C31 Variable selection in high-dimensional models	C31.1 - C31.6	59
C32 Longitudinal data analysis II	C32.1 - C32.5	60
C33 Relative and net survival	C33.1 - C33.6	61
C34 Methodology	C34.1 - C34.6	63
C35 The biostatistician's toolbox II	C35.1 - C35.6	65
C36 Issues in multiple testing	C36.1 - C36.6	66

14:00-15:30

Invited session:

S2 The power of data sharing: advancing research for everyone's benefit? (Panel discussion)	S2.1	68

Contributed sessions:

C37 Causal inference from observational studies I	C37.1 - C37.5	68
C38 Patient-centered outcomes	C38.1 - C38.5	70
C39 Multistate models and competing risks I	C39.1 - C39.5	71
C40 Model performance evaluation	C40.1 - C40.5	72
C41 Survival analysis II	C41.1 - C41.5	73
C42 Poly-omics studies & Systems Biology	C42.1 - C42.5	75

16:00-17:30

Invited session:

I6 Statistical methods for poly-omics studies	I6.1 - I6.3	77

Contributed sessions:

C43 Causal inference from observational data II	C43.1 - C43.4	78
C44 Validation of prediction models	C44.1 - C44.4	79
C45 Multistate models and competing risks II	C45.1 - C45.5	80
C46 Multiple imputation	C46.1 - C46.5	81
C47 Special types of censored data	C47.1 - C47.4	82
C48 Drug development	C48.1 - C48.5	83

Thursday 28 August 2014

9:00-12:30

Minisymposia:

M1 Statistical challenges in the epidemiology of aging	M1.1 - M1.4	86
M2 Genomics-based Personalized Medicine	M2.1 - M2.4	87

Conference course:

Post-conference Course	Course 7	88

Poster Sessions

Monday, 25th August 2014

15:30-16:00

Poster session P1:

P1.1 Bayesian methods in biostatistics	P1.1.1 - P1.1.164	89
P1.2 Design and analysis of clinical trials	P1.2.5 - P1.2.195	92
P1.3 Pharmacoepidemiology and drug development	P1.3.51 - P1.3.136	100

Tuesday, 26th August 2014

10.30-11.00

Poster session P2:

P2.1 Longitudinal data analysis	P2.1.54 - P2.1.161	102
P2.2 Methodological issues and case studies in epidemiology	P2.2.16 - P2.2.185	104
P2.3 Methods for handling missing data	P2.3.2 - P2.3.196	109
P2.4 Penalized methods in high- and in low-dimensional regression analyses	P2.4.23 - P2.4.56	111
P2.5 Statistical methods for systems biology and genetics	P2.5.40 - P2.5.194	111
P2.6 Software aspects of efficient statistical analyses	P2.6.26 - P2.6.73	113

Wednesday, 27th August 2014

11.00-11.30

Poster session P3:

P3.1 Survival analysis, multistate models and competing risks	P3.1.19 - P3.1.191	114
P3.2 Diagnostic studies	P3.2.91 - P3.2.154	118
P3.3 Analysis of electronic health records	P3.3.49 - P3.3.152	120
P3.4 Comparative effectiveness and outcomes research	P3.4.25a - P3.4.181	121
P3.5 Development and validation of clinical prediction models	P3.5.3 - P3.5.175	123

15.30-16.00

Poster session P4:

P4.1 Meta-analysis and network meta-analysis	P4.1.4 - P4.1.190	128
P4.2 Modeling infectious diseases	P4.2.24 - P4.2.102	131
P4.3 Novel designs and methods for simulations	P4.3.13 - P4.3.70	132
P4.4 Observational studies and causal inference methods	P4.4.14 - P4.4.192	132
P4.5 Other	P4.5.6 - P4.5.187	135

Author Index 141

PROGRAMME OVERVIEW

Sunday 24 August 2014

	Hörsaal 26	Hörsaal 34	Hörsaal 16	Hörsaal 24
09:00 10:30	**Course 1**: Handling missing outcome data in clinical trials	**Course 2**: Data analysis with competing risks and multiple states	**Course 3**: Extension of frailty models for recurrent or clustered survival data with prediction	**Course 4**: Statistical methods in Systems Medicine
10:30 11:00	Coffee break			
11:00 12:30	continued	continued	continued	continued
12:30 14:00	Lunch			
14:00 15:30	continued	continued	**Course 6**: Interaction analysis	**Course 5**: Data and Safety Monitoring Board workshop
15:30 16:00	Coffee break			
16:00 17:30	continued	continued	continued	continued

Monday 25 August 2014

	Auditorium Max	Hörsaal 7	Hörsaal 21	Hörsaal 41	Hörsaal 34	Hörsaal 16	Hörsaal 31
09:00 10:30	**I1**: On trial: integrated care pathways, EBM and EHR	**C01**: Randomized clinical trials	**C02**: Missing data	**C03**: Regression modelling in epidemiology	**C04**: Dose finding studies	**C05**: High-dimensional data analyses I	**C06**: The biostatistician's toolbox I
10:30 11:00	Coffee break						
11:00 12:30	Welcome address **IP**: President's invited speaker: Tom Louis						
12:30 14:00	Lunch						
14:00 15:30	**I2**: Beyond R packages	**C07**: Clinical trials	**C08**: Prediction models: case studies	**C09**: Latent variable methods	**C10**: Genome-wide association studies	**C11**: Functional data analysis	**C12**: Vaccine studies and infectious diseases
15:30 16:00	Coffee break & **P1** Poster session						
16:00 17:30	**I3**: Inverse probability weighting techniques	**C13**: Personalized and stratified medicine I	**C14**: Meta-analysis	**C15**: Longitudinal data analysis I	**C16**: High-dimensional data analysis II	**C17**: Adaptive designs I	**C18**: Binary and count data analysis
19:00	Welcome Reception City Hall						

Tuesday 26 August 2014

	Auditorium Max	Hörsaal 7	Hörsaal 21	Hörsaal 41	Hörsaal 34	Hörsaal 16	Hörsaal 31
09:00 10:30	**I4**: New methods to control for unmeasured confounding	**C19**: Development of prediction models	**C20**: Individual patient data meta-analysis	**C21**: Survival analysis and competing risks	**C22**: Surrogate and composite endpoints	**C23**: Design and analysis of clustered studies	**C24**: Group-sequential designs
10:30 11:00	Coffee break & **P2** Poster session						
11:00 12:30	**I5**: Prediction to support clinical decision making	**C25**: Personalized and stratified medicine II	**C26**: Network meta-analysis	**C27**: Survival analysis I	**C28**: Marginal structural models	**C29**: Clinical trial designs	**C30**: Adaptive designs II
12:30 14:00	Lunch						
14:00	Conference Tours						
19:00	Young Researchers Lounge "Heuriger" in Nußdorf						

Wednesday 27 August 2014

	Auditorium Max	Hörsaal 7	Hörsaal 21	Hörsaal 41	Hörsaal 34	Hörsaal 16	Hörsaal 31
09:00 11:00	**S1**: STRATOS initiative: first results & future steps	**C31**: Variable selection in high-dimensional models	**C32**: Longitudinal data analysis II	**C33**: Relative and net survival	**C34**: Methodology	**C35**: The biostatistician's toolbox II	**C36**: Issues in multiple testing
11:00 11:30	Coffee break & **P3** Poster session						
11:30 12:30	Annual General Meeting						
12:30 14:00	Lunch					12:35 - 13:15 STRATOS public meeting (Elise-Richter-Saal)	
14:00 15:30	**S2**: The power of data sharing (Panel discussion)	**C37**: Causal inference from observational studies I	**C38**: Patient-centered outcomes	**C39**: Multistate models and competing risks I	**C40**: Model performance evaluation	**C41**: Survival analysis II	**C42**: Poly-omics studies & Systems Biology
15:30 16:00	Coffee break & **P4** Poster session						
16:00 17:30	**I6**: Statistical methods for poly-omics studies	**C43**: Causal inference from observational data II	**C44**: Validation of prediction models	**C45**: Multistate models and competing risks II	**C46**: Multiple imputation	**C47**: Special types of censored data	**C48**: Drug development
19:00	Conference Dinner Palais Ferstel						

Thursday 28 August 2014

	Hörsaal 7	Hörsaal 21	Hörsaal 24
09:00 10:30	**Mini-Symposium 1**: Statistical challenges in the epidemiology of aging	**Mini-Symposium 2**: Genomics-based personalized medicine (joint with the IGES)	**Course 7**: Designing adaptive clinical trials
10:30 11:00	Coffee break		
11:00 12:30	MS1 continued	MS2 continued	continued
	IGES Conference (Imperial Riding School Renaissance Hotel)		

Sunday, 24th August 2014 – Pre-conference Courses - Full-day

Conference courses

Course 1

Handling missing outcome data in clinical trials

I White[1], S Seaman[1]
[1]*MRC Biostatistics Unit, Cambridge, United Kingdom*

This course aims to provide practising statisticians with the necessary practical skills to handle missing data in their analyses, and in particular to move beyond the use of complete-case analysis and last observation carried forward analysis.
Four sessions will each consist of a lecture followed by a short discussion exercise:
1. An introduction to missing data in randomised trials: popular ways to analyse trials with missing data, focussing on their assumptions; the intention-to-treat principle.
2. Mixed models analysis of incomplete data: how it should be implemented in randomised trials, including issues arising from missing baseline data.
3. Multiple imputation: a brief introduction, and how to use it in randomised trials. 4. Sensitivity analysis to departures from assumptions: principled sensitivity analysis, and a suggestion of how to implement it.
We focus on trials with quantitative outcomes, and also consider binary outcomes but not time-to-event outcomes.

Course 2

Data analysis with competing risks and multiple states

R Geskus[1], H Putter[2]
[1]*Academic Medical Center, Amsterdam, The Netherlands,* [2]*Leiden University Medical Center, Leiden, The Netherlands*

In the end we all die, but not all from the same cause, nor with the same life histories. This course will be devoted to the analysis of different types of events that can occur either exclusively (competing risks) or sequentially (multi-state models).
The morning session is devoted to competing risks analysis. Competing risks take the spectrum of event types into account. The main difficulty is the choice of the correct quantity to be estimated. When do we need a competing risks analysis? Do we want to estimate cause-specific or subdistribution hazards? What do we need to assume with respect to the censoring by the competing event? The actual analysis is much easier because software to perform the computations has become readily available.
Multi-state models are the topic of the afternoon session. They extend competing risks models to the analysis of what happens beyond some first event, by allowing individuals to progress through different states. Estimation in multi-state models is reasonably straightforward, at least as long as all transitions are observed without uncertainty. Arguably the most interesting use of multi-state modles is for dynamic prediction. This aspect will be discussed in some detail; dynamic prediction using multi-state models will be contrasted with more recent developments such as landmarking.

Sunday, 24th August 2014 – Pre-conference Courses - Half-day - Morning

Conference courses

Course 3

Extension of frailty models for recurrent or clustered survival data with prediction

V Rondeau[1]
[1]*Department of Biostatistics at the National Health and Medical Research Institute of the University Bordeaux Segalen, Bordeaux, France*

Simple shared frailty models have been largely developed and applied for recurrent or clustered survival data in the literature. However, extensions of frailty models are less common in publications and are not well represented in classical software. We are aiming at filling this gap by considering extensions of frailty models (such as additive frailty models, nested frailty models or joint frailty
models) and by presenting an implementation of these models using the R package **frailtypack**. Particular interest will be given to joint frailty models in order to jointly analyse recurrent events such as cancer relapses and a dependent terminal event (death or loss to follow-up). Prediction tools associated with this package will be presented, too.
The first part of this course will introduce general frailty models, the estimation methods and the research questions they may address.
The second part of this course will be dedicated to the joint frailty models with illustration on real data. The estimation and the predictive dynamic tools that can be derived from them will be exposed, with methods to evaluate their performance.
Emphasis is given, via examples on real data, of the ability of extended frailty models to describe a very broad range of practical situations. Each concept will be illustrated through implementation of these models using the R package **frailtypack**.

Course 4

Statistical methods in Systems Medicine

H Fröhlich[1]
[1]*University of Bonn, Bonn, Germany*

Discovery of prognostic and diagnostic biomarker signatures for diseases, such as cancer, is seen as a major step towards a better personalized medicine. During the last decade various methods have been proposed for inferring such signatures from high dimensional molecular data (e.g. genomics, transcriptomics, proteomics and metabolomics profiles). However, one important obstacle for making molecular signatures a standard tool in clinical diagnosis is the typical low reproducibility of these signatures combined with the difficulty to achieve a clear biological interpretation. For that purpose in the last years there has been a growing interest in approaches that employ biological background knowledge. In addition, the increasing availability of different -omics profiles for the same patient now raises the question on how to integrate these data.
The purpose of this course is to shed light on current integrative modeling efforts that combine different -omics entities and/or biological background knowledge in order to achieve higher robustness, stability and interpretability of molecular biomarker signatures.

Sunday, 24th August 2014 – Pre-conference Courses - Half-day - Afternoon

Conference courses

Course 5

Data and Safety Monitoring Board workshop

T Jaki[1], L Hampson[1]
[1]*Lancaster University, Lancaster, United Kingdom*

Data and Safety Monitoring Boards (DSMBs) are a common feature of long-term clinical studies in serious and life-threatening conditions. This workshop describes the remit and composition of DSMBs, and how their work relates to other parties involved in the study, such as the sponsor, the study project team, the investigators, the Steering Committee and the data management centre. The importance of pre-trial preparation by the DSMB is stressed. Consideration is given to the nature and purpose of safety and efficacy data reports presented to the DSMB, and the balance between the timeliness and the accuracy of the data available is discussed. Statistical problems inherent in repeatedly making multiple treatment comparisons are highlighted, and formal stopping guidelines based on repeated safety analyses are presented. The role of the DSMB in trials with pre-specified interim efficacy analyses will be discussed.
The workshop is structured around group discussions in which participants will play the roles of DSMB members and will discuss realistic trial reports of interim safety and efficacy.

Course 6

Interaction analysis

T VanderWeele[1]
[1]*Departments of Epidemiology and Biostatistics at the Harvard School of Public Health, Boston, United States*

This workshop will provide a relatively broad introduction to the topic of interaction between exposures. We discuss interaction on both additive and multiplicative scales using risks, and we discuss their relation to statistical models (e.g. linear, log-linear and logistic models). We discuss and evaluate arguments that have been made for using additive or multiplicative scales to assess interaction.
We describe inferential procedures for interaction when logistic models are fit to data but when additive and not just multiplicative measures of interaction are desired. We discuss issues of confounding for interaction analyses and how whether control has been made for only one or both of two exposures affects whether interaction estimates can be interpreted as causal interaction between the two exposures or only as effect heterogeneity.
We further discuss conditions under which interaction gives evidence of synergism within the sufficient cause framework, when interaction is robust to unmeasured confounding, interaction for time-to-event outcomes, case-only estimators of interaction, and power and sample size calculations for additive and multiplicative interaction. Illustrations will be given from environmental, genetic, and infectious disease epidemiology. Software code will be provided.

Monday, 25th August 2014 – 9:00-10:30

Invited session

I1 On trial: integrated care pathways, evidence based medicine and EHRs

Organizer: Els Goetghebeur

I1.1
Integrated care pathways, evidenced based medicine and electronic health records: an overview

MJ Campbell[1]
[1]*ScHARR, University of Sheffield, Sheffield, United Kingdom*

Integrated care pathways are structured multidisciplinary care plans which detail essential steps in the care of patients with a specific clinical problem. They have been proposed as a way of encouraging the translation of national guidelines into local protocols and their subsequent application to clinical practice.
However, guidelines may be *eminence* based rather than *evidence based,* and may be biased or wrong. One of the barriers to integrated care has been the problem of assessing exactly what treatments patients actually receive when their care is split between various providers, particularly between primary and secondary care. In the UK, a new electronic health record (EHR) *care.data* is being set up to try and provide linked information about the care received from all of the different parts of the health service, including hospitals and general practitioners. This has had numerous teething problems, particularly with failure to reassure patients about confidentiality. Existing general EHRs in the UK include the Hospital Episodes Statistics (HES) and the Clinical Practice Research Database (CPRD), and there are a number of disease specific ones, including the DAFNE database for patients with Type 1 diabetes.
This talk will summarise the evidence for the efficacy of integrated care pathways, and describe some of the worst excesses of clinical guidelines. The use of HES and the DAFNE data base in particular for evaluating hospitals and care of patients with Type 1 diabetes respectively will be discussed. The prospects of research with *care.data* will be touched upon.

I1.2
From idealized to realized: learning about dynamic treatment regimens using electronic medical records?

EEM Moodie[1], DA Stephens[1]
[1]*McGill University, Montreal, Canada*

Due to the cost and complexity of conducting a sequential multiple assignment randomized trial, learning about optimal treatment from non-experimental data is highly attractive -- particularly with the use of routinely collected data on a large number of people.
However the use of electronic medical records presents several challenges. For example, visit times are patient-driven and are frequently related to outcomes. Simulations can therefore provide an important first step into estimating optimal rules, and understanding possible biases that may arise in an electronic medical record setting.
In this presentation, I will present a simulation design for a complex, continuous dosing problem, and discuss ongoing work in which we relax idealized assumptions and move towards more realistic scenarios.

I1.3
A causal inference approach to optimising care pathways in type 2 diabetes using electronic health records

RA Emsley[1]
[1]*Centre for Biostatistics, The University of Manchester, Manchester, United Kingdom*

Electronic health records such as CPRD provide a rich data source for identifying personalised care pathways in conditions such as diabetes. Marginal structural models can be used to estimate these optimal dynamic treatment regimes, but one significant challenge to their application is that measurement times in CPRD are largely patient-driven, and so follow-up times vary in frequency and duration by each patient, and are related to the outcome being considered.
In the UK, the prevalence of diabetes has increased dramatically and affects approximately 4.6% of the population; in 2012, 2.6 million people were newly diagnosed with type 2 diabetes mellitus. Many of these will develop serious disease-related outcomes and microvascular complications can serve as an early marker for later complications. Diabetes management is based on NICE (National Institute for Health and Care Excellence) guidelines based on repeated glycated haemoglobin (HbA_1c) values.
We wish to personalise the treatment pathway for patients dependent on their characteristics, allowing for the variable measurement occasions. We will illustrate the methods with an inception cohort of 41,825 newly diagnosed type 2 diabetes patients initiating an antidiabetic medication (ADM) within CPRD. Of these, over 12,000 proceed onto dual therapy regimens and 3,500 develop microvascular complications. We describe the patterns and choice of the first ADM, consider its effect on HbA1c levels and identify the optimal choice of dual therapy to control hbA_1c and delay onset of microvascular complications.
This work is part of the MRC Health eResearch Centre of the Farr Institute for Health Informatics Research.

Contributed sessions

C01 Randomized clinical trials

C01.1
Analysis of clinical trials requiring rescue medication

GK Rosenkranz[1]
[1]*Novartis Pharma AG, Basel, Switzerland*

Clinical trials in some indications like asthma or diabetes require to administer rescue medication in case a patient does not sufficiently respond to investigational treatment. The application of additional treatment on an as needed basis causes problems to the analysis and interpretation of the results of these studies since the effect of the drug under study can be confounded by the additional medication. Following-up all patients until study end and capturing all data are not fully addressing the issue.
The paper introduces an analysis that takes care of the fact that rescue is a study outcome and not a covariate by considering potential outcomes. This approach allows to clearly define a causal estimand. For normally distributed longitudinal data a practically unbiased causal effect estimator for the randomized treatment can be obtained. The results are compared to those from an ITT analysis and an analysis on all patients not receiving rescue.

C01.2
Assessment of chronological bias in randomized clinical trials

M Tamm[1]
[1]*RWTH Aachen University, Aachen, Germany*

In clinical trials patients are usually recruited sequentially over time. Often the recruitment takes place over several years. Especially in clinical trials studying rare diseases, low accrual rates and thus long recruitment phases are common. As a result of the prolonged recruitment time, time trends are suspected to occur. This can be due to several reasons, such as changes in the recruitment policy or learning effects in the application of the new methods. If treatment effects are confounded with time trends, this can result in the so called chronological bias.
Even in randomized clinical trials, time trends may impact the results, for instance if a long series of consecutive patients are assigned to the same treatment. To account for this, the ICH E9 guideline recommends the use of randomization in blocks. However, one major drawback of permuted block randomization with short blocks is the increase in the risk of selection bias.
Using different time trend models, we evaluate and compare the extent of chronological bias under the random allocation rule and the permuted block randomization with different block sizes. We present theoretical results regarding the extent of bias in the treatment effect estimate under different time trends. Further, the bias in statistical hypothesis testing is discussed considering the empirical type I error rate in simulations. An assessment of worst-case-scenarios as well as an overall assessment considering the choice of randomization is given.
Acknowledgement: This research is part of the IDeAl EU-FP7 project, Grant-Agreement No. 602 552.

C01.3
An analytic framework for randomised trial designs that take patient preferences into account

S Walter[1]
[1]*McMaster University, Hamilton, Canada*

Patients in clinical trials often prefer one of the treatments under investigation, and such preferences (which are usually unmeasured) may have affect study outcomes. Several trial designs have been proposed to investigate the impacts of patient preferences, something that is not possible in the usual parallel group design.
We will describe a model framework to represent the effects of treatment preferences on study outcomes. In particular, we consider the selection effect, which measures how expected outcomes differ between participants who would select one treatment or the other, if they were free to do so. Additionally we can investigate the difference in outcomes for participants who do or do not receive their preferred treatment, which we designate as a preference effect.
We will review several alternative trial designs including: (1) the two-stage design, in which some randomly sampled participants are allowed to choose their treatment; (2) the fully randomised preference design, where preferences are known for all participants, but treatments are nevertheless always randomised; and (3) the partially randomised preference design, where only those participants who are indifferent between treatments are randomised. Based on the model framework, we then determine which effects are estimable with each design. Examples of each of these designs can be found in the medical literature.
Preference designs offer potentially greater insight than the conventional parallel group design, by informing investigators about potentially important effects of preferences on patient outcomes, effects which may sometimes exceed the direct effect of treatment itself.

C01.4
Covariate adjustment has similar benefits in small and large randomised controlled trials

DD Thompson[1], HF Lingsma[2], EW Steyerberg[2]
[1]*University of Edinburgh, Edinburgh, United Kingdom,* [2]*Erasmus Medical Center, Rotterdam, The Netherlands*

Aim: Covariate adjustment is a standard statistical approach in the analysis of randomised controlled trials, but the benefit in small versus larger studies is not well known. Specifically, chance imbalance in prognostic factors is greater in small trials than in large. We aimed to determine whether the benefit of covariate adjustment on statistical significance and power differed between small and large trials.
Methods: We repeatedly (500,000 times) drew random samples of sizes 300 and 5000 per arm from a large trial (GUSTO-I, N=30,510) and simulated 30-day mortality using the control arm. We empirically determined the treatment effects required to fix power at 80% for the unadjusted analyses and calculated the joint probabilities in the discordant cells (i.e., $p<0.05$ vs. $p\geq0.05$) when cross-classifying adjusted and unadjusted results from logistic regression models.
Results: The power gained from an adjusted analysis for small (OR=0.27 for simulated treatment effect) and large (OR=0.77) samples was approximately 5% (from 80% to 85%). Similar discordance irrespective of sample size was noted, with a 2% change from significant unadjusted to non-significant adjusted treatment effect and a 7% change from non-significant unadjusted to significant adjusted. The Type I error was close to 5% with unadjusted and adjusted analyses, with discordance balanced at 2% for small and large trials.
Conclusions: The proportions of change in statistical significance from covariate adjustment were the same for small and large trials with similar gains in statistical power to detect treatment effects. Covariate adjustment is hence equally recommendable in small and large trials.

C01.5
Long-term evaluation of different designs of a series of phase III clinical trial in rare cancers: a simulation study

M-A Bayar[1], G Le Teuff[1], S Michiels[1], D Sargent[2], M-C Le Deley[1]
[1]Institut Gustave Roussy, Department of Biostatistics and Epidemiology, Villejuif, France, [2]Mayo Clinic, Rochester, United States

Background: In rare diseases, randomized trials designed with typical 5% α-level and 80%-power are unfeasible. A previous simulation study by Le Deley et al. suggested performing a series of small trials with relaxed α-levels; the treatment effect of each new treatment was characterized by the hazard ratio drawn from a certain distribution, yielding benefit accumulation trial after trial, which is a questionable hypothesis. We reviewed and extended this simulation framework.
Methods: We simulated a series of two-treatment superiority trials over a 15-year research period. Trial parameters examined included the α-level and the number of trials run over the 15-year period (thus, the trial sample size). Each design was evaluated for different disease scenarios and accrual rates. In our simulation study, the treatment effect of each new treatment was defined by the hazard rate, avoiding benefit accumulation. Different assumptions of how treatments improve over time were considered.
Results: Our simulation study shows that performing a series of small trials with relaxed α-levels leads on average to larger survival gains over a long research horizon than performing larger trials with typical 5% α-level while controlling the risk of selecting a worse treatment at the end of the research period. The survival gains were substantially lower than those shown in the previous work but the pattern is similar and recommendations remain valid.
Conclusion: Designs aiming to maximize the expected survival gain over a long research horizon across a series of trials are worth discussing in the context of rare diseases.

C02 Missing data

C02.1
A comparison of multiple imputation methods for hierarchical data when there is whole cluster non-response

K DiazOrdaz[1], MG Kenward[1], M Gomes[1], R Grieve[1]
[1]LSHTM, London, United Kingdom

Background: Missing data are common in cluster randomised trials and need to be handled appropriately.
When using multiple imputation (MI), we must recognise the data structure.
Methods: We conducted a simulation study to compare complete-case analysis, multilevel MI, fixed-effects MI, and single-level MI, when the analysis model is a linear mixed-model.
Clustered data, consisting of normally-distributed outcomes with baseline individual and cluster-level covariates, were simulated with different levels of clustering, measured by the intra-cluster correlation coefficient (ICCs), number and size of clusters and probability of non-response.
Missing data were introduced according to missing-at-random mechanisms driven by: (i) an individual-level covariate, (ii) cluster-level covariate, (iii) treatment, (iv) cluster size and included cluster non-response.
For each missing data scenario, 1000 sets were simulated. Empirical bias of the treatment effect and confidence interval (CI) coverage were obtained for each method.
Results: Complete-case analyses resulted in biased effect estimates when treatment was associated with missingness, while multilevel MI produced estimators with negligible bias across all the missingness mechanisms considered.
Fixed-effects MI over-estimated the variance resulting in CI coverage in excess of nominal levels (up to 99.8%), especially for datasets with low ICC. Moreover, it resulted in biased estimates for scenarios where a non-negligible number of clusters were missing and a cluster-level variable was associated with the missingness mechanism.
Conclusion: The validity of inferences may depend on how clustering is accommodated in the imputation step. Multilevel MI performed well across all settings considered and is theoretically appropriate for studies that have a hierarchical design.

C02.2
Pattern mixture models applied to clinical trials for chronic pain

G Thoemmes[1], J Abellan-Andres[1], M Janssens[2]
[1]Grunenthal GmbH, Aachen, Germany, [2]Hasselt University, Hasselt, Belgium

Missing data are a common problem in chronic pain trials due to premature termination of subjects experiencing side effects or having insufficient analgesia over the course of a long-term trial. Pattern mixture models have shown to provide a flexible and transparent framework for handling missing data. In practice they can be implemented using multiple imputation. Over the last years the combination of pattern mixture models and multiple imputation has been recognized as one of the most promising approaches for analyzing trials with missing values. A number of pattern mixture models will be presented and some notes regarding implementation details will be given. A case study based on a pooled dataset of several three-arm, parallel-group, active and placebo controlled chronic pain trials will presented. The different pattern mixture models will be applied to this dataset and compared with a direct likelihood analysis using a mixed model for repeated measures (MMRM), and also with well-known single imputation methods, e.g. LOCF and BOCF.

C02.3 *Conference Award for Scientists*
CD4+ counts in a 3-arm longitudinal clinical trial with substantial missing data: a sensitivity analysis

AC Grobler[1], G Matthews[2], G Molenberghs[3]
[1]CAPRISA, Durban, South Africa, [2]University of KwaZulu Natal, Durban, South Africa, [3]I-BioStat, Universiteit Hasselt and KU Leuven, Hasselt, Belgium

The SAPiT trial was an open label, randomised controlled trial in HIV-tuberculosis co-infected patients. Patients were randomised to three arms; each initiating antiretroviral therapy at a different time. CD4 count was measured 6-monthly for 24 months. The assumption that missing data are missing completely at random (MCAR) was not supported by the observed data. We performed a range of sensitivity analyses under both missing at random (MAR) and missing not at random (MNAR) assumptions.
Under MAR assumptions Bayesian analysis, multiple imputation and maximum likelihood analyses (mixed model repeated measures with arm, time and the interaction between arm and time) were done.
Under MNAR assumptions several pattern mixture models (PMMs) were fitted. These included analysis using the CCMV and NCMV identifying restrictions, PMMs using random effects mixed models and PMMs using multiple imputation. Selection models were fitted using Bayesian methods. All these methods are based on different unobserved assumptions about the missing data processes; thus the importance of a sensitivity analysis.
More than one third of participants were lost to follow-up. Results are given and contrasted from all these methods allowing conclusions taking the missing data into account. MAR analyses showed a larger difference between the treatment arms than the MNAR analyses. The conclusion is that mean CD4+ count increased more in the early and late integrated

treatment arm than in the sequential treatment arm. This is an encouraging finding, since it means that we can recommend integrated treatment, which reduces mortality, without the risk of worsening HIV outcomes.

C02.4
Proposition of a multiple imputation approach for MNAR mechanism using Heckman's model

J-E Galimard[1], M Resche-Rigon[1], S Chevret[1]
[1]INSERM UMR 1153, Université Paris 7, Paris, France

Standard implementations of multiple imputation (MI) approaches provide unbiased inferences assuming underlying MAR mechanisms. However, in presence of missing data generated though MNAR mechanisms, although the MAR assumptions can be approached by collecting more explanatory variables, MI is not satisfactory and difficult to validate. Coming from econometric statistics, the Heckman's method, also called the sample selection method, deals with selected sample using two joined linear equations, namely the *selection equation* and the *outcome equation,* respectively. The Heckman´s method has been successfully applied to missing outcomes in presence of MNAR mechanism. Nevertheless, such a method deals with missing outcomes only, and this is a strong limitation in clinical epidemiology settings where covariates are also often missing.
We propose to extend the validity of MI to MNAR mechanisms by using the Heckman´s model as the imputation model, using a two-step estimation process. This will provide a solution that can be used in a MI by chained equation framework to impute missing variables (outcomes or covariates) resulting either from a MNAR or a MAR mechanism.
This approach will be validated by a simulation study. We will first consider MNAR missing outcomes, and evaluate the robustness of the approach in case of model miss-specification. Then it will be evaluated in presence of missing outcome and missing covariates either under MNAR or MAR mechanisms.

C02.5
Allowing for nonignorable missingness in HIV status using multiple imputation with delta-adjustment: applications to causal mediation analysis and prevalence estimation

FP Leacy[1,2], IR White[1], TA Yates[3], S Floyd[4], HM Ayles[4,5], P Godfrey-Faussett[4]
[1]MRC Biostatistics Unit, Cambridge, United Kingdom, [2]University of Cambridge, Cambridge, United Kingdom, [3]University College London, London, United Kingdom, [4]London School of Hygiene & Tropical Medicine, London, United Kingdom, [5]ZAMBART Project, University of Zambia, Lusaka, Zambia

Aims: The primary goal of this analysis is to assess the mediating influence of HIV status on the relationship between socioeconomic status and prevalent tuberculosis in Zambia, allowing for nonignorable missingness in the HIV test result variable. Anticipated sensitivity of HIV prevalence estimates to subgroup-specific departures from MAR will also be investigated.
Methods: Utilising data from the Zambia South Africa Tuberculosis and AIDS Reduction (ZAMSTAR) trial, we first present results from a complete case causal mediation analysis. We then analyse the data using multiple imputation under MAR, comparing results from two imputation models that differ only with regard to inclusion of information on self-reported HIV status. We proceed to perform a sensitivity analysis to departures from MAR using multiple imputation with delta-adjustment. We extend the standard procedure by allowing the value of the adjustment, δ, to vary across subgroups defined by participants' self-reported HIV status.
Results: While estimates of conditional natural direct effects, conditional natural indirect effects and conditional total effects on the odds ratio scale exhibited significant sensitivity across the investigated δ range, estimates of the corresponding marginal quantities were relatively insensitive to departures from MAR. Subtle variations in HIV prevalence estimates by gender were observed across the range of missingness mechanisms considered.
Conclusion: Multiple imputation with delta-adjustment offers a transparent and flexible means to allow for nonignorable missingness in HIV status. This method may represent a particularly important tool for sensitivity analysis in contexts such as mediation analysis where multiple subcomponent models must be fitted to the data.

C03 Regression modelling in epidemiology

C03.1
Interpretation of linear regression coefficients under mean model misspecifications

W Brannath[1], M Scharpenberg[1]
[1]University of Bremen, Faculty 3, Bremen, Germany

Linear regression is an important and frequently used tool in medical and epidemiological research. However, its validity and interpretability relies on strong model assumptions. While robust estimates of the coefficients' covariance matrix extends the validity of hypothesis tests and confidence intervals, a clear and simple interpretation of the regression coefficients is lacking when the mean structure of the model is misspecified. To overcome this deficiency, we suggest a new mathematical rigorous interpretation that is independent from specific model assumptions. The idea is to quantify how much the (unconditional) mean of the dependent variable Y can be changed by changing the distribution of an independent variable X in the population. We show that with a suitable standardization of the distributional changes, the maximum change in the mean of Y is well defined and equals zero if and only if the conditional mean of Y given X is independent of X. Restriction to linear functions for the distributional changes in X provides the link to linear regression. It leads to a conservative approximation of the newly defined and generally non-linear measure of association. The conservative linear approximation can then be estimated by linear regression. We show how the new interpretation can be extended to multiple regression analysis with the goal of adjusting for confounding covariates. We illustrate the utility (and limitations) of the new interpretation and point to perspectives for new analysis strategies.

C03.2
Simulation study to assess and compare strategies for modelling two continuous covariates with a spike at zero

C Jenkner[1], E Lorenz[2], H Becher[2], W Sauerbrei[1]
[1]University Medical Center Freiburg, Freiburg, Germany, [2]University of Heidelberg, Institute of Public Health, Heidelberg, Germany

In epidemiology and clinical research, a common goal is to estimate the effect of predictors on an outcome using appropriate regression models. Such predictors often consist of an amount of individuals with a value of zero while the distribution of the remaining ones is continuous (variables with a spike at zero). Examples in epidemiology are smoking or alcohol consumption. Since the risk for a certain disease may be substantially different between unexposed and exposed individuals, it is important to allow a separate estimate for the unexposed and a continuous function for those exposed.
A strategy for the univariate case was proposed in Royston et al, Becher et al. For a logistic regression model, theoretical odds ratio functions for selected bivariate distributions were calculated (Lorenz et al., submitted). Four possible methods how to include information of the zero values in the bivariate case using fractional polynomials were proposed.

We aim to assess and compare the properties of these strategies and identify their strengths and weaknesses. First results of a simulation study for a continuous response variable in linear regression with two covariables with a spike at zero will be presented. Different aspects, such as how the distribution of zero and nonzero values influences the model fit, will be investigated.
We measure the accuracy of the data fit calculating mean squared errors and R^2 separately in 4 categories of observations determined by the spike variables.
References:
Royston et al. Stat. Med. 2010; 29: 1219-27.
Becher et al. Biom. Journal 2012;54: 4. 686-700.

C03.3
A new measure of association based on non-linear regression
M Scharpenberg[1], W Brannath[1]
[1]*University of Bremen, Bremen, Germany*

Kernel smoothers and Spline methods are popular non-linear regression techniques that often lead to a better understanding of the regression dependencies than linear regression but are mainly used for descriptive purposes only. In order to derive some type of inferential conclusion from the non-linear regression curve they are often supplemented by confidence bands. However, there is no simple way to quantify the overall association between the variables from such type of non-linear regression analysis. In this talk, we present a new non-linear measure of association (called "mean impact") which enables us to quantify the overall association between the target and independent variable when fitting a non-linear regression curve. The idea is to consider the maximum change in the population mean of the target variable when the distribution of the covariates is changed in a suitably standardized way. We show that linear and non-linear regressions provide conservative estimates for our new model-independent measure of association. Furthermore, we derive confidence intervals for the new association parameter based on normal approximations as well as bootstrap based confidence intervals. The method is illustrated with examples and investigated in a simulation study.
References:
Brannath and Scharpenberg (2014). Interpretation of linear regression coefficients under model-misspecification. In preparation.
Martin Scharpenberg (2012). A population-based approach to analyse the influence of covariates. Diploma Thesis in Mathematics. University of Bremen.

C03.4
Weighted mean impact analysis
S Burger[1], W Brannath[1]
[1]*University of Bremen, Bremen, Germany*

Linear regression analysis is a popular tool that is often applied in medicine and epidemiology. It relies on strong model assumptions that are rarely satisfied. To overcome this difficulty Brannath, Scharpenberg (2014) and Scharpenberg (2012) proposed a new population based interpretation of linear regression coefficients. The idea is to quantify how much the unconditional mean of the dependent variable Y can be changed by changing the distribution of the independent variable X, the maximum change is called "mean impact". They show that linear regression can be used to obtain conservative estimates of the mean impact and other population based association measures. This provides a clear interpretation of linear regression coefficients also under miss-specifications of the mean structure. A disadvantage of the new association measures is their dependency on the distribution of the independent variables in the specific study population. Hence, it may be difficult to compare the results between different studies with differing covariate distributions. To overcome this difficulty we develop a method to transfer the "mean impact" from one study population to another by reweighting the observations. Accordingly, we call the resulting estimates "weighted mean impact". We derive the asymptotic distribution and develop bootstrap confidence intervals for the weighted mean impact, and we illustrate the utility of the new method by examples and results from a simulation study.
References:
Brannath, Scharpenberg (2014). Interpretation of linear regression coefficients under model-misspecification. In preparation.
Martin Scharpenberg (2012). A population-based approach to analyse the influence of covariates. Diploma Thesis. University of Bremen.

C03.5
Non-parametric self controlled case series method
Y Ghebremichael-Weldeselassie[1], H Whitaker[1], P Farrington[1]
[1]*The Open University, Milton Keynes, United Kingdom*

The self-controlled case series (SCCS) method is an alternative to cohort and case control study designs used to investigate potential associations between vaccine or other drug exposures and adverse events (side effects). It requires information only on cases, individuals who have experienced the adverse event at least once, and automatically controls all fixed confounding variables that could modify the true association between exposure and adverse event. Time-varying confounders (such as age, season), on the other hand, are not automatically controlled.
The SCCS method has been extended by modelling only the age effect or only the time-varying exposure effect using splines while representing the other by a piecewise constant step function. In these two extensions, there is a need to pre-specify exposure groups or age groups a priori. Misspecification of these groups may lead to biased association between exposure and adverse event.
In this talk, we propose a non-parametric SCCS method in which both age and exposure effects are represented as linear combinations of cubic M-splines at the same time. Cubic M-splines are piecewise polynomials of degree 3. To avoid a numerical integration of product of two spline functions in the likelihood function of the SCCS method we defined the first second and third integrals of I-splines based on the definition of integral of M-splines.
Simulation studies showed that the new method performs well. This new method is applied to a data on paediatric vaccines.

C04 Dose finding studies

C04.1
Dose finding methods based on longitudinal ordinal data: Realistic prior hypotheses identified from 49 phase I studies
X Paoletti[1], E Rizzo[2], L Collette[2]
[1]*Institut Curie, Paris, France,* [2]*EORTC Headquarters, Brussels, Belgium*

Context: Recently, dose finding designs have been extended to incorporate ordinal graded toxicity with the proportional odds (PO) model, or repeated toxicity measurements with mixed effect models.
In a very large data warehouse of 63 studies of single targeted agents from the EORTC DLT-TARGETT initiative, we explored (i) the PO assumption for the relation between graded toxicity and both the dose and the cycle of treatment using residuals and score tests, (ii) the variability and the intra-patient serial correlation in longitudinal measurements using mixed-effect models and markov chain modeling in cumulative logistic regression.

Results: We selected 49 trials (1735 patients treated at 4919 cycles) in which mild, moderate and severe grades were observed. The PO assumption for dose effect at cycle 1 was rejected at the 5% level only in 4 trials. On repeated cycles, PO were observed for both the dose and the cycle variables in 3 and 5 out of the 44 studies respectively.
In mixed effect models, variance of the random-intercept ranged from 1.5 to 13. Markov chain modeling revealed significant association between the risk of severe toxicity at a cycle and a moderate or mild toxicity at the previous cycle. In studies with no time effect detected from PO mixed model, common pattern of serial correlation was observed across several trials.
Conclusions: PO assumption is reasonable and can be implemented in single agents phase I trials. Large inter-patient variability can be expected and modeled using mildly informative priors in dose-finding trials incorporating repeated toxicity measurements.

C04.2
A Bayesian approach to oncology combination dose-finding

H Shen[1], M Whiley[2], B Neuenschwander[2]
[1]China Novartis Institute of Biomedical Research, Shanghai, China, [2]Novartis Pharma AG, Basel, Switzerland

In oncology the primary objective of Phase I trials for new compounds is to determine the maximum tolerated dose (MTD) or recommended Phase II dose (RP2D) based on severe safety events called dose limiting toxicities (DLTs). Bayesian approaches have been increasingly used in this setting and the flexibility they bring to dose escalation study design enables us to better balance statistical, clinical and operational considerations. However, efforts to further improve the efficiency and effectiveness of drug development in early phase oncology have led to increasingly complex trials, and adapting statistical methods to support such complex development strategies is challenging.
We will present a practical and comprehensive Bayesian approach used in a dual-agent combination dose escalation study. In this approach a 5-parameter logistic regression model is used to characterize the dose-toxicity relationship, and Bayesian inference is used to guide the dose escalation process. We will discuss the use of historical data from previous single agent dose escalation studies to inform the combination dose escalation. We will further demonstrate how this methodology can be used to address emerging drug development issues. In our example we will show how to support both a change to the dosing schedule and to the drug formulation during the course of the dose escalation. In each case we will demonstrate how the Bayesian approach allows current study information to inform the continued dose escalation.

C04.3
Bayesian optimal clinical trial design for monoclonal antibodies

MR Lange[1], H Schmidli[1]
[1]Novartis Pharma AG, Basel, Switzerland

Monoclonal antibodies are increasingly and successfully used for the treatment of many chronic diseases. Semi-mechanistic nonlinear models are needed to adequately describe the dose-time-response relationship for such drugs. We consider the optimal allocation of patients to doses in a planned clinical trial for a monoclonal antibody, in order to learn most about the nonlinear dose-time-response model. To characterize a design as optimal, Lindley (Ann. Statist. 1956) introduced an information theoretic approach which aims to maximize the expected Kullback-Leibler difference between prior and posterior of the model parameters. Later, Bernardo (Ann. Statist. 1979) showed that this criterion can also be justified within a decision theoretic framework and emphasized its compatibility with the concept of Bayesian inference. Despite the strong theoretical motivation, this criterion has not often been used in practice due to the heavy computational challenges for complex models. We will demonstrate the applicability of these concepts for finding optimal clinical trial designs for monoclonal antibodies. For illustration, we will discuss the design of a clinical trial in patients with urticaria.

C04.4
Dose-escalation strategies which utilise subgroup information

A Cotterill[1], T Jaki[1]
[1]Lancaster University, Lancaster, United Kingdom

In determining the maximum tolerated dose (MTD) of a drug, phase I oncology trials commonly assume that the patient population is homogeneous. A single MTD is therefore identified for the entire population. Strict inclusion criteria can be used to justify this assumption but may lead to a treatment effect being missed in the excluded patients. Conversely, inclusion of a subgroup in which the treatment is inefficacious could mask a treatment effect in the remaining population. Removing the underlying assumption of a homogeneous population by investigating suspected subgroup effects in dose-escalation to identify, when necessary, an MTD in each subgroup could therefore be beneficial.
A simple example involves a drug approved in adults and to be tested in children. In the adult trial, patients with a certain biomarker appeared more likely to experience side-effects than the remaining population. This information is of interest here as it could be used to aid development of patient-specific dosing in the paediatric population.
A Bayesian decision theoretic approach assuming a homogeneous population is used as the baseline for comparison of three designs which account for a subgroup effect. The first design escalates independently within subgroups while in the second design information is shared across subgroups.
Finally, an approach using sparsity inducing priors is presented. Simulations comparing these designs were carried out in the setting of a phase I study of temozolomide in children and adolescents with recurrent solid tumours. Simulation results indicate that accounting for potential differences in tolerance between subgroups can be beneficial.

C04.5
Continual reassessment method for dose escalation clinical trials in oncology: a comparison of prior approaches using AZD3514 data

G James[1], S Symeonides[2], J Marshall[2], J Young[2], S Smith[2], G Clack[2]
[1]PHASTAR, London, United Kingdom, [2]AstraZeneca, Macclesfield, United Kingdom

Background: The continual reassessment method (CRM) is considered more efficient and ethical than standard methods for dose-escalation trials in oncology, but requires an underlying model of the dose-toxicity relationship ("priors") and there is limited guidance of what this should be when little is known about this association.
Aim: Compare the impact of applying the CRM with different prior approaches and the 3+3 method in terms of ability to determine the true maximum tolerated dose (MTD) and number of patients allocated to sub-optimal and toxic doses.
Methods: Post-hoc dose-escalation analyses on real-life clinical trial data on an early oncology drug (AZD3514) using the 3+3 method and CRM using five prior approaches: conservative, aggressive, dose-linear, sigmoidal and O'Quigley. The prior probability values were further examined by adding 10% to each prior within each method. Dose limiting toxicity (DLT) was retrospectively defined as moderate or greater nausea/vomiting.
Results: All methods correctly identified the true MTD. The 3+3 method allocated six patients to both sub-optimal and toxic doses. All CRM approaches allocated four patients to sub-optimal doses. No patients were

allocated to toxic doses from sigmoidal, two from conservative and five from other approaches. Adding 10% to the prior probabilities had little effect.

Conclusion: The CRM outperformed the 3+3 method. The underlying model of the dose-toxicity relationship influences the number of patients allocated to toxic doses. The sigmoidal approach was optimal by these criteria, although the two patients receiving the higher dose in the conservative approach may give more confidence in the MTD.

C05 High-dimensional data analyses I

C05.1
Assessment of positivity in ELISPOT assays based on FDR-type and mixture procedures

I James[1]
[1]*Murdoch University, Perth, Australia*

ELISPOT (enzyme-linked immunosorbent spot**)** assays are widely used in immunotherapy trials to assess whether T-cells of patients respond to immune stimulus. Reactive cell responses lead ultimately to spots on plates which can be counted and compared with the numbers of background spots resulting from corresponding unstimulated cells. Typically there will be a small number of replicates of both stimulated and unstimulated wells for each patient, with the assay repeated across a number of patients. Assessment of a "positive" response requires determination of a real excess of responses or spots in the stimulated wells compared with their background. A number of *ad hoc* and statistical methods have been proposed to indicate positivity.

Assessment of ELISPOT response is similar to the issue of false-discovery rate analysis in multiple testing where the aim is to determine those p-values associated with real effects. Here we consider adaptation of the FDR-type approach to ELISPOT analyses and compare it with an approach based on a novel mixture method in which the numbers of samples responding positively are estimated directly without a binomial sampling assumption. The latter method appears to provide some robustness to the estimation over a range of parameters encountered in this setting.

C05.2
Estimating individual peptide effects from pooled ELISPOT data

P Ström[1], NJ Borthwick[2], T Dong[2], T Hanke[2], M Reilly[1]
[1]*Karolinska Institutet, Stockholm, Sweden,* [2]*Oxford University, Oxford, United Kingdom*

Background: In studies of the immune system, interferon-γ ELISPOT assays are used to detect the T cell responses to peptides of interest. The peptides are placed in the wells of 96-well plates, either individually or in groups ("pools") and after incubation with the blood cells to be tested, the responding cells appear as spots, with a high number of spots indicating that the peptide(s) stimulated a strong immune response.

Objective: To develop and implement a method for estimating individual peptide responses from the ELISPOT responses from peptide pools and investigate its performance in a study where individual peptide responses were validated.

Materials and methods: By regarding the individual responses of peptides in a pool as "unobserved" data, the problem can be solved using the EM algorithm. We model the response from any pool as a Poisson random variable whose expected value is the sum of the intensities of the individual peptides that constitute the pool. We apply the method to data from 4 volunteers in a HIV-vaccine trial where 199 individual peptides were assayed in 80 overlapping pools and compare our estimates to the observed responses from the individual confirmatory assays.

Results: The EM equations can be solved by simple matrix operations to yield the maximum likelihood estimates of the individual peptide responses. Applying the method to the pooled responses from the 4 volunteers, we demonstrated excellent agreement between the estimates and the individual validations.

C05.3
Estimation of antibody concentration from multiple dilutions data

Y Xu[1], YB Cheung[1]
[1]*Duke-NUS Graduate Medical School, Singapore, Singapore*

In medicine and chemistry, measurement of concentrations usually involves calibration that maps the observed response to the underlying concentration level using the inversion of a standard curve. The Enzyme-linked ImmunoSorbent Assay (ELISA) is one such method that is commonly used to measure antibody concentration. A problem in this type of technology is that an accurate measurement is obtainable only if the observations fall within the optimal, near-linear range of the standard curve. It is common to conduct a series of doubling or tripling dilutions of the samples, so that at least some of the diluted samples are within the optimal range. A single dilution may then be selected for statistical analysis. This common practice does not fully utilize the data from multiple dilutions and reduces accuracy. We consider a weighted average estimation approach for fully utilizing the information from multiple dilutions. Simulation results demonstrated the superiority of this weighted estimation approach over the conventional approach of analyzing a single selected dilution. We apply the methods to an experimental study of vaccine candidates.

C05.4
Mixed models for the analysis of brain magnetic resonance imaging data

KJ Lee[1,2], DK Thompson[1,3], PJ Anderson[1,2], LW Doyle[1,4], JL Cheong[1,4], JB Carlin[1,2]
[1]*Murdoch Children's Research Institute, Melbourne, Australia,* [2]*University of Melbourne, Melbourne, Australia,* [3]*Florey Neuroscience Institutes, University of Melbourne, Melbourne, Australia,* [4]*Royal Women's Hospital, Melbourne, Australia*

Advanced Magnetic Resonance Imaging (MRI) of the brain often involves measuring specific quantities, for example volume or diffusion parameters, in a number of different regions in both the left and right hemispheres of the brain. We explore approaches to the analysis of such data when we are interested in a comparison of brain metrics between two groups of individuals. We compare the standard approach, where each region in each hemisphere is analysed separately, to a mixed model approach, where we combine all of the regional data into a single regression model including a random effect to allow for the multiple measurements on each individual. Within the mixed model approach we compare the use of a single or region-specific random effect and single or region-specific error term. The advantage of the mixed model approach is that it jointly estimates the differences in the quantity of interest across the regions. This is more efficient than conducting separate comparisons and enables an assessment of overall patterns of differences across the regions, thereby reducing the multiple testing and increasing the power to find a difference between the groups compared with the standard approach. It also provides a direct assessment of whether the effect of group varies across the regions i.e. whether there is a region-specific effect of group. We apply these approaches to data from the Victorian Infant Brain Study where we wish to compare the regional brain volumes at term-equivalent age between infants born very preterm and those born at term.

C05.5
Comparison of analysis approaches for multi-level vascular imaging data
JR Forman[1], SJ Bond[1], KM Mäki-Petäjä[2], IB Wilkinson[1,2]
[1]Cambridge Clinical Trials Unit, Cambridge, United Kingdom, [2]Clinical Pharmacology Unit, University of Cambridge, Cambridge, United Kingdom

Vascular PET-CT imaging studies produce multi-level data structures, but at present there is no consensus on how best to analyse such data. We present our analyses applied to data observed from a cohort of patients with rheumatoid arthritis and a matched control cohort with stable cardiovascular disease on whom vascular PET-CT scans were taken (UK ethics ID 08/H0305/19). The primary endpoint is change in vascular inflammation, as measured by PET-CT. Each participant receives a PET-CT scan at baseline and 8 weeks later. Each scan covers up to five vessels per participant, and each vessel is analysed in "slices", generating a three-level data structure: participants, vessels, slices.
Here, we present and compare three analysis approaches that have been used in previous studies. 1: Data are pooled by vessel and by patient, by considering the change in the mean (or max) inflammation. 2: An index vessel is selected at baseline and followed-up in each patient. The index vessel is selected as the vessel with highest inflammation at baseline. 3: A multi-level model incorporates the complete data set and data structure. Through comparing these analysis approaches, we aim to identify a method which provides an accurate measure of the treatment effect, and a straightforward interpretation.

C06 The biostatistician's toolbox I

C06.1
Translational statistics and dynamic nomograms
J Newell[1], A Jalali[1], A Alvarez-Iglesias[1], M O'Donnell[1], J Hinde[1]
[1]National University of Ireland, Galway, Ireland

Translational Medicine promotes the convergence of basic and clinical research disciplines and the transfer of knowledge on the benefits and risks of therapies. In an analogous fashion we propose the concept of Translational Statistics to facilitate the integration of Biostatistics within clinical research and enhance communication of research findings in an accurate manner to diverse audiences (e.g. policy makers, patients and the media).
One example of this knowledge transfer is in Classification problems, commonly presented to statisticians, typically involving a binary outcome. The usual summary is the Odds Ratio. It has been argued that, when possible, a summary quoting the underlying probabilities is more informative than one based on ratios of odds or indeed of probabilities. As statistical inferential methods become more computational the models arising are increasingly complex and difficult to interpret.
Nomograms can be used as calculators of a predicted response for different values of the explanatory variables but can become cumbersome as the model becomes more complex. Tree based models allow prediction on a probabilistic scale but as the classifier becomes more complex (e.g. random forests) a simple classification rules is unavailable.
In this presentation dynamic nomograms are introduced that can be created automatically from any glm model object in R. In theory any model appearing in a scientific publication can be accompanied by a URL directing the 'user' to the accompanying dynamic nomogram from which the results of the models are directly translational and the robustness of the model verified through automatically generated model diagnostic plots.

C06.2
Dynamic graph generation and data analysis of complex data: a web-application based on R and shiny
L Lusa[1], C Ahlin[1]
[1]University of Ljubljana, Ljubljana, Slovenia

R statistical environment includes facilities for data display and analysis that are extremely flexible. It recently became rather straightforward to create interactive web applications and interactive graphics based on code written in R, using the shiny package and Scalable Vector Graphics (SVG). We illustrate, through an example, the feasibility of developing a user friendly web application that incorporates a variety of interactive graphical displays and tools for the analysis of complex data.
The medplot application was developed to help medical researchers to explore and analyse longitudinal data, where numerous variables are recorded for each patient over time. Several interactive graphical displays allow an easy exploration of the data. The analyses tools evaluate the association of the variables with other characteristics of the patients, taking into account the multiple testing problem, the repeated measurements and the possibility of non-linear associations between the covariates and the outcomes.
The application can be used by users that are not familiar with R or other statistical programs. It can be used through a web-browser and it does not require the installation of any program. Template spread sheets for data preparation are also provided, together with example data from a clinical study including patients with erythema migrans, where the variables are the presence and intensity of numerous symptoms recorded over time.

C06.3
The grammar of parametric boxplots
R Vonthein[1]
[1]Universität zu Lübeck, IMBS, ZKS, Lübeck, Germany

Boxplots were introduced as distribution-free summary plots of distributions shown side by side. They may be defined for specific distributions, of an assumed model, say. Then, numbers counted and numbers measured need to be treated differently. One may accomodate censoring.
Following the grammar of graphics approach, there may be several graphems or layers to a plot. First, all data may be plotted neatly side by side; then summaries are added. Summaries may be nonparametric or parametric, e.g. estimated quantiles, densities or both. Color and symbols add more information or help discern data from summary. You may decide between joint and conditional distributions and you may pick one out of several theoretical distributions. Distributions of counts may be drawn to emphasize discreteness or to hint at an approximate continuous distribution by confluent graphems. There are several ways to illustrate the fit of the model to the data.
Applications to real data of an R-function available from the author will illustrate possibilities and limitations. Especially the dotplot layer needs special attention in case of very small or large samples or widely differing sample sizes. Parametric boxplots look focused on central tendency rather than extreme quantiles when compared to distribution or survival functions. Boxplots readily show joint distributions with many categories when density and distribution functions get unreadable.

C06.4
Structure and interpretation of classification and regression trees

RJ Marshall[1]
[1]*Auckland University, Auckland, New Zealand*

Since Breiman's seminal work in 1984 the use of classification and regression trees in many areas of statistical analysis, including medicine, has become widespread. Extensions to "forests of trees" and associated B-word refinements (bagging, Bayes, boosting, bootstrapping) have been proposed. However, the issue of the nature of tree structures, and classifiers generated by them, is seldom raised. The nature of the tree partitions is generally ignored; emphasis is nearly always placed on performance measures. Yet in medical context the nature of tree structures is important, for the combinations of signs, symptoms, and other prognostic factors that make up a tree node, need to make clinical sense. In this presentation, the nature of tree structures will be discussed. Examples of published trees in the medical literature are presented and examined. It is concluded that the hierarchical nature of trees is restrictive and will invariably output awkward and difficult to interpret factor combinations.

C06.5
Classification and regression trees for moderator effects in clinical trials

R Gueorguieva[1], W-M Tsai[1], R Wu[1], H Zhang[1], SS O'Malley[1]
[1]*Yale University, New Haven, United States*

Classification and regression trees are a powerful and systematic method for ascertaining combinations of predictors of good outcome in medical studies. This approach allows one to consider a large pool of predictor variables, to derive empirically the strongest predictors, to consider interactive effects and to present the results in the form of decision trees that are easily interpreted by clinicians. Classical applications of this approach focus on identification of predictors of outcome regardless of treatment. However, in treatment studies identification of subject characteristics associated with good outcome on a particular treatment is of primary interest. We extend the method for automatic tree growing of Zhang et al. (2011) to identify subgroups of subjects who respond more favorably to one treatment than another based on their baseline characteristics. An automatic pruning step is incorporated, and a novel validation method is also proposed and evaluated. Terminal nodes in the constructed tree are associated with better outcome on a particular treatment, thus can be used to inform personalized treatment decisions. The approach is evaluated on simulated data and illustrated with an application from a clinical trial of alternative pharmacological treatments in alcohol dependence. The approach is also compared to alternative methods for identification of subgroups with enhanced treatment effects and advantages and disadvantages of the different approaches are discussed.

Acknowledgement: The project described was supported by Grants R01AA017173, K05 AA014715, and K23 AA020000 from the National Institute on Alcohol Abuse and Alcoholism.

Monday, 25th August 2014 – 11:30-12:30

President's invited speaker

IP President's invited speaker

Organizer: A. Zwinderman (ISCB President)

IP.1

Bayes, why bother?

TA Louis[1,2]

[1]Johns Hopkins Blooomberg SPH, Baltimore, United States, [2]U. S. Census Bureau, Washington, United States

The use of Bayesian-based designs and analyses in biomedical, environmental, political and many other applications has burgeoned, even though its use entails additional overhead. Consequently, it is evident that statisticians and collaborators are increasingly finding the approach worth the bother. In the context of this escalating incidence, I highlight a subset of the potential advantages of the formalism in study design ("Everyone is a Bayesian in the design phase"), conduct, analysis and reporting. Approaches include designs and analyses with required frequentist properties (Bayes for frequentist) and for fully Bayesian goals (Bayes for Bayes). Examples are drawn from sample size estimation, design and analysis of cluster randomized studies, use of historical controls, frequentist CI coverage, evaluating subgroups, dealing with multiplicity, ranking and other nonstandard goals.

The Bayesian approach is by no means a panacea. Valid development and application places additional obligations on the investigative team, and so it isn´t always worth the effort. However, the investment can pay big dividends, the cost/benefit relation is increasingly attractive, and in many situations it is definitely worth the bother.

Monday, 25th August 2014 – 14:00-15:30

Invited session

I2 Beyond R packages: getting our methods into standard software

Organizer: Georg Heinze

I2.1
Writing and developing statistical software: the statistical methodologist´s view

IR White[1]
[1]*MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, United Kingdom*

This talk describes my experience of writing and developing statistical software. Much of it has been universally ignored, but some of it has been widely used. The key to wide uptake is to tackle an important general problem rather than to focus on a specific solution.
I will discuss the importance of choosing a syntax which is both easy to use and flexible, and of using the software extensively oneself ("what happens if I try this?"). Most important in my experience is establishing two-way communications with users, since no developer can anticipate the range of uses to which a good piece of software will be put. Such two-way communication also often raises interesting methodological issues.
I will illustrate these ideas using my experience of developing software in Stata for multiple imputation (with Patrick Royston) and for multivariate meta-analysis. The former was ultimately superceded by a package in "official" Stata, which reproduces much of our functionality.

I2.2
Beyond wild horses: developing innovation at Cytel

Y Jemiai[1]
[1]*Cytel Statistical Software and Services, Cambridge, United States*

Henry Ford once said, "If I'd asked customers what they wanted, they would have told me, 'A faster horse!'"
With the power and flexibility of R, statisticians all over the world have been able to breed their very own "horses," but how does one make a "car?" This is what we do at Cytel. To develop innovative statistical software products that people love, we ask ourselves three key questions:
1. Of all the new statistical methods being proposed on a daily basis, which ones will be useful to the statistical community?
2. How do we create an exceptional user experience that helps spread the use of innovative statistical methods?
3. How do we establish trust in the results produced by the software?
Answering these questions is a complex undertaking. The process by which we answer these questions relies on a combination of frequent interactions with our customer base, strategic consulting engagements to solve our clients' most pressing and challenging statistical problems, and internal methodological research.
This talk describes how Cytel develops unique and innovative software products in an effort to serve the statistical community and promote the application of better statistical techniques in the scientific community.

I2.3
An inside perspective on the development of SAS statistical software

RN Rodriguez[1]
[1]*SAS Institute Inc., Cary, United States*

This presentation provides a behind-the-scenes look at how SAS/STAT software is developed, beginning with the qualifications we consider in recruiting research statistician developers, and progressing through the stages of design, programming, testing, documentation, and user support. Our decisions about which methods to implement are based on customer requirements, which are driven by the increasing value, complexity, and size of data, coupled with advances in methodology and technology. These decisions are also informed by discussions of promising methodology between developers and researchers.
In order to deliver versatile methods that work robustly across many areas of statistical practice, developers must often extend the literature on theory and methods to handle issues such as unbalanced data, large numbers of random effects, and missingness. Again, this presents opportunities for interacting with researchers. Our development process also emphasizes syntax that is common across procedures, clear and consistent output, high-performance computing, numerical accuracy, and helpful documentation. Examples drawn from recent releases illustrate these principles.

Monday, 25th August 2014 – 14:00-15:30

Contributed sessions

C07 Clinical trials

C07.1

Dealing with challenges in design and analysis of clinical trials in long-term care

L Thabane[1]
[1]McMaster University, Hamilton, Canada

Conducting trials in long-term care (LTC) can present many implementation, methodological, ethical and analytical challenges. First, the population is quite frail which can present serious methodological and ethical challenges; second, the environment is hard to work in because of the overcommitted staff and this can create major implementation problems; third, outcome assessment can be challenging; fourth, determining the appropriate unit of randomization, analysis or inference can be complicated by several practical factors. In this presentation, I will briefly describe some of the key challenges and use

· Our experience from a scoping review of trials of hip-protectors in LTC residents; and

· the ViDOS (Vitamin D and Osteoporosis) trial—a cluster randomized controlled trial of 40 LTC homes in Ontario—designed to determine the feasibility and effectiveness of a multi-faceted knowledge translation intervention aimed at improving vitamin D supplementation and other evidence-based osteoporosis/fracture prevention strategies in LTC, to illustrate the issues and to suggest potential solutions.

C07.2

Estimating efficacy and effectiveness using data retrieved after treatment non-compliance

A-K Leuchs[1]
[1]BfArM, Bonn, Germany

Treatment non-compliance (e.g. treatment discontinuation, switch or augmentation) and missing data are common issues in clinical trials. In the presence of non-compliance, differentiating between the true benefit of the treatment (efficacy), the effect if the medication is taken as observed (effectiveness), and different estimands for both is essential.

In addition to data on non-compliance itself, the retrieval of patients irrespective of their protocol adherence provides useful information on effectiveness.

In longitudinal neuropsychiatric disease trials, in which missing data and non-compliance are frequent, mixed models for repeated measurements (MMRM) are favoured methods to estimate treatment effects. However, depending on the data included (inclusive/exclusive of retrieved data), MMRMs address either efficacy or effectiveness. In various different settings, we oppose and compare these standard methods to piecewise linear mixed effects models also including retrieved data regarding their ability to appropriately estimate efficacy and effectiveness, primarily focusing on type 1 errors. Different assumptions on the trend while complying and non-complying are made.

It becomes evident that retrieving and including data after protocol violation is crucial to obtain unbiased estimates of effectiveness. A simple MMRM including all data may be severely biased and hence inappropriate, especially if data are only partly retrieved after non-compliance. In some settings, however, MMRMs may still be adequate to estimate efficacy.

In conclusion, it is paramount to precisely define the estimation's objective and the corresponding estimand, to choose an appropriate analysis method addressing the trial's target estimand and, if effectiveness is of interest, to make any effort to retrieve data.

C07.3

Including historical data in the analysis of clinical trials using the modified power prior: practical and theoretical issues

J van Rosmalen[1], D Dejardin[2], E Lesaffre[1,2]
[1]Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands, [2]L-Biostat, KU Leuven, Leuven, Belgium

Including historical data in the analysis of clinical trials may improve the precision of the results and reduce the required sample size. In the Bayesian context, Ibrahim and Chen (2000) proposed the power prior to combine historical and current data, where the likelihood of the historical data is raised to a power α (with $0 \leq \alpha \leq 1$). Using the modified power prior (MPP; see Duan et al. 2006), α is estimated in an adaptive way: if the current data and historical data are commensurate, α will be high, whereas in case of substantial discrepancy, α will be close to 0, effectively discarding the historical data.

We give a methodological overview of the MPP and discuss the relationships with other methods for including historical data. Also, we show how the MPP can be used to account for the common situation where the historical data differ substantially from the current data with respect to nuisance parameters (e.g. the baseline hazard in survival models). An important practical issue is the calculation of the normalizing constant of the MPP, which cannot be done using standard MCMC samplers. We propose new algorithms for computing the posterior results, based on Laplace approximation and path sampling (Friel and Pettitt 2008).

We illustrate the usefulness of the MPP using data from two randomized controlled trials for progression-free survival in patients with metastatic breast cancer. We find that the MPP is a promising method for incorporating historical data in clinical trials.

C07.4

A robust Bayesian meta-analytic-predictive approach to borrow strength from historical information in thorough QT studies

K Meiser[1], H Schmidli[1]
[1]Novartis Pharma AG, Basel, Switzerland

Thorough QT (tQT) trials are key studies to evaluate the cardiac safety of new test drugs. The primary endpoint is the QT interval measured by electrocardiography (ECG). A prolongation of the QT interval is associated with serious cardiac events (ICH E14 guidance, 2005). In tQT studies, the test drug is compared to placebo and an active control known to prolong the QT interval. As these studies are routinely done, many historical studies with healthy subjects are available, providing information on the active control. Borrowing strength from this historical information to reduce the number of patients randomized to the active control would be desirable from an ethical and efficiency perspective.

We propose a robust Bayesian meta-analytic-predictive approach to derive an informative prior on the active control in a new tQT study from the historical trials. This approach essentially assumes exchangeability of the active control parameter in the historical and the new trial. However, the possibility that the active control parameter is systematically different from the historical trials is taken into account by adding a weakly informative mixture component to the meta-analytic-predictive prior. This provides robustness of the approach to prior-data conflicts. A tQT study will be used to illustrate the methodology.

C07.5
Power and sample size of trials with a partially nesting design for binary outcomes
C Roberts[1], E Batistatou[1], S Roberts[1]
[1]University of Manchester, Manchester, United Kingdom

Introduction: Partial Nesting describes the situation where some subjects are in clusters and others are not. This design may arise in individually randomized trials of non-pharmacological interventions where patients are clustered in one treatment arm due to treatment but not in the comparator. We consider sample size and power for binary outcomes of trials of this design.
Methods: Formulae for determining power and sample size on the scale of proportions, log-odds and using an arc-sine transformation are presented. These are compared with empirical power estimated using an adjusted test of proportions, a summary measures test, a logistic model with random intercept and a logistic GEE model. Empirical power is determined using simulation with 40,000 replications for each scenario and a range of study sizes and proportions appropriate for estimation of small to medium treatment effects.
Results: There is some under- and over-estimation of the formula determined power relative to empirical power up to a maximum of 6%. For methods on the scale of proportions increased variance of the test statistics inflates the type II error thereby reducing empirical power relative to the formula. Power determined using the arc-sine formula tends to perform better than the proportions method. For logistic models, the empirical power appears to be increased relative to the proposed formulae by small sample bias.
Conclusions: All methods of calculating power gave approximations to empirical power adequate for many practical situations. There were nevertheless differences between methods of sample size and power determination and between determined and empirical power.

C08 Prediction models: case studies

C08.1
Multidimensional assessment of the predictive ability of a Trichotomous-outcome model in the NICHD Collaborative Pediatric Critical Care Research Network (CPCCRN)
R Holubkov[1], MM Pollack[2], AE Clark[1], T Funai[1], JM Dean[1]
[1]University of Utah School of Medicine, Salt Lake City, United States, [2]University of Arizona College of Medicine, Phoenix, United States

The CPCCRN Trichotomous Outcome Prediction in Critical Care (TOPICC) study is applying generalized logistic regression to a cohort of 10,000 critically ill children, predicting three outcome states (dead, alive with new morbidity, or good outcome at hospital discharge). We have implemented various graphical and analytic approaches to describe the model´s diagnostic performance over the three outcome levels.
Three-dimensional plots of the ROC hypersurface are informative when rotated to show individual pairwise ROC curves on the surface borders; the hypersurface can also be "sliced" to show diagnostic ability for two outcomes conditional on a particular correct prediction rate for the third. The Volume under the Surface (VUS) statistic may be intuitively presented as a brute-force integration of the ROC hypersurface, or as proportion of all possible one-patient-per-outcome triplets correctly classified (Mossman, MedDecisMaking 1999). However, clinicians do not have applied experience with the VUS and its properties, such as its 1/6 rather than 1/2 value under a model without any discriminative ability. Example hypothetical ROC hypersurfaces under "pathological", poor/perfect prediction settings were helpful when presenting our model´s performance.
Extensions of the *c*-index summarizing performance over pairs of outcomes are more interpretable to clinicians than VUS as numerical summaries of model performance. For example, we have found the ordinal *c*-index of van Calster et al (BiometricalJ 2012) to be a useful integration of VUS and pairwise approaches.
We will describe how the above approaches, and setting-specific modifications, were used in TOPICC to summarize diagnostic ability of our three-level outcome model.

C08.2
Developing dynamic prediction models for acute diseases
KL Phung[1], M Wolbers[1,2]
[1]Oxford University Clinical Research Unit, Ho Chi Minh City, Viet Nam, [2]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Dynamic prediction models which incorporate longitudinal covariate data and allow for temporal updating of predictions are increasingly popular. Unlike models based only on baseline information, such models use data efficiently and are able to capture the dynamic evolution of the disease in each individual. Several methods for developing such models have been proposed including extensions of classical regression models (Cox regression with time-dependent covariates, landmarking) and joint models. These approaches have been developed and applied mainly in chronic diseases, such as cancer or cardiovascular diseases, where the disease progresses slowly over a long time period. However, dynamic and updatable prediction models are also very attractive to physicians working in the acute setting, where diseases evolve rapidly over a short period of a few hours or days.
We will summarize differences between chronic and acute diseases with respect to the types of outcomes of primary interest, expected associations between longitudinal predictors and outcomes, and the amount of longitudinal data available. We then discuss implications of these differences for the development of dynamic prediction models. Our recommendations for dynamic modeling in the setting of acute diseases will be illustrated with a dataset of 2614 hospitalized children with dengue infection, an acute disease commonly seen in tropical regions, where the outcome of interest is the occurrence of shock. We will compare several possible modeling approaches and discuss how time-dependent accuracy measures and reclassification methods can be used to compare dynamic models and models based on baseline information only.

C08.3
A patient-specific predicting tool for functional recovery after stroke
A Douiri[1], JJ Grace[1], C McKevitt[1], AG Rudd[1], CDA Wolfe[1]
[1]King's College London, London, United Kingdom

Predicting recovery over time at various stages of rehabilitation after stroke could potentially allow for sequential monitoring of patients and early identification and management of patients with poor recovery.
This study aims to develop and validate a patient-specific tool for predicting recovery trajectories post-stroke. Data from 495 patients after first-ever stroke between 2002-2004 were determined from the ongoing population-based South London Stroke Register for the model development. Functional recovery was assessed using Barthel Index (BI) at 1, 2, 3, 4, 6, 8, 12, 26 and 52 weeks post-stroke. The selected model incorporates reliable prognostic factors which are prevalent in stroke and which have good clinical accessibility. Penalized iteratively reweighted nonlinear least squares for generalized linear mixed models were adapted to develop recovery curve models. The internal prediction error of the model was assessed using leave-one-out cross-validation. Temporal validations were conducted using two different cohorts between the period of 2005-2011 to evaluate the overall accuracy of the recovery curves, and to assess the

prognostic accuracy and utility for the classification of patients with poor recovery at 3 and 12 months. Predictive accuracy of the recovery curves was acceptable, with a root mean square deviation of 3.32 BI points. The prognostic accuracies to predict poor recovery at both 3 and 12 months were also satisfactory (94%, 95% CI [90.9-96.9] and 89%, 95% CI [84.4%-93.2] respectively). This presentation will describe and discuss the different statistical and computational methodologies adopted in the development and validation of the final recovery curve model.

C08.4
A tree based model for thyroid cancer prognostication
M Banerjee[1], D Muenz[1], M Haymart[1]
[1]University of Michigan, Ann Arbor, United States

Thyroid cancer is becoming an increasingly common cancer and yet little is known about the prognostic factors associated with survival. Controversy also exists over appropriate treatment for thyroid cancer. Prognostic models are needed to determine correlates of overall survival and identify subgroups of patients with poor prognosis who may benefit from earlier therapeutic intervention. In this talk we present a tree-based model for thyroid cancer prognostication using data from the US National Cancer Database. Trees are conceptually simple yet powerful, and are being increasingly used in biomedical studies for analyzing censored survival data where the primary goal is prognostication of patients. To gain accuracy in prediction and address instability in a single tree, an ensemble of trees is typically grown and the predictions are averaged across the trees in the ensemble. In this talk, we present a methodology for identifying the most representative tree from the ensemble based on several tree distance metrics. Out of bag error based on the cumulative hazard estimate is computed for the representative tree. For the thyroid cancer data, the representative tree from the ensemble was able to identify four distinct prognostic groups, defined by age, gender, lymph node involvement, tumor size, and metastasis status. Five year survival rates in these groups ranged between 64% and 99%. The prognostic groups derived can provide guidance for patient management, clinical trial design, and future treatment policy. The representative tree itself can be used as a decision making tool in the clinical setting.

C08.5
Stratified weighted regression for subgroup signatures from prognostic models with molecular data
V Weyer[1], H Binder[1]
[1]University Medical Center Mainz, IMBEI, Mainz, Germany

In the analysis of high dimensional molecular data with time-to-event endpoints, developing subgroup signatures is one way for taking individual heterogeneity into account. We propose an alternative method to subgroup analysis based on weighted regression. As an application we consider RNA-Seq data from acute myeloid leukemia (AML) patients with different cytogenetic risk profiles, where a survival gene expression signature for cytogenetic low risk patients is to be developed while taking information from high risk patients into account. For signature development we use automated variable selection by componentwise boosting with a weighted Cox regression partial log-likelihood, allowing for different baseline hazards in the different subgroups by strata. Thus, the partial likelihood takes the form of a weighted product of terms, one for each stratum. Further we propose two approaches for automatically choosing the weights which are based on resampling methods. For evaluation, we consider model stability and prediction performance. In a simulation study and in the real data set of AML patients the stratified approach is compared to an unstratified variant. The stratified approach is seen to perform well in terms of identifying of important factors as well as with respect to prediction performance. Automated selection of weights is seen to adequately identify situations where information from the respective other subgroup is useful. Thus stratified weighted regression with automated weight selection seems promising when subgroups need to be taken into account for signature development in a time-to-event setting.

C09 Latent variable methods

C09.1
Treatment effect estimation in latent variable models with structural misspecification
A Kifley[1,2]
[1]University of Sydney, Sydney, Australia, [2]Macquarie University, Sydney, Australia

Latent variable models are sensitive to misspecifications of the nature of the relationships between observed variables and unobserved underlying latent variables that may be of primary interest. However misspecifications of this type are likely to occur in practice. In this study, we evaluate the performance of reflective latent variable models in estimating treatment or exposure effects when presented with observed item measures that include a mixture of formative and reflective item types. Reflective models assume that observed items serve merely as indicators of the status of the underlying latent variables, while formative items in fact affect the latent variables directly. We explore the sensitivity of global treatment or exposure effect estimates to levels of direct, indirect and mediated effects of treatment, under a range of different conditions. We find a tendency toward overestimation of treatment effects by the reflective model if, in truth, the treatment affects formative items present in the assessment with little or no direct treatment effect on the latent variable of interest. We find a weaker tendency toward underestimation of treatment effects by the reflective model if, in truth, the treatment directly affects the latent variable but does not affect potentially formative items that are included. Problems in estimation are substantially greater if the assessment is predominantly formative and the formative items share strong similarities with each other. Our simulation studies were motivated by issues arising in analysis of health-related quality of life data, but are relevant to many other applications of latent variable modelling.

C09.2
Joint modeling of longitudinal binary and continuous responses
E Kurum[1], R Li[2], S Shiffman[3], W Yao[4]
[1]Istanbul Medeniyet University, Istanbul, Turkey, [2]The Pennsylvania State University, University Park, United States, [3]University of Pittsburgh, Pittsburgh, United States, [4]Kansas State University, Manhattan, United States

Motivated by an empirical analysis of ecological momentary assessment data (EMA) collected in a smoking cessation study, we propose a joint modeling technique for estimating the time-varying association between two intensively measured longitudinal responses: a continuous one and a binary one.
A major challenge in joint modeling these responses is the lack of a multivariate distribution. We suggest introducing a normal latent variable underlying the binary response and factorizing the model into two components: a marginal model for the continuous response, and a conditional model for the binary response given the continuous response. We develop a two-stage estimation procedure and establish the asymptotic normality of the resulting estimators. We also derived the standard error formulas

for estimated coefficients. We conduct a Monte Carlo simulation study to assess the finite sample performance of our procedure.
The proposed method is illustrated by an empirical analysis of smoking cessation data, in which the important question of interest is to investigate the association between urge to smoke, continuous response, and the status of alcohol use, the binary response, and how this association varies over time.

C09.3
Joint modelling of longitudinal and time-to-event data: a comparison between shared random-effect and latent class model

M Bastard[1], J-F Etard[1,2]
[1]Epicentre, Paris, France, [2]UMI 233 TransVIHMI, IRD, Université Montpellier 1, Montpellier, France

In clinical research, a growing interest is to investigate the effect of a longitudinal marker on the occurrence of an event. Joint models have recently been developed to investigate this association within two approaches: the shared-random effect (SRE) and the latent class approach (LC). We propose to present, apply and compare the two approaches.
The SRE consists in including the random effects of the mixed model used for longitudinal data as covariates in the model for the event to model the correlation between the marker profile and the event. The LC makes the assumption that the population is composed of several latent subpopulations in which the profile of the longitudinal marker and the risk of the event are different, using a discrete latent variable to link the marker profile and the event.
To illustrate the two models, we apply them to a cohort of HIV-infected patients supported by Médecins sans Frontières to study the effect of longitudinal CD4 profile on survival. Both approaches show consistent results, with a higher risk of death for patients with low CD4 profile.
SRE allows a straightforward interpretation of the effect of longitudinal marker on the event, but is numerically intensive and assumes homogeneity in marker evolution. LC is particularly useful when population is heterogeneous and provides an easier interpretation for clinicians but requires multiple fits. SRE and LC approaches are two powerful tools to model jointly longitudinal and time-to-event data and should be used carefully according to the research question of interest.

C09.4
Longitudinal patterns of stages of change and lifestyle intervention outcomes - a latent class analysis with distal outcomes

L Jiang[1], S Chen[1], J Beals[2], CM Mitchell[2], SM Manson[2], Y Roubideaux[3]
[1]Texas A&M University, College Station, United States, [2]University of Colorado Anschutz Medical Campus, Aurora, United States, [3]Indian Health Service, Rockville, United States

Stages of change measure the readiness to change a health behavior. To examine the transition patterns of stages of change for regular exercise over time and to investigate the association between longitudinal patterns of SoC and lifestyle intervention outcomes, we analyzed data from a lifestyle intervention program to prevent diabetes among American Indian and Alaska Natives (AI/ANs). Latent class analysis (LCA) was conducted to identify the longitudinal patterns of SoC for regular exercise reported at the three time points. LCA with distal outcomes was performed to investigate the associations between latent class membership and behavioral changes after the intervention. Traditionally, LCA with distal outcomes were estimated using maximum probability classification followed by standard regression approach (classical three-step approach). Yet, simulations have shown that this approach could underestimate the associations of interest. Two new methods have been proposed recently: a one-step approach and an improved three-step approach. In the current study, various estimation approaches were used to estimate the parameters and standard errors for the LCA with distal outcomes models and the results were compared. We identified three latent classes: Pre-action, Transition, and Maintenance classes. The participants in the Transition class moved from pre-action stage at baseline to action or maintenance stage post-intervention. Compared to the other two classes, those in the Transition class had the greatest improvements in physical activity and weight outcomes at both time points post-baseline.

C09.5
Joint latent class model for longitudinal data and competing interval-censored events: application to the study of Alzheimer's disease

A Rouanet[1], H Jacqmin-Gadda[1]
[1]INSERM U897 - ISPED, Bordeaux, France

Alzheimer´s disease is a chronic illness characterized by a continuous cognitive degradation process and a progressive loss of autonomy. This work aims at developing a descriptive model of the natural history of Alzheimer´s disease, specifically the cognitive decline before diagnosis, considering the competing risk of death. The cognitive decline, measured with repeated psychometric tests, is jointly modeled with the risk of dementia in a latent class model. The risk of death is also considered because the population under consideration corresponds to elderly people with high risk of death and most of the risk factors of dementia are associated to death too.
Moreover, the cohort data used in these analyses are interval censored because dementia can be diagnosed at visit times only. The exact date of onset of dementia is thus unknown and a subject who becomes demented and dies between two visits is not diagnosed as demented. To consider both the competing risk of death and interval censoring, it is necessary to use a multi-state Illness-Death model and to calculate the likelihood accounting for interval censoring. The transition intensities depend on age but the Dementia-Death transition can possibly depend on the time spent in the demented state, in a semi-Markov model.
In this work, we propose an Illness-Death joint model for competing interval-censored events and repeated measures of a marker. This model is applied to the French Paquid cohort, which includes 3777 patients, older than 65, followed every 2 or 3 years during 20 years.

C10 Genome-wide association studies

C10.1
Entropy-based statistics to detect gene-gene interactions

PG Ferrario[1,2]
[1]Universität zu Lübeck, Lübeck, Germany, [2]Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Germany

Despite the success of genome-wide association studies (GWAS) in the identification of genetic regions associated with complex diseases, an important proportion of the assumed heritability is as yet unexplained for many traits. To fill this gap, it has been proposed not only to analyse if genes act to a certain phenotype but also if genes interact between each other.
In the last ten years different methods have been introduced in order to detect gene-gene interactions. Different approaches from the data mining, machine learning, Bayes-statistics and from the multidimensional reduction (MDR) were applied on genome-wide data and the results have been presented regularly in the literature.

New emerging approaches are the so called entropy-based statistics, supported on the Shannon definition of entropy and the whole information theory.
The notions of (conditional) entropy, synergy, (conditional) mutual information are well-posed and enable to give a measure for gene-gene interactions and describe also non-linear dependencies between genotypes and phenotypes.
In this presentation we give an overview of the different entropy-based statistics, underlining their strengths and weaknesses. Furthermore, we show that there are more open issues and suggest which directions can be interesting for future work.

C10.2
Network Approach to Identify Gene-by-Secondary Phenotype Interactions in GWAS

AN Vidyashankar[1], G Diao[1], B Etain[2], S Katsahian[1]
[1]George Mason University, Fairfax, United States, [2]Institut Universitaire d'Hématologie, Université Paris VII, Paris, France

Aims: Analysis of secondary phenotypes in Genome-Wide Association studies has recently received much attention in the literature. Motivated by an application to a recent psychiatric study, the primary objectives of the study are the following: (a) to identify significant associations between the primary phenotype and secondary phenotype by gene interactions and (b) establish an interaction network and use it to identify groups of genes associated with the primary phenotype.
Methods: Regression models will be used to estimate gene by secondary phenotype interactions at each marker. An efficient score approach with adjusted p-values is then used to test for the significance of the interaction effects after controlling for family-wise error rate (FWER). These results are then analyzed using network based methods. Network wide metrics are then applied to identify groups of secondary phenotype by gene interactions that are associated with the primary phenotype.
Results: Use of network wide metrics allows for a principled approach to cluster genes that are significantly associated with the primary phenotype through their interactions with the secondary phenotypes. This facilitates identifying interaction peaks amongst secondary phenotypes.
Conclusions: Identification of gene-by-secondary phenotype interactions associated with the primary phenotypes helps in identifying different treatment options for varying subgroups of patients.

C10.3
Gene-gene interaction analysis of correlated phenotypes

T Park[1], I-S Huh[1], M-S Kwon[1]
[1]Seoul National University, Seoul, Republic of Korea

Despite of many successful results from genome-wide association studies (GWAS), only a small number of genetic variants tend to be identified and replicated given a stringent genome-wide significance criterion. Furthermore, in many GWAS, one binary trait is commonly used which is derived from multiple quantitative traits. The use of this summary phenotype may decrease power due to the loss of information about phenotypes. Therefore, we propose a multivariate approach which uses all information about correlated phenotypes. Especially, we focus on identification of gene-gene interaction effects on the correlated phenotypes. Generalized multifactor dimensional reduction (GMDR) method has been commonly used in identifying gene-gene interactions. We propose a multivariate GMDR approach in order to identify gene-gene interaction for the multiple phenotypes. Our proposed multivariate GMDR method uses multiple phenotypes simultaneously. We applied the multivariate GMDR approach to a GWA data dataset of 8,842 Korean individuals.

C10.4
Estimating the rediscovery rate for assessing the validation of genome-wide association studies

D Lee[1], W Lee[2], A Ganna[3], Y Pawitan[3]
[1]Department of Statistics, Ewha Womans University, Seoul, Republic of Korea, [2]Department of Statistics, Inha University, Incheon, Republic of Korea, [3]Department of MEB, Karolinska Institutet, Stockholm, Sweden

Validation studies are often used to confirm that the observed findings in the training studies are not artifact due to chance or uncontrolled bias. These replicated studies increase the generalizability of the results and decrease the possibility to report false positive findings. Although the importance of replication or validation studies is well recognized, there seems to be less awareness of the factors that influence the reproducibility of significant findings. Indeed, the selection of the validation study is more often driven by data availability rather than study design. In this study we aim to investigate the factors influencing the proportion of significant findings from a training sample that are replicated in a validation sample. We quantify this by introducing a measure called rediscovery rate (RDR), and show how to estimate it nonparametrically from the training dataset. This RDR estimate can be used to design and to assess the validation study.
Furthermore, we discuss the meaning of local RDR to interpret and measure the reproducibility of each significant SNP outcome. We use simulated data and real examples from genome-wide association studies to illustrate the application of the RDR and local RDR concept in high-throughput data analyses.

C10.5
A multivariate method for meta-analysis of multiple outcomes in genetic association studies

NL Dimou[1], PG Bagos[1]
[1]University of Thessaly, Lamia, Greece

In this work we present a simple, yet powerful approach for performing multivariate meta-analysis of genetic association studies when multiple outcomes are assessed. The key element of our approach is the analytical calculation of the within-studies covariances.
We propose a model based on summary data, uniformly defined for both discrete and continuous outcomes (using log odds-ratios or mean differences). The within-studies covariances can be calculated using the cross-classification of the genotypes in both outcomes, which are retrieved using a log-linear model using the iterative proportional fitting algorithm under the assumption of no three-way interaction.
As an example, we assess the association of MDR1 C3435T polymorphism with two exclusive outcomes (Ulcerative colitis and Crohn's disease), as well as the association of GNB3 C825T polymorphism with two non-exclusive dichotomous outcomes (diabetes and hypertension). We also present an application using continuous outcomes (diastolic and systolic blood pressure). We show the applicability and the generality of the method performing the analysis assuming the genetic model beforehand or following a genetic model-free approach. The method is simple and fast, it can be extended for several outcomes and can be fitted in nearly all statistical packages. There is no need for individual patient data or the simultaneous evaluation of both outcomes in all studies.
We conclude that the proposed method constitute a useful framework for performing meta-analysis for multiple outcomes within the context of genetic association studies. Connections to other similar models presented in the literature, are discussed, as well as potential extensions to future work.

C11 Functional data analysis

C11.1
Functional data analysis of temporal glucose curves compared with gold standard measurements of insulin sensitivity and beta-cell function

KF Frøslie[1,2], K Færch[3], J Røislien[1]
[1]University of Oslo, Oslo, Norway, [2]Oslo University Hospital, Oslo, Norway, [3]Steno Diabetes Center A/S, Gentofte, Denmark

A recent study of blood glucose curves in pregnant women, applying functional principal component analysis (FPCA) of oral glucose tolerance test (OGTT) data, identified the general glucose level and the timing of the glucose peak as the two main modes of temporal variation between individuals. The latter was a significant predictor of gestational diabetes later in pregnancy. Glucose curve trajectories may also reflect distinct physiological processes.
The aim of the present study was to extract glucose curve characteristics from OGTTs in healthy, non-pregnant individuals, and compare these to gold standard measurements of physiological features.
OGTT glucose data from 20 participants in the Inter99 study were analysed by FPCA. Glucose curve characteristics were compared with measurements of insulin sensitivity and beta-cell function obtained from euglycaemic hyperinsulinemic clamps and intravenous glucose tolerance tests.
The first two functional principal components (FPCs) explained 65% and 19%, respectively, of the variance in glucose curves. The first FPC (FPC1) represented the general postprandial glucose levels during the OGTT and the timing of the postprandial glucose peak. High FPC1 scores (high levels and late peak) were associated with low insulin sensitivity (r=-0.43) and low first-phase insulin response (r=-0.41). FPC2 represented the "flatness" of the curve, with high scores (flat curve, higher than average postprandial glucose values in the later half of the OGTT) being associated with high first-phase insulin response (r=0.43), but low insulin sensitivity (r=-0.22).
The curve characteristics derived by FPCA adds to the understanding of the various physiological mechanisms that are mirrored in glucose curves.

C11.2
Unsupervised classification of functional data based on covariance structures

F Ieva[1], AM Paganoni[2], N Tarabelloni[2]
[1]Università degli studi di Milano, Milano, Italy, [2]Politecnico di Milano, Milano, Italy

We address the problem of performing an unsupervised classification of samples from two populations that differ in terms of variability, rather than location. This is of interest in biomedical context, where the dichotomy between physiological and pathological features often shows an interesting pattern in change of variability. To this aim, we recur to a proper notion of distance between covariance structures, which is the basis of our new classification method. We formulate it in an abstract setting, suitable for both standard finite-dimensional, and functional data analysis. In order to detect the two groups of samples, we search, inside the set of possible recombinations of samples, the labelling that maximizes the distance between the induced covariance structures, under the hypothesis that the true subdivision realizes this maximum. We identify the members of the two estimated populations as those who fulfil the two maximally-distant covariance structures. Special care is taken of traslating this procedure into a heuristic algorithm, able to restrain the explosive complexity of a naive, greedy search. We also propose an exploratory technique to early assess the successful application of our method. This is based on information drawn from an approximate permutation test performed on data, in which our target is the distribution of the covariances´ distance under different recombinations of samples. We also point out that, when analysing functional samples, best results are obtained by improving the estimation of covariances with a shrinkage approach. We test our method first on synthetic data, and then on real data.

C11.3
Functional data analysis in radiobiology and radiation epidemiology

MA Benadjaoud[1,2,3], H Cardot[4], F de Vathaire[1,2,3]
[1]Radiation Epidemiology Group CESP - INSERM 1018, Villejuif, France, [2]Institut Gustave Roussy, Villejuif, France, [3]Université Paris Sud, Le Kremlin-Bicêtre, France, [4]Institut de Mathématiques de Bourgogne, Dijon, France

In many fields, each observation consists of discrete measurements collected over a continuum. These data points can be thought as discrete sampling from an underlying smooth process. The functional data analysis (FDA) aims to analyze data providing by curves, (hyper)surfaces…etc as opposed to a point or a finite-dimensional vector and extract additional information contained in the functions and their derivatives, not normally available through traditional methods.
We propose to illustrate the performance of the FDA technics through two applications from the ionizing radiations field.
The first application concerns the normal tissue complication probability (NTCP) modeling in external radiotherapy. The most widely NTCP model is the Lyman-Kutcher-Burman (LKB) model based on the reducing of the dose distribution to an equivalent uniform dose (EUD) using a power law. In this example, we propose a non-parametric NTCP model where the weights dose values in EUD are estimated flexibly using logistic model based on the scores of functional principal component analysis conducted on the estimated dose probability density functions. The estimated parameter function leads to a better understanding of the dose-volume effect.
The second application focus on the analysis of the temporal response of H2AX, an important marker of a dangerous radio-induced DNA lesion: the double strand breaks (DSBs). In these data, each subject is a cluster wherein three H2AX temporal curves are measured.
We used a multilevel functional principal component analysis to quantify intra-and-inter subject variability and investigate the association between the H2AX temporal dynamics and the risk of radiation-induced second malignancies.

C11.4
Use of finite mixture models for dietary patterns analysis

N Sauvageot[1], A Alkerwi[1], A Albert[2], M Guillaume[2]
[1]CRP Santé, Luxembourg, Luxembourg, [2]University of Liege, Liege, Belgium

Free-living individuals have multifaceted diets and consume foods in numerous combinations. The effect of the overall diet beyond that of single foods can be studied with dietary pattern analysis. Furthermore, the dietary pattern approach reduces data-dimensionality and alleviates problems of model over-fitting and residual confounding that occur with the statistical analysis of many food items.
Most recent dietary pattern analyses have used factor and cluster analysis. We describe a finite mixture modelling (FMM) approach for dietary pattern analysis and show its advantages over previous ones.
First, FMM allows estimating pattern prevalence directly from the model parameters as opposed to the subjective joint classification of the factors. Moreover, in contrast to ‚hard' assignment of clustering methods, FMM also produces posterior cluster membership probabilities for each subject providing measures of uncertainty of the associated classification.
Second, it allows problems in determining the number of clusters and

choosing an appropriate clustering method to be recast as statistical model choice problems.
Third, it allows for covariates adjustment simultaneously with the fitting process and the size of pattern to depend on a set of concomitant variables. Additionally, FMM is invariant to linear transformation, for example standardization.
We discuss these advantages and illustrate the approach with an analysis of the NESCAV (Nutrition, environment and cardiovascular health) dataset (Alkerwi et al, 2010) and show how identified dietary patterns and their associated uncertainty can be used to predict disease.

C11.5
A statistical model of breast cancer tumour growth with estimation of screening sensitivity as a function of mammographic density
L Abrahamsson[1], K Humphreys[1]
[1]Karolinska Institutet, Stockholm, Sweden

Understanding screening sensitivity and tumour progression is important for designing and evaluating screening programs for breast cancer. Relevant variables, such as tumour size, are typically only observable at time of diagnosis. How can one estimate tumour growth when the size of each tumour is only measured once? There exists information in differences between tumours found at screening and tumours found symptomatically. Stochastic models of cancer development and detection can therefore be constructed, which yield the distribution of observable variables at diagnosis. Several approaches for estimating tumour growth rates have been described, some of which simultaneously estimate (mammography) screening sensitivity. None of these approaches have incorporated mammographic density, although it is known to have a profound influence on mammographic screening sensitivity.
We describe a new approach for estimating breast cancer tumour growth which builds on recently described continuous tumour growth models and estimates mammographic screening sensitivity as a function of tumour size and mammographic density.

C12 Vaccine studies and infectious diseases

C12.1
Taking into account strains heterogeneity in the estimation of vaccine efficacy against seasonal influenza
A Benoit[1], P Lebrun[2], W Dewé[3], C Legrand[1]
[1]ISBA, Université Catholique de Louvain, Louvain-la-Neuve, Belgium, [2]Arlenda, Mont-Saint-Guibert, Belgium, [3]GSK Biologicals, Rixensart, Belgium

Influenza is an infectious disease caused by several virus strains whose repartition varies between regions and seasons. Typically, a vaccine contains 3 or 4 strains and the antigen content is annually reconsidered based on the WHO recommendation. For the same vaccine formulation, pharmaceutical regulations only require efficacy again clinical disease to be shown for a single season, which is performed through a large phase III trial. Subsequent annual modifications of the strain related portion of the vaccine only have to be validated through immunogenicity trials.
Classically, influenza vaccine efficacy (VE) trials take place over one season but over several regions assuming common VE. However, depending on the circulating strains characteristics such as their immunogenicity and their matching levels with the vaccine strains, the vaccinal protection level may vary from one season/region to another.
We argue that not taking this into account provides incomplete and unreliable response as for the benefit of the vaccine in the future. We therefore propose to run phase III VE trials over several regions and seasons in order to characterize the VE heterogeneity. We consider VE as the sum of a common quantity to all clusters (season and region) and of a random cluster-specific part. VE is reported based on a tolerance interval, providing insight on the range of future VE across seasons and regions. Our model parameters and the tolerance interval for the cluster-specific VE are estimated using Bayesian statistics. Our work will be illustrated by discussing real data examples and simulation results.

C12.2
Estimating the effects of time-since-exposure using case-control data, motivated by a study of vaccine efficacy over time
RH Keogh[1], P Mangtani[1], P Nguipdop Djomo[1], L Rodrigues[1]
[1]London School of Hygiene & Tropical Medicine, London, United Kingdom

It is sometimes of interest to study whether the effect of an exposure on an event rate depends on time since exposure. Our motivation is an investigation of whether the effect of the BCG vaccination on occurrence of tuberculosis (TB) wears off as time since vaccination increases.
In prospective studies the effects of exposure duration can be estimated using a proportional hazards model, allowing exposure effects to differ across time bands or by modelling the exposure effect as a function of time. For a variety of reasons it is of course common to study exposure effects using a case-control study. We discuss the challenges of estimating the effects of exposure duration from population-based case-control studies, which have not typically been used to incorporate information of exposure duration due to their retrospective nature and sampling which disregards time. In the motivating example, cases were historical TB cases and controls were sampled from the underlying population using frequency matching on birth cohort. Individuals and their parents were interviewed, and medical records examined, to establish whether or not they had received the BCG vaccination, and when.
Methods for making efficient use of controls at multiple times-since-vaccination are discussed and different methods of analysis considered. These include fitting a series of logistic regression models within time bands based on time-since-vaccination. Another possibility is to view the case-control sample as a form of case-cohort study and to apply a modified proportional hazards analysis. The appropriateness and efficiency of different methods are compared using simulation studies.

C12.3
Integrative analysis of high-dimensional data in clinical trials: an example in HIV vaccine development
R Thiébaut[1], B Hejblum[1], Y Levy[2]
[1]INSERM U897, Bordeaux, France, [2]Vaccine Research Institute, Creteil, France

Background: Because there is no definitive surrogate endpoint, many immunological markers are measured in HIV vaccine trials. Furthermore, the availability of high throughput assays leads to the constitution of high-dimensional data. We present a modelling strategy in 2 steps to analyse all available information in order to identify the gene signature of the observed viro-immunological response.
Methods: DALIA is a trial evaluating the response to an *ex vivo* generated DC loaded with HIV-lipopeptides in 19 HIV patients on antiretroviral therapy (ART). Gene expression in whole blood was measured by microarrays (Illumina HumanHT-12) at 14 time points. Post vaccination immune responses were evaluated using various assays. In step 1, a Time-course Gene Sets Analysis (TcGSA) was performed using hierarchical models

allowing heterogeneity in predefined gene sets and implemented in an R package. Statistical properties of this approach have been studied through simulations. Association between abundance of genes selected in step 1, immune responses at w16 and viral replication after ART interruption was analysed in step 2 using sparse-Partial Least Square.
Results: TcGSA simulations showed good statistical properties (type I and type II errors). In DALIA, although the vaccine elicited strong immunological responses, no differential expression was found with a gene-by-gene analysis. Using TcGSA, we found 69 genesets out of 260 that varied significantly during vaccination. In step 2, we show relationships between HIV-specific responses, gene expression and viral replication.
Conclusions: The new proposed approach allowed in this example to detect relevant genesets associated with the immune response and viral dynamics.

C12.4
Assessing vaccine effectiveness using observational data in the presence of hidden confounders

LR Rodgers[1], N Lin[1], W Henley[1]
[1]University of Exeter, Exeter, United Kingdom

Electronic healthcare databases, such as the Clinical Practice Datalink, have the potential to provide a wealth of information on vaccine efficacy. The problem with using observational data is that the lack of randomisation can lead to bias in estimates of treatment effect due to hidden confounders. A novel quasi-experimental design, the prior event rate ratio, incorporates information from a period prior to treatment. In these methods the ratio of period before and after treatment reflects the combined effect of known and unknown confounders. These novel designs along with traditional models are applied to an analysis of vaccine effectiveness using observational data. Our approach is to design a simulation study alongside the data analysis to test the validity of model assumptions.
We illustrate the methods using the influenza vaccination. The efficacy of this vaccination in the elderly, particularly due to lack of randomised controlled trials, is a subject of debate. The quasi-experimental methods and established models are applied to a dataset of antibiotic prescriptions before and after an influenza vaccination in elderly patients. The simulation study is designed to replicate key features of the motivating data set. Confounders, continuous and binary, are generated to test the impact of imbalance between treatment and control groups and the influence study period on treatment effect estimates. The sensitivity of the methods to the effects of hidden confounders is explored. Of particular interest is the impact of subgroups within the data, a feature in studies of the elderly due to the potential for immune senescence.

C12.5
Modeling reporting delays for outbreak detection of infectious diseases

A Noufaily[1], Y Weldeselassie[1], P Farrington[1], D Enki[2], P Garthwaite[1], N Andrews[3], A Charlett[3]
[1]The Open University, Milton Keynes, United Kingdom, [2]Plymouth University, Plymouth, United Kingdom, [3]Public Health England, London, United Kingdom

The delay that necessarily occurs between the emergence of symptoms and the identification of the cause of those symptoms affects the timeliness of detection of emerging outbreaks of infectious diseases, and hence the ability to take preventive action. We propose a new method to analyse reporting delays using a continuous time spline-based model for the hazard, along with an associated proportional hazards model. This allows analysis of both long and short delays. The delay distributions for laboratory-based surveillance data from the UK are found to have extremely long tails, the hazard at longer delays being roughly constant, suggestive of a memoryless process, though some laboratories appear to stop reporting after a certain delay. We use these findings to inform outbreak detection of infectious diseases based on laboratory reports.

Monday, 25th August 2014 – 16:00-17:30

Invited session

I3 Inverse probability weighting techniques

Organizers: Ronald Geskus and Karen Leffondré

I3.1
Iterative inverse probability weighting

R Geskus[1,2]
[1]*Academic Medical Center of Amsterdam, Amsterdam, The Netherlands,* [2]*Public Health Service of Amsterdam, Amsterdam, The Netherlands*

Inverse probability weighting (IPW) is used in many areas of statistics in order to correct for systematic or random imbalance in observed data. Examples of systematic imbalance are 1) confounding in causal inference based on observational data, 2) selection bias due to informative censoring in survival analysis, 3) differential sampling fractions in surveys. The mechanism that generates imbalance can be visualized via arrows in a directed acyclic graph. The purpose of IPW is to remove these arrows in the analysis by weighting individuals according to the observed imbalance generating mechanism. It eliminates bias in case of systematic imbalance and increases efficiency in case of random imbalance.

There is a bias-variance tradeoff in the choice of model for the weights. Rarely is a simple saturated model sufficient to eliminate imbalance. We may have continuous variables or a large amount of variables that we need to correct for. However, using a too flexible weight model may cause near or complete violation of the positivity assumption. This will generate heavy tailed weight distributions and increase bias and variance. Weight truncation has been suggested as a way to tradeoff bias and variance.

We describe an algorithm that performs IPW iteratively (IIPW). We examine performance of IIPW in a couple of point treatment simulation studies with confounding. Compared to IPW estimators, IIPW estimators 1) suffer substantially less from small sample bias, 2) are less variable, and 3) are more robust against model misspecification. We give heuristics why the method improves upon standard IPW.

I3.2
Bias-variance trade off in IPCW: Is it possible to hear the curse of dimensionality in a random forest?

TA Gerds[1]
[1]*University of Copenhagen, Department of Biostatistics, Copenhagen, Denmark*

Estimating the nuisance parameter of a semiparametric model is more efficient than not to, even if the nuisance parameter is known. This somewhat surprising result is the heart of inverse of the probability (of censoring) weighting technique. In the context of a clinical study with survival endpoint, it implies that even if the censoring mechanism is known to be independent of the covariates it is advantageous to use a working regression model to estimate the conditional censoring distribution. But, a bias is introduced into the IPCW estimate if the working model is misspecified. This is particularly delicate if the nuisance parameter is a conditional distribution function (given covariates) because the curse of dimensionality seems to prohibit a purely non-parametric approach to estimation of the weights.

In this talk I will provide an intuitive understanding of the efficiency gain, illustrate the bias-variance trade off in real data applications and by using simulations, and propose to estimate the weights based on a machine learning approach: We shall see to what extend a random forest can absorb the curse of dimensionality.

I3.3
Inverse probability weighting methods for biomarker evaluation with case cohort studies

T Cai[1]
[1]*Harvard University, Boston, United States*

Identification of novel biomarkers for risk prediction is important for both disease prevention and optimal treatment selection. However, discovering which biomarkers are useful for prediction will require the use of stored biological samples from large assembled cohorts, and thus the depletion of a finite and precious resource. To preserve these samples, the case cohort (CCH) design provides a resource-efficient sampling design, especially when the outcome is rare. However, existing methods for CCH designs focus on efficient inference of the relative hazard parameters from the Cox regression model, or have considered measures of predictive accuracy of only a single biomarker. In this talk, we will discuss inverse probability weighted approaches to deriving robust risk prediction rules under general survival models. A major obstacle for making inference under two phase studies is due to the correlation induced by the finite population sampling which prevents standard inference procedures such as the bootstrap from being used for variance estimation. We propose a novel resampling procedure to obtain p-values and confidence intervals for parameters of interest. The proposed procedure will be applied to a Danish case-cohort study of novel lipid markers for prediction cardiovascular risks.

Contributed sessions

C13 Personalized and stratified medicine I

C13.1
Interaction of treatment with a continuous variable: simulation study of significance level and power for several methods of analysis

W Sauerbrei[1], P Royston[2]
[1]University Medical Center Freiburg, Freiburg, Germany, [2]University College London, MRC Clinical Trials Unit, London, United Kingdom

Interactions between treatments and covariates in RCTs are a key topic. Standard methods for modeling treatment-covariate interactions with continuous covariates are categorization or linear functions.
Spline based methods and multivariable fractional polynomial interactions (MFPI) have been proposed as an alternative which uses full information of the data. Four variants of MFPI, allowing varying flexibility in functional form, were suggested.
In order to work toward guidance strategies in the sprit of the STRATOS initiative we have conducted a large simulation study to investigate significance level and power of the MFPI approaches, versions based on categorization and on cubic regression splines. We believe that the results provide sufficient evidence to recommend MFPI as a suitable approach to investigate interactions of treatment with a continuous variable. If subject-matter knowledge gives good arguments for a non-monotone treatment effect function, we propose to use a second-degree fractional polynomial (FP2) approach, but otherwise a first-degree fractional polynomial (FP1) function with added flexibility (FLEX3) has a power advantage and therefore is the method of choice. The FP1 class includes the linear function and the selected functions are simple, understandable and transferable.

C13.2
Comparing a marker based stratified treatment strategy with the standard treatment in a randomized clinical trial

H Sun[1], F Bretz[2], O Gerke[3], W Vach[1]
[1]Department of Medical Biometry and Statistics, Freiburg, Germany, [2]Novartis Pharma AG, Basel, Switzerland, [3]Nuclear Medicine, Odense University Hospital, Odense, Denmark

The increasing emergence of successful molecularly targeted agents (MTAs) raises the question of how to study treatment strategies suggesting a variety of different (combination) therapies based on multiple marker information.
Here we consider the situation where there already exists a stratified treatment strategy being dependent on a marker pattern, and dividing the whole population into small subpopulations. This strategy has to be compared with the standard treatment in an randomized clinical trial. Due to limited knowledge about the MTAs, we expect no benefit from the new strategy in some subpopulations. In such situation the objective should be trying to demonstrate a treatment effect for a subset of subpopulations, instead of each single subpopulation.
We consider a wide class of methods to approach this situation, allowing to select from a variety of significant subsets an optimal one according to different criteria. We present a framework to compare various methods of this class, aiming on measuring not only power (i.e. the probability to find at least one significant subset) but also the actual gain in average improvement of the outcome. Using the framework, we can observe substantial differences between the methods, which allows to give first recommendations on the choice of adequate methods.

C13.3
A framework for comparing methods for marker-based selection of treatment change

M Kechel[1], W Vach[1]
[1]Center for Medical Biometry and Medical Informatics, Freiburg, Germany

Today, biomarkers often promise to assist in choosing between two different therapeutic alternatives. Clinical trials randomizing all patients to the two alternatives and measuring the biomarker in all patients allow to check such a promise by establishing a (qualitative) interaction. Moreover, they allow to determine a cut point or a more complex decision rule to decide on the treatment for each patient. Many statistical approaches have been made to determine such cut points. In comparing such methods, often only the power of the methods, i.e. the probability to come to a (differential) decision rule, is considered. In our talk, we develop a more general framework to compare such methods. The framework is suitable for RCTs as described above, when the choice is to be made between a standard therapy and a new therapy, e.g. an add on taking into account the information given by the biomarker. In this framework we take also the size of the treatment effect in each patient into account, and we define quantities like the expected overall gain, the gain in shifters, the unused gain in stayers, etc. We apply the framework to compare different approaches based on linear and quadratic models for the marker dependent treatment effect and the use of simultaneous and pointwise confidence bands. We demonstrate that the framework is useful to obtain more insights than just considering the power.

C13.4
Analyzing treatment-by-subgroup interactions in time-to-event data - comparison of two multivariate approaches

H Sommer[1], A-S Stoehlker[1]
[1]University of Freiburg, Freiburg, Germany

For determining the differential effects of therapy dependent on individual patient characteristics, most of the common methods investigate treatment-by-covariate interactions only for one single predictor. We introduce and compare two approaches to examine the influence of several covariates simultaneously in context of survival data.
Firstly, we focus on a tree-based subgroup identification procedure called Interaction Tree proposed by Su et al. (Applied Statistics, 2011). Perfectly homogeneous subgroup-allocation is assumed implicitly but does not hold naturally. By additionally incorporating specific weights in a locally weighted Cox-regression according to a weighting scheme of Simon (StatMed, 2002) we try to mitigate this problem with a second procedure based on the subgroups from the former. For comparison, this second procedure is also applied to subgroups resulting out of another method.
Treatment selection is performed via Selection Impact curves proposed by Song and Pepe (Biometrics, 2004), which we apply on the Interaction Tree subgroups. For this purpose, sufficient heterogeneity of the treatment effect between these subgroups is necessary, which is tested with Cochran´s Q commonly used in meta-analyses.
Finally, the approaches are illustrated with data of a randomized clinical trial and we investigated comprehensive simulation study covering various realistic scenarios to examine strengths and weaknesses.

C13.5

Bounds for causal interaction

A Sjolander[1], W Lee[2], H Kallberg[1], Y Pawitan[1]
[1]Karolinska Institutet, Stockholm, Sweden, [2]Inha University, Incheon, Republic of Korea

Interaction in statistical and epidemiological literature appears in at least two forms: (i) statistical interaction as deviation from additive models, and (ii) causal interaction as latent classes in the population that describe how subjects causally respond to risk factors, e.g. a class of people that develops a disease if and only if two risk factors are present. Almost all analyses of interaction in the literature are of the first type, which is conceptually problematic since statistical interaction is scale-dependent. For example, lack of interaction in the multiplicative scale -- such as the logistic model -- must mean the existence of additive interaction in the linear probability scale. In contrast, causal interaction is invariant to the choice of scale, but has the disadvantage that the latent classes are not estimable from the observed data. A well-known solution is simply to test the presence of the causal interaction, but this does not tell us its magnitude. In this work we solve the problem by providing lower and upper bounds for the causal interaction between two risk factors with arbitrary number of levels. The magnitude is well captured when the bounds are tight. In a real data example of rheumatoid arthritis, we observe these tight bounds for two genetic risk factors when we further assume that they have monotone effects. In conclusion, the concept of causal interaction is a useful general data-analytic concept, complementary to the standard statistical interaction, and can be practically assessed in commonly available datasets.

C14 Meta-analysis

C14.1

Addressing continuous missing outcomes in pair-wise and network meta-analysis

D Mavridis[1,2], IR White[3], JP Higgins[4,5], A Cipriani[6], G Salanti[2]
[1]University of Ioannina, Department of Primary Education, Ioannina, Greece, [2]University of Ioannina, School of Medicine, Ioannina, Greece, [3]MRC Biostatistics Unit, Cambridge, United Kingdom, [4]University of Bristol, Bristol, United Kingdom, [5]University of York, York, United Kingdom, [6]University of Oxford, Oxford, United Kingdom

Missing outcome data may affect results of individual trials and their meta-analysis by reducing precision and, if the missing-at-random (MAR) assumption does not hold, by introducing bias in the estimated treatment effects.
We propose a pattern-mixture model to estimate meta-analytic treatment effects for continuous outcomes when these are missing for some of the randomised individuals. Our model is applicable to both pairwise and network meta-analysis (NMA) and makes explicit assumptions about parameters in the unobserved data conditional on observed data. Specifically, in each study we quantify departures from the MAR assumption via a missingness parameter that relates the outcome means in the observed and missing data. This leads to an adjusted estimate of the effect size and uncertainty in this estimate is estimated using either Monte Carlo methods or a Taylor series approximation. The adjusted effect size accounts properly for the fact that some of the outcome data are missing.
We illustrate the suggested methodology using a meta-analysis of studies comparing mirtazapine to placebo for depression and a NMA involving nine antidepressants. Summary mean difference of mirtazapine relative to placebo decreases from -2.34(95% -4.67,0) to -2.66 (95% -4.90,-0.41) as we depart from the MAR assumption.When we account for missing outcome data, study weights depend on the missingness rate, and summary results may change if missing rates vary considerably across studies. As we depart from the MAR assumption, within-study uncertainty increases but between-study heterogeneity decreases, and changes in summary estimates depend on the trade-off between these two sources of variability.

C14.2

The impact of choice of heterogeneity estimator in meta-analysis

D Langan[1], J Higgins[1], M Simmonds[1]
[1]University of York, York, United Kingdom

In meta-analyses, effect estimates from different studies usually vary above and beyond what would be expected by chance alone, due to inherent variation in the design and conduct of the studies. This type of variance is known as heterogeneity and is most commonly estimated using a moment-based approach described by DerSimonian & Laird. However, this method has been shown through simulation studies to produce biased estimates. Alternative methods to estimate the heterogeneity variance include proposals from Paule & Mandel and Hartung & Makambi, and estimators derived from maximum likelihood and restricted maximum likelihood approaches. This presentation compares these methods and the impact they have on the results of a meta-analysis using 12,894 meta-analyses extracted from the Cochrane Database of Systematic Reviews. The methods are compared in terms of: (1) the extent of heterogeneity, expressed in terms of the I^2 statistic; (2) the overall effect estimate; (3) the 95% confidence interval of the overall effect estimate; and (4) p-values testing the hypothesis of no effect. Results suggest that in some meta-analyses, I^2 estimates can differ by more than 50% when different heterogeneity estimators are used. Conclusions based naively on statistical significance (at a 5% level) were discordant for at least one pair of estimators in 7.4% of meta-analyses, indicating that the choice of heterogeneity estimator can affect the conclusions of a meta-analysis. These findings highlight the need for a greater understanding of why heterogeneity estimates disagree and the need for guidance on alternatives to the DerSimonian & Laird method.

C14.3

A re-analysis of the Cochrane Library data: the dangers of unobserved heterogeneity in meta-analyses

E Kontopantelis[1], D Springate[1], D Reeves[1]
[1]University of Manchester, Manchester, United Kingdom

Background: Heterogeneity has a key role in meta-analysis methods and can greatly affect conclusions. However, true levels of heterogeneity are unknown and often researchers assume homogeneity.
Methods and findings: We accessed 57,397 meta-analyses, available in the Cochrane Library in August 2012. Using simulated data we assessed the performance of various meta-analysis methods in different scenarios. The prevalence of a zero heterogeneity estimate in the simulated scenarios was compared with that in the Cochrane data, to estimate the degree of unobserved heterogeneity in the latter. We re-analysed all meta-analyses and assessed the sensitivity of the statistical conclusions.
Levels of unobserved heterogeneity in the Cochrane data appeared to be high, especially for small meta-analyses. A bootstrapped version of the DerSimonian-Laird approach performed best in both detecting heterogeneity and in returning more accurate overall effect estimates. Re-analysing all meta-analyses with this new method we found that in cases where heterogeneity had originally been detected but ignored, 17-20% of the statistical conclusions changed.
Conclusions: When evidence for heterogeneity is lacking, standard practice is to assume homogeneity and apply a simpler fixed-effect meta-analysis. We find that assuming homogeneity often results in a misleading analysis, since heterogeneity is very likely present but undetected. Our new method represents a small improvement but the problem largely

remains, especially for very small meta-analyses. One solution is to test the sensitivity of the meta-analysis conclusions to assumed moderate and large degrees of heterogeneity. Equally, whenever heterogeneity is detected, it should not be ignored.

C14.4
Estimating heterogeneity when pooling proportions in a meta analysis

A Benedetti[1], R Platt[1], D Thomas[1]
[1]*McGill University, Montreal, Canada*

The volume of medical research has exploded over the last several decades, increasing the focus on evidence based medicine, but also posing challenges on how to synthesize that evidence. Meta-analytic methods sit at the crossroads of these trends.
We consider pooling proportions via a generalized linear mixed model (GLMM) as in (Hamza et al. J Clin Epidemiol. 61(1):41-51). Contrary to an inverse variance weighted pooling (e.g. der Simonian Laird) one of the advantages of the GLMM is the ability to deal with proportions that are 0 or 1 without applying a continuity correction. Our interest lies in estimating the overall pooled proportion and heterogeneity as defined by the interstudy variability in situations when there are proportions that are 0 or 1.
Via simulation study, we consider varying the study sizes, number of studies, true interstudy heterogeneity, and true pooled proportion. We compare estimating the pooled proportion and interstudy variability, via a GLMM, estimated via penalized quasi-likelihood (PQL) and adaptive Gaussian hermite quadrature (AGHQ), as well as Bayesian approaches. We further compare our estimates of heterogeneity to the I2, as estimated from a der Simonian and Laird random effects model with a continuity correction. We demonstrate our methods in a real life example.
Preliminary results suggest that interstudy heterogeneity is underestimated when the model is estimated via PQL or AGHQ - though to a lesser degree when AGHQ is used when the level of heterogeneity is high. The degree of underestimation increased as the number of proportions that were 1 increased.

C14.5
Meta-analysis and the surgeon general's report on smoking and health

M Schumacher[1], G Rücker[1], G Schwarzer[1]
[1]*Medical Center-University of Freiburg, Freiburg, Germany*

Although first meta-analyses in medicine have already been conducted at the beginning of 20th century, its major breakthrough came with the activities of the Cochrane Collaboration during the 1990s. It is less known that the landmark report on "Smoking and Health" to the Surgeon General of the Public Health Service, published fifty years ago, makes substantially use of meta-analyses performed by statistician William G. Cochran.
Based on summary data given in the report we reconstructed meta-analyses of seven large, prospective studies that were initiated in the 1950s by concentrating on overall and lung cancer mortality (N Engl J Med 370; 2: 186-188; January 9, 2014). While visualization of results including confidence intervals was largely neglected in the report we are able to give a vivid impression of the overwhelming evidence on the harmful effects of cigarette smoking. We will put William G. Cochran's contribution into the context of the so-called lung cancer controversy in which other prominent statisticians, e.g. Sir Ronald Fisher, played a major role. In contrast to the latter who selected a specific study that supported his personal view, Cochran took an impartial systematic approach for evaluation and followed the major steps of a modern systematic review including an appraisal of risk of bias based on sensitivity analysis. For that he used state-of-the-art statistical methodology while neglecting visualization of results. Although substantially contributing to an important public policy issue this work is often overlooked and deserves much more attention.

C15 Longitudinal data analysis I

C15.1
Longitudinal models with outcome dependent follow-up times

I Sousa[1]
[1]*Universidade do Minho, Guimarães, Portugal*

In longitudinal studies individuals are measured repeatedly over a period of time. In observational studies individuals have different number of measurements assessed at different times. In standard longitudinal models (Diggle et al. 2002), the follow up time process is assumed deterministic, the follow-up time process is noninformative about the outcome longitudinal process of interest. However, in medical research a patient is usually measured according to their clinical condition. Therefore, follow-up time process is considered dependent of the longitudinal outcome process and it should not be considered deterministic. The classical longitudinal analysis does not consider the dependence that can exist between the follow-up time process and the longitudinal outcome process. In this work we propose to joint model the longitudinal process and the follow-up time process.
We propose a model where the follow-up time process is stochastic. The model is described through the joint distribution of the observed process and the follow-up time process. Estimation of model parameters is through maximum likelihood, where a Monte Carlo approximation is necessary. We conducted a simulation study of longitudinal data where model parameter estimates are compared, when using the model proposed and the model in Lipsitz et al. (2002). Finally, the model proposed is applied to data of an observational study on kidney function.

References:

Diggle, P.J., Liang, K-Y., Zeger, S.L. 2002: Analysis of Longitudinal Data, Oxford: Clarendon Press.
Lipsitz, S.R., Fitzmaurice, G.M., Ibrahim, J.G., Gelber, R., Lipshultz, S. 2002: Parameter estimation in longitudinal studies with outcome-dependent follow-up, Biometrics, 58, 50 - 59.

C15.2
Joint modelling of longitudinal and survival data incorporating delayed entry: application to longitudinal mammographic breast density and breast cancer survival

MJ Crowther[1], TM-L Andersson[2], PC Lambert[1,2], K Humphreys[2]
[1]*University of Leicester, Leicester, United Kingdom,* [2]*Karolinska Institutet, Stockholm, Sweden*

This work is motivated by a study in breast cancer, where in particular, we are interested in how changes in mammographic density are associated with survival. This type of question, where a longitudinally measured biomarker is associated with the risk of an event, can be addressed using joint modelling. The most common approach to joint modelling combines a longitudinal mixed effects model for the repeated measurements with a proportional hazards survival model, where the models are linked through shared random effects.
The analysis in the motivating study is complicated by the fact that patients are only included if they have at least two density measurements. Therefore patients do not become at risk of the event of interest until the time of second measurement, which is by definition after the baseline of diagnosis. Consequently, delayed entry needs to be incorporated into the modelling framework. The extension to delayed entry requires a second set of numerical integration, beyond that required in a standard joint model.
We therefore implement two sets of fully adaptive Gauss-Hermite quadrature with nested Gauss-Kronrod quadrature, conducted simultaneously, to evaluate the likelihood. We evaluate the importance of accounting for

delayed entry through a simulation study. To maintain flexibility in modelling the breast cancer data, we use restricted cubic splines to model the baseline hazard function, and model the longitudinal trajectory using fractional polynomials. User friendly Stata software is provided.

C15.3
Dynamic time process model for the association among two longitudinal markers in the presence of survival: application to healthABC cohort

D Geva[1], DR Shahar[1], TB Harris[2], S Tepper[1], G Molenberghs[3], M Friger[1]
[1]Ben Gurion University, Beer Sheeba, Israel, [2]Laboratory of Epidemiology, Demography, and Biometry, Bethesda, United States, [3]Katholieke Universiteit Leuven, Leuven, Belgium

Background: Further expansion of joint models to include several markers led to the formulation of the *joint model with multiple longitudinal-markers. D*ue to the integral over the random effects, estimation of such models is complex. Thus, in practice a model to study the impact of an additional marker on the joint longitudinal-survival outcome is lacking. The aim of our work is to offer an appropriate model to study the impact of a single marker measured repeatedly over time on longitudinal outcome in the present of survival.
Methods and Results: Founded on the shared parameter joint model (SREM), we propose the *Dynamic Time Process Model(DTPM)* as follows: For each time cut-points t=1,2,3…T obtain the trajectory of the marker by heterogynous latent class mixed model for the period up to t. Then, use the trajectory class in the SREM for time period beyond t. The model result in a T frames dynamically capturing the marker trajectory up to t and the joint longitudinal-survival projection beyond t. We wrote an R function, dtpm() and illustrate the model on the Health- ABC cohort data and again explore its boundaries using simulation.
Remark: *The proposed model* provides an elegant analytical framework to study the impact of marker- trajectories on longitudinal outcome in the presence of survival. The unique feature is the dynamic progression before and after time cut-point. Although, the proposed model does not have a natural expansion to multiple outcomes, it is valuable for testing new hypotheses for the joint survival-longitudinal setting.

C15.4 *Student Conference Award*
Combined dynamic predictions using joint models of multiple longitudinal outcomes and competing risk data

E-R Andrinopoulou[1], D Rizopoulos[1], JJM Takkenberg[1], E Lesaffre[1]
[1]Erasmus Medical Center, Rotterdam, The Netherlands

Nowadays there is increased medical interest in personalized medicine thereby tailoring decision making to the needs of the individual patient. In this work we focus on the statistical methodology for providing subject-specific predictions of survival probabilities used in this context. Our developments find their motivation in a Dutch study at the cardiothoracic department of the Erasmus Medical Center. It is our aim to utilize the available follow-up measurements of the current patients to predict both survival and freedom from re-intervention for future patients. Since the human tissue has limited durability due to calcification and degeneration resulting in valvular dysfunction, it is of interest to physicians to predict the need for future re-interventions using all available repeated echo measurements.
To analyze the data and obtain subject-specific risk predictions we use the joint modeling framework. In this work we extend the concept of prediction to multiple longitudinal combined with competing risk survival outcomes and we derive the dynamically updated cumulative incidence functions. Moreover, we investigate whether different features of the longitudinal processes would change significantly the prediction for the events of interest. Our final contribution focuses on optimizing the quality of the derived predictions. In particular, instead of choosing one final model over a list of candidate models, we propose to suitably combine predictions from all considered models using Bayesian model averaging (BMA). The advantage of using BMA in this setting is that predictions are tailored to each individual patient because the model weights are both subject- and time-dependent.

C15.5
Development and validation of individualized dynamic predictions based on repeated biomarker data according to scenarios of new treatments

M Sène[1], JM Taylor[2], H Jacqmin-Gadda[1], C Proust-Lima[1]
[1]INSERM U897, Bordeaux, France, [2]University of Michigan, Ann Arbor, United States

With the emergence of rich information on biomarkers after treatments, new types of prognostic tools are being developed: dynamic prognostic tools that can be updated at each new biomarker measurement. Such predictions are of particular interest in oncology where after an initial treatment, patients are monitored with repeated biomarker data. However, in such setting, patients may receive second treatments to slow down the progression of the disease, which is not currently handled by dynamic predictive tools.
This paper aims to develop and validate dynamic individual predictions that allow the possibility of a new treatment in order to help understand the benefit of initiating new treatments during the monitoring period. The prediction of the event in the next x years is done under two scenarios: (1) the patient initiates immediately a second treatment, (2) the patient does not initiate any treatment in the next x years.
The dynamic predictions are derived from joint (shared random-effect) models. The predictive accuracy of the dynamic predictions is evaluated with two measures (the Brier score and the prognostic information) for which approximated cross-validated estimators that correct the usual over-optimistic predictive performances are proposed.
Applied to the monitoring data of more than 2300 men initially treated by radiation therapy for a localized prostate cancer, different specifications for the dependence between the PSA repeated measures, the initiation of a second treatment (hormonal therapy) and the risk of clinical recurrence are investigated and compared.

C16 High-dimensional data analysis II

C16.1
A comparison of prediction models for gene expression data by resampling techniques

S Yilmaz Isikhan[1], E Karabulut[1], CR Alpar[1]
[1]Hacettepe University, Ankara, Turkey

Gene set methods aim to assess the overall evidence of association of a set of genes with a phenotype, such as dose values or a quantitative trait. The purpose of this study is to compare the effects of such frequently used methods of generalization as Bootstrap, Cross Validation on some machine learning regression methods.
To assess the Root Mean Squared Error and Coefficient of Determination performance of these generalization methods for different regression methods, an extensive simulation study was completed in which the scenarios varied according to: sample size, number of genes associated with the phenotype, regression coefficients, correlation between expression of genes within a gene set and the original correlation structure of

the gene sequence. Generalization methods based on resampling such as Bootstrap and Cross Validation methods were considered. The regression methods used are Support Vector Regression (SVR) and Decision Trees Regression (DTR). As a result of simulation studies, actual performance of the regression techniques for gene data was approximated by use of Bootstrap and Cross Validation methods.
Overall, when results are examined for each simulation senario, it appears that the bootstrap method yields a lower error of estimation than Cross Validation.

C16.2
An application of sequential meta-analysis to gene expression studies

PW Novianti[1], VL Jong[1,2], I van der Tweel[1], KCB Roes[1], MJC Eijkemans[1]
[1]University Medical Center Utrecht, Utrecht, The Netherlands, [2]Erasmus Medical Center, Rotterdam, The Netherlands

Most of the discoveries from gene expression data are driven by a single study, claiming an optimal subset of genes that play a key role in a specific disease. Results from a differentially expressed genes (DEGs) analysis may be used in drug development. An optimal new drug that is based on the results of the DEGs analysis is potentially hard to achieve, due to false-positive findings.
Meta-analyzing the available datasets potentially helps in getting concordant results so that a real life application may be more successful. Sequential meta-analysis (SMA) is an approach for combining studies in chronological order by preserving the type I error and pre-specifying the statistical power to detect a given effect size. This study focuses on the application of SMA (following Whitehead's triangular test boundaries approach) to find gene expression signatures across microarray experiments in acute myeloid leukemia (AML). Seven raw datasets on AML patients versus healthy controls fulfilled our predefined search criteria and were downloaded from the ArrayExpress repository. The boundaries in the triangular test were constructed for a pre-specified effect size θ_R=0.8, a type 1 error α=0.5% and the power 1-β=80% . The between-study variance was estimated by the Paule-Mandel method. We found 169 DEGs, based on the cumulative information of the seven experiments. Meanwhile, Bonferroni correction of α=5% to α=0.0007% yielded 24 DEGs.
This study shows whether there is enough evidence at a certain time point to draw a conclusion for a particular gene or to hold the conclusion until the evidence is adequate.

C16.3
Ensemble classifiers in the high-dimensional setting with class-imbalanced data

R Blagus[1], L Lusa[1]
[1]University of Ljubljana, Ljubljana, Slovenia

The goal of biomedical studies is often to develop a rule (classifier) to predict the class-membership of new samples based on the values of some measured variables. Boosted classifiers combine the votes of a base classifier trained on modified versions of the training data; typically boosting improves the accuracy of the base classifier and reduces its variance. However, the usefulness of boosting remains questionable when data are high-dimensional, where the number variables greatly exceeds the number of samples.
We consider AdaBoost.M1, gradient boosting and logistiboost and use classification trees as base classifiers. On simulated and real high-dimensional data boosting algorithms often do not improve upon their base classifier; the best performance is achieved by stochastic gradient boosting, while AdaBoost.M1 and gradient boosting can perform very poorly with small samples.
We propose a straightforward, yet efficient, modification of the AdaBoost.M1 algorithmthat can perform well also in these settings.
It is known that high-dimensionality exacerbates the class-imbalance bias, where most samples are assigned to the majority class unless the differences between the classes are large; so far the performance of boosting on imbalanced high-dimensional data was not investigated. Our results show that boosting can increase the class imbalance bias of its base classifier. We show that this problem can be avoided by using boosting on previously down-sized training set, or by using more complex ensembles that combine boosting with bootstrap aggregating.

C16.4
Combining techniques for screening and evaluating interaction terms on high-dimensional time-to-event data

I Hoffmann[1], M Sariyar[2], H Binder[1]
[1]Universitätsmedizin Mainz, Johannes-Gutenberg Universität, Mainz, Germany, [2]Universitätsmedizin Berlin, Charité, Berlin, Germany

When linking high-dimensional molecular covariates to some clinical endpoint, e.g., when using gene expression measurements for prognosis, sparse regression techniques are destined to provide a short list of marginal or main effects.
While interactions are highly likely to be present in molecular applications, it is still very challenging to identify interactions terms that should be considered together with potential main effects for predicting a clinical outcome.
Additionally it is well known that gene expression data is highly correlated. To address this, we present a strategy based on the combination of a regularized regression approach for fitting prognostic models, and different approaches for interaction screening.
We specifically consider componentwise likelihood-based boosting to select main effects for a prognostic model in a time-to-event setting.
Random survival forests and logic regression are considered for preselecting the potential interaction terms. [h1]
Specifically, the screening step considers permutation accuracy importance and pairwise inclusion frequencies.
The benefits and limits of the different interaction screening approaches are evaluated in a simulation study with respect to prediction performance and sensitivity concerning main effects and interactions.
We consider scenarios with different relative main effect and interaction effect sizes, and with different correlation structures.
The proposed strategy for interaction screening and prognostic model building is further illustrated with gene expression data from patients with diffuse large B-cell lymphoma.

C16.5
Comparing models of location and scale for genome-wide DNA methylation data

S Wahl[1], N Fenske[2], S Zeilinger[1], K Suhre[1,3], C Gieger[1], A Peters[1], M Waldenberger[1], H Grallert[1], M Schmid[2,4]
[1]Helmholtz Zentrum München, Neuherberg, Germany, [2]Ludwig-Maximilians-Universität München, Munich, Germany, [3]Weill Cornell Medical College in Qatar, Doha, Qatar, [4]Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany

With the help of methylome-wide association studies, increasing knowledge on the role of DNA methylation in disease processes is obtained. In terms of statistical analysis, specific challenges arise from the characteristics of methylation data. First, they represent proportions with skewed and heteroscedastic distributions.
Traditional strategies assuming a normally distributed response might therefore be inappropriate. Second, recent evidence suggests that not only mean differences but also variability in site-specific DNA methylation

is involved in disease processes, including cancer.
The purpose of this study was to compare different modeling strategies for methylation data in terms of model performance and performance of downstream hypothesis tests. Specifically, we used the generalized additive models for location, scale and shape (GAMLSS) framework to compare beta regression with Gaussian regression on raw, logit2 and arcsine square root transformed methylation data, with and without modeling a covariate effect on the scale parameter.
Using simulated and real data, we show that model performance is improved in models of location and scale, specifically on logit2-transformed methylation values, as compared to traditional models of location only. Our results further suggest that models of location and scale are specifically sensitive towards violations of the distribution assumption and towards outliers in the methylation data. Therefore, a resampling procedure is proposed as a mode of inference and shown to diminish type I error rate in practically relevant settings. We apply the proposed method to genome-wide data from the large population-based KORA study and reveal biologically relevant phenotypic associations with methylation variability.

C17 Adaptive designs I

C17.1
Dose-escalation using safety and biomarker data: a Bayesian adaptive approach

D Sabanes Bove[1], G Palermo[1]
[1]F. Hoffman-La Roche Ltd., Basel, Switzerland

In early clinical dose-escalation studies typically the target is to find a dose with a certain toxicity probability, say between 20 and 35%. Therefore, the dose-escalation is only driven by safety data, ignoring potential biomarkers for efficacy. This strategy relies on the assumptions that the efficacy increases monotonically with the dose, and that such levels of toxicity can actually be reached. However, for targeted monoclonal antibody therapies it is often the case that no dose-limiting toxicity is observed, such that dose selection cannot solely rely on safety, but must take into account pharmacodynamics (PD) data. Therefore we propose a Bayesian adaptive dose escalation framework that also uses a continuous biomarker to find the dose with maximum PD effect within certain safety constraint. Our approach builds on the work by Bekele and Shen (Biometrics, 2005), which uses the probit model to transform the binary safety outcome into a continuous variable, allowing to model safety and biomarker data by a bivariate normal distribution. We compare our approach with alternative dual endpoint designs, and illustrate the performance with simulation results.

C17.2
Bayesian adaptive dose-escalation procedures utilizing a gain function with binary and continuous responses

WY Yeung[1], T Jaki[1], J Whitehead[1], B Reigner[2], U Beyer[2], C Diack[2]
[1]Lancaster University, Lancaster, United Kingdom, [2]Hoffmann-La Roche Ltd., Basel, Switzerland

One of the main aims in early phase clinical trials is to identify a relatively safe dose with an indication of therapeutic benefit to administer to patients in further studies. Therefore, dose-limiting events (DLEs) and efficacy responses of subjects should be included in the dose-escalation procedure. Several methodologies have been suggested for incorporating both DLEs and efficacy responses in escalation trials in oncology.
In the presentation, we describe and evaluate a dose-escalation procedure for use in non-oncology trials that utilizes measures of efficacy and safety. This is a Bayesian adaptive approach based on one binary response (occurrence of a DLE) and one continuous response (a measure of potential efficacy) per subject. A logistic regression and a linear log-log relationship are used respectively to model the binary DLEs and the continuous efficacy responses. A gain function concerning both the DLEs and efficacy responses is used to determine the dose to administer to the next cohort of subjects. Stopping rules are proposed to enable efficient decision making. Simulation results shows that our approach performs better than one that takes account of DLEs responses only. To assess the robustness of the approach, scenarios where the efficacy responses of subjects are generated from an Emax model, but treated as coming from a linear log-log model are also considered. This evaluation shows that the simpler log-log model leads to robust recommendations even under model misspecification.

C17.3
Adaptive dose-finding designs to identify multiple doses that achieve multiple response targets

S Bond[1,2], A Mander[2], J Todd[3], L Wicker[3], F Waldron-Lynch[3]
[1]Cambridge Clinical Trials Unit, Cambridge, United Kingdom, [2]MRC Biostatistics Unit Hub for Trials Methodology Research, Cambridge, United Kingdom, [3]JDRF/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge, United Kingdom

The objective of the "Adaptive study of IL-2 dose on regulatory T cells in type 1 diabetes (DILT1D)" (NCT 01827735) was to identify doses of interleukin-2 that achieve targeted increases in the T regulatory cell population in recently diagnosed type 1 diabetes participants. DILT1D aimed to identify a minimally and a maximal effective dose in a limited number of participants (40) that may be repeatedly administered in future late phase trials.
The dose was administered subcutaneously so can be chosen from a continuous range up to a maximum determined by tolerability. The design has an initial learning phase where pairs of patients were assigned to five pre-assigned doses. The next phase was fully sequential with an interim analysis after each patient to determine the choice of dose based on the optimality criterion to minimise the determinant of the covariance of the estimated target doses. The dose-choice algorithm assumes that a specific parametric dose-response model is the true relationship, and so a variety of models were considered at the interims and dose determining committee approved all treatment decisions.
The estimated dose-response curves and the estimated target doses from the final study data are presented. We consider the statistical lessons learnt during the simulations performed pre-trial and practical lessons learnt whilst conducting the trial and assigning doses.

C17.4
Bayesian adaptive designs for biomarker trials with biomarker discovery

J Wason[1]
[1]MRC Biostatistics Unit, Cambridge, United Kingdom

Response to treatment is highly heterogeneous in many diseases and, in particular cancer. Increased availability of biomarkers and targeted treatments has led to the urgent need for new trial designs to efficiently test treatments in patient subgroups. In this presentation I propose a novel Bayesian adaptive randomisation (BAR) design for use in multi-arm phase II trials where there are biomarkers that are thought to be predictive for the effect of different treatments. This design is motivated by a phase II neo-adjuvant breast cancer trial. The proposed design starts by using pre-specified 'pairings' of linked biomarkers and experimental treatments, with patients randomized to the control treatment or to experimental treatments that are paired with biomarkers they are positive for. At interim analyses, the results of patients assessed so far are used to update the allocation probabilities. If the linked treatments are truly effective, the allo-

cation to them remains high; on the other hand if they are ineffective, the allocation changes over the course of the trial to ones that are effective. The proposed design has a high power to recommend treatments that work well in subgroups, especially if the initial pairings were suitable. Also considered is a biomarker discovery step, where a new biomarker can be substituted in during the trial. This can lead to increased power when the new biomarker is truly predictive for one of the treatments.

C17.5
Design of telehealth trials - introducing adaptive approaches
LM Law[1], J Wason[1]
[1]*MRC Biostatistics Unit, Cambridge, United Kingdom*

Telehealth is the use of technology to allow communication of information between patient and care-provider whilst the patient is outside the clinical environment, e.g. in their own home. The range of telehealth is broad, from the self-monitoring of blood glucose levels in diabetics to patients of mental illness receiving therapy treatment online. The field of telehealth and telemedicine is expanding as the need to improve efficiency of health care becomes more pressing. The decision to implement a telehealth system can be an expensive undertaking that impacts a large number of patients and other stakeholders. It is important that the decision is fully supported by accurate evaluation of telehealth interventions. Numerous reviews of telehealth have described the evidence base as inconsistent. In response they call for larger, more rigorously controlled trials, and trials which go beyond evaluation of clinical effectiveness alone. Adaptive designs could be ideal for addressing these needs. This presentation discusses various options of adaptive design, which have so far been applied in drug trials only. These include sample size reviews to address uncertain parameters, group sequential and multi-arm multi-stage trials to improve efficiency, and enrichment designs to target the patient population that responds best to the intervention. The presentation will then focus in on an example of a telehealth study, using simulated data to demonstrate the benefit of employing an adaptive design over a standard design.

C18 Binary and count data analysis

C18.1
Multiple comparisons of treatments with highly skewed ordinal responses
T-Y Lu[1], W-Y Poon[2], SH Cheung[2]
[1]*China Jiliang University, Hangzhou, China,* [2]*The Chinese University of Hong Kong, Hong Kong, China*

Clinical studies frequently involve the comparisons of treatments with ordinal responses. The Wilcoxon-Mann-Whitney test and its modified versions based on the proportional odds assumption are popular methods used to compare treatments with ordinal responses. However, it has long been recognized that the validity of these methods depends heavily on the equal variance assumption. A recently proposed latent normal model has been shown to be a better alternative when treatments are having heterogeneous variances. However, for highly skewed ordinal data, the latent normal model that relies on the assumption of symmetric underlying distributions does not perform satisfactorily. To remedy the problem, we propose a new approach for treatment comparisons for highly skewed ordinal responses, with the adoption of the latent Weibull model for multiple comparisons, including multiple comparisons with a control and pairwise comparisons. Our findings indicate that this new approach is superior to the latent normal model. Data from clinical studies are also used to illustrate our proposed procedure.

C18.2
Calculating confidence intervals for risk differences by means of MOVER-R
R Bender[1], RG Newcombe[2]
[1]*Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany,* [2]*Cardiff University, Cardiff, United Kingdom*

In Cochrane reviews as well as in the GRADE system absolute estimates of treatment effect are frequently calculated by using relative risk (RR) estimates based on a meta-analysis in combination with an independent baseline risk (BR) estimate. Spencer et al. (*BMJ* 2012; 345: e7401) pointed out that GRADE and all other systems for rating confidence in absolute treatment effect estimates do not fully address uncertainties in BR estimates. If BR and RR are estimated from different independent sources, confidence limits for the corresponding RD can be calculated from those for BR and RR by a procedure called method of variance estimates recovery (MOVER-R) according to Newcombe (*Stat. Methods Med. Res.* 2013). This method is explained and applied to examples. The resulting confidence intervals are compared to those obtained by the method currently used in Cochrane reviews, and to those obtained by the naive method of directly combining the confidence limits for RR and BR. It is shown that a simple and effective method is available to calculate confidence intervals for the absolute treatment effect from independent interval estimates of BR and RR taking both sources of uncertainty into account. This method should be applied in practice.

C18.3
Misspecified Poisson regression models for large-scale registry data: problems with "large n and small p"
R Grøn[1], TA Gerds[1], PK Andersen[1]
[1]*Section of Biostatistics, University of Copenhagen, Copenhagen K, Denmark*

Poisson regression based on registry data is an important tool in applied epidemiology which is used to study the association between exposure and event rates. In this talk we will illustrate problems related to "small p and large n" where p is the number of available covariates and n the sample size.
Specifically, we are concerned with modeling options when there are multiple timescales and time-varying covariates which can have time-varying effects. One problem is that tests for proportional hazard assumptions, interactions of exposure with other observed variables, and linearity of the exposure effects have large power due to the large sample size, and will often indicate statistical significance even for numerically small deviations that have no interest for the subject matter. In practice this insight may lead to simple working models (which are then likely misspecified and potentially confounded).
To support and improve conclusions drawn from such models, we shall discuss the use of robust standard errors, the choice of time-scales, and sensitivity analysis. The methods are illustrated with data from the Danish national registries.

C18.4

Bayesian analysis of zero-inflated beta regression models with application to quality of life and functional outcomes

L Sharples[1], C Jackson[2]
[1]Clinical Trials Research Unit, Leeds, United Kingdom, [2]MRC Biostatistics Unit, Cambridge, United Kingdom

Methods for zero-inflated Poisson and negative-binomial distributions are established and incorporated into standard software. Zero-inflated beta models have received less attention although they arise in several contexts. For example, renal transplantation recipients listed for re-transplant can be classed as unreactive (frequency=0) or reactive to a proportion of the potential donor pool (frequency $\in (0,1)$). In clinical trials quality of life and utility are often measured on a scale bounded above and below, with a substantial proportion of patients exhibiting a floor (or ceiling) effect. In both these cases interest surrounds estimation of model parameters including patient and treatment effects.

Although the Beta distribution is very flexible its coverage does not include zero or one, and so it cannot be used to model zero-inflated data. We show how the response can be modelled using a mixture in which (i) the non-zero responses arise from a (suitably-parameterised) Beta distribution, (ii) zero observations are Bernoulli random variables with probability *p*. Covariates can be included for both the probability of a zero response and the level of (non-zero) response, using regression. Multiple responses per individual can be incorporated using random effects. MCMC implementation is straightforward and flexible enough to accommodate missing data under a *missing at random* assumption.

However, the Bayesian paradigm requires careful specification of priors. Competing models can be compared using the DIC and goodness of fit assessed by comparing observations with model-predictions. We discuss these issues through applications to renal transplantation and quality of life outcomes in RCTs.

C18.5

Count data analysis in nutrition clinical trials

Y Yavuz[1], S Swinkels[1]
[1]Danone Nutricia Research, Biometrics, Utrecht, The Netherlands

In Early Life Nutrition division of Danone Nutricia Research, the analysis of count data such as the number of infections, number of hospital visits, number of doctor-diagnosed diarrhea, is of interest in many clinical trials. Poisson regression models would be the simplest standard framework for the analysis of such data.

However, in real life, count data do not always meet the assumption of equal variance-mean relationship induced from Poisson distribution leading to over-(or under-)dispersion. The source of the over- (or under-)dispersion could be due to a higher than expected occurrence of zero counts. A toddler may have no infection either because of his/her resistance to the infection, or simply because no disease spores have landed on him/her. This is the distinction between structural zeros, which are inevitable, and sampling zeros, which occur by chance. Another source of over-dispersion might be the fact that having an infection might make individuals more vulnerable for a second one.

We demonstrate the use of four different models for over-dispersed count data with certain levels of zero inflation: Poisson, negative binomial, zero-inflated Poisson and zero-inflated negative binomial models. We discuss the performance of the models using data from a clinical trial on the number of infections in toddlers during a 12 month period.

Tuesday, 26th August 2014 – 9:00-10:30

Invited session

I4 New methods to control for unmeasured confounding

Organizer: Michal Abrahamowicz

I4.1
Bias sensitivity analysis of unmeasured confounding

RH Groenwold[1]
[1]*UMC Utrecht, Utrecht, The Netherlands*

Observational epidemiologic research is prone to confounding bias, particularly due to unmeasured confounding. Bias sensitivity analysis can help to guide discussions on the possible direction, magnitude, and impact of unmeasured confounding. Bias sensitivity analysis of unmeasured confounding often focuses on a single unmeasured confounder, which may summarize the information of multiple weak unmeasured confounders. Although easy to evaluate, the downside of such a summary confounder is that it is not very intuitive and may be hard to conceptualize. Alternatively, the impact of multiple (possibly weak) confounders can be evaluated, in which case the correlation between those confounders is a key driver of the magnitude of unmeasured confounding.
In this presentation, we will review the literature on bias sensitivity analysis of unmeasured confounding. Multiple bias sensitivity analyses will be illustrated with a study of ascorbic acid intake. The substantial mortality reduction associated with high ascorbic acid intake found in a observational study was not replicated in an RCT, which may be the result of unmeasured confounding in the observational study. Using this example, we will focus on the distinction between sensitivity analysis of unmeasured confounding due to a single (summary) confounder and multiple confounders.

I4.2
A Bayesian perspective on unmeasured confounding in large administrative databases

LC McCandless[1], P Gustafson[2], JM Somers[1]
[1]*Simon Fraser University, Burnaby, Canada,* [2]*University of British Columbia, Vancouver, Canada*

Confounding creates terrible problems in observational studies using large administrative databases. The massive sample size crushes p-values and standard errors to zero that are calculated from standard analytic adjustment. While this may delight health researchers who discover that everything is significant, it obscures the role of bias, including unmeasured confounding.
The Bayesian approach to statistics provides an appealing way forward because uncertainty about bias can be funneled into the analysis using prior distributions. The posterior distribution for model parameters incorporates uncertainty from bias in addition to the usual random sampling error. In this talk I will discuss Bayesian approaches to adjustment for unmeasured confounding in large administrative database studies.
I will focus on the example of causal mediation analysis with confounding in the mediator-outcome relationship. The Bayesian method is illustrated in a mediation analysis of mortality among offenders with mental illness in British Columbia.

I4.3
New statistical methods for using validation subsamples to adjust for unmeasured confounders in survival analysis

M Abrahamowicz[1], R Burne[1]
[1]*McGill University, Montreal, Canada*

Observational studies of the effects of treatments on clinical outcomes typically rely on large administrative databases, which often lack information on important confounders such as clinical and lifestyle characteristics. However, such confounders are often recorded in smaller clinical 'validation' datasets. Recently, a few methods have been proposed which use validation data to control for unmeasured confounding, however only Propensity Score Calibration (PSC) (Stürmer et al., Am. J. Epi. 2005) can be easily implemented in survival analysis.
We propose a new method specifically designed for application in time-to-event analyses which makes use of such validation datasets in order to impute values for the missing confounders in the large databases. Our approach uses martingale residuals as an indication of lack of fit in the multivariable Cox proportional hazards (PH) model due to the unmeasured confounders.
First, from a Cox model that includes only the measured confounders we obtain the martingale residuals, which can then be used within an imputation model for the unmeasured confounders. We expect the martingale residual to be informative about the value of the missing confounders. Thus, including the martingale residual in an imputation of the missing confounder(s) may improve the accuracy of the imputation. We assess this method in simulations under a variety of assumptions, altering the strength and direction of confounding and the censoring mechanism.
The results are compared to (i) a conventional method which does not adjust for unmeasured confounders, (ii) PSC and (iii) standard imputation.

Contributed sessions

C19 Development of prediction models

C19.1

Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement

GS Collins[1], JB Reitsma[2], DG Altman[1], KG Moons[2]
[1]University of Oxford, Oxford, United Kingdom, [2]UMC Utrecht, Utrecht, The Netherlands

Prediction models are developed to aid healthcare providers in estimating the probability that a specific outcome or disease is present (diagnostic models) or will occur in the future (prognostic models), to inform their decision-making. Clinical prediction models are abundant in the medical literature.
Some disease areas show an overwhelming number of competing prediction models (sometimes even >100) for the same outcome or target population. Only when full information on all aspects of a prediction model study are clearly reported can risk of bias and potential usefulness of the prediction model be adequately assessed. Many reviews have shown that the quality of published reports on the development, validation and updating of prediction models, is very poor. The **T**ransparent **R**eporting of a multivariable prediction model for **I**ndividual **P**rognosis **O**r **D**iagnosis (**TRIPOD**) Initiative therefore developed a set of consensus-based recommendations for the reporting of studies developing, validating or updating a prediction model, whether for diagnostic or prognostic purposes. The development was based on systematic reviews of the literature, web-based surveys and a 3-day expert meeting among methodologists, healthcare professionals and journal editors. The TRIPOD checklist includes 22 items deemed essential for transparent reporting of a prediction model study.
The development and contents of the TRIPOD checklist will be presented and illustrated, along with empirical evidence and rationale for their inclusion. The TRIPOD statement intends to improve the transparency and completeness of reporting of studies that report the development, validation, or updating of a diagnostic or prognostic prediction model.

C19.2

The multi-split testing approach for choosing between 2 prediction strategies

P Blanche[1], M van de Wiel[2], TA Gerds[1]
[1]University of Copenhagen, Department of Biostatistics, Copenhagen, Denmark, [2]VU University, Dept. of Epidemiology & Biostatistics, Amsterdam, The Netherlands

Due to a growing interest in personalized medicine, the demand for new prediction tools is currently strongly increasing. While numerous works have proposed promising statistical models and strategies that develop prognostic tools, in practice, it remains challenging to choose among them.
For choosing between 2 prediction strategies, a commonly applied strategy is to split the data once into two data sets: a "learning sample", used to train the 2 prediction tools, and a "validation sample" used to compare them. Unfortunately, the results usually depend strongly on how the data were split. Recently, van de Wiel et al. (Biostatistics, 2009) proposed a test based on multiple splits of the data. The key idea of the method is to aggregate the p-values obtained by several different random splits, to obtain a conclusion that does not depend on the choice of any specific split.
From a practical point of view, the strengths of the approach are its computational ease and universality, enabling one to compare arbitrary prediction strategies. It is also general with respect to the prediction accuracy criterion, and thus extensions to right censored data and situations with competing risks are readily available, as is shown in this talk.
We provide new insights regarding type one error control and power of the original testing procedure and also discuss how to test alternative hypotheses.
The ideas are motivated and illustrated by a real data analysis of cardiovascular risk prediction models.

C19.3 *Cancelled*

The impact of events per variable on the predictive performance of the Cox model

EO Ogundimu[1], DG Altman[1], GS Collins[1]
[1]University of Oxford, Oxford, United Kingdom

Sample size requirements for developing multivariable prediction models using Cox regression are routinely based on the rule of thumb derived from simulation studies ranging from a minimum of 5 to 20 events per variable (EPV). However, a common design feature of these simulation studies is the small sample size and limited scenarios, and that only one binary predictor was included in the models. The effects of multiple binary predictors with varying degrees of prevalence, reflecting clinical practice, have not been investigated. Furthermore, emphasis in these studies has focussed on the accuracy and precision of regression coefficients, and not on the predictive accuracy of the fitted model, which ultimately characterises the predictive ability of the model.
We therefore conducted extended simulation studies using a large general practice dataset (THIN), comprising over 2 million anonymised patient records to examine the sample size requirements for prediction models developed using Cox regression. Investigating both fully specified models and models derived using variable selection, we examine the stability and precision of regression coefficients and their impact on the apparent model performance (e.g. c-index, D-statistic, R^2) as well as subsequent performance in an external validation dataset. We also present results examining models containing low prevalence binary predictors and the impact in terms of sample size on the predictive accuracy of the model.
We will demonstrate that more events are needed to achieve precise measures of predictive accuracy in situations where `many´ low prevalence binary predictors are included in the model.

C19.4

The number of events per variable needed to build logistic prediction models in clustered data: a simulation study

L Wynants[1,2], W Bouwmeester[3], S Van Huffel[1,2], B Van Calster[4], Y Vergouwe[5]
[1]KU Leuven Dept of Electrical Engineering / ESAT-STADIUS, Leuven, Belgium, [2]iMinds Medical Information Technologies, Leuven, Belgium, [3]MediQuest, Utrecht, The Netherlands, [4]KU Leuven Dept of Development & Regeneration, Leuven, Belgium, [5]Public Health Dept, Erasmus MC, Rotterdam, The Netherlands

Researchers increasingly combine data from several centers to develop clinical prediction models for diagnosis or prognosis. Guidelines for the required sample size of such multicenter studies are lacking. We studied the impact of the number of events per variable (EPV) on the estimation of regression coefficients and the performance of the resulting prediction model. We performed a simulation study to investigate the influence of the amount of clustering (the intraclass correlation or ICC), backward variable selection, the number of centers, center size, and the total sample size.
A high EPV increased the accuracy of the regression estimates and the performance of the prediction model, while the ICC did not meaningfully influence estimation or performance. In addition to EPV, also the total

sample size positively influenced estimation and model performance. However, the composition of the sample does not influence the results, given the EPV and total sample size: with a few large clusters, estimation and prediction performance is as good as with many small clusters. Stepwise variable selection led to a substantial bias in one of the regression coefficients, but this did not worsen predictive performance. Our findings demonstrate the limited importance of the amount of clustering. In line with several studies dealing with unclustered data, we recommend at least ten EPV for predefined models, although up to fifty EPV may be needed when variable selection is performed.

C19.5

Review and evaluation of penalised likelihood methods for risk prediction in data with few events

M Pavlou[1], G Ambler[1], S Seaman[2], R Omar[1]
[1]University College London, London, United Kingdom, [2]MRC Biostatistics Unit, Cambridge, United Kingdom

Prognostic regression models typically use multiple predictors to predict an outcome. When the number of events is small compared to the number of regression coefficients, the danger of model overfitting is particularly pronounced. Traditional guidance suggested the ´rule of 10´ to mean than at least 10 events per estimated regression coefficient (Events Per Variable-EPV) are necessary for the development of reliable risk models. An overfitted model tends to demonstrate poor calibration and predictive accuracy when applied to new data. In this work we review penalised likelihood methods for binary outcome. We consider Ridge and Lasso, both of which shrink coefficient estimates (Lasso can provide parsimonious models by also omitting some of the predictors). Additionally, we consider extensions of these (e.g. Elastic Net and Adaptive Lasso), their Bayesian analogues and Bayesian approaches based on ´spike and slab´ priors. We evaluate the predictive performance of the methods in comparison to standard MLE in simulated data derived from real datasets. Several features of the data are varied, namely the EPV, the strength of predictors, the number of ´noise´ predictors and the correlation between predictors. Simulation and real data analyses suggest that MLE tends to produce overfitted models with poor predictive performance in scenarios with few events. Penalised methods offer significant improvement. The choice of method depends on the features of the particular data. Elastic Net performed well overall, while the Bayesian approaches were also found to be useful for prediction.

C20 Individual participant data meta-analysis

C20.1

How to appraise Individual Participant Data (IPD) meta-analysis in diagnostic and prognostic risk prediction research

KGM Moons[1], T Debray[1], M Rovers[2], RD Riley[3], JB Reitsma[1]
[1]UMC Utrecht, Utrecht, The Netherlands, [2]Radboud University Medical Center, Nijmegen, The Netherlands, [3]University of Birmingham, Birmingham, United Kingdom

Background: The development and (external) validation of diagnostic and prognostic prediction models is an important aspect of contemporary epidemiological research. Unfortunately, many prediction models perform more poorly than anticipated when tested or applied in other individuals, and interpretation of their generalizability is not straightforward. During the past decades, evidence synthesis and meta-analysis of individual participant data (IPD) has become increasingly popular for improving the development, validation and eventual performance of novel prediction models. Also, IPD meta-analysis lead to a better understanding in the generalizability of prediction models across different populations. There is, however, little guidance on how to conduct an IPD meta-analysis for developing and validating diagnostic or prognostic prediction models.
Objective and Methods: We provide guidance for both authors and reviewers in appraising IPD meta-analyses that aim to develop and/or validate a prediction model using multiple IPD datasets. Furthermore, we demonstrate why and how IPD meta-analysis of risk prediction research differs from IPD meta analysis of intervention research. Finally, we provide methodological recommendations for conducting an IPD meta-analysis for risk prediction research, and illustrate these with a clinical example.
Conclusions: Whereas meta-analytical strategies for intervention research have been well described during the past few decades, evidence synthesis in risk prediction research is relatively new. Appropriate methods for conducting an IPD meta-analysis in risk prediction research have become available during the past few years, and clearly differ from their counterparts in intervention research.

C20.2

Being PRO ACTive - what can a clinical trials database reveal about ALS?

N Zach[1], R Kueffner[2], A Shui[3], A Sherman[4], J Walker[4], E Sinani[4], I Katsovskiy[4], D Schoenfeld[3], G Stolovitzky[5], R Norel[5], N Atassi[4], J Berry[4], M Cudkowicz[4], M Leitner[6]
[1]Prize4Life, Herzliya, Israel, [2]Helmholtz Zentrum, Munich, Germany, [3]MGH Biostatistics Center, Massachusetts General Hospital, Boston, United States, [4]Neurological Clinical Research Institute, MGH, Charlestown, United States, [5]The DREAM Project, IBM, Yorktown Heights, United States, [6]Prize4Life, Boston, United States

Understanding a given patient population is a necessary step in advancing clinical research and clinical care and conducting successful and cost-effective clinical trials. To overcome the challenge of gathering a large enough cohort of patients in rare diseases such as ALS. We developed the Pooled Resource Open-access ALS Clinical Trials (PRO-ACT) platform. The PRO-ACT database consists of 8600 ALS patients who participated in 17 clinical trials. The dataset includes demographic, family history, vital signs, clinical assessment, lab-based, treatment arm, and survival information. The database was launched open access on December 2012, and since then over 225 researchers from 25 countries have requested the data.
Several assessments were made to start understanding the value of the PRO-ACT in addressing pivotal questions in ALS clinical research. One such initiative included a crowdsourcing effort-the ALS Prediction Prize challenge- to develop improved methods to accurately predict disease progression at the individual patient level. The challenge brought in 1000+ registrants and led to the creation of multiple novel disease progression algorithms.
Other highly important insights from the database include newly identified predictive features, definitive support for previously proposed predictive features based on smaller samples, and newly identified stratification of patients based on their disease progression profiles.
These results demonstrate the value of large datasets for developing a better understanding of ALS natural history, prognostic factors and disease variables. Such critical questions include patient stratification, associations with disease co-morbidities and concomitant medications, identification of biomarkers, and potentially new ways to enhance clinical practice and clinical trials.

C20.3
Missing data in individual patient data meta-analysis

M Quartagno[1], J Carpenter[2]
[1]London School of Hygiene and Tropical Medicine, London, United Kingdom, [2]London School of Hygiene & Tropical Medicine, London, United Kingdom

Aim: Individual Patient Data (IPD) meta-analyses rely not only on results from different studies, but aim to directly combine individual data from each study, in order to better account for the statistical heterogeneity between studies.
Missing Data are common in clinical datasets and therefore also in meta-analyses of such datasets.
This can take two forms: missing data within studies and important variables not being available from some studies. We aim to define good strategies to handle issues raised by the presence of missing data in IPD meta-analyses.
Methods: Multiple Imputation (MI) provides a natural approach to missing data in this context. There are, broadly, two approaches to MI: Full Conditional Specification (FCS) and Joint Modelling (JM). We argue that JM provides a more natural approach for multilevel MI, and thus for IPD meta-analysis. However REALCOM, the only current software available for JM-MI, is too inefficient to handle the size and complexity of IPD Meta-Analysis datasets. We then present a new software that efficiently programmes JM and demonstrate its feasibility vis-a-vis FCS.
Results: preliminary analyses show how our software is computationally feasible and competitive with FCS. We are going to investigate the main characteristics of JM imputation in IPD meta-analyses through simulations.
We will then analyse real datasets, exploring advantages and issues raised by the use of JM in this setting. This includes comparing within-study imputation with stratified imputation and finding if it is better to combine results from the analysis of different imputed datasets before the meta-analysis or vice versa.

C20.4
Multiple imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE

TPA Debray[1], S Jolani[2], H Koffijberg[1], S van Buuren[3], KGM Moons[1]
[1]UMC Utrecht, Utrecht, The Netherlands, [2]Utrecht University, Utrecht, The Netherlands, [3]TNO Quality of Life, Leiden, The Netherlands

Background: Individual participant data meta-analyses (IPD-MA) are increasingly used for developing and validating multivariable (diagnostic or prognostic) risk prediction models. Unfortunately, some predictors or even outcomes may not have been measured in each study and are thus systematically missing in the IPD-MA. As a consequence, it is no longer possible to evaluate between-study heterogeneity and to estimate study-specific predictor effects, which severely hampers the development and/or validation of novel prediction models.
Methods: Here, we describe a novel approach for imputing systematically missing data and adopt a generalized linear mixed model to allow for between-study heterogeneity. This approach can be viewed as an extension of Resche-Rigon´s method (Stat Med 2012), but relaxes assumptions regarding variance components and allows imputation of linear (e.g. continuous) and non-linear (e.g. categorical) predictors.
Results: We illustrate our approach in a case study with the IPD from 13 studies for predicting the presence of Deep Venous Thrombosis. We compare the results after applying various imputation methods, and make recommendations about their implementation.
Conclusions: Our approach improves the estimation of predictor effects and between-study heterogeneity, thereby facilitating the development and validation of novel prediction models from an IPD-MA.

C20.5
Analysis of repeated ordinal measurements and trial planning in a rare neurological disorder

L Tanadini[1,2], J Steeves[3], A Curt[2], T Hothorn[1]
[1]Division of Biostatistics, IFSPM, University of Zurich, Zurich, Switzerland, [2]Spinal Cord Injury Center, Balgrist University Hospital, Zurich, Switzerland, [3]ICORD, University of British Columbia, Vancouver, Canada

Despite three decades of methodological developments in the field of ordinal data analysis, clinical studies confronted with ordinal outcomes often resort to statistical methods that assume more refined measurement scales. The field of neurology, where ordinal scales are ubiquitous, provides concrete examples of this practice.
In spinal cord injury, the neurological status of a patient is assessed through a sequential testing of multiple key muscles, each one being graded on a six-point ordinal scale. The current approach of analyzing the total sum of all key muscles scores is questionable on a number of levels.
We provide a statistical framework for the analysis of this type of neurological assessments that takes into account the ordinal nature of each muscle score. We show that a proportional odds model conditioning on the interaction between baseline level of lesion and distance from lesion provides an adequate description of the motor scores distribution at six months after injury. This holds true even when compared to a cumulative link mixed model explicitly incorporating the repeated measurements (multiple muscles) on the same patient. We further investigate the structuring of our initial, simplistic linear predictor via model-based boosting. Additionally, we simulate several clinical trials scenarios to provide benchmark data for powering future trials.
In addition to foster the use of correct methodology in neurology and related disciplines, the proposed analysis framework is likely to be a missing, but essential element in the currently lagging translation of promising preclinical results into clinical therapies for humans.

C21 Survival analysis and competing risks

C21.1
Stagewise pseudo-value regression for time-dependent effects on the cumulative incidence

H Binder[1], D Zöller[1], I Schmidtmann[1], A Weinmann[2], T Gerds[3]
[1]University Medical Center Mainz, IMBEI, Mainz, Germany, [2]University Medical Center Mainz, 1. Med, Mainz, Germany, [3]University of Copenhagen, Department of Biostatistics, Copenhagen, Denmark

The cumulative incidence describes the absolute risk of an event as a function of time in a competing risks setting. For regression analysis, one can either choose to model all competing events by separate cause-specific hazard models, or directly model the association between covariates and the cumulative incidence of one of the events. With a suitable link function, the direct regression models allow for a straightforward interpretation of covariate effects on the cumulative incidence. In practice where data can be right-censored, they are implemented using a pseudo-value approach. For a grid of time points the possibly unobserved binary event status is replaced by a jackknife pseudo-value based on the Aalen-Johansen method. We combine a stagewise regression technique with the pseudo-value approach to provide variable selection while allowing for time-dependent effects. This is implemented by coupling variable selection between the grid times, but determining estimates separately. The effect estimates are regularized to allow for model fitting also with a low to moderate number of observations. The technique is illustrated in an application to clinical

cancer registry data from hepatocellular carcinoma patients.
The results are contrasted to traditional hazard-based modeling, in particular highlighting the difference in interpretation. In addition to more straightforward interpretation, identification of time-dependent covariate effect patterns on the cumulative incidence is seen to be feasible with a moderate number of observations, when using the proposed technique.

C21.2
Imputing missing covariate values in presence of competing risks

M Resche-Rigon[1], IR White[2], S Chevret[1]
[1]INSERM UMR 1153, Université Paris 7, Paris, France, [2]MRC Biostatistics Unit, Institute of Public Health, Cambridge, United Kingdom

Due to its flexibility, practicability and efficiency, multiple imputation by chained equations is widely used to impute missing data. To avoid bias in the substantive model, it is well known that the imputation model must include all the variables from the analysis model, including the outcome. In survival analysis, the outcome is defined by an event indicator D and the observed event or censoring time T. In 2009, White and Royston showed that when a Cox model is used for the analysis, the imputation model for each covariate should include the event indicator and the cumulative baseline hazard estimated by the Nelson-Aalen estimator (White IR, Royston P, Stat Med. 2009).
In the competing risks setting, multiple imputation has been proposed only to impute missing information on the cause of failure, and has mostly been used in analyses of cumulative incidence functions. We extend the work of White and Royston to impute missing covariates in a competing risks setting, where the substantive model is either a cause specific proportional hazards model or a sub-distribution proportional hazards model: we show that the event indicators and cumulative baseline hazards of all the competing events should be included in the imputation model. Consequently, even in a standard survival analysis framework, the cumulative baseline hazard of being censored should be included in the imputation model.
These approaches will be evaluated by a simulation study, and then applied to a sample of 278 adult patients with acute myeloid leukaemia.

C21.3
Evaluation of a peritoneal dialysis program using semiparametric multi-state models in the presence of competing risks

L Teixeira[1], C Cadarso-Suarez[2], A Rodrigues[3,4], D Mendonça[1,5]
[1]ICBAS-UP, Porto, Portugal, [2]Department of Statistics and Operations Research, USC, Santiago de Compostela, Spain, [3]CHP-HGSA, Porto, Portugal, [4]UMIB/ICBAS-UP, Porto, Portugal, [5]ISPUP, Porto, Portugal

Chronic kidney disease is becoming a major public health problem with a growing number of patients in need of replacement therapy, such as peritoneal dialysis (PD). As the trajectory of PD patients is complex, characterized by the presence of several transient and absorbing states, the evaluation of such programs may be addressed using a multi-state approach taking competing risks into account.
The present study has as main objectives:
(i) to discuss the use of flexible regression models like Structured Additive Regression (STAR) models in a multi-state competing risk framework, expressing results of continuous covariates in terms of hazard ratio curves taking a specific covariate value as the reference; and
(ii) to adapt the definition of time-dependent ROC curves to a multi-state competing risk framework to assess the predictive accuracy of the STAR model.
The methodologies discussed were applied to explore the effects of major clinical covariates such as age, sex and diabetes in a PD data. These methods revealed to be very relevant for this type of real clinical data as the developed models were an informative tool for the evaluation of the patients and consequently for the medical decision process. The use of STAR models complemented with the use of temporal ROC curves in clinical context allowed to identify relevant factors associated with each one of the specific transitions. The identification of these factors, which could not have been obtained with standard survival models, contributes for a better knowledge of patient trajectories resulting in better management of treatment programs.

C21.4
Predicting optimal cumulative doses for breast cancer chemotherapy via competing risks regression models

G Cortese[1]
[1]University of Padua, Padua, Italy

In breast cancer, the risk of cardiotoxicity due to chemotherapy increases with the cumulative dose of treatment over time. Therefore, it is of interest to estimate an optimal cumulative dosage over time that guarantees a low risk for cardiotoxicity, while controlling the competing risk for mortality by maximizing the antitumor effect.
For this purpose, we consider a competing risks regression model with two events, cardiotoxicity and death. The aim is to predict optimal cumulative doses along a given treatment time that keep the cumulative risk for cardiotoxicity below a certain threshold (e.g. <5%).
Data from breast cancer patients, treated with chemotherapy during follow-up, were analysed with different direct regression models for competing risks. The cumulative dose, predetermined according to given time schedules, was included as time-dependent covariate in addition to other risk factors. The cumulative incidence function for cardiotoxicity over a certain time window [s, t], $P_c(s,t; X(s))$, e.g. a one-year prediction t=s+1, was treated as a function of cumulative dose at time s, X(s).
The direct regression models allow finding a one-to-one relationship between $P_c(s,t; X(s))$ and X(s). Then, the optimal cumulative doses at a sequence of landmark time points, were found by inverting the one-year prediction of a cardiotoxicity risk equal to 5%. Confidence intervals for the doses were estimated by inverting pointwise confidence intervals for $P_c(s,t; X(s))$.
To control also for increased risk of dying, we finally predict optimal cumulative doses by minimizing a combination of the two cumulative risks of cardiotoxicity and death.

C21.5
The liability-threshold model for case-control family studies applied to censored time to event data

L Cederkvist[1,2], KK Holst[1], T Scheike[1]
[1]Copenhagen University, København K, Denmark, [2]Danish Cancer Society Research Center, København Ø, Denmark

In case-control family studies, familial aggregation of a disease is investigated using families collected via case or control probands, who are chosen based on their disease status. By comparing the correlation for different family members, the presence and magnitude of familial aggregation of the disease can be assessed.
A model which is often used is the liability-threshold model where the disease status outcome is defined from a normal distributed latent variable, the so-called liability. The disease status outcome and the unobserved continuous liability are linked using the Probit function. A threshold on the liability scale determines whether an individual is affected by the disease or not. If an individual´s liability exceeds the threshold he or she is affected. The variance of the latent liability can be decomposed into genetic and environmental components; a process that requires specification of

within-family correlations for both the genetic and the environmental components. The heritability defined as the relative contribution of the genetic components to the total variance may also be estimated. No time-aspect is included in the liability-threshold model and thus the model does not account for censoring, which may lead to severe bias.
In this study, we focus on extending the liability-threshold model for case-control family data to take censoring into account using weighting. We illustrate our work using Danish case-control family data on cancer.

C22 Surrogate and composite endpoints

C22.1
A new audit strategy to detect possible bias in the evaluation of progression free survival

J Link[1], U von Wangenheim[1], F Fleischer[1]
[1]*Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riß, Germany*

In oncology trials, an endpoint of increasing use is progression-free survival (PFS), i.e. the time until objective tumour progression or death. To mitigate potential evaluation bias, its assessment is often performed via a blinded independent central review (BICR) in addition to the local evaluation by the investigator. However, BICR is not only time consuming, costly and operationally demanding, it may also introduce informative censoring of PFS.
Thus, specific audit methods have been developed to detect potential bias by only using a subset sample for BICR (Zhang et al 2013 [1]) If no bias is present in that subset one can omit further BICR evaluations. While examining these methods several disadvantages become apparent. For example: One approach cannot be applied before study end, which causes practical problems like time delay. Additionally, the magnitude of the bias is not taken into consideration if only the treatment effect is large enough. To overcome these constraints, we developed a new strategy based on equivalence testing in a group-sequential approach whereby focusing on testing the potential difference between BICR and local assessments.
This approach is able to detect the presence of bias on an ongoing basis during trial conduct, no matter how strong the treatment effect is. Its performance is analysed via simulations and an example using real study data shows that it can be used efficiently in practice.

Reference:

[1] Jenny J Zhang, Lijun Zhang, Huanyu Chen. Assessment of audit methodologies for bias evaluation of tumor progression in oncology clinical trials.Clinical Cancer Research, 19, 2637-2645, 2013

C22.2
A causal inference/mediation analysis based approach for assessing pseudo end-points applied to to ovarian cancer trials

T Lange[1]
[1]*University of Copenhagen, Copenhagen, Denmark*

The use of pseudo end-points in clinical trials is desirable both from a feasibility and ethic point of view as replacing the final end-point by an earlier and/or easier to measure end-point, can assist in speeding up treatment development. Of course these benefits can only be harvest if a suitable pseudo end-point can be identified and validated.
Until now the validation of a pseudo end-point has been quite cumbersome and the obtained measures of quality of the pseudo end-point not very intuitive, see eg. Eisenhauer ("Optimal assessment of response in ovarian cancer", Annals of Oncology 22 (Supplement 8): viii49-viii51, 2011).
In this work it is explored how to formulate the pseudo end-point problem in a counterfactual framework.
In addition it is proposed to employ mediation analysis in the validation of a pseudo end-point since this will provide a direct parameterization of the degree to which the effect of treatment is fact captured in the pseudo end-point.
The talk will include both theoretical aspects (which assumptions are made and how can we devise a counterfactual based framework) and practical advice on implementation. All results are applied to ovarian cancer trials.

C22.3
Bias assessment of surrogate threshold effects in simplified correlation based validation approaches

C Schürmann[1], W Sieben[1]
[1]*Institute for Quality and Efficiency in Health Care (IQWiG), Köln, Germany*

A well established approach of surrogate endpoint validation is the correlation based meta-analysis as outlined in the seminal work of Buyse et al. (Biostat 2000). Surrogacy can be assumed if high values of individual and of study level correlation can be demonstrated.
Alternatively, if a true endpoint is to be predicted from a surrogate endpoint, the surrogate threshold effect (STE, Burzykowski and Buyse (PharmStat 2006)) can be used. In practice, as individual patient data are hard to obtain, often only aggregated data are used and simplified analyses are performed. We are interested in how much simplified analyses are biased compared to the full model with individual patient data. To this end we conduct a simulation study with individual patient data and compute STEs with full and simplified analyses in various data situations (study sizes, correlations, variances etc.) with respect to bias.
Comparison of the results will help us decide to what extent STEs of the different approaches are suitable if an effect on a true endpoint is to be predicted.

C22.4
Extension of win-ratio: analyzing a composite endpoint considering the clinical importance order among components

S Witte[1], G Dong[2]
[1]*Novartis Pharma AG, Basel, Switzerland,* [2]*Novartis Pharmaceuticals, East Hanover, United States*

A composite endpoint consists of multiple endpoints combined in one outcome, and is frequently used as the primary endpoint in randomized clinical trials. During past decade, there are discussions on pros and cons of using a composite endpoint. The event rate in the composite endpoint is higher, therefore the sample size needed for a clinical trial may be reduced, subsequently length of the study can be shortened and costs can be saved. In addition, with a composite endpoint as a single endpoint, multiplicity issue and competing risk problem may be avoided.
However, in the conventional analyses, all components are treated equally important; and in time-to-event analysis, the first event considered may not be the most important component. Recently Pocock [1] published the win ratio method to address these disadvantages. In this new method, they proposed two approaches: matched pair and unmatched pair. In the unmatched pair approach, the confidence interval is constructed based on bootstrap re-sampling, and the hypothesis testing is based on the generalized Wilcoxon test. We extend the unmatched pair approach of Pocock's win-ratio method to perform hypothesis testing and construct the confidence interval for win ratio based on its asymptotic distribution. This asymptotic distribution is derived via U-statistics following Wei [2].
We illustrate our method with an example from a liver transplant study,

which shows that the confidence interval based on our derivation can be narrower than the bootstrapping confidence interval.

References:

1. Pocock SJ et al. Eur Heart J. 2012;33:176-182
2. Wei LJ and Johnson W.E. Biometrika,1985;27:359-364

C22.5

Weighted comparisons of composite endpoints

M Wolbers[1,2], A Nguyen Duc[1]
[1]Oxford University Clinical Research Unit, Ho Chi Minh City, Viet Nam, [2]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Composite endpoints are widely used as primary endpoints of randomized controlled trials (RCTs) in many clinical disciplines. For example, a systematic review of published RCTs in cardiovascular medicine and surgery showed that 37% of them reported a composite endpoint with a median of 3 individual component outcomes. One limitation of composite endpoints is that they frequently pool component outcomes of varying clinical importance. Thus, several authors have suggested replacing or complementing the standard analysis of composite endpoints which just analyzes whether a subject experiences any component outcome or not by an analysis which weights each component with respect to its clinical importance or associated cost.

We suggest interpretable test statistics based on weighted linear combinations of absolute risk differences of component outcomes for between-group comparisons of both multinomial and time-to-event composite endpoints. Considerations for choosing component weights are reviewed and it is shown that there is often a conflict between choosing weights that lead to powerful tests and weights that are clinically relevant.

One problem of weighted comparisons is that elucidation of quantitative component weights is difficult in practice. However, it is often possible to rank components according to their relative importance. To address this, we introduce methods which control the family-wise error rate across all non-negative weights or across all sets of weights satisfying an order constraint, respectively.

C23 Design and analysis of clustered studies

C23.1

Methods for observed-cluster inference when cluster size is informative

SR Seaman[1], M Pavlou[2], AJ Copas[3]
[1]MRC Biostatistics Unit, Cambridge, United Kingdom, [2]University College London, London, United Kingdom, [3]MRC Clinical Trials Unit at University College London, London, United Kingdom

Clustered data commonly arise in epidemiology. We assume each cluster member has an outcome Y and covariates X. When there are missing data in Y, the distribution of Y given X in all cluster members (`complete clusters´) may be different from the distribution just in members with observed Y (`observed clusters´).

Often the former is of interest, but when data are missing because in a fundamental sense Y does not exist (e.g. quality of life for a person who has died), the latter may be more meaningful (quality of life conditional on being alive). Weighted and doubly weighted generalised estimating equations and shared random-effects models have been proposed for observed-cluster inference when cluster size is informative, i.e. the distribution of Y given X in observed clusters depends on observed cluster size. We show these methods can be seen as actually giving inference for complete clusters and may not also give observed-cluster inference. This is true even if observed clusters are complete in themselves rather than being the observed part of larger complete clusters: here methods may describe imaginary complete clusters rather than the observed clusters. We show under which conditions shared random-effects models proposed for observed-cluster inference do actually describe members with observed Y. A psoriatic arthritis dataset is used to illustrate the danger of misinterpreting estimates from shared random-effects models.

C23.2

Generalised estimating equation methods for analysing continuous outcomes when cluster size is informative

LN Yelland[1,2], TR Sullivan[2], JB Carlin[3,4]
[1]Women's and Children's Health Research Institute, North Adelaide, Australia, [2]The University of Adelaide, Adelaide, Australia, [3]Murdoch Children's Research Institute, Parkville, Australia, [4]University of Melbourne, Melbourne, Australia

Generalised estimating equations (GEEs) are a popular method for analysing clustered data but parameter estimates may be biased when cluster size is related to the outcome. This type of informative clustering is a common problem in perinatal trials when infants from both single and multiple births are included, since infants from multiple births tend to have worse health outcomes.

A cluster weighted GEE approach has been proposed for handling informative cluster size, which estimates parameters with a cluster-level interpretation. Alternative methods of analysis are also available, including individually weighted GEEs to estimate individual-level parameters and GEEs with adjustment for cluster size, but these have received limited attention. In this presentation, I will report the results of a study comparing these three approaches for analysing clustered continuous outcomes in terms of their theoretical properties, interpretation and finite sample performance.

I will show why these methods often produce different unadjusted results and demonstrate that adjusting for cluster size does not always solve the problem of informative cluster size.

The relative merits of choosing a cluster-level or an individual-level approach in the informative cluster size setting will be discussed and recommendations will be made for dealing with informative cluster size in the context of perinatal trials with multiple births.

C23.3

Choosing covariates and the effects of covariate adjustment in the analysis of CRTs

N Wright[1]
[1]Blizard Institute, Queen Mary University of London, London, United Kingdom

Research on the effects of covariate adjustment in the analysis of randomised trials has mainly focused on trials in which individuals are randomised to treatment arms. This has led to published guidance on choosing covariates in the analysis of randomised trials.

In cluster randomised trials (CRTs) pre-existing groups (clusters) of individuals are randomised to treatment arms. A valid analysis of a CRT must take into account the additional data structure imposed by cluster randomisation, for example by using a mixed effects model. We can adjust for covariates in these models, just as in fixed effects models, by including variables and parameters for covariates in the linear predictor. There is limited published research on the effects of covariate adjustment in the analysis of CRTs, especially in the analysis of binary outcome variables.

We firstly review the published guidance for choosing covariates in randomised trials, in the context of analysing CRTs.

We then present a selection of results from simulation studies on the effects of covariate adjustment in the analysis of CRTs. Simulations included

comparisons of the use of individual and cluster level covariates, cluster aggregation of covariate data, and the use of separate individual and cluster level covariate effect parameters. To conclude this work, we suggest some additional guidance on choosing covariates that is specific to the analysis of CRTs.

C23.4
Sample size and analysis considerations for cluster randomised crossover trials with unbalanced cluster sizes and binary data

AB Forbes[1], M Akram[1], R Bellomo[1,2]
[1]Monash University, Melbourne, Australia, [2]Australian and New Zealand Intensive Care Society, Melbourne, Australia

Cluster randomised crossover trials are a class of multiple-period cluster designs that have been increasingly used in clinical and public health research. These trials gain efficiency by incorporating treatment crossover across observation periods within each cluster.
However, the development and assessment of these designs to date has been fairly limited. In this presentation we report on our recent design and analysis work:
We present expressions for the variance of treatment effect estimators which take into account period effects, within- and between-period intracluster correlations, as well as unbalanced within-and between-period cluster sizes. Using these expressions, we present sample size formulae for unbalanced cluster sizes as would typically occur in practice. We then discuss extensions for use with binary outcomes. Using an underlying marginal model for binary outcomes, we present results of a simulation exercise with varying numbers of clusters, cluster sizes and outcome prevalences, and discuss problematic parameter configurations.
We illustrate all methods with an application involving a proposed large cluster randomised crossover trial to evaluate interventions to reduce mortality in the intensive care research setting. We also discuss the potential for extension to multiple-period-multiple-treatment designs and conditions for their feasibility in particular research settings.

C23.5
A multidisciplinary approach to benefits and drawbacks of the stepped wedge cluster randomized design

E de Hoop[1], H Koffijberg[1], R van der Graaf[1], JB Reitsma[1], I van der Tweel[1]
[1]University Medical Center Utrecht, Utrecht, The Netherlands

Stepped wedge cluster randomized designs are increasingly being used over the last couple of years. Some reviews have listed reasons for the use of this design. Furthermore, benefits and drawbacks have been mentioned in several papers.
However, there is extensive debate whether this design is useful or not. Uncertainties concern the usefulness of this design for different types of research questions, but also various design issues such as risk of bias, efficiency, ethical issues, data analysis and economic evaluations are topics of debate. Although these issues have been mentioned before, it remains unclear whether and to what extent these are in favour or against the use of the stepped wedge design. Therefore, a comprehensive overview of characteristics, benefits and drawbacks of the stepped wedge design is required.
We took a multidisciplinary approach where input from statisticians, methodologists, ethicists and health economists led to such an overview. For this overview we compared the stepped wedge design to the parallel group clustered trial design. We will present our findings and explain which aspects are unique to the stepped wedge design and what its advantages and disadvantages are compared to a parallel group cluster randomized trial.
Besides, we will illustrate our findings with a trial that used the stepped wedge design. This will help researchers and research ethics committees to decide on the appropriateness of a stepped wedge design in a cluster randomized trial for their research question.

C24 Group-sequential designs

C24.1 *Student Conference Award*
Group sequential monitoring of response-adaptive randomised clinical trials with censored survival data

HY Liu[1]
[1]Queen Mary University of London, London, United Kingdom

Methods for combining group sequential tests with a response-adaptive randomisation design in clinical trials with immediate responses have been studied. However, application of this combined approach to survival responses with inherent right-censoring has not yet been investigated. Such an approach that does not require the number of interim analyses to be pre-specified is applied to censored survival times. It utilises an error spending function to compute the critical values at any point during the course of the trial.
The approach is based on the canonical joint distribution of the sequence of test statistics, which generalises to group sequential response- adaptive randomisation designs. Since the design is adaptive, both the sample size and the treatment allocation proportions are random at interim analyses.
In this paper, the mean of the treatment allocation proportions and the corresponding standard deviations for two response-adaptive randomisation designs are compared with complete randomisation under group sequential monitoring. Simulation results show that the combined approach can not only reduce the total number of patients, but also increase the power compared with that of a group sequential non-adaptive randomised design.
Moreover, more patients are assigned to the more promising treatment using the response-adaptive randomisation designs. But the difference in the means and standard deviations of the treatment allocation proportions between the two response-adaptive designs is generally less than 1%. In conclusion, the combined approach can achieve higher power and have ethical benefits.

C24.2
Group-sequential designs for cross-over trials

MJ Grayling[1], J Wason[1], A Mander[1]
[1]MRC Biostatistics Unit, Cambridge, United Kingdom

Group-sequential procedures have assisted greatly in reducing the expected sample size of parallel clinical trial designs, which remain the conventional means by which to estimate the treatment effect of an experimental drug in many settings. However, particularly within the context of treatments for chronic diseases, the cross-over trial remains the design of choice; exploiting the opportunity to treat patients with multiple experimental drugs to substantially reduce the variance of the estimated treatment effects. Theoretically, group-sequential approaches to cross-over trials promise to bring the same advantages as in a parallel setting. Here, we discuss our work to date on establishing a framework for such designs. By determining the joint distribution of the test statistics, within a linear mixed model setting for data analysis, optimal designs in-terms of minimising the expected sample size or expected number of observations, subject to type-I and type-II error constraints, can easily be determined using a simple search over possible sample sizes at each stage of

the trial. Early stopping for both futility and efficacy will be discussed, for trials of varying numbers of drugs and stages. Moreover, controlling the Family Wise Error Rate of such designs will also be explored. Using results from the four-treatment four-period TOMADO trial for sleep apnoea devices, the performance of the group sequential cross-over designs will be examined.

C24.3
Optimal sequential clinical trials in small populations

S Nikolakopoulos[1], KCB Roes[1], I van der Tweel[1]
[1]UMC Utrecht, Utrecht, The Netherlands

A sequential clinical trial will on average require fewer patients to reach a conclusion, possibly resulting in superior treatments becoming available faster. The vast majority of the developed and most commonly used sequential methods rely on asymptotic distributions of the test statistics. Recent guidelines and published research suggest the use of sequential methods, among other options, as an alternative for the design of clinical trials in orphan diseases.
In this work we thus investigate the behavior of sequential tests in small to very small sample sizes. We explore their operational characteristics and point out - usually overlooked - simple corrections that can preserve type 1 errors accurately. In addition we look into the case where the sample size has a given maximum due to, e.g., a very rare disease. When exploring such a case in a fixed-sample size, design and sample size considerations might become irrelevant and, for a given type I error level, there is one design with a respective power function. In the sequential case however, different design ingredients - e.g. timing and number of interim analysis , efficacy and futility boundaries - can lead to designs with different operational characteristics and the choice will matter in terms of power and efficiency.
We suggest an optimization rule taking into consideration a maximum sample size and prior belief of the treatment effect. We illustrate our method with a real example using a clinical trial for a rare disease.

C24.4
Group sequential designs for verifying whether effective drug concentrations are similar in adults and children

L Hampson[1], T Jaki[1], R Fisch[2]
[1]Lancaster University, Lancaster, United Kingdom, [2]Novartis Pharma AG, Basel, Switzerland

New medicines for children should be subject to rigorous testing while avoiding unnecessary experimentation in children. In particular, paediatric dosing recommendations should be informed by existing relevant data. If the effective concentration of a drug can be assumed to be similar in adults and children, an appropriate paediatric dosing rule may be found by 'bridging', that is, conducting pharmacokinetic studies in children to find doses that produce concentrations therapeutic in adults. However, this strategy may result in children receiving an ineffective or hazardous dose if, in fact, effective concentrations differ between adults and children. When there is uncertainty about the equality of effective concentrations, some pharmacokinetic-pharmacodynamic (PK-PD) data may be needed in children to verify whether differences between adults and children are small. In this presentation, we develop adaptive procedures that can be used to verify this assumption efficiently. Asymmetric inner wedge group sequential tests are constructed which permit early stopping to accept or reject an assumption of similar effective drug concentrations in adults and children. Asymmetry arises because the consequences of under- and overdosing may differ. Using exact calculations we compare the efficiency of error spending inner-wedge tests with optimal designs which minimise the expected sample size needed to reach a conclusion. If there is time, we will show how stopping rules based on predictive tail area probabilities can be derived for testing whether observed paediatric PK-PD data are consistent with an assumption of similar effective concentrations in adults and children.

C24.5
Group-sequential strategies when considering multiple outcomes as co-primary in clinical trials

T Hamasaki[1], K Asakura[1], SR Evans[2]
[1]Osaka University Graduate School of Medicine, Suita, Japan, [2]Harvard School of Public Health, Boston, United States

Many recent clinical trials, especially in pharmaceutical drug development, have utilized more than one endpoint as co-primary, thus evaluating the intervention's multidimensional effects.
In this paper, we consider group sequential strategies in clinical trials with multiple co-primary endpoints when appropriate planning for the varying number of analyses and information space, is conducted, where the trial is designed to evaluate if the intervention is superior to the control on ALL endpoints. Then we investigate operating characteristics of group sequential strategies in terms of overall power, Type I error rate and sample sizes. Based on the investigations, selecting different numbers of analyses for each endpoint with equal/unequal information space among the endpoints may not decrease the overall power and will not inflate the overall Type I error rate. Strategic selection regarding the testing procedure and the number of analyses with equal/unequal information space may reduce the average sample number. Early interim evaluations should be carefully evaluated as they can provide higher power, but larger average sample numbers.
These results are useful when constructing efficient group-sequential strategies in clinical trials with multiple co-primary endpoints.

Tuesday, 26th August 2014 – 11:00-12:30

Invited session

I5 Prediction to support clinical decision making

Organizer: Alessandra Nardi

I5.1

What price Cox regression? Ranking predictions from semiparametric and parametric hazard regression models via focused information criteria

NL Hjort[1]
[1]*University of Oslo, Oslo, Norway*

Consider survival data of the familiar kind, involving possibly censored survival times along with a covariate vector for each individual. The standard methods for analysing such data are based on Cox´s proportional hazard regression model, involving an unspecified baseline hazard function multiplied with a log-linear component, featuring regression coefficients etc. I shall explore fully parametric alternatives to Cox´s semiparametric model, where also the baseline hazard is modelled parametrically. I give results detailing how much the semiparametric methods may lose in terms of precision of estimates of the most relevant parameters, like survival curves and median survival time for given covariates, and also develop model and variable selection methods via suitable focused information criteria (FIC). This is different from the FIC apparatus developed in Hjort and Claeskens (JASA, 2006), in that these previous methods concentrated on selecting covariates inside the semiparametric regression framework; the present task also involves comparing and ranking models of both semiparametric and parametric types. The methods are being illustrated on real survival data.

I5.2

Individualized predictions of event times using joint longitudinal-survival models

JM Taylor[1]
[1]*University of Michigan, Ann Arbor, United States*

Following radiation therapy treatment for prostate cancer patients are monitored by regular measurements of prostate specific antigen (PSA), a simple blood test. Increasing trends in PSA are suggestive that the cancer may be regrowing and that clinical recurrence of a detectable tumor may be imminent. The patient may choose to start hormone therapy if the risk is perceived to be high. Thus for each patient it would be useful to be able to calculate the risk of clinical recurrence in the next short period of time under two conditions, either do start hormone therapy or don't start hormone therapy. Using a large training dataset we build a joint longitudinal model for the PSA values and survival model for the clinical recurrences. The longitudinal model involves random effects and the survival model involves a proportional hazards model with PSA as a time-dependent covariate. Markov chain Monte Carlo methods are used for estimation. To provide individualized predictions for a new patient, the posterior distribution from the training dataset is used as a prior for the data from the new patient. Calculation of the probability of recurrence in the next three years for the new patient involves a second MCMC algorithm. In this talk I will discuss estimation from the training data and of the predictions, how validation might be performed and the interpretation of a treatment of effect hormonal therapy in this setting.

I5.3

Model selection and ensemble predictive performance

R Henderson[1]
[1]*Newcastle University, Newcastle, United Kingdom*

With the increase in size of data sets now routinely available for analysis, traditional concepts such as significance tests or confidence intervals have become less central. Tiny differences are statistically significant and confidence intervals can be very narrow. Instead, as statisticians we need to think about practical differences between competing models say, which can only be done in close collaboration with subject specialists. It also means that the role of predictive performance is attracting more attention.

This talk considers predictive performance and model comparison in the area of event history analysis, including standard survival analysis as a special case. We look at measures of fit for complex repeated event data and examine the role and potential for ensemble prediction, as used in meteorology. There is no interest in model selection per se, but rather in combining predictions from competing models, with an appropriate weighting and a carefully chosen performance measure.

We will consider prediction in the context of dynamic models in which subject-specific non-predictable histories are informative for the future. A particular question in event history analysis is the length of follow-up needed before heterogeneity between patients can reliably be quantified.

Contributed sessions

C25 Personalized and stratified medicine II

C25.1
Mastering variation: variance components and personalised medicine

S Senn[1]
[1]CRP Santé, Strassen, Luxembourg

It ought to be clear to any statistician that there are at least four potential sources of variation in clinical trials: the main effect of treatment, the main effects of patients, treatment-by-patient interaction and within-patient variation. It should also be obvious that identification of interactive effects requires replication at the level at which interaction is claimed. Hence treatment-by-patient interaction is only fully identifiable in multi-period cross-over trials, or, which amounts to the same thing, series of n-of-1 trials.
The medical literature, however, pays scant attention to these realities and it is plausible that much of the belief that the personal component in response is important is based on a misunderstanding that apparent observed difference in outcome *must* reflect differences in the effectiveness of treatment.
Ironically, there is a fifth important source of variation: difference in medical practice that is nearly always overlooked.
It will be argued here that the key to improving the treatment of patients is to master variation and that this involves the following elements.

1. Better communication of the problems by statisticians to their colleagues (some graphical approaches will be suggested).
2. Application of decision analysis to determine when personalisation is worth pursuing..
3. Appropriate design for teasing out components of variation.
4. Application of random effect methodology for improving estimates.
5. Translating from additive to relevant scales.
6. Application of Deming´s ideas to understanding the system.
7. Realistic monitoring and feedback.

Some suggestions for addressing these issues are given.

C25.2
Subgroup analyses: time to be specific about their goals

J Tanniou[1,2], I van der Tweel[1], S Teerenstra[2,3], KCB Roes[1,2]
[1]Julius Center, Department of Biostatistics, UMC, Utrecht, The Netherlands, [2]Medicines Evaluation Board, Utrecht, The Netherlands, [3]Department of Biostatistics, Radboud UMC, Nijmegen, The Netherlands

The growing interest in personalised medicines and targeted therapies increases the attention for subgroup analysis, illustrated by the draft EMA's "guideline on the investigation of subgroups in confirmatory clinical trials".
In light of the need to further develop our understanding and improve methodology and practice, a literature review of existing statistical and methodological methods on subgroup analysis was undertaken. At least five different objectives for subgroup analysis could be considered:

1. Confirm consistency of the treatment benefits across all subgroups,
2. Identify safety problems limited to one or few subgroups,
3. Identify subgroups with larger effect when the study reaches an overall significant effect,
4. Check specific subgroups that *a priori* are suspected to show less or no treatment effect, and
5. Identify a statistically positive subgroup in case of an overall non-significant effect.

Based on this classification, most papers we reviewed fall in more than one category, but most authors were not explicit about subgroup analyses objectives when presenting methodology.
Moreover, limited attention was given to objectives 2, 3 and 4 illustrating that thinking about subgroup analyses in terms of distinctive objectives is not commonplace. Thus, research to improve statistical and methodological aspects depending on the subgroup's objective is still clearly needed.
Subgroup analysis methodology are too often undertaken without prior thoughts or knowledge about their goals, and are therefore inadequately incorporated into trial designs and methodologies.
Adequate and efficient trials should be designed not only for the main analyses but also for subgroup analyses depending of their objective.

C25.3
On the evaluation of predictive biomarkers with dichotomous endpoints: a comparison of the linear and the logistic probability models

N Heßler[1], A Ziegler[2]
[1]Inst of Med Biom & Stat, Lübeck University, Lübeck, Germany, [2]Inst of Med Biom & Stat & Center for Clinical Trials, Lübeck University, Lübeck, Germany

The standard statistical approach for analyzing dichotomous endpoints is the logistic regression model which has major statistical advantages. However, some researchers prefer the linear probability model over the logistic model in randomized trials for evaluating predictive biomarkers. The main reason seems to be the interpretation of effect estimates as absolute risk reductions which can be directly related to the number needed to treat. In the first part of our presentation, we provide a comprehensive comparison of the two different models for the investigation of treatment and biomarker effects.
Using the logistic regression model, Kraft et al. (2007, Hum Hered) showed that the combined 2 degrees of freedom (2df) gene, gene-environment interaction test should be the test of choice for testing genetic effects. In the biomarker treatment setting a gene corresponds to the treatment and environment to biomarker. Using this analogy we extend the study of Kraft et al. in the second part of our presentation.
We compare several test statistics including the 2df combination test using the linear probability model. The pros and cons of the combined test are discussed in detail. We demonstrate substantial power loss of the combination test in comparison with either the test for treatment or the test for treatment-biomarker interaction in many scenarios. Although the combination test has reasonable power in all situations considered, its power loss compared to a specialized 1df test can be large.
Therefore, the combined test cannot be recommended as the standard approach in studies of treatment-biomarker interaction.

C25.4
Design dilemmas in the multi-drug, genetic-marker-directed, non-comparative, multi-centre, multi-arm phase II National Lung Matrix Trial

L Billingham[1,2], L Crack[1], K Brock[1], S Popat[3], G Middleton[4]
[1]CRUK Clinical Trials Unit, University of Birmingham, Birmingham, United Kingdom, [2]MRC Midland Hub for Trials Methodology Research, Birmingham, United Kingdom, [3]Royal Marsden Hospital and Imperial College, London, United Kingdom, [4]School of Cancer Sciences, University of Birmingham, Birmingham, United Kingdom

Stratified medicine aims to tailor treatment decisions to individual patients, typically using molecular information to predict treatment benefit. The potential impact to benefit patients is considerable and recognised as strategically important. Cancer Research UK has made major investments into their Stratified Medicine Programme which provides a significant step

in making targeted therapies available for people with cancer in the UK, with the National Lung Matrix Trial forming the next major phase in the agenda.
The trial consists of a series of parallel, multi-centre, single-arm phase II trials, each arm testing an experimental targeted drug in a population stratified by multiple pre-specified target biomarkers. There are currently 7 targeted drugs and 20 different drug-biomarker combinations. The aim of statistical analysis is to determine whether there is sufficient signal of activity in any drug-biomarker combination to warrant further investigation. Due to the complexity of the trial, we have chosen to use a Bayesian adaptive design that gives a more realistic approach to decision-making and flexibility to make conclusions without fixing the exact sample size. The design allows early stopping of recruitment to any drug-biomarker combinations that do not show sufficient promise at an interim analysis to warrant continuation. There have been many dilemmas over the design and, for example, choices had to be made regarding whether to use a Bayesian approach, what priors to use and what outcome measure and criteria should be used for the decision to proceed. The paper will present the dilemmas and rationale for the final design.

C25.5

A new framework using G-estimation for placebo-controlled randomized phase 3 trials with extensive crossovers for biomarker-driven molecularly targeted oncology agents

S Nomura[1,2], T Shinozaki[3], C Hamada[2]
[1]National Cancer Center, Chiba, Japan, [2]Tokyo University of Science, Tokyo, Japan, [3]University of Tokyo, Tokyo, Japan

Placebo-controlled randomized phase 3 trials for biomarker-driven molecularly targeted agents (BMTAs) are playing increasingly important role in cancer drug development. In such phase 3 trials, crossover of treatment is often occurred because most of recent BMTAs are known to be overwhelmingly effective in trials of biomarker-enriched patient populations and therefore this evidence drives patients to hesitate the participation of the crossover-prohibited trials.
While intention-to-treat (ITT) analysis is the most recommendable approach in superiority trials, it is evident that the ITT analysis could not evaluate the causal survival benefit that would have been obtained had all patients complied protocol therapies in the presence of extensive non-random crossovers. To overcome this issue, we use a randomization-based G-estimation method with rank preserving structural accelerated failure time (RPSFT) models and incorporate it into phase 3 trial designs. In this proposed method, we first re-construct the potential overall survival (OS), defined as the time to death if the patients had received the assigned therapy throughout the study duration, using RPSFT models and then develop a G-test-like decision rule instead of the ITT log-rank test. Considering the three copula-type dependence structures between time to treatment switch and potential OS, we compared the performance between the proposed and ITT analysis with simulation studies.
It is shown that the proposed method could increase the power with a slight inflation of type I error. We will show the performance of two-stage design which adaptively select the analysis strategies (proposed or ITT) from the first stage data.

C26 Network meta-analysis

C26.1

A multi-state Markov model for network meta-analysis of studies with missing data

O Efthimiou[1], S Leucht[2], G Salanti[1]
[1]University of Ioannina, Ioannina, Greece, [2]Technische Universität München, Munich, Germany

Background: Missing data constitute a serious threat for the validity and precision of inferences from trials and their quantitative synthesis via meta-analysis. Various imputation methods have been proposed for the case that individual patient data (IPD) are available to systematic reviewers but their performance highly depends on the, typically unknown, missing mechanism.
Methods: We propose a multi-state Markov model for network meta-analysis with a dichotomous outcome. The model combines IPD and aggregated study data on multiple competing treatments and incorporates possible differences in study duration while accounting for patients dropping out without making any imputation for their outcomes. Three states are included in the model: response to the treatment, non-response, and study discontinuation. Our model takes into consideration the exact time of each observation, thus allowing for time-dependent relative treatment effects.
Results: We apply our model to compare the effectiveness of treatments for schizophrenia. Our model produces a joint estimation of the relative treatment effects and the dropout rates as a function of time and increases precision compared to popular imputation methods
Conclusions: The suggested model constitutes a viable candidate for performing network meta-analysis in the presence of non-ignorable missing data, with studies reporting on multiple time points and in different formats.

C26.2

Investigating consistency of mixed treatment comparisons by approximating sub-networks

J König[1], U Krahn[1], H Binder[1]
[1]University Medical Center Mainz, IMBEI, Mainz, Germany

In network meta-analysis, evidence of different studies is pooled, each comparing only few treatments. The results are network-based effect estimates for all pairs of treatments, taking indirect evidence into account. These mixed treatment comparisons are based on linear combinations of effect estimates, with coefficients that form a network of flows in a weighted directed acyclic graph. Consistency is crucial for the validity of these network based effect estimates. However, global assessment of consistency ignores which part of the network informs specific treatment comparisons. We show up a way how to tailor investigation of consistency to a specific comparison. We therefor construct approximating sub-networks that capture the bulk of evidence relevant to the comparison at hand, and define the approximating evidence proportion and a residual evidence based pseudo effect for assessing goodness of fit. As sub-networks we consider independent path decompositions, obtained via shortest path algorithms, which are amenable to a forest plot display, and more generally structured sub-networks obtained by selection based on evidence weights. The methods can be used both in fixed and random effects models. When applying them to networks of antidepressants and thrombolytics, we achieve approximating evidence proportions of over 90% while discarding half of the designs. The corresponding visualizations highlight how evidence along a few short paths is pooled into a given mixed treatment comparison. In particular, we are able to identify mixed treatment comparisons that rely on evidence distributed evenly and consistently over different independent sources and thus might be more reassuring than others.

C26.3
Visualisation of networks in meta-analysis
G Rücker[1]
[1]Medical Center, University of Freiburg, Freiburg, Germany

In network meta-analysis, the network of treatments and comparisons should be visualised. In principle, there are innumerable ways to draw a complex network, all representations of the underlying graph in the plane. Several criteria exist for optimising this representation. Often a star-shaped presentation is chosen, where all vertices lie on a circle. This is easy to implement, but typically comes at the price of many crossings. Instead, it might be desired to avoid crossings as far as possible. For planar graphs, crossings can be completely avoided. Finally, a perspective view (i.e., a 2D projection of a 3D object) might be desired, particularly for regular geometrical objects such as a cube. Hall [Hall 1970] proposed to produce 2D representations of networks using as coordinates eigenvector coefficients of the Laplacian matrix, in order to select projections in which the distances of neighboured edges are minimised. For regular planar networks, this often results in a nice perspective view when using the eigenvectors to the second and third smallest eigenvalues, whereas it results in star-shaped representations when using the eigenvectors to the largest eigenvalues. However, unfortunately, for irregular real networks this method does not work well. Alternative network visualisation algorithms aim at, e.g., keeping pre-specified `ideal´ distances between vertices, or minimising the number of crossings. We discuss the pros and cons of several criteria and approaches and give examples how to realise them using R. We argue that for sake of clarity a planar or perspective network representation is preferable to a star-shaped representation.

C26.4
Bayesian network meta-analysis for cluster-randomized trials
L Uhlmann[1], K Jensen[1], M Kieser[1]
[1]University of Heidelberg, Heidelberg, Germany

Cluster-randomized trials are used when randomization of single study participants is not possible. In the analysis of data from cluster-randomized trials the correlation within clusters has to be taken into account, otherwise the type I error rate may be inflated. To combine the effects in a meta-analysis of cluster-randomized trials accordingly, either the correlation must be taken into account in the analysis of the single trials, or an adjustment of the effects or the variances must be performed.
The variance inflation caused by the correlation can be taken into account in various ways. For classical pairwise meta-analysis, methods were proposed and their performance characteristics were compared. In our contribution we extend these approaches to network meta-analyses. In a first step, we illustrate how pairwise meta-analyses including cluster-randomized trials can be conducted using a Bayesian approach. Furthermore, the models are extended such that they can be used to conduct multiple comparisons in a meta-analysis. With these models, network meta-analyses of cluster-randomized trials can be performed.
By use of simulation studies we evaluate the type I error rate to compare the derived methods. The results show that, in contrast to the unadjusted approach, our models do not lead to an inflation of the type I error rate. Finally, results of an investigation of the power characteristics are presented.

C26.5
Precision of the estimates from a network meta-analysis model and their role in planning future studies
A Nikolakopoulou[1], D Mavridis[1,2], G Salanti[1]
[1]University of Ioannina, School of Medicine, Ioannina, Greece, [2]University of Ioannina, Department of Primary Education, Ioannina, Greece

When there are multiple competing interventions for a healthcare problem the design of new studies could be based on the entire network of evidence as reflected in a network meta-analysis (NMA). There is a practical need to answer how many (if any) studies are needed, of which design (the treatments being compared) and with what sample size to infer conclusively about the relative treatment effects of all competing treatments and their relative ranking.
We have previously addressed these questions based on the conditional power of NMA. Here we present methodology that approaches the same questions from a different angle. We consider the precision in the results obtained from NMA: the precision in the joint distribution of the estimated basic parameters of the model and the precision in the treatment ranking. We quantify the precision in the estimated effects by considering their variance-covariance matrix and estimate the precision in ranking by quantifying the dissimilarity of the density functions of summary estimates.
Then, based on a target improvement in precision we calculate the required sample size for each possible study design and number of study arms and we present visual tools that can help trialists select the optimal study design. We used a published network of interventions for the treatment of hepatocellular carcinoma to illustrate the suggested methodology.
Results show that precision gain depends on the type of treatment comparisons tested in new studies. The presented methodology can aid investigators making informed and evidence based decisions about planning new studies.

C27 Survival analysis I

C27.1
Survival probability with non-reversible time varying treatment indicator: theoretical quantities and nonparametric estimators
L Antolini[1], DP Bernasconi[1], S Iacobelli[2], MG Valsecchi[1]
[1]Università Milano Bicocca, Monza, Italy, [2]Università di Roma "Tor Vergata", Roma, Italy

Inference on survival according to a non-reversible time varying Treatment is often performed by applying the Cox model or the Mantel-Byar's test while a reliable non-parametric description of the survival experience is still not fully established.
Simon-Makuch (SM) curves were derived generalizing the Kaplan-Meier (KM) formula, initially classifying patients at a landmark time and dynamically updating the risk-sets for subsequent time-points. The curve for patients switching treatment has been criticized, since it is unclear which quantity it estimates. The time-scale originates from start of standard treatment and the switch is considered as delayed entry into the alternative treatment. SM estimates at time t from landmark are based on a mixture of individual hazards, heterogeneous with regard to time from switch. An alternative approach consists in using a clock-back scale where the survival is estimated at time from switch, i.e. considering homogeneous risks sets on this time-scale, but heterogeneous on the original one. Both curves estimate the counterfactual survival of patients under alternative

treatment from beginning, whereas the survival of patients under standard treatment represents the counterfactual net survival where treatment switch is removed.
Motivated by an international study on acute lymphoblastic leukemia with patients undergoing chemotherapy and with possible additional stem cell transplant, we aim to identify survival probabilities and suitable nonparametric estimators.
In particular, we will compare the SM survival curve as well as the KM "clock back" curve, discussing the assumptions (e.g. Markov property) of both methods.
Alternatives derived in the framework of multi-state models will be considered.

C27.2

Using pseudo-values for comparing long-term survival after stem-cell transplantation (SCT) with long-term survival after chemotherapy

U Pötschger[1,2], H Heinzl[2], MG Valsecchi[3], M Mittlböck[2]
[1]Children's Cancer Research Institute, Vienna, Austria, [2]Center for Medical Statistics, Medical University, Vienna, Austria, [3]Center of Biostatistics for Clinical Epidemiology, Monza, Italy

Aims: Allogeneic SCT is a therapeutic option in high-risk leukemia whose ability to improve long-term survival is still under discussion. From a methodological perspective, when investigating long-term survival rates, it is unclear and challenging how to adjust for waiting-time until donor identification without relying on proportional hazards.
Methods: With non-proportional hazards, the pseudo-value regression provides an attractive approach. An adaptation can be introduced that allows for a time-dependent covariate. Pseudo-values for survival expectations of patients with transplantation are generated which address the cumulative hazards before and after transplantation. These are compared in a generalised linear model to pseudo-values for baseline survival rates without SCT. Real data in leukaemia and a simulation study illustrate the practical value and the statistical properties of the novel approach.
Results: The simulations show unbiasedness of the estimated parameters. In the real data example, SCT has worse early survival due to toxicities but better survival later on. Consequently Cox-regression is not suitable (HR=0.91, p=0.39) due to non-proportional hazards. However the pseudo-value approach shows that the cumulative hazards at 5-years under SCT is favorable to no SCT (HR=0.67, p=0.02). Additionally no dependence between waiting time and resulting 5-years survival was observed.
Conclusion: The proposed novel approach allows investigating the impact of a binary time-dependent covariate on long-term survival without relying on proportional hazards.

C27.3

Assessing effects of treatment change on survival when the measurement pattern of covariates and events are dependent

I Schmidtmann[1], A Weinmann[2], A Schritz[1,3], D Zöller[1], H Binder[1]
[1]IMBEI, Universitätsmedizin Mainz, Mainz, Germany, [2]I. Medizinische Klinik, Universitätsmedizin Mainz, Mainz, Germany, [3]Hochschule Koblenz, Rhein-Ahr-Campus, Remagen, Germany

Time-to-event data arising from clinical cancer registries often involve competing risks or other complex event patterns, such as time-dependent confounding, which may occur when treatment changes are driven by time-varying covariates that both influence prognosis and depend on previous treatment. Extensions of the Cox regression model allow for such complex event structures. A further complication arises when the pattern of measuring such time-varying covariates is irregular and updates are more frequent depending on patient health status or imminent treatment change.
We investigate the effect of a treatment change in a hepatocellular carcinoma registry by considering conditional survival at several update time points after start of initial treatment. Covariates considered here include treatment (changed / not yet changed), baseline covariates, and covariate values prior to the update time points. The influence of the measurement pattern of the time-varying covariates is examined by fitting regression models using alternatively all available covariate values and covariate values from artificially coarsened measurement patterns. As illustrated for the Mainz hepatocellular carcinoma registry, the difference in regression coefficients then indicates the potential extent of bias due to event-associated measurement patterns. Use of a propensity score approach is considered for reducing this bias while still retaining the information from all measurements.

C27.4

Clustering for treatment effect on recurrent events

C Schramm[1,2], G Diao[3], S Katsahian[1,3]
[1]INSERM UMRS 1138, Team 22, Paris, France, [2]INSERM U955, Team 1, Créteil, France, [3]Department of Statistics, George Mason University, Fairfax, United States

Recurrent events occur when a subject experiences repeated occurrences of the same event over follow-up time. Rate of events gives information about disease gravity. A specific treatment must reduce this rate and improve the quality of life of patients. However, treatment benefit could vary according to patients' characteristics. Identifying a subpopulation of patients that could most benefit from a treatment is interesting in order to target that treatment to those patients.
The current work aims to define positive and negative treatment responders according to the treatment effect on occurrence of recurrent events.
The proposed method supposes that patients are followed-up before and after treatment initiation. We propose to use a Cox proportional hazards mixed model with the treatment as a time-dependent covariate. Gaussian frailty terms are associated to the intercept and the treatment effect. Then the random components of each patient are predicted and put in entry of non-supervised clustering algorithms in order to build clusters of patients.
A simulation study is performed to examine the performance of the proposed methodology. Several non-supervised algorithms combined with several distances are envisaged. Methods are evaluated on the percentage of correctly classified patients and on the estimation of treatment effect in each cluster.
Finding cluster of patients with a potential benefit from a treatment is essential in order to find biomarkers of a treatment efficacy.

C27.5

A total time approach for the simulation of recurrent event data when planning a clinical trial

A Jahn-Eimermacher[1], K Ingel[1], A-K Ozga[2], H Binder[1]
[1]University Medical Center, Mainz, Germany, [2]University of Applied Sciences, Koblenz, Germany

Simulation techniques are an important tool for planning clinical trials with complex time-to-event structure. We specifically consider simulation of recurrent events, which can follow two different general types of event-generating processes. In a gap time approach the hazard for experiencing events is defined on the time since last event, and simulation of the identically distributed inter-event-times is straightforward. However, in many clinical settings the hazard depends on the time since study start, e.g. the time since beginning treatment of some progressive disease. This calls for a total time approach.
Accordingly, we propose a method for simulating recurrent event data for an arbitrary hazard function defined on total time. We identify the distributions of inter-event-times conditional on the times of previous events

by deriving their conditional hazard functions. We recursively simulate inter-event-times using the conditional cumulative hazards and adopt the univariate approach of Bender et. al. (StatMed 2003). We extend our methods to incorporate fixed and random covariates into a proportional hazards model.
Methods are illustrated by simulating clinical trial data with recurrent events under different total time models (weibull, log-normal, step-function). This is used to obtain the empirical power of an Andersen-Gill analysis depending on censoring, shape of hazards and between-subject-variation. We find that censoring most affects the power of a study if hazards vary with total time and between subjects. This highlights the usefulness of our simulation approach for planning clinical trials, as different potentially relevant scenarios, e.g. for progressive diseases, can easily be investigated.

C28 Marginal structural models

C28.1
Addressing measurement error in time-varying covariates through the use of simex-adjusted marginal structural models

RP Kyle[1], EEM Moodie[1], M Abrahamowicz[1]
[1]McGill University, Montreal, Canada

Background: The assumption of no unmeasured confounding is fundamental to unbiased estimation of causal parameters from marginal structural models (MSMs). While substantial measurement error is known to induce residual confounding in unweighted regression models, few authors have addressed this issue in the context of the treatment model used to generate inverse probability weights (IPW) for use in MSMs.
Objectives: Our primary aims were to validate a novel application of the simulation-extrapolation (SIMEX) procedure to reduce the impact of measurement error in MSMs, and to demonstrate its utility by application to an analysis of empirical data.
Methods: The SIMEX is a simulation-based method for reducing measurement error when the measurement error variance of a given covariate is known precisely or may be well-estimated (Cook & Stefanski, JASA 1994). In a series of simulation studies, we examined varying degrees of measurement error in a single time-varying covariate in the treatment model from which IPW were obtained. We modified several simulation parameters, including sample size, treatment and covariate effect sizes, and assessed the robustness of error correction given differing assumptions of error variance. Following analyses of simulated data, we fit SIMEX-adjusted MSMs to data from the Multicenter AIDS Cohort Study.
Conclusions: Correcting measurement error via SIMEX in MSMs improves covariate balance, and is a useful tool for reducing bias and improving precision in treatment estimates from MSMs.

C28.2
A caution on the use of stabilized weights in marginal structural models

D Talbot[1,2], J Atherton[1], AM Rossi[3], SL Bacon[3,4], G Lefebvre[1]
[1]Université du Québec à Montréal, Montréal, Canada, [2]Université Laval, Québec, Canada, [3]Concordia University, Montreal, Canada, [4]Montreal Behavioural Medicine Centre, Montreal, Canada

Marginal structural models (MSMs) are commonly used to estimate the causal effect of a time-varying treatment in presence of time-dependent confounding. When fitting a MSM to data, the analyst must specify both the treatment model for the inverse-probability-of-treatment weights and the marginal structural model for the outcome. With MSMs, the use of stabilized weights is recommended since they are generally less variable than the standard weights. In this work, we are concerned with the use of the common stabilized weights when the structural model is specified to only consider partial treatment history, such as the current or most recent treatments. We present various examples of settings where these stabilized weights yield biased inferences while the standard weights do not. These issues are first investigated on the basis of simulated data and subsequently exemplified using data from the Honolulu Heart Program. In conclusion, we suggest replacing the common stabilized weights with basic stabilized weights that do not share the problems of the former.

C28.3
Dialysis, catheter use and mortality: challenges in applying marginal structural models to data from a clinical registry

J Kasza[1], R Wolfe[1], K Polkinghorne[1]
[1]Monash University, Melbourne, Australia

With aging populations and growing rates of obesity, the burden of kidney disease is increasing in many countries. Dialysis, the most frequently-used treatment for end-stage kidney disease, is undertaken using one of three modalities: haemodialysis, classified by location of delivery (home/facility), or peritoneal dialysis. Haemodialysis requires a vascular access be in place, and there are three alternative access types: central venous catheter, arteriovenous graft, or arteriovenous fistula.
Using data from the Australian and New Zealand Dialysis and Transplant Registry (ANZDATA), we seek to determine the effect of dialysis modality, with haemodialysis sub-classified by vascular access type and location, on mortality. Complicating the estimation of this effect is that throughout the course of a patient's treatment, their dialysis modality and vascular access type may change. Further, co-morbid conditions such as coronary artery disease are time-dependent confounders affected by dialysis history.
We use marginal structural models to estimate the effect of dialysis modality on mortality. However there are several difficulties posed by the structure of ANZDATA that are typical of clinical registries. In particular, we will discuss approaches to accounting for the clustering of patients in dialysis treatment centres when not all dialysis modalities are available at all centres. We also explore the bias in the estimation of treatment effects when modality changes are recorded as they occur (as is optimal), but vascular access and comorbidities are only recorded annually. We also discuss methods for dealing with large stabilised inverse probability of treatment weights, which are an issue in this context.

C28.4
Non-specific effects of vaccines on child morbidity examined with a marginal structural model for recurrent events.

AKG Jensen[1,2], H Ravn[2], PK Andersen[1]
[1]University of Copenhagen, Copenhagen, Denmark, [2]Statens Serum Institut, Copenhagen, Denmark

Non-specific vaccination effects denote a possible boost or attenuation of the immune system following a vaccination. Observational studies from Guinea-Bissau where even ordinary infectious diseases can be fatal have shown a dramatic effect on child survival due to non-specific effects.
To examine non-specific effects in a high-income setting based on Danish registries where infant morbidity (infectious hospitalizations) is the outcome and recurrent events are common we introduce a marginal structural model for recurrent events. In this model the recurrent outcome is associated with future risk of infectious hospitalization and future vaccination-status. In addition the recurrent outcome is affected by past vaccinations and hence also plays the role as time-dependent confounder affected by previous treatment in the marginal structural model.

The effect measure for this model is based on the mean number of events in the study period for a given treatment regime. This number is obtained by integration over hazards of hospitalization in discretized risk-periods. It allows a flexible modelling with multiple hazard-ratio parameters compared to the proportional hazard assumption in the standard marginal structural Cox model.
Weights based on continuously updated vaccination probabilities adjust for the time-depending confounding. The weights are estimated using poisson regression models.
We illustrate our approach in a multistate setup with 5 different vaccination states allowing causal comparison, on a day to day basis, between a variety of vaccination schedules for children aged 15-24 months.

C28.5
Robustness and efficiency in instrumental variable models with covariates
V Didelez[1], S Vansteelandt[2]
[1]Maths, University of Bristol, Bristol, United Kingdom, [2]Ghent University, Ghent, Belgium

Instrumental variables provide an approach for consistent inference on causal effects even in the presence of unmeasured confounding. Such methods have for instance been used in the context of Mendelian randomisation, as well as in pharmaco-epidemiological contexts. In these and other applications, it is common that covariates are available, even if deemed insufficient to adjust for all confounding. As IVs allow inference when there is unobserved confounding, it appears that often the analyst assumes that even observed confounders / covariates do not need to or should not be taken into account.
However, this is not generally the case. With view to the role of covariates, we here contrast two-stage least squares estimators, generalized methods of moment estimators and variants thereof with methods more common in biostatistics using G-estimation in so-called structural mean and distribution models. When using covariates, there are structural aspects to be considered, e.g. whether the covariates are prior to or potentially affected by the instruments. But in addition, one has to worry even more about efficiency versus model misspecification when modelling covariates. We discuss this for the IV procedures mentioned above, especially for linear and loglinear instrumental variable models.
Our results motivate adaptive procedures that guarantee efficiency improvements through covariate adjustment, without the need for covariate selection strategies. Besides theoretical findings, simulation results will be shown to provide numerical insight.

C29 Clinical trial designs

C29.1
Patient-oriented randomization - a new clinical design
C Schulz[1], J Timm[1]
[1]University of Bremen, Bremen, Germany

The "gold standard" for clinical studies is a controlled, randomized and double-blinded trial usually comparing specific treatments. However, this procedure is far away from the physician´s daily routine. From an ethical point of view, the physician should be involved in the decision concerning treatment, taking risks and healing opportunities of each patient into account. Therefore, the question arises: is there a way to combine randomization and patient-oriented decisions in a clinical trial design?
The answer is yes if strategies instead of specific treatments are compared. The idea is to randomize the strategies and let the physician decide between treatments within these strategies.
An example is the clinical trial NeSSy [1] with a randomized design comparing efficacy and safety of the strategies to use either conventional or newer antipsychiotic drugs in patients suffering from schizophrenia.
The new idea of randomization is generalized to the case of two different strategies with an arbitrary number of treatments within each strategy. Preliminary results will be displayed showing the behavior of this innovative design with respect to balance between strategies and between single treatments within each strategy. Main results cover the influence of treatment number in each strategy, number of patients and centres and the decision rules of the physicians. Results have been generated by theoretical consideration as well as simulation studies. Furthermore, results regarding the study design of the NeSSy study will be presented.
Reference:
[1] The Neuroleptic Strategy Study - NeSSy, funded under the program of clinical studies of the BMBF

C29.2
A novel modified standard-gamble task to measure patients' preferences for biomarker-led care
J Harrington[1,2], M Morgan[2], A Cronin[3,4], R Hilton[3], S Sacks[4], M Hernandez-Fuentes[1,4,5], I Rebollo-Mesa[4]
[1]NIHR Biomedical Research Centre, King's College London, London, United Kingdom, [2]D.Health and Social Care Research, King's College London, London, United Kingdom, [3]Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom, [4]MRC Centre for Transplantation, King's College London, London, United Kingdom, [5]GAMBIT Consortium, www.GAMBITstudy.co.uk, London, United Kingdom

Background: The success of personalized medicine is directly dependent upon the development of accurate clinical prediction models. However, predictive markers are rarely 100% accurate. Varying levels of accuracy are associated with different levels of risk in the subsequent clinical decision. Incorporating patients' preferences into the development and validation of clinical prediction models is vital for a seamless translation into clinical practice.
Aims:
1) To develop a novel modified Standard-Gamble task (MSG) to measure patients' preferences for biomarker-led care (BLC), as a function of the marker's sensitivity and specificity;
2) To use the MSG to obtain the Minimum Tolerated Specificity (MTSp) and Minimum Tolerated Sensitivity (MTSe) in a study of biomarkers of tolerance in kidney transplant recipients (KTR); and 3) To study the association between MTSp, MTSe and patients' symptom burden.
Methods: Participants: 87 KTR with stable function, and 13 operationally tolerant. Measures: 1) MSG task, 2) symptom burden questionnaire and 3) qualitative interview.
Results: Preliminary results (N=57) show that KTR require a highly specific test to accept BLC (Median=0.91). Sensitivity is of less concern (Median=0.67). No significant relationship was found between attitude towards risk and symptom burden, although immunosupression related symptoms tend to be higher among high-risk takers.
Conclusion: Our findings show that the MSG can be used to measure patients' treatment choice as a function of a (bio)marker's specificity and sensitivity, and the risks associated with it. Further research is necessary to understand the factors associated with patients' choice, and incorporating them in the process of (bio)marker validation.

C29.3
Response adaptive randomization in large phase III confirmative clinical trials with binary outcomes - benefits are unlikely
W Zhao[1], VL Durkalski-Mauldin[1]
[1]*Medical University of South Carolina, Charleston, United States*

In the recent decade, response adaptive randomization (RAR) has been advocated for its benefits in study subject ethics (assigning a higher percentage of subjects to the so far better performing arm) and trial efficiency (power). Literature on the benefit and cost of using RAR in a real trial is minimal. Based on pure theoretical analysis with conceptual scenarios of trials with binary outcomes and a fixed sample size, it is indicated that using RAR can minimize the total number of failures and maximize the power. However, our computer simulation studies based on various large confirmative phase III trials in a frequentist setting reveal that the efficiency benefit is trivial; and the ethical benefit is obtained at the cost of efficiency. More importantly, under the condition of fixed power, using RAR will more likely increase, not decrease, the total number of failures. This result contradicts what many investigators expect from RAR. Further studies demonstrate that when a time trend exists in the trial, using RAR may cause noticeable inflation or deflation of type I error, which will make interim analysis more complex and ultimately reduce the interpretability of the trial. Therefore, we recommend not to use RAR for large confirmative phase III clinical trials in a frequentist setting.

C29.4
Incorporating feasibility assessment in the design of clinical studies
T Jaki[1], LV Hampson[1]
[1]*Lancaster University, Lancaster, United Kingdom*

Many publicly funded clinical trials fail to meet their recruitment timelines, with the consequence that these trials then require an extension of funding in order to complete recruitment. To avoid this scenario, there is a movement by funders towards requiring that larger Phase II and Phase III clinical trials incorporate a feasibility stopping rule, with the aim of establishing early on whether recruitment targets can be met within the planned time frame.
The feasibility evaluation is usually based upon factors that are not of primary interest to the trial (i.e., do not concern the endpoint of direct clinical interest) and allows for three different actions: continue as planned; adapt recruitment procedures; or abandon the trial. Efficacy data collected during the feasibility phase of the trial contribute towards the final analysis of efficacy. In this presentation, we will show how ideas from the adaptive designs literature can be used to incorporate feasibility evaluations into the main trial design to ensure that the required type I error rate for testing efficacy is maintained and power is maximised. Simulations are used to illustrate the potential gains in power that follow from using our proposed approach. Optimal boundaries for the feasibility stopping rule are derived which minimise the expected overrun of the trial beyond its planned duration subject to controlling the probabilities of incorrectly allowing a trial to proceed when the recruitment rate is insufficient, and incorrectly abandoning a trial that would have gone on to complete in a timely manner.

C29.5
Some novel alternatives to parallel group designs for pragmatic clinical trials
R Hooper[1], L Bourke[1]
[1]*Queen Mary University of London, London, United Kingdom*

A before-after comparison in the same participants is a powerful way to evaluate the effect of an intervention, but a clinical trial requires a concurrent control - for example a parallel groups design with baseline and follow-up assessments in both intervention and control groups.
One alternative which exploits the power of the before-after comparison is the cross-over design, but this assumes the treatment effect from the first period has disappeared by the time the effect is measured in the second period. Cross-over trials are therefore problematic for interventions whose effects are maintained. In a trial where the comparator is routine care, however, it is often reasonable to assume the effect of a treatment introduced at some period of time following randomisation is independent of that period of time. In this case there are a variety of alternatives to parallel groups and cross-over designs.
One approach - the stepped wedge design - has been used extensively. Stepped wedge designs come with a heavy burden of assessment, however, and require a model for how treatment effects are maintained in order to analyse repeated assessments after introduction of the intervention. An intriguing alternative is to reduce the schedule of assessments in the different randomised groups to a much sparser arrangement. These incomplete unidirectional cross-over designs (the simplest being the recently published 'dog-leg' design) offer the remarkable possibility of more power with fewer assessments than a parallel groups design. Dog-leg designs are likely to be particularly useful for cluster-randomised trials involving repeated cross-sections.

C30 Adaptive designs II

C30.1
Estimation after blinded sample size reassessment
F Klinglmueller[1], M Posch[1], F König[1], F Miller[2]
[1]*Medical University of Vienna, CeMSIIS, Vienna, Austria,* [2]*Stockholm University, Dep. of Statistics, Stockholm, Sweden*

When comparing the means of normally distributed endpoints the sample size to achieve a target power typically depends on nuisance parameters as the variance. It has been shown that superiority trials where the sample size is reassessed based on blinded interim estimates of the nuisance parameter achieve the target power regardless of the true nuisance parameter and the sample size reassessment has no relevant impact on the type I error rate.
While previous work has focused on the control of the type I error rate, we investigate the properties of point estimates and confidence intervals following blinded sample size reassessment. We show that the maximum likelihood estimates for the mean and variance may be biased and quantify the bias in simulations. Furthermore, we provide a lower bound for the bias of the variance estimate and show by simulation that the coverage probabilities of confidence intervals may lie below their nominal level, especially when first stage sample sizes are small. Finally, we discuss the impact of the findings for blinded sample size reassessment in clinical trials.

C30.2
A two-stage adaptive design for small clinical trials

S Nikolakopoulos[1], KCB Roes[1], I van der Tweel[1], C Jennison[2]
[1]UMC Utrecht, Utrecht, The Netherlands, [2]University of Bath, Bath, United Kingdom

Patients suffering from rare conditions should be entitled to the same quality of treatment as other patients. Clinical trials for orphan drugs are conducted with moderate to very small sample sizes since the pool of possible study participants is by definition small. In addition, such trials may be conducted with little to no prior information on expected effects and their variation, thus making the assumptions made at the design stage prone to misjudgment. Therefore, flexible designs that have the ability to respond to information gathered within the trial seem an appealing alternative to traditional trial designs. We propose a two-stage adaptive design for the situation where the total sample size has an assumed maximum which cannot be exceeded. In combination with the lack of information about the parameters of interest at the design stage, standard sample size calculations fall short in such cases. To optimize the trade-off between sample units and information gathered, a utility-based approach is suggested. In such a way, success of the clinical trial is taken under consideration as a frequentist concept, as it is most of the times required by drug regulators. In addition, utilities are defined for the treatment effect to be detected on the patient level, as well for the cost of sampling.

C30.3
Adaptive designs for time-to-event trials

D Magirr[1], T Jaki[1], F König[2], M Posch[2]
[1]Lancaster University, Lancaster, United Kingdom, [2]Medical University of Vienna, Vienna, Austria

Mid-study design modifications are becoming increasingly accepted in confirmatory clinical trials, so long as appropriate methods are applied such that error rates are controlled. It is therefore unfortunate that the important case of time-to-event endpoints is not easily handled by the standard theory. We analyze current methods that allow design modifications to be based on the full interim data, i.e., not only the observed event times but also secondary endpoint and safety data from patients who are yet to have an event. We show that the final test statistic may ignore a substantial subset of the observed event times. Since it is the data corresponding to the earliest recruited patients that is ignored, this neglect becomes egregious when there is specific interest in learning about long-term survival. An alternative test incorporating all event times is proposed, where a conservative assumption is made in order to guarantee type I error control. We examine the properties of our proposed approach using the example of a clinical trial comparing two cancer therapies.

C30.4
Adaptive designs for two candidate primary time-to-event endpoints

G Rauch[1], M Kieser[1]
[1]University of Heidelberg, Heidelberg, Germany

Composite endpoints combine several time-to-event variables of interest within a single time-to-first-event analysis. The motivation for the use of a composite endpoint is to increase power by increasing the number of expected events.
However, in some situations a particular component that was exclusively added to the composite in order to increase the effect in fact decreased the composite effect. The CAPRICORN Trial [1] is a very illustrative example for this situation. Another possible scenario would be that the main component which is the most relevant for the patient (e.g. time-to-death) shows a higher effect than originally anticipated. In this situation it might be feasible to base sample size recalculation on the main component in order to improve the interpretation of the trial.
In both situations, an adaptive design that allows sample size recalculation during the interim analysis based on the larger observed effect of two candidate endpoints would be helpful.
We propose different adaptive design strategies to face the above problems and evaluate and compare them in terms of power and type I error using Monte-Carlo simulations. Applications are illustrated by a clinical study example.

References:

[1] The CAPRICORN Investigators. Effect of carvedilol on outcome after myocardial infarction inpatients with left-ventricular dysfunction: the CAPRICORN randomised trial. *Lancet* 2001; 357: 1385-1390.

C30.5
Backward image confidence intervals for adaptive group sequential trials

S Solanki[1], N Deshpande[1]
[1]Cytel Statistical Software and Services Pvt. Ltd., Pune, India

An adaptive trial is defined as any clinical trial which uses accumulating data, possibly combined with external information, to modify aspects of the design without undermining the validity and integrity of the trial. Müller and Schäfer provided a methodology for conducting an adaptive trial which guaranteed control of type 1 error while providing maximum flexibility.
However, corresponding solutions for the equally important and related problem of parameter estimation at the end of the adaptive trial have not been completely satisfactory. In their paper "Exact inference for adaptive group sequential designs" (April 2013), Ping Gao, Lingyun Liu and Cyrus Mehta introduce a method called Backward Image Confidence Intervals (BWCI) which is based on mapping the final test statistic obtained in the modified trial into a corresponding backward image in the original trial. It computes a two-sided confidence interval having exact coverage, along with a point estimate that is median unbiased for the primary efficacy parameter in a two-arm adaptive group sequential design.
This method will be discussed here with the help of several examples. Along with it, we will discuss advantages of this procedure over previously available methods, which either produced conservative coverage or no point estimates or provided exact coverage for one-sided intervals only. We will use simulation results generated by Cytel's software East® for this purpose.

Wednesday, 27th August 2014 – 9:00-11:00

Invited session

S1 STRATOS (Strengthening Thinking about Analyses of Observational Studies) initiative: first results & future steps

Organizers: Willi Sauerbrei and Harbajan Chadha-Boreham

S1.1
Setting the stage with initial data analyses

M Huebner[1], S Le Cessie[2], W Vach[3], M Blettner[4], D Bodicoat[5]
[1]Michigan State University, East Lansing, United States, [2]Leiden University Medical Center, Leiden, The Netherlands, [3]Universitaet Freiburg, Freiburg, Germany, [4]Universitaet Mainz, Mainz, Germany, [5]University of Leicester, Leicester, United Kingdom

The importance of initial data analyses in observational studies has been recognized but is often neglected. Careful data preparation and description is crucial before embarking on complex analyses to avoid spurious results.

Obtaining high quality data starts far before data collection and includes a careful database design with variable definitions, within and between variable plausibility checks and date checks. Data should be cleaned systematically and carefully, especially when integrating multiple data sources. Changes to the data such as corrections, transformations, definitions of categories, or treatment of missing data need to be integrated in the programming code for reproducibility, rather than changes to the raw dataset. Data cleaning can take as much as 80% effort of the analysis, and the process may need to be automated for large data sets. The inclusion and exclusion criteria in the process of selecting the subset of data to be analyzed in the study should be described with an overview of missing measurements and follow-up data. Numerical and graphical descriptions include table summaries, illustrating correlations and confounding factors, or examining distributions of variables and homogeneity of groups, keeping in mind the objectives of the study.

It is especially important that the complete initial data analysis process is transparent and that researchers document all steps for reproducibility. STRATOS topic group 3 aims to provide guidance on this process based on an overview of existing literature with examples and feedback from experienced statisticians.

S1.2
Evaluation of incremental value of a marker: a historic perspective on the Net Reclassification Improvement

EW Steyerberg[1], P Macaskill[2], AV Vickers[3]
[1]Erasmus MC, Rotterdam, The Netherlands, [2]University of Sydney, Sydney, Australia, [3]Memorial Sloan Kettering Cancer Center, New York, United States

The net reclassification improvement (NRI) is an increasingly popular measure for evaluating improvements in risk predictions. In a recent review, 67 publications were considered from high-impact general clinical journals that considered the NRI. Incomplete reporting of NRI methods, incorrect calculation, and common misinterpretations were found. To aid improved applications of the NRI, the article elaborated on aspects of the computation and interpretation in various settings [1].

An accompanying Editorial emphasized conceptual problems (such as weighting reclassification inappropriately), and statistical problems (such as artificial inflation of NRI values and Type I error). It doubted that anything would be gained by reporting the NRI, either overall or in its components. The Editorial suggested that investigators should move away from statistical abstractions, such as the NRI, and illustrate the consequences of using a marker or model in straightforward clinical terms [2].

In this presentation, we will discuss the use of NRI and related performance measures from a historical perspective and suggest directions for improvement, particularly with respect to the use of decision analytic measures.

References:

1. Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. Ann Intern Med. 2014 Jan 21;160(2):122-31.
2. Vickers AJ, Pepe M. Does the net reclassification improvement help us evaluate models and markers? Ann Intern Med. 2014 Jan 21;160(2):136-7.

S1.3
Review of methods used in recent observational epidemiological studies to select variables and their functional forms [STRATOS Task Group 2]

M Abrahamowicz[1], RP Kyle[1]
[1]McGill University, Montreal, Canada

The over-arching STRATOS goal is to improve statistical methodology used in real-life observational studies, focusing on selected 'generic' issues. Task Group 2 deals with selection of independent variables, and functional forms for continuous variables, in multivariable explanatory models. To convince end-users to adapt more sophisticated statistical methods, and demonstrate the weaknesses of 'conventional', currently applied methods, we reviewed methods in 50 papers published in 2013 in high-ranking epidemiology and clinical journals, which focused on some continuous variable(s).

Whereas several studies selected independent variables *a priori,* on substantive grounds, many used arbitrarily selected data-dependent criteria or procedures, and failed to account for, or even mention, their impact on the estimation and inference. As expected, most studies imposed *a priori* linearity of the effects of continuous variable(s), and did not test or evaluate this assumption. Review of additional papers revealed that linearity was often imposed even for associations consistently demonstrated to be non-linear, in previous flexible analyses (e.g. BMI versus mortality; age at diagnosis versus recurrence or death in various cancers). Studies that did consider possibly non-linear relationships employed different flexible methods (polynomials, fractional polynomials, and various spline-based approaches), and used different criteria to assess (non-)linearity.

In conclusion, current applied research will benefit from evidence-based guidance, and a systematic comparison of methods, for selection of variables and their functional forms.

S1.4
Causal questions and principled answers: a guide through the landscape for practicing statisticians

E Goetghebeur[1], E Moodie[2], I Waernbaum[3], S Le Cessie[4]
[1]Ghent University, Gent, Belgium, [2]Mc Gill University, Montreal, Canada, [3]Umea University, Umea, Sweden, [4]Leiden University Medical Center, Leiden, The Netherlands

Causal inference came a long way over the past decade. New methodological approaches casting assumptions, models and results in terms of potential outcomes find their way into the clinical literature. This opens great potential for deepened understanding, but also for misunderstandings if subtle assumptions or interpretation of results are misunderstood. When choosing the causal analysis method or synthesizing evidence from different approaches one should be clear about the specific questions they aim to answer and about what can (not) be achieved from available

observations. The choices lead to a focus on direct or indirect, conditional or marginal effects of a particular type of exposure for specific (sub)populations over a time horizon. We examine what is kept fixed and let loose for the practical question `what if exposure had been different'.

The choices come with different meaning and distinct technical challenges. We consider how practical questions and answers differ under the no unmeasured confounders assumption relying on outcome and/or propensity score models and/or matching; or the instrumental variables assumption in a marginal or (double) conditional set-up including principle strata. We point to tutorials on the separate methods, line up overlap and differences in a principled fashion and by example. We zoom in on effectiveness research aiming to learn about drug effects (such as statins) from electronic health records and on the evaluation of quality of care in terms of hospital outcomes.

Presented by Els Goetghebeur for STRATOS TG7 including Saskia Le Cessie, Erica Moodie, Ingeborg Waernbaum et al.

Wednesday, 27th August 2014 – 9:00-10:48

Contributed sessions

C31 Variable selection in high-dimensional models

C31.1 *Student Conference Award*

An extension of the lasso penalization to reduce false positive selection in high-dimensional Cox models

N Ternès[1,2], F Rotolo[1], S Michiels[1,2]
[1]*Institut Gustave Roussy, Department of Biostatistics and Epidemiology, Villejuif, France,* [2]*University Paris Sud, Le Kremlin-Bicêtre, France*

Introduction: An increasing interest is being devoted to select the right prognostic biomarkers among multiple candidates. Regression with LASSO penalization is a popular variable selection method, but strongly depends on the penalization parameter λ. Usually, λ is chosen via maximum cross-validated log-likelihood (max-*cvl*). Yet, this choice often detects too many false positives.
Methods: We propose an AIC-like penalized *cvl* (*pcvl*), a function of the number of non-null regression parameters in the model, trading off the goodness of fit (small λ) and the parsimony of the model (big λ). According to this extension, the optimal λ is greater or equal to the max-*cvl* one, and selects fewer biomarkers.
We evaluate the false discovery (FDR) and false negative rate (FNR) in a simulation study by varying sample size, number and prevalence of binary biomarkers, number of active biomarkers, correlation and censoring. Finally, we apply these methods on two publicly available mutation and gene expression data sets in non-small cell lung cancer from The Cancer Genome Atlas database.
Results: In null scenarios (i.e. no active biomarker), no difference was observed between the two methods in terms of FDR, however, *pcvl* selected on average fewer biomarkers. In alternative scenarios, the FDR was systematically lower for *pcvl*.
The FNR was low and comparable for both methods, although slightly higher for the *pcvl* with a small sample size and a high number of active and non-active biomarkers.
Conclusion: Maximum *pcvl* yields much less false positive biomarkers with lasso penalization in high-dimensional Cox regression models.

C31.2

Biomarker discovery: controlling false discoveries in high dimensional situations

B Hofner[1]
[1]*FAU Erlangen-Nürnberg, Erlangen, Germany*

Modern biotechnologies often result in high-dimensional data sets with much more variables than observations ($n \ll p$). These data sets pose new challenges to statistical analysis: Variable selection becomes one of the most important tasks in this setting. Recently, Meinshausen and Bühlmann (JRSSB, 2010) proposed a flexible framework for variable selection called stability selection, which was refined by Shah and Samworth (JRSSB, 2013).
By the use of resampling procedures, stability selection adds a finite sample error control to high dimensional variable selection procedures such as Lasso or boosting.
We consider the combination of boosting and stability selection and present results from a detailed simulation study that presents insights on the usefulness of this combination. Limitations will be discussed and guidance on the specification and tuning of stability selection will be given. The results will then be used for the detection of metabolic biomarkers for autism.
All methods are implemented in the R package mboost (http://cran.r-project.org/package=mboost).

C31.3

Deviance residuals based sparse PLS and sparse kernel PLS regression for censored data

P Bastien[1], F Bertrand[2], N Meyer[3], M Maumy-Bertrand[2]
[1]*L'Oreal R&I, Aulnay, France,* [2]*CNRS, Université de Strasbourg, Strasbourg, France,* [3]*INSERM, Faculté de Médecine, Strasbourg, France*

There has been a vast literature in the last decade devolved to relating gene expression profiles to subject survival or to time to cancer recurrence. The proportional hazard regression model suggested by Cox, 1972, to study the relationship between the time to event and a set of covariates in the presence of censoring is the model most commonly used for the analysis of survival data.
However, like multivariate regression, it supposes that there are more observations than variables, complete data, and variables not strongly correlated between them. In practice when dealing with high-dimensional data, these constraints are crippling. Collinearity gives rise to issues of over-fitting and model mis-identification. Variable selection can improve the estimation accuracy by effectively identifying the subset of relevant predictors and enhance the model interpretability with parsimonious representation. In order to deal with both collinearity and variable selection issues, many methods based on Lasso penalized Cox proportional hazard have been proposed since the seminal paper of Tibshirani, 1997. Regularization could also be performed using dimension reduction as is the case with PLS regression. We propose two original algorithms named sPLSDR and its non linear kernel counterpart DKsPLSDR, by using sparse PLS regression (sPLS) based on deviance residuals. We compared their predicting performance with state of the art algorithms based on reference benchmarks and simulated datasets.
Results: sPLSDR and DKsPLSDR compare favorably with other methods in their computational time, prediction and selectivity. The R-package plsRcox is available on the CRAN and maintained by Frédéric Bertrand.

C31.4

Weibull regression with Bayesian variable selection to identify prognostic biomarkers of breast cancer survival

PJ Newcombe[1], H Raza Ali[2,3,4], FM Blows[5], E Provenzano[6], PD Pharoah[4,5,7], C Caldas[2,4,5], S Richardson[1]
[1]*MRC Biostatistics Unit, Cambridge, United Kingdom,* [2]*Cancer Research UK, Cambridge, United Kingdom,* [3]*Department of Pathology, University of Cambridge, Cambridge, United Kingdom,* [4]*Cambridge Experimental Cancer Medicine Centre, Cambridge, United Kingdom,* [5]*Department of Oncology, University of Cambridge, Cambridge, United Kingdom,* [6]*NIH Cambridge Biomedical Research Centre, Cambridge, United Kingdom,* [7]*Strangeways Research Laboratory, Cambridge, United Kingdom*

As large, data-rich medical datasets are becoming routinely collected, there is a growing demand for regression methodology that facilitates feature selection over a large number of predictors. Bayesian variable selection algorithms offer an attractive solution, whereby a sparsity inducing prior allows inclusion of sets of predictors simultaneously and inference of those which are most important. Since predictors are included simultaneously, effect estimates are adjusted for one another and issues around multiple testing are avoided. Furthermore, uncertainty in the subset of important predictors and their effect estimates is naturally captured.

We present a new implementation of Bayesian variable selection for survival analysis under the Weibull regression model which is based on a Reversible Jump MCMC algorithm. In a realistic simulation study, we demonstrate superior power and specificity in comparison to an alternative LASSO based feature selection strategy. Subsequently we present a real data application in which 119 protein-based biomarkers are explored for association with breast cancer survival in a case cohort of 2,287 patients with ER-positive disease. Our method outperformed alternative strategies to provide evidence for three independent prognostic biomarkers of survival, one of which is novel.

C31.5
Approximate Bayesian model selection with the deviance statistic

L Held[1], D Sabanés Bové[2]
[1]University of Zurich, Zurich, Switzerland, [2]Roche, Basel, Switzerland

Bayesian model selection poses two main challenges: the specification of parameter priors for all models, and the computation of the resulting Bayes factors between models. There is now a large literature on automatic and objective parameter priors, which unburden the statistician from eliciting them manually in the absence of substantive prior information. One important class are g-priors, which were recently extended from linear to generalized linear models.
We show that the resulting Bayes factors can conveniently and accurately be approximated by test-based Bayes factors using the deviance statistic. For the estimation of the hyperparameter g, we show how empirical Bayes estimates correspond to shrinkage estimates from the literature, and propose a conjugate prior as a fully Bayes alternative. Considerable computational gains are obtained which enable an exhaustive evaluation of the model space in moderate size variable selection problems without the need to employ MCMC methods.
We illustrate the methods with the development of a clinical prediction model for 30-day survival in the GUSTO-I trial, and with variable and function selection in Cox regression for the survival times of primary biliary cirrhosis patients.
This is joint work with Daniel Sabanès Bovè.

C31.6
A novel variable selection method for Monte Carlo logic regression

M Malina[1], F Frommlet[1]
[1]Medical University of Vienna, CeMSIIS, Vienna, Austria

Logic regression is a promising approach to detect epistatic effects in genetic association studies. A Bayesian version of logic regression called Monte Carlo Logic Regression (MCLR) is based on Markov Chain Monte Carlo (MCMC) search, where the relevance of a specific epistatic term is assessed by the frequency of appearance of this interaction among all models visited by the MCMC search.
One problem with this approach is that the number of possible logic expressions is growing rapidly with the number of observed genetic markers. Therefore the MCMC search might not visit relevant models often enough to provide reliable estimates of model posterior probabilities, and alternative estimates are desirable. The aim of this talk is to compare MCLR with a novel variable selection approach based on model posterior estimates using Laplace approximation like in Schwarz BIC, but in combination with the more appropriate geometric prior distribution for the model size.
Results of a systematic simulation study and real genetic data analysis are presented, which illustrate the properties and benefits of the newly proposed selection method to detect epistasis.

C32 Longitudinal data analysis II

C32.1
Mixed-effects location scale model for time to event data

DS Courvoisier[1,2], TA Walls[3], D Hedeker[4]
[1]University Hospitals of Geneva, Geneva, Switzerland, [2]Harvard School of Public Health, Boston, United States, [3]University of Rhode Island, Kingston, United States, [4]University of Illinois at Chicago, Chicago, United States

Compared to mixed-effects models with only a random intercept (i.e., mixed-effects location models), mixed-effects location scale (MELS) model have several advantages. First, these models allow researchers to include covariates not only on the random (between-subject) intercept, but also on the random within subject residual variance. Including covariates on the between-subject variance allows researchers to, for example, test whether drug A has a more similar action over all individuals (blood pressure decreases by, on average, 10mmHg +/- 5mmHg) than drug B (blood pressure decreases by, on average, 10mmHg +/- 20mmHg). Furthermore, including covariates on the within-subject variance allows researchers to test hypotheses about how well specific measurements can be estimated within each individual. A second advantage concerns the estimation of a random scale, which captures the unexplained variation in within-individual variability. MELS have been developed for outcomes that follow a normal or ordinal distribution. In this presentation, we extend this model to time to event data, allowing for censoring. We illustrate the model with data on time to first cigarette measured every day for 7 days in smokers randomized to a group asked to keep smoking or another group asked to stop.

C32.2
Bayesian growth mixture models to distinguish hemoglobin value trajectories in blood donors

K Nasserinejad[1], JV Rosmalen[1], D Rizopoulos[1], WD Kort[2], M Baart[2], KV den Hurk[2], E Lesaffre[1,3]
[1]Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands, [2]Sanquin Blood Supply, Nijmegen, The Netherlands, [3]L-Biostat, KU Leuven, Leuven, Belgium

Blood donation leads to a temporary reduction in the hemoglobin level necessitating a period after donation for the hemoglobin value to recover to its pre-donation level. A premature invitation may therefore result in a too low hemoglobin value, a deferral from donation and may demotivate the candidate donor. All in all this implies an inefficient planning of the donation process.
The trajectory of the hemoglobin level after donation and the duration of the recovery period may differ between individuals. Here, we aim to classify the longitudinal hemoglobin values measured in blood donors. More specifically, we wish to detect early which donors will run the risk of a too low hemoglobin value (8.4 mmol/l for men, 7.8 mmol/l for women). For this, we used a random sample of 2000 new-entrant whole-blood donors who have donated repeatedly during 2005-2012. This data set was collected by Sanquin Blood Supply (the Netherlands).
To capture the unobserved heterogeneity of hemoglobin profiles, we implemented a Bayesian growth mixture model. This model assumes that each donor belongs to one of several latent classes. Within each class, the hemoglobin trajectory follows a linear mixed model. In addition we let the latent class membership depend on the age and hemoglobin value at first visit.
Our fitted growth mixture model suggests different classes of hemoglobin trajectories. This model gives some insight in the donation process and is a start to better predict for which donors care needs to be exercised not to produce a too low hemoglobin level.

C32.3
Pairwise residuals and diagnostic tests for misspecified dependence structures in models for binary longitudinal data

N Breinegaard[1], S Rabe-Hesketh[2], A Skrondal[3]
[1]*University Hospital of Copenhagen, Copenhagen, Denmark,* [2]*University of California, Berkeley, United States,* [3]*Norwegian Institute of Public Health, Oslo, Norway*

Maximum likelihood estimation of models for binary longitudinal data is inconsistent if the dependence structure is misspecified. Unfortunately, there are currently no diagnostics specifically designed for detecting misspecified dependence structures in longitudinal models. Traditional goodness-of-fit tests for categorical data that compare expected and observed frequencies often suffer from two fundamental problems: (1) sparseness invalidating the assumed null distributions and (2) low power since the tests are non-targeted. To address these problems, tests based on marginalized tables have been proposed for log-linear, latent class, and item response models.
We introduce these ideas to a longitudinal setting and extend the methods to handle covariates. For exploratory diagnostics, we recommend inspecting pairwise residuals based on second-order marginal tables. Diagnostic tests based on such residuals can be targeted to specific types of model violation. We consider the important case where a random-intercept model is misspecified because of serial dependence that decays as the time-lag between pairs of observations increases. For this situation, adjacent-pair concordance statistics are shown to have substantially greater power than tests based on all pairwise residuals. The methods proposed in this paper are straightforward to implement.

C32.4
Evaluation of LRT in joint modelling of repeated time-to-event and longitudinal data using nonlinear mixed effects models

M Vigan[1], F Mentré[1]
[1]*IAME, UMR 1137, INSERM, Univ Paris Diderot, Paris, France*

Joint modelling is used to describe the relationship between the evolution of biomarkers, and events, repeated or not. The Stochastic Approximation Expectation Maximization (SAEM) algorithm implemented in Monolix has been extended and assessed for joint model.
In the present study, we aim to evaluate, by simulation, the properties of the Likelihood Ratio Test (LRT) for the assessment of biomarker evolution on the occurrence of events. Simulation settings are inspired from a real clinical study. Evolution of biomarkers is defined by an exponential decrease nonlinear mixed effects model and the repeated time-to-event by a frailty model with an exponential hazard baseline function. Various scenarios are studied: i) no, mild or strong association between biomarker and events, ii) different probability of events, iii) different frequency of biomarkers measurements and iv) no or some independent dropout. For each scenario, we simulate 500 datasets with 200 patients. Estimations were performed using the Stochastic Approximation Expectation Maximization (SAEM) algorithm implemented in Monolix 4.3.0, with 3 Markov Chains, and the likelihood was evaluated by Importance Sampling (IS) with 20000 chains.
We evaluate the type I error and the power of the Likelihood Ratio Test according to the different scenarios. For all scenarios, type I error was close to 5%. Powers were influenced by dropout and number of events. SAEM in Monolix and LRT with likelihood computed using IS gave good results.

C32.5
Estimation of the linear mixed integrated Ornstein-Uhlenbeck stochastic model

R Hughes[1], J Sterne[1], K Tilling[1]
[1]*University of Bristol, Bristol, United Kingdom*

Background: Longitudinal biomarker data (e.g. CD4 counts) are commonly analysed using a linear mixed model (LMM). For continuous data Taylor, Cumberland and Sy proposed a LMM with an added integrated Ornstein-Uhlenbeck (IOU) non-stationary stochastic process (LM-IOU model), which allows for autocorrelation and estimation of the degree of derivative tracking. Due to lack of available software, the LM-IOU model is rarely used.
Methods: We have implemented the LM-IOU model in Stata. Using simulations we assessed the feasibility and practicality of estimating the LM-IOU model by restricted maximum likelihood. We compared different (1) optimization algorithms, (2) parameterizations of the IOU process, (3) data structures and (4) random-effects structures.
Results: The Newton-Raphson (NR) algorithm achieved convergence with fewer iterations and the computations were faster compared to a combination of the Fisher-Scoring and NR algorithms, and the Average-Information and NR algorithms. The combined algorithms did not provide additional robustness to starting values. When there was a strong degree of derivative tracking convergence depended upon the parameterization of the IOU process. With respect to bias of the point estimates, a dataset of 500 subjects each with 20 measurements was preferable over a dataset of 1000 subjects each with 10 measurements. In some cases, LM-IOU models with random effects other than the random intercept failed to converge due to competition for the same source of stochastic variation.
Conclusion: The LM-IOU model can be fitted using standard software to balanced and unbalanced datasets, but LM-IOU models with two or more random-effects may be impractical.

C33 Relative and net survival

C33.1
Flexible modeling of continuous covariates in Net Survival: additive vs multiplicative model

A Mahboubi[1], L Remontet[2], M Abrahamowicz[3], C Binquet[4], R Giorgi[5], C Quantin[4,6]
[1]*Dijon University Hospital, Dijon, France,* [2]*Hospices Civils de Lyon, Lyon, France,* [3]*McGill University, Montreal, Canada,* [4]*Inserm, U866, Univ de Bourgogne, Dijon, France,* [5]*SESSTIM, Marseille, France,* [6]*CHRU, Service de Biostatistique et d'Informatique Médicale, Dijon, France*

Accurate assessment of the effects of continuous prognostic factors requires flexible modeling of both time-dependent (TD) and non-linear (NL) effects. To address this issue, two alternative flexible extensions of the Estève et al model[a] have been developed[b,c]. Both models use cubic regression splines to estimate the TD and NL effects but differ in that the TD and NL effects of the covariate on the log-excess hazard are assumed to be: additive[b] or multiplicative[c]. Specifically, the disease-specific hazards are written, respectively as:
$I_c(t|z)=\exp(g(t))*\exp(a_i(z_i)+b_i(t)*z_i)$ and $I_c(t|z)=\exp(g(t))*\exp(a_i(z_i)*b_i(t))$
where: $g(t)$ represents the baseline log-hazard and $a_i(z_i)$ and $b_i(t)$ represent, respectively, the NL and TD effects of the continuous covariate z_i. However, the impact of the differences in the assumptions underlying alternative models on the resulting estimates is unknown.
To investigate the implications of these analytical differences, we applied both models to real-life datasets of cancers from registry-based studies.
Results obtained with the two models were compared, in terms of estimated hazards, TD and NL effects of age at diagnosis, and their significance

was tested with likelihood-ratio-tests. Both models were found to be flexible enough to reproduce the main features of the data structure and led to almost identical interpretation.

References:

[a]Estève et al, Stat-Med 1990;9:529-538
[b]Remontet et al, Stat-Med 2007;26:2214-2228
[c]Mahboubi et al, Stat-Med 2011;30:1351-1365

C33.2

An excess hazard model adjusting for lack of additional life table variables

C Touraine[1,2], N Grafféo[1,2], R Giorgi[1,2,3]
[1]INSERM, UMR912 (SESSTIM), Marseille, France, [2]Aix Marseille University, UMR_S912, IRD, Marseille, France, [3]APHM, Hôpital Timone, BIOSTIC, Marseille, France

Excess hazard model is commonly used in population-based cancer studies to estimate the real impact on the excess mortality of prognostic factors that influence overall mortality. In this model, the mortality observed is usually decomposed into the sum of the overall mortality and the excess mortality due to cancer. Overall mortality is obtained from population life tables stratified by sociodemographic variables (typically age, sex, calendar year). However, some additional variables are known to impact overall mortality. They could have a potential effect on excess mortality and are often absent in life tables (for example, ethnicity or deprivation). It has been shown that the use of a life table that lacks stratification by such a variable can lead to a biased estimate of its effect and of the other covariate effects on excess mortality.

In this work, we propose an excess hazard model that adjusts for additional variables in order to reduce this bias. We extended a model proposed by Cheuvart and Ryan for grouped data in a clinical framework. In our model, the overall mortality is allowed to differ from the one from life tables by a scale parameter assuming a proportional effect. Estimates are obtained from individual data using a maximum likelihood approach. A likelihood ratio test allows testing the significance of the scale parameter. The performance of the model was evaluated by simulations considering different scenarios. The interest of the model is illustrated using a population-based dataset on colon cancer with life tables stratified or not by ethnicity.

C33.3

Generalization of a log-rank type test to compare net survival distributions

N Grafféo[1,2], F Castell[3], A Belot[4,5,6], R Giorgi[1,2,7]
[1]INSERM, UMR912 (SESSTIM), Marseille, France, [2]Aix Marseille Université, UMR_S912, IRD, Marseille, France, [3]Aix Marseille Univ, CNRS, Centrale Marseille, I2M, UMR 7373, Marseille, France, [4]Hospices Civils de Lyon, Service de Biostatistique, Lyon, France, [5]Université Lyon 1, UMR 5558 Laboratoire Biostatistique-Santé, Villeurbanne, France, [6]Institut de Veille Sanitaire, DMCT, Saint-Maurice, France, [7]APHM, Hôpital Timone, BIOSTIC, Marseille, France

Net survival is the survival that would be observed, in a hypothetical world, if the disease under study were the only possible cause of death. In cancer research, by removing the effect of death from causes other than cancer, net survival allows us to compare cancer survival between different groups. Pohar-Perme et al. proposed a non-parametric consistent estimator of net survival. However, to the best of our knowledge, there is no statistical test for the comparison of Pohar-Perme net survival functions for more than 2 groups.

Our purpose is to build a generalized log-rank-type test for comparing net survival functions of several groups. Following the approach used in our previous work in the context of two groups, we expressed the log-rank type test in the counting process framework. As done in the Pohar-Perme estimator, we introduced the inverse probability weighting procedure in the counting and the at risk processes. Covariance matrix of our test statistic was obtained thanks to the martingale theory. We proved the asymptotic distribution of our test statistic under the null. Simulation studies were performed to evaluate the performance of our test in terms of type I error and power. We generated survival times depending on age, sex, and a covariate X defining the groups to compare. Different effects of X were considered to obtain similar groups or not regarding net survival. Results obtained under different scenarios show that our log-rank type test performs well in terms of type I error and power.

C33.4

Additive relative survival multistate semi-Markov model

F Gillaizeau[1,2,3], E Dantan[1], M Giral[2,3], Y Foucher[1,2]
[1]Université de Nantes, Nantes, France, [2]Centre Hospitalier Universitaire de Nantes, Nantes, France, [3]INSERM CR1064 ITUN, Nantes, France

Medical researchers are often interested to investigate the relationship between explicative variables and times-to-events like disease progression or death. Such multiple times-to-events can be studied using multistate models. For chronic diseases, it may be relevant to consider semi-Markov multistate models because the transition intensities between two clinical states more likely depend on the time already spent in the current state than on the chronological time. When the cause of death for a patient is unavailable or not totally attributable to the disease, it is not possible to specifically study the associations with the excess mortality related to the disease. Relative survival allows an estimate of the net survival in the hypothetical situation where the disease under study would be the only possible cause of death.

We propose here a new semi-Markov additive relative survival (SMRS) model that combines the multistate and the relative survival approaches. Using simulated data, we highlight the effectiveness of the SMRS model whose results tend to those obtained if the different causes of death are known. Regardless the parameter considered, absolute biases were lower than 0.04, and coverage rates greater than 92% (proportion of samples in which the 95% confidence intervals includes the theoretical value). The usefulness of the SMRS model is illustrated for a cohort of kidney transplant recipients. We have developed a package in R for the analysis of semi-Markov additive relative survival models.

C33.5

Diagnostic tools for model building in net survival: use and comparison of two methods to test the proportional hazards assumption

C Danieli[1,2], N Bossard[1,2], L Roche[1,2], A Belot[1,2,3], Z Uhry[1,2], L Remontet[1,2]
[1]Hospices Civils de Lyon, Service de Biostatistique, Pierre-Bénite, France, [2]Université Lyon 1, CNRS, UMR5558, LBBE, Villeurbanne, France, [3]Institut de Veille Sanitaire, Saint-Maurice, France

Net survival is the most relevant indicator to compare cancer survivals between countries or periods. It can be obtained using the Pohar-Perme estimator or an excess mortality hazard model including the demographic variables that define the expected mortality (usually age, sex and year of diagnostic). The latter solution involves a complex model-building strategy that requires diagnostic tools. The main assumptions to check concern the baseline distribution, the link function, the functional form and the proportional effect of the covariate of interest.

We focus on the only two methods developed to check the assumption of proportionality within the net survival context: the Stare method based on the partial residuals (similar to Schoenfeld residuals) and the Cortese method, extended from Lin method and adapted to the semi-parametric excess hazard model based on partial score processes.

The objectives of the present work are:
1) to adapt the Cortese method to parametric models;
2) to compare, using a simulation study, the performance of the method thus adapted with that of the Stare method;
3) to illustrate the use of these two methods on real data (none is currently used in practice).
The performance criteria will be the ability to detect the non-proportionality of the effect of a covariate taking or not into account its linear or non-linear effect. Several scenarios were considered with changes in several modeling aspects, mainly the baseline distribution and the amplitudes of the parameters. Simulation results and an application to real cancer survival data from French registries will be presented.

C33.6
Oblique decision trees for spatial clusters detection of net cancer survival rates

N Grafféo[1,2], J Gaudart[1,2,3], K Ndiaye[1,2], R Giorgi[1,2,3]
[1]INSERM, UMR912 (SESSTIM), Marseille, France, [2]Aix Marseille Université, UMR_S912, IRD, Marseille, France, [3]APHM, Hôpital Timone, BIOSTIC, Marseille, France

Net survival is the survival that would be observed, in a hypothetical world, if the disease under study was the only cause of death. Because of geographical variations in factors impacting on patients' net cancer survival, spatial study is of particular interest. However, it relied on pre-specified administrative maps, which are not always appropriate in the case of epidemiological research. The goal of our work was to propose a method providing potential spatial clusters which could contain patients with similar net cancer survival rates at a given time without pre-specified boundaries.
We extended to net survival analysis an oblique decision trees approach which had been developed for counted data. This non-parametric regression model eliminates the need to define any specification of geographical areas, shapes, or sizes of the clusters, provides potential aggregates with oblique partitions of the space, and allows adjusting on covariates. In this work, we used the Pohar-Perme estimator at time *t*, which yields consistent estimates of net survival. First, the algorithm splits the geographic area into two adjacent partitions by maximizing the statistic of a Z-test, comparing net survival estimates between each potential split of angular sectors. Second, the algorithm goes on recursively until one of the proposed stopping rules is reached and the oblique decision tree is completed.
Simulation studies will be used to investigate the performance of the proposed algorithm. Then, this method will be illustrated by a cancer population-based study. This approach could be useful to examine geographical variations in net survival rates.

C34 Methodology

C34.1
Constrained ordination analysis with an increased number of bell-shaped response functions with applications in metagenomics

Y Zhang[1], O Thas[1,2]
[1]Ghent University, Ghent, Belgium, [2]University of Wollongong, Wollongong, Australia

Ecologically meaningful bell-shaped responses of species to ecological gradients is a fundamental assumption of most current analytical methods in community ecology. However, statistical methods often make no distinction between convex and concave response functions. The analysis output is therefore misleading and the conclusion are prone to errors. We identify this problem in classical model-based method, such as constrained ordination analysis (COA), by means of several diagnostic graphical tools. To solve the issue a penalty term similar to L1-penalization is proposed so as to penalize convexity in the likelihood ratio criterion. A fast method of determining tuning parameter is also introduced.

C34.2
Multivariate statistical process control for mixed-type data: an overview and a simulation study

G Vidmar[1], N Majdič[1], R Blagus[2]
[1]University Rehabilitation Institute, Ljubljana, Slovenia, [2]Institute for Biostatistics and Medical Informatics, Ljubljana, Slovenia

Multivariate statistical process control (MV SPC) based on mixed-type data (i.e., when some of the variables describing the process are numeric and some categorical) is a relatively new and undeveloped field. The usual approach to MV SPC addresses measurement data by constructing a Shewhart chart based on the Hotelling´s T^2 statistic. We review the possibilities for MV SPC with mixed-type data and identify three main approaches: multivariate outlier detection for mixed data; dimensionality reduction (via PCA, MDS or ICA) yielding numeric dimensions followed by T^2 (or multivariate EMWA or multivariate CUSUM) control charts; and measuring distances between mixed-data points using Gower´s distance (i.e., Gower´s dissimilarity coefficient, Gower´s index or Gower´s general coefficient of similarity) and then constructing T^2 charts, D^2 charts (based on support vector data description, SVDD) or K^2 charts (based on *k*-nearest neighbours data description, *k*NN). The control limits for the D^2 and K^2 charts are established via bootstrapping, whereby distances from the whole phase I sample (global) or just from the *k*NN (local) are considered. We present a simulation study comparing the Gower´s-distance-based approaches and the T^2 approach with categorical variables coded as binary indicator variables. The highly realistic simulations are based on a planned setup for health-care quality monitoring in the field of rehabilitation after lower-limb amputation. The results indicate that the Gower´s distance approach improves as the number of categorical variable increases and that the local Gower´s-distance-based K^2 chart outperforms the global one.

C34.3

Case-wise diagnostics for the multinomial logit-link regression model

L Blizzard[1], DW Hosmer[2], S Quinn[3], JD Canary[1]
[1]Menzies Research Institute Tasmania, Hobart, Australia, [2]University of Massachusetts, Amherst, United States, [3]Flinders University, Adelaide, Australia

For nominal outcomes with more than two attributes, odds ratio estimates are obtained by fitting a multinomial logistic regression model. Several summary measures of goodness-of-fit provide a global test of the adequacy of a fitted multinomial model, and a variety of diagnostic quantities can be used to identify observations that influence the estimated coefficients of the fitted model and/or its predicted probabilities. These case-wise diagnostics are a natural adaption of those proposed by Pregibon (1981) for the binary logistic model, and were extended to the multinomial logistic model by Lesaffre andAlbert (1989).
The multinomial diagnostics have not been implemented in statistical packages. Hosmer and Lemeshow (Hosmer, Lemeshow and Sturdivant, 2013) continue to recommend that until they are, the fit of a multinomial logistic regression model can be investigated by assessing the fit of separate binary models fitted to the data. Whilst this approach due to Begg and Gray (1984) is generally sound, we show by demonstration and data simulations that it may fail to detect a lack of fit that would otherwise be revealed by the multinomial diagnostics.
The use of the multinomial diagnostics in combination with graphical plots is demonstrated, some troublesome cases in which different diagnostics provide conflicting results are highlighted, and guidance in the interpretation of their values is offered with tentative guidelines for identifying outlying and influential observations.

C34.4

Interpreting small differences in mean z-scores in sick populations: does dichotomisation help?

J Peacock[1], O Sauzet[2], J Lo[1]
[1]King's College London, London, United Kingdom, [2]Universität Bielefeld, Bielefeld, Germany

Background: A recent study in ex-preterm children, ie 'sick' individuals, observed a small difference in lung function mean z-scores which though statistically significant was of uncertain clinical importance. However, the corresponding difference in proportion at high risk was substantial. We explore this issue by comparing effects of a small shift in mean z-score in 'normal' and 'sick' (Rose IJE 1985) populations.
Methods: We assumed a comparison of sample mean z-scores in two 'normal' (mean z near 0) and two 'sick' populations (mean z near-1) similar to data observed. We defined high risk as z-score<-1.96, <-1.64, <-1.28 (2.5^{th},5^{th},10^{th} centiles respectively in a 'normal' population). We used a distributional approach to calculate proportions at high risk assuming z-score was Gaussian (Peacock StatMed 2012). This approach provides differences in proportions with the same precision as differences in means.
Results: For 'normal' populations a difference of 0.25 in mean z-score equates to a difference of 1.9 percentage points in individuals with z-score<2.5^{th} centile. This contrasts with a difference of 7 percentage points for 'sick' populations. Results for percentage of individuals <10^{th} centile show a small mean difference of 0.25 equates to a large difference of 10 percentage points in 'sick' populations.
Conclusions: Small differences in mean z-scores equate to larger differences in proportions at high risk in 'sick' compared to 'normal' populations. Hence reporting means alone is potentially misleading; we recommend a dual approach reporting differences in means and proportions at high risk calculated using the distributional approach.

C34.5

Confidence bounds for monotone dose-response relationships

C Baayen[1], P Hougaard[1]
[1]H. Lundbeck A/S, Valby, Denmark

An important aim of drug trials is to characterize the dose-response relationship of a new compound. Such a relationship can often be described by a parametric (non-linear) function that is monotone in dose. To establish proof of concept, or find the minimal effective dose, it is of interest to know the uncertainty of the estimated dose-response curve. It is well known that Wald confidence intervals are based on linear approximations and may be unsatisfactory in nonlinear models. They can be unreasonable in the sense that the lower confidence limit of the difference to placebo can be negative even when the overall test shows significant positive effect under a monotonicity assumption. In nonlinear models, profile likelihood based confidence intervals for the parameters have been shown to have better coverage. In this work we use a similar approach to compute confidence intervals for the dose-response curve. These confidence bounds have a more reasonable shape (as function of dose) than Wald confidence intervals. Finally, the method is robust when there is poor information (few doses, or irregular choice of doses) for estimating the dose response curve.

C34.6

Simpler is better: a comparison of methods for construction of fetal reference charts

D Nevo[1], M Mandel[1], E Ein-Mor[2], O Chen[2], E Daniel-Spiegel[3,4], S Yagel[2]
[1]The Hebrew University of Jerusalem, Jerusalem, Israel, [2]Hadassah University Hospital-Mount Scopus, Jerusalem, Israel, [3]Ha'Emek Medical Center, Afula, Israel, [4]Technion Israel Institute of Technology, Haifa, Israel

Reference charts for fetal measures have been developed over the years in order to estimate gestational age and fetal weight. These reference charts have also been used for early detection of pregnancies that should be monitored closely, since values in the tail of measure´s distribution are associated with fetal defects and disorders. Construction of reference charts is essentially an estimation of quantiles of a distribution as function of the gestational age.
Existing methods were developed under various modelling assumptions, typically by fitting a polynomial regression for certain functionals of the distributions (e.g., mean, standard deviation, quantiles). We relax the assumptions of a parametric polynomial link between the distribution parameters and the age. We consider nonparametric regression and discretization of the age in order to allow more flexible models. We use a large cross sectional data with repeated measures to compare between the various existing and suggested methods. The question of homogeneity of reference charts is also of interest. Curves built using the same method but using data from different hospitals, located 100km from each other, are compared.
We conclude that simple methods should be preferred, provided enough data is available, and that reference charts should be constructed separately for different subpopulations.

C35 The biostatistician's toolbox II

C35.1
Pharmacodependence: new graphical representations

E Peron[1,2], F Feuillet[2,3], R Morgane[1], C Victorri-Vigneau[1,2], J-B Hardouin[2,3]
[1]Service de Pharmacologie Clinique, Nantes, France, [2]EA 4275 SPHERE, Nantes, France, [3]Plateforme de Biométrie, Nantes, France

Introduction: In France, a network of 13 Centers for Evaluation and Information on Pharmacodependence (CEIP) monitors substance abuse risk. This network is coordinated by the French medicines agency.
To assess abuse and dependence potential of drugs, CEIPs record cases of substance abuse and dependence arising from health professionals. Each substance reported is evaluated by the pharmacodependence gravity score.
This is an 8-item score (each item is binary rated: positive or negative) divided into 2 different paths:

- The first path assesses physical and compulsive signs: tolerance, withdrawal syndrome, dose taken in larger amounts or over a longer period than was originally expected, desire to cut down.
- The second path assesses harmful consequences of the pharmacodependence: a great deal of time is spent, interpersonal problems, consummation persistence despite health problems, transgression behavior.

At the end of this assessment, we obtain for each substance a score on 8 consists of 2 sub-scores on 4 for each path.
Methods: This work aimed to develop new graphical representations and indexes to compare substances' pharmacodependence profile using the CEIP electronic records.
For instance, 3 substances which have very different pharmacodependence profile will be shown: buprenorphine, heroine and a control: paroxetine - a non-potential pharmacodependence substance.
Results: Graphical representations and indexes allow distinguishing between substances using the score and the 2 sub-scores. Their interpretation will be explained.
Conclusion: The distinguished feature of the pharmacodependence score and the 2 sub-scores is demonstrated. This work open up new prospects of methodological tools development.

C35.2
Constructing robust confidence intervals for drug utilization time series data

A Corkum[1], Y Zhang[1], P Cabilio[1]
[1]Acadia University, Wolfville, Canada

For a drug safety and effectiveness study, we are interested in constructing confidence intervals in drug utilization time series to investigate the level/trend over a signal or several time periods. The current methods either assume that the data follow an independent process, or rely on the explicit knowledge of underlying process distribution and its dependence structure, which is not applicable.
In this paper, we generalize properties of the sign and Wilcoxon signed rank statistics to time series data, and develop new nonparametric methods to construct confidence intervals when the data are dependent. Specifically, we use the theory of U-statistics for mixing processes to develop asymptotic normality theorems for these statistics, which are then used to approximate the confidence interval. The variances of the sign and Wilcoxon signed rank statistics are keys in computing such confidence intervals.
We consider both the unrealistic case in which we know the underlying process distribution and its dependence structure and are able to compute the variances exactly, and the more realistic situation in which the variance must be estimated due to lack of information. We implement three methods for the variance estimation: block bootstrapping; sieve bootstrapping; and the empirical distribution method, and then compare them with the exact confidence interval coverage through extensive simulations.
The results show that the proposed methods are effective and robust for time series data. Finally, we illustrate our methods with time series data on a heartburn medication to investigate the effect of a provincial insurance policy change.

C35.3
Using constraints to compare state structures in cost-effectiveness decision models

H Thom[1], C Jackson[2], N Welton[1], L Sharples[3]
[1]University of Bristol, Bristol, United Kingdom, [2]MRC Biostatistics Unit, Cambridge, United Kingdom, [3]University of Leeds, Leeds, United Kingdom

Cost-effectiveness decision models are used to estimate expected costs and effects of interventions for the management of disease and thus guide the decision making of national health services. These are often multistate models with states corresponding to categorizations of disease status. A common difficulty is the choice of states for the multistate model and it may be unclear whether there are sufficient data to inform the transition probabilities. Similar states could be merged but different structures can give different decision recommendations.
Our aim is to compare different structures by balancing fit and parsimony based on the available data, and to quantify the associated decision uncertainty in terms of the expected value of perfect information about the model and its parameters. Models with different states are difficult to compare by standard statistical methods because they are fitted to different data and the corresponding likelihoods are on different scales.
However, we will show that models with coarsened state structures are practically equivalent to special cases of a single, sufficiently flexible, model. These special cases are defined by constraints on the model parameters and can be compared statistically as they are fitted to the same data. The expected value of perfect information of the extra parameters required to relax the constraints therefore represents the decision uncertainty associated with the model structure.
We shall illustrate this approach for a variety of state transition patterns used in cost-effectiveness models, and present an application to a model for diagnostic testing strategies for coronary artery disease.

C35.4
Optimal and maximin sample sizes for multicentre cost-effectiveness trials

A Manju[1], MJJM Candel[1], MPF Berger[1]
[1]Maastricht University, Maastricht, The Netherlands

This paper deals with the optimal sample sizes for a multicentre trial in which the cost-effectiveness of two treatments in terms of net-monetary benefit is studied. The optimal sample sizes concern the number of centres and the number of individuals per centre in each of the treatment conditions.
These numbers maximize the efficiency or power for given research costs or minimize the research costs at a desired level of efficiency or power. Information on several model parameters of a bivariate linear mixed model and sampling costs are required to calculate these optimal sample sizes. In case of limited information on relevant model parameters, sample size formulas are derived for so-called maximin sample sizes which guarantee a desired power level at the lowest study costs. Four different maximin sample sizes were derived based on signs of the lower bounds of the correlation between random slopes for costs and effects and individual level correlation between costs and effects, where one case is worst compared

to other three cases.
We studied numerically the efficiency of the worst case maximin sample sizes instead of using others. Finally, an expression is derived that enables calculating optimal and maximin sample sizes that yield sufficient power to test the cost-effectiveness of two treatments.

C35.5 *Student Conference Award*
Correcting for bias in the detection and validation of informative diagnostic tests
DS Robertson[1], J Bowden[1], AT Prevost[2]
[1]MRC Biostatistics Unit, Cambridge, United Kingdom, [2]King's College London, London, United Kingdom

When developing a new diagnostic test for a disease, there are often multiple candidate classifiers to choose from, and it is unclear if any will offer an improvement in performance compared to current technology. A two-stage design can be used to select a promising classifier (if one exists) in stage one for definitive validation in stage two.
However, estimating the true properties of the chosen classifier is complicated by the first stage selection rules. In particular, the usual maximum likelihood estimator (MLE) that combines data from both stages will be biased high. Consequently, confidence intervals and p-values flowing from the MLE will also be incorrect.
Building on the results of Pepe et al. (SIM **28**:762-779) and others, we derive the most efficient conditionally unbiased estimator and exact confidence intervals for a classifier's sensitivity in a two-stage design with arbitrary selection rules; the condition being that the trial proceeds to the validation stage. We apply our estimation strategy to data from a recent family history screening tool validation study by Walter et al. (BJGP **63**: 393-400), and are able to identify and successfully adjust for bias in the tool's estimated sensitivity to detect those at high risk of breast cancer.

C35.6
Modelling and choice of cutoff in meta-analysis of diagnostic studies with varying cut-off value
D Böhning[1]
[1]Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, United Kingdom

Meta-analysis of diagnostic studies is often done on the basis of one pair of sensitivity and specificity per study. For this kind of situation the summary receiver operating charactistic has been suggested as a sutiable approach sicne it accommodates a varying cut-off value across studies. This approach becomes less favourable if the cut-off value itself is the parameter of primary interest.
We suggest a range of generalized linear models as natural extension of Youden index, diagnostic odds ratio or the likelihood ratio. Some case studies will illustrate the suggested methodology.

C36 Issues in multiple testing

C36.1
An informative modification of the fallback procedure
S Schmidt[1], W Brannath[1]
[1]University of Bremen, Bremen, Germany

The fallback procedure is an extension of the hierarchical test allowing for a more flexible alpha allocation. It can be applied for example in dose finding studies. If interest is in extending the fallback procedure to simultaneous confidence intervals, one may use the construction proposed by Strassburger and Bretz (Stat. Med. 2008; 27: 4914--4927).
However, these intervals are not optimal in the sense that non-informative rejections may arise. This means that the confidence interval of a rejected null hypothesis may contain all parameters of the alternative and thus gives no useful information about the true value of the effect parameter. Guilbaud (Biometrical Journal 2009; 51: 721--735) exploited the fact that the fallback procedure is not alpha-exhaustive in order to improve this deficiency. However, a positive probability for non-informative rejections remains.
We will present a modification of the fallback procedure with corresponding simultaneous confidence intervals which is informative in every case where a hypothesis is rejected. Our method is a straightforward extension of a former approach with respect to the hierarchical test. The main idea consists of a continuous parameter dependent level splitting after rejection of a null hypothesis to test a nested family of informative hypotheses. We will explain our idea, illustrate it by a graphical description and compare it to the approach of Guilbaud by simulations in the context of a clinical trial.

C36.2
Confirmatory testing for a beneficial treatment effect in dose-response studies using MCP-Mod
F König[1], B Bornkamp[2], F Bretz[2], E Glimm[2]
[1]Medical University of Vienna, CeMSIIS, Vienna, Austria, [2]Novartis Pharma AG, Basel, Switzerland

The MCP-Mod approach from Bretz et al. (2005) has attracted attention in the recent years due to its potential to increase the efficiency of selecting the "right" dose. The testing part of MCP-Mod was originally developed to significant dose response signal conduct proof-of-concept (PoC) tests, i.e., to demonstrate that the dose response relationship of the test drug is not flat. But it is not appropriate to make a claim that the drug has a positive effect at some specific dose.
In this presentation we extend the MCP-Mod approach by using the closed testing procedure from Marcus et. (1976) to obtain confirmatory p-values for dose response signal detection as well as for the pairwise comparisons of individual doses against placebo. The proposed test uses two-sided optimal contrasts tests based on a-priori information about plausible dose response shapes available at the planning stage of a clinical trial. However, by using two-sided contrast tests only weak Type I error rate control can be achieved when testing superiority for individual doses. We show suitable restrictions for the contrasts are needed to achieve strong Type I error rate control. The operating characteristics of the proposed method will be evaluated for certain dose-response profiles.

C36.3
Graph based multiple testing strategies for confirmatory adaptive enrichment designs

T Sugitani[1], M Posch[1], F Bretz[2], F König[1], F Klinglmueller[1]
[1]Medical University of Vienna, CeMSIIS, Vienna, Austria, [2]Novartis Pharma AG, Basel, Switzerland

An important objective in the development of targeted therapies is the identification of a population where the treatment has a clinically relevant effect. Recently, adaptive enrichment designs have been proposed that allow to restrict enrolment to a subpopulation after an interim analysis. In addition to the selection of a subpopulation, the sample sizes in the subgroups may be adapted.
In this work we apply the recently proposed adaptive graph based multiple testing procedures to the analysis of adaptive enrichment designs. These procedures control the familywise error rate for hypothesis tests in multiple (sub-)populations and can be easily extended to cover also tests of multiple endpoints. The definition of the testing procedure by a graph allows to map the difference in importance and the logical structure of the tested hypothesis to the testing procedure. In addition, the graph provides a convenient tool to communicate the testing procedure to the clinical study team. In a simulation study we assess the operating characteristics of the graphical adaptive testing approaches and compare them to testing procedures based on group sequential tests. Furthermore, we investigate how the power of the adaptive testing procedures can be optimized by stratification of the hypotheses tests to adjust for the heterogeneity of treatment effects in subgroups. Finally, the application of the adaptive graph based testing strategy is illustrated with a case study for the development of a targeted therapy.

C36.4
Likelihood ratio tests for multiple nonlinear models

G Gutjahr[1], B Bornkamp[2]
[1]University of Bremen, Bremen, Germany, [2]Novartis Pharma AG, Basel, Switzerland

Consider a set of nonlinear models that predict a mean vector of normally distributed observations and the hypothesis that at least one of these models fits the data significantly better than a constant model.
For a single ``sufficiently smooth´´ model, Hotelling showed that the likelihood ratio test statistic is a monotonous function of the correlation between the observations and the maximum likelihood prediction from the model; using methods from differential geometry, the exact null distributions of this statistic can be obtained. For multiple models, the best prediction from the multiple models is used in the likelihood ratio test statistic. The null distribution is determined by volumes of tubular neighborhoods on the unit sphere. We describe how such volumes can be approximated numerically.
This approach can also be used to calculate the distribution under alternative hypotheses and it does not required that the models are smooth. We compare the power of the likelihood ratio test with locally most powerful tests and with multiple-contrast tests and apply it to data from a dose-response clinical trial.

C36.5
Are multiple outcomes analysed appropriately in randomised controlled trials? A systematic review

V Vickerstaff[1], G Ambler[1], R Omar[1]
[1]University College London, London, United Kingdom

Many procedures for addressing multiplicity in clinical trials have been introduced in the literature; however, the techniques are rarely used in practice. Investigators often analyse multiple primary outcomes using several independent tests with no adjustments. Reporting several unadjusted p-values can increase the probability of erroneously rejecting at least one true null hypothesis.
We performed a review to quantify how many trials analysed multiple primary outcomes and how many analysed them appropriately. We reviewed all randomised controlled trials published July 2011-June 2013 in top neurology and psychiatry journals: American Journal Psychiatry, JAMA Psychiatry, Psychotherapy and Psychosomatics, Lancet Neurology and Neurology. Typically in these areas, data on multiple correlated outcomes are collected. We focused on the results in the abstract, methods used for sample size calculation and statistical analysis.
We identified 154 randomised controlled trials of which 70 analysed multiple primary outcomes. Among these, 55/70 did not adjust for the multiple comparisons. If multiplicity was addressed, the significance of the results and trial conclusions would have changed in several papers.
Of the 15/70 papers which accounted for multiplicity; 5 performed MANOVA, 6 used Bonferroni's correction and 4 used other correction methods. Nine trials provided a sample size calculation which considered multiplicity.
Our review shows that multiple primary outcomes are commonly analysed in clinical trials and are often inadequately handled. Further methodological research is necessary to assess the appropriateness of existing methods for addressing multiplicity in different scenarios, particularly when the outcomes are correlated and to provide guidance on their use in practice.

C36.6
A multiple testing procedure for three primary endpoints

R Ristl[1], F Frommlet[1], M Posch[1]
[1]Medical University of Vienna, CeMSIIS, Vienna, Austria

When efficacy of a treatment is measured by co-primary endpoints, efficacy is claimed only if for each endpoint an individual statistical test is significant at a local level α. While such a strategy controls the family-wise error rate (FWER) at level α, it may be strictly conservative and have low power. We improve the test of three co-primary endpoints to allow inference also in settings where only two out of the three show a significant result at the local level. While the test does not allow to reject an elementary null hypothesis in this case, it rejects an intersection hypothesis such that an effect in at least one of the endpoints can be inferred and the trial still serves as a proof of principle.
We show under the assumption of multivariate normal test statistics with arbitrary correlation matrix that the procedure controls the FWER at level α in the strong sense. Besides the application to tests for co-primary endpoints the result uniformly improves the Rüger test in the setting of trivariate normal test statistics. The latter rejects if two out of three hypotheses are significant at level 2α/3 but controls the type 1 error rate at level α without the assumption of multivariate normality.
We investigate the power of the improved test procedure and compare it to hierarchical and Bonferroni tests for co-primary endpoints. The test procedure is illustrated with a clinical trial for a rare disease. An application of the procedure in the assessment of diagnostic tools is discussed.

Wednesday, 27th August 2014 – 14:00-15:30

Invited session

S2 The power of data sharing: advancing research for everyone's benefit? (Panel discussion)

Organizers: Martin Posch and Franz König
Panelists: Hans-Georg Eichler, Senior Medical Director, EMA; Simon Day, CTCT; Ulrich Burger, Roche; Trish Groves, BMJ; Stephen Senn, CRP-Santé

It is widely recognized that the current publication practice of clinical trial results is deficient. A large number of trials are never published, and if they are published, essential information is often not included in the published manuscripts. Potential consequences are impaired meta-analyses, clinical trials that are unnecessarily repeated and a lack of transparency regarding the statistical analysis. To address these issues, recently, academia, journal editors, regulators and the pharmaceutical industry proposed a variety of approaches to data sharing and the European Medicines Agency plans to finalise its draft policy on the publication and access to clinical-trial data by June 2014. Of particular interest is wether and how access to patient level data is granted.
In this panel discussion key stakeholders from academia, regulators and industry will explore the wide range of opportunities and discuss the opportunities and risks that arise with the implementation of comprehensive data sharing commitments.

Contributed sessions

C37 Causal inference from observational studies I

C37.1

Adjusted survival curves by using inverse probability of treatment weighting: the comparison of three adapted log-rank tests

F Le Borgne[1,2,3], M Giral[2], A-H Querard[1,2,4], Y Foucher[1,2]
[1]Department of Biostatistics EA 4275, Nantes, France, [2]Transplantation, Urology and Nephrology Institute (ITUN), Nantes, France, [3]IDBC/A2com, Pace, France, [4]Departmental Hospital Center of Vendée, La Roche sur Yon, France

In observational studies, the presence of confounding factors is common and the comparison of different groups of subjects requires adjustment. In presence of survival data, this adjustment can be achieved with a multivariate model (usually a Cox model) allowing to validate a difference observed from crude survival curves. However, by using such type of regression, the effect of the factor under interest is often summarized by the hazard ratio (HR). This loss of information is so damaging that most research projects in biology or medicine present both crude survival curves (biased but illustrating precisely the differences in survival) and adjusted HR (not biased but too synthetic). A recent solution is the use of adjusted survival curves and log-rank test based on inverse probability of treatment weighting (IPTW). However, three adaptations of the log-rank test are found in the literature without any comparison of the performances in terms of type I and II errors. We performed a simulation study in order to (i) evaluate if the performances of these adjusted log-rank tests are acceptable compared to the Cox model used classically, and (ii) choose the most powerful of these three IPTW approaches. For illustration, we also propose to study the patient and graft survival of kidney transplant recipients according to the expanded donor criteria (ECD). Among the three approaches, the one proposed by Xie and Liu (2005) should be preferred in future studies to compare adjusted survival curves. Nevertheless, the results show that the Cox model remains the most efficient approach.

C37.2

Inverse probability weighting of overmatched nested case-control data to enable estimation of main effects and interactions

B Delcoigne[1], E Colzani[1], K Czene[1], M Reilly[1]
[1]Karolinska Institutet, Stockholm, Sweden

Introduction: Matched nested case-control designs are generally analyzed with conditional logistic regression. However, breaking the matching in such data and using inverse probability weighting offers a way to exploit the data for additional research questions. Our aim is to use this method for nested case-control data to overcome a problem generated by overmatching and enable us to address a research hypothesis involving interaction terms.
Methods: A nested case-control study was conducted of lung cancer in women who had radiation therapy for a previous breast cancer diagnosed in Sweden during 1958-2001. Cases were individually matched to controls on age, calendar period of diagnosis and region. We broke the matching and analyzed the data with weighted Cox proportional hazards regression, investigating the role of radiotherapy as a potential risk factor for lung cancer and its effect modification by smoking.
Results: The study included 1525 breast cancer patients, of whom 731 were lung cancer cases. Overmatching was apparent, with 75% of the

matched sets consisting of cases and controls with the same treatment, and a conditional logistic regression identified no significant treatment effect. From the weighted Cox regression, we estimated a significant risk associated with radiotherapy and evidence of effect modification, with a hazard ratio of 2.71 (1.37 - 5.36) in smokers.
Conclusion: Inverse probability weighting provides a way to exploit nested case-control data in real settings to overcome weaknesses in design and address new research questions with existing data.

C37.3
Performance of targeted maximum likelihood estimation in point-exposure studies using high-dimensional covariate data

M Pang[1], T Schuster[1,2], KB Filion[1,3], M Eberg[1], RW Platt[2,4]
[1]Lady Davis Institute, Centre For Clinical Epidemiology, Montreal, Canada, [2]McGill University, Dept of Epidemiology and Biostatistics, Montreal, Canada, [3]McGill University, Division of Clinical Epidemiology, Montreal, Canada, [4]Research Institute of the McGill University Health Centre, Montreal, Canada

Double robust targeted maximum likelihood estimation (TMLE) has been proposed for estimating marginal causal effects, allowing specification of both treatment and outcome models. While inverse probability weighting (IPW) methods are known to be sensitive to violation of the positivity assumption, the consequences of such violation in the TMLE framework have not been widely investigated. As non-positivity is frequently present in high-dimensional covariate settings, a better understanding of the mechanism of TMLE is of particular interest in pharmcoepidemiological studies using large databases.
Using plasmode simulation, we evaluated the performance of TMLE compared to that of IPW estimator based on a typical point-exposure drug effect cohort study of statin use post-myocardial infarction and the 1-year risk of all-cause mortality from the Clinical Practice Research Datalink. A variety of model specifications were considered inducing different degrees of non-positivity.
Our simulations showed that the performance of TMLE and IPW estimator was comparable when the dimension of the treatment model was modest; however, they diverged when a large number of covariates were considered. In some cases, we found irregular bias, large standard errors, and non-convergence results with TMLE even with a correctly specified but saturated treatment model. IPW estimator showed slightly better MSE performance with high-dimensional model specifications.
In conclusion, TMLE and IPW estimator using the same modeling can perform differently due to their different sensitivity to the positivity violation. Although TMLE is appealing for its double robustness property, a near violation of the positivity assumption in a high-dimensional covariate setting might be problematic.

C37.4
The impact of pCR after neoadjuvant chemotherapy in patients with large operable breast cancer on survival outcomes: a causation analysis

A Efendi[1], L Slaets[1], S Vandenberghe[2], J Bogaerts[1]
[1]EORTC Headquarters, Brussels, Belgium, [2]Dept. of Applied Mathematics, Universiteit Gent, Gent, Belgium

Purpose: Pathological complete response (pCR) is a well-used endpoint in neoadjuvant breast cancer trials. Although it is established as a predictive marker for long-term outcome (PFS/OS), surrogacy was not demonstrated (Cortazar et al., The Lancet 2014). Commonly used statistical analyses considered pCR as one of the independent covariates in a multivariate model, along with baseline data. However, pCR is measured after treatment exposure and dependent on baseline characteristics. This study examines the causal effect of pCR on PFS/OS independent of baseline covariates, using causal modelling.
Methods: Inverse Probability Weighting (IPW) enables us to account for potential confounding between pCR and baselines. IPW creates a pseudo-population, a re-weighted version of the original population, in which the measured association between baseline characteristics and pCR is removed. A causal effect is then obtained by fitting the weighted Cox regression for PFS/OS with pCR as the sole covariate. Landmarking will be used to account for lead-time bias, that is, the analysis is restricted to patients that are event-free at time of pCR assessment.
Results: The estimation of the effect of pCR on PFS and OS independent of baseline covariates will be presented and compared to the classical approach (multivariate modeling) in a large neoadjuvant breast cancer trial (N=1856).
Conclusion: In clinical trials**,** causal modeling can provide meaningful estimates of the added predictive value of endpoints or markers, assessed during/after treatment exposure (e.g. pCR, toxicity), on long-term outcome. We observe that pCR indeed has a causal effect on PFS, independent of treatment and baseline.

C37.5
Causal mediation analysis in a clinical survival trials - can statistics help to understand treatment mechanisms?

S Strohmaier[1], K Røysland[1], T Lange[2], Ø Borgan[1], T Pedersen[1], O Aalen[1]
[1]University of Oslo, Oslo, Norway, [2]University of Copenhagen, Copenhagen, Denmark

When it comes to clinical survival trials, regulatory restrictions usually require the application of methods that solely utilize baseline covariates covariates and the intention-to-treat principle. Thereby a lot of potentially useful information is lost, as collection of time-to-event data goes hand in hand with collection of information on other internal time-dependent covariates and patients deviate from assigned treatment plans.
Whereas considerable effort has been put into developing methods for dealing with treatment deviations, less attention is paid to a secondary objective, to employ those repeated measurements to shed more light on underlying treatment mechanisms.
We have data from a large-scale secondary prevention trial available, that compared how different statin treatment strategies would effect the risk of cardiovascular disease among patients with history of acute myocardial infarction, comprising repeated measures of internal makers.
To gain a better understanding about how treatment effects evolve over time, we adopt the model of analysis on dynamic path analysis, a model that can be viewed as an extension of classical path analysis and the concept of directed acyclic graphs (DAGs) to settings that involve time-to-event outcomes and time-dependent covariates. Additionally, we suggest a definition of direct, indirect and total effects that allows a causal interpretation and discuss other causal aspects that arise in this particular setting, where we obtain treatment effect estimates on the mediator whilst conditioning on survival.

C38 Patient-centered outcomes

C38.1

The design of diagnostic studies - another case for STRATOS?

W Vach[1]
[1]*Clinical Epidemiologiy, IMBI, University of Freiburg, Freiburg, Germany*

Recently, a new initiative called STRATOS appeared. STRATOS abbreviates "STRengthening Analytical Thinking for Observational Studies", and it aims to provide guiding documents at different levels for topics related ot the design and analysis of observational studies. In this talk I discuss some first ideas how STRATOS may contribute to the field of designing diagnostic studies.
Methodological standards for diagnostic studies are rapidly changing in the last years. Accuracy studies have been the cornerstone in diagnostic research for many years, but they are today often regarded as insufficient as they do not aim in measuring directly a patient benefit. RCTs with patient centered outcomes are often recommended as alternatives, but there is little experience so far in designing, planing and analysing such studies. Actually, such RCTs would evaluate a combination of diagnostic tests and subsequent treatment and management processes, i.e. complex interventions. So their analysis may still include many elements of observational studies. Accuracy studies will probably still also play an important role in future, but there is a need for better design, reporting and analyses taking the benefit aspect into account.
In such a state of change, it may be hard to develop guiding documents. Nevertheless, there are some key issues independent of the study type, which I discuss in my talk: Clear definitions of the clinical target situation and the target population, discrepancy between target and study population due to recruitment, need for a sufficient, but not artificial standardization, choice of external reference test or patient centered outcomes.

C38.2

Methodological issues in developing scores and cut-offs of rheumatoid arthritis activity

O Collignon[1]
[1]*CRP Santé, Strassen, Luxembourg*

Rheumatoid arthritis (RA) is a systemic disease which occurs in about 1% of the world population and triggers joints inflammations that may worsen patients' quality of life. In order to define treatment strategy and to evaluate response to therapy, disease activity may be measured via several scores using several bio-clinical variables, as the Disease Activity Score involving 28 joint counts (DAS28), the Simplified Disease Activity Score (SDAI) and the Clinical Disease Activity Score (CDAI). Furthermore cut-offs for these scores have been designed to help physicians classify patients into disease activity categories.
However some methodological issues were neglected when the scores were built, leading potentially to inaccurate classification of patients and thus inappropriate choice of therapy. Also, problems like inter-physician variability in evaluating disease activity, choice of the clinical parameters, methods of validation of the scores are highlighted. A strategy to develop a relevant surrogate to disease activity and cut-offs using penalized logistic regression and bootstrap internal validation is then proposed.
As long as the issues reviewed in this presentation are not addressed, results of studies based on such disease activity scores should be considered with caution.

C38.3

Determining optimal fractional factorial designs of discrete choice experiments using d-efficiency: application in addiction services

T Vanniyasingam[1], C Cunningham[1], G Foster[1], A Niccols[1], L Thabane[1]
[1]*McMaster University, Hamilton, Canada*

In a discrete choice experiment (DCE), individuals are asked to choose the most preferred alternative among a set of alternatives. We conducted a DCE where participants were asked a series of questions and had to choose one of three options presented in each scenario. A scenario was comprised of three levels, one from three different attributes. We explored individual preferences of 16 four-level attributes for a survey designed to elicit stated preferences for professional development, by addiction service providers and administrators, for the enhancement of addiction services. Typically DCE surveys are generated based on a fractional experimental design. An issue of survey design arises when determining which of the various combinations of attributes and levels of attributes should be presented within options in a scenario, and how many attributes will minimize participants' response burden - whilst ensuring an optimal design.
The objective of this talk is to present our results of how the optimality of the design, measured using d-efficiency, is affected by: the number of attributes, number of levels, number of scenarios, number of overlapping attributes between scenarios, and number per scenario. We will use simulations to create the fractional designs to evaluate the d-efficiency under the various conditions listed above.

C38.4

Developing robust scoring methods for use in child assessment tools

P Gichuru[1], G Lancaster[1], A Titman[1]
[1]*Lancaster University, Lancaster, United Kingdom*

Earlier and more sensitive diagnosis of disability reduces its detrimental effect on children. We therefore seek to develop robust scoring methods for Child Assessment tools which will ensure more timely intervention of detected delayed development to reduce stress on the child and its family. Most of the current development scores are dependent on age hence a key objective is to develop methods that correct or account for age.
Generally, regardless of implementation medium, two main scoring approaches are usually used; item by item scoring creates score norms for each item and total scoring uses all the responses of the child to give one score across the entire domain. Using data from 1,446 normal children from the recent Malawi Development Assessment Tool (MDAT) study, we review classical total scoring methods including simple scoring, Log Age Ratio methods and Item Response Models under different assumptions to derive normative scores in this child development context using binary responses only. While evaluating the pros and cons of each method, we also suggest extensions to current total scoring methods including smoothing methods and using more flexible models within various scoring algorithms. Preliminary results show that weighting simple scores is important as a lack of response to all items does not necessarily imply a lack of ability. Further, smoothing of score values is beneficial when variability in certain age groups is high. The more complex methods produce more reliable and generalizable normative scores. The sensitivity analysis showed that simple methods perform well in ideal situations.

C38.5
Random effect models for quality of life analysis in oncology
A Barbieri[1,2], C Mollevi[1], C Lavergne[2]
[1]*Institut régional du Cancer de Montpellier (ICM), Montpellier, France,* [2]*Institut de Mathématiques et de Modélisation de Montpellier, Montpellier, France*

In Oncology, the Health-related Quality of Life (QoL) has become an essential criterion in clinical trials. However, the longitudinal analysis of this criterion is complex and non-standardized. Indeed, the observations are obtained through self-questionnaires (Patient-Reported Outcomes) and there are both multiple responses, repeated and ordinal ones. From a statistical standpoint, QoL is not directly measurable and is considered as a latent trait which is accessible through responses to items. To evaluate QoL in most cancer clinical trials, the QLQ-C30 questionnaire has been used. Nowadays, the statistical analysis is done on a score from the EORTC recommendations, corresponding to the average of item responses.
Longitudinal competing models are exploited such as a linear mixed model (LMM) classically used for score modelling and generalized linear mixed models (GLMM) employed for ordinal categorical data. The latter model family builds on the Item Response Theory (IRT) and allows considering raw data (item responses). Regarding the longitudinal analysis, the IRT models are proposed as an alternative to LMM and extended to take into account the clinical covariates and data characteristics.
These presented models were compared through the analysis of a dataset from a clinical trial and then a simulation study was performed. The IRT model for polytomous data is quite complex and fastidious to estimate the regression coefficients and to predict the random effects. Finally, a less complex approach of linearization advanced by Schall in 1991 is proposed to estimate these GLMM in order to complete the simulation study.

C39 Multistate models and competing risks I

C39.1
Illness death models and their applications in cancer research
M Yu[1]
[1]*University of Wisconsin-Madison, Madison, United States*

Cancer studies frequently deal with non-terminal (T1) and terminal (T2) events. In many cases, T1 is disease progression and T2 is death. When T2 occurs first, it censors T1, but not vice versa. In practice, instead of considering such bivariate outcomes, a composite outcome known as progression free survival, defined as the minimum time to either of the two events, is frequently used.
We illustrate problems with such approach by using the well-known illness-death models. We therefore advocate data analysis using illness-death models to account for possible dependent censoring of T1 by T2 and to improve prediction of T2 using T1. We propose flexible random effects to capture heterogeneous correlation structures that are usually present in real data.
Our model also represents a generalization of the popular shared frailty models. We use Bayesian computation for analysis that can utilize existing software packages. The approach is demonstrated on both simulation and breast cancer data sets.

C39.2
Multi-state model for analysis of modified Rankin Scale in acute stroke trials: a new approach with a twist
YY Palesch[1], W Zhao[1]
[1]*Medical University of South Carolina, Charleston, United States*

Ordinal response outcomes in clinical trials are one of the more difficult types of outcome to analyze. In Phase III trials of acute therapy for stroke, modified Rankin Scale (mRS) score at 90 days from randomization is a standard primary outcome measure assessing the subjects' functionality. The 7-point ordinal scale of mRS ranges from 0 (normal function) to 6 (death). The most common method for analysis is to dichotomize the mRS (0-1 vs 2-6 or 0-2 vs 3-6 as good vs bad outcome) which yields clinically meaningful statistics (relative risk/benefit or odds ratio). Dichotomization has been criticized to be inefficient, leading to more recent approaches that use the full ordinal scale, such as proportional odds model and assumption-free Cochran-Mantel-Haenszel test.
The former maintains the advantage of yielding a clinically meaningful interpretation using the common OR; however, it is sensitive to the proportionality of odds assumption. The latter has limitations in the number covariates and difficulty in clinical interpretation.
We propose a multi-state modeling as an alternative approach to analyzing the mRS. The method utilizes the full ordinal scale, covariates can be accommodated, and transitional probabilities are clinically interpretable. One caveat is that mRS is unavailable at baseline, and we have explored a surrogate measure (NIH Stroke Scale) that yields good sensitivity and specificity against the mRS. Data from two recently completed large Phase III trials are re-analyzed and compared with the published results.

C39.3
The illness death model under left truncated and right censored data
B Vakulenko-Lagun[1], M Mandel[1]
[1]*The Hebrew University of Jerusalem, Jerusalem, Israel*

Left truncated data arise when a lifetime variable T and an independent truncation variable L are observed only if L<T. There are several ways to perform statistical inference under this setting. One can (i) condition on the event {L<T} only, (ii) condition on the event {L<T} and on L, or (iii) condition on the event {L<T}, on L and on all the history up to time L. When all covariates are time independent, the latter two approaches are exactly the same.
However, the situation becomes more complicated when multi-state models are considered, as approaches (ii) and (iii) differ. More specifically, in the illness-death model, conditioning on all the history up to the truncation time leads to loss of important information as subjects truncated in the illness state do not contribute to estimation of functionals related to the healthy state. This information can be exploited using approaches (i) and (ii), but they require new and more complicated estimation methods. We discuss estimation for non-parametric and regression models under various assumptions and show that estimators obtained in the framework of (i) and (ii) perform better than estimators obtained under (iii). The methods are applied to ICU data collected in a cross-sectional design, where the illness state corresponds to blood-stream infection.

C39.4
Multi-state models for treatment success after stem cell transplantation
LC de Wreede[1,2], J Schetelig[2,3], CJM Halkes[4], H Putter[1]
[1]Medical Statistics, Leiden University Medical Center, Leiden, The Netherlands, [2]Clinical Trials Unit, DKMS, Dresden, Germany, [3]Medical Dept. I, University Hospital Carl Gustav Carus, Dresden, Germany, [4]Hematology, Leiden University Medical Center, Leiden, The Netherlands

The use of multi-state models to model complex disease histories has been advocated for over a decade; however, its use in clinical applications has been limited so far. This is especially striking in the field of hematopoietic stem cell transplantations, since many examples in the statistical literature on this topic come from this field. We will show two examples where we tried to bridge the gap between statistical methodology and clinical questions.
Our main outcome of interest is the probability of treatment succes over time. This outcome is both influenced by baseline characteristics and by intermediate events.
Two related models will be used to analyse treatment success in 2 different transplantation settings. The first dataset describes a multi-center cohort of transplanted CLL (chronic lymphocytic leukemia) patients collected by the European Society for Blood and Marrow Transplantation. The second dataset gives detailed information about a cohort of acute leukemia patients transplanted in Leiden UMC. In the first model, the impact of (baseline) covariates is considered. Dynamic prediction methods are applied to update the estimate of the probability of treatment success.
All analyses will be performed by means of the 'survival' and 'mstate' packages in R.
Our examples show the potential of multi-state models in the analysis of clinically meaningful outcomes. The models are flexible and can easily be adjusted to different clinical questions. However, careful consideration of clinical aspects, data quality and limitations by small sample size are necessary to make applications successful.

C39.5
Comparing multistate approaches for reducing the bias of relative risk estimates from cohort data with missing information due to death
N Binder[1,2], M Schumacher[1]
[1]Center for Medical Biometry and Medical Informatics, Freiburg, Germany, [2]Freiburg Center for Data Analysis and Modeling, Freiburg, Germany

In clinical and epidemiological studies information on the outcome of interest (e.g. disease status) is usually collected at a limited number of follow-up visits. The disease status can often only be retrieved retrospectively in individuals who are alive at follow-up, but will be missing for those who died before. Restricting the analysis to the survivors yields biased hazard ratio estimates of a potential risk factor, and the bias can be in either direction.
We focus on two approaches that use the same likelihood contributions derived from an illness-death multistate model for reducing this bias by including the death cases into the analysis: first, a penalized likelihood approach by Leffondré et al. (Int J Epidemiol, 2013) and second, an imputation based approach by Yu et al. (Biom J, 2010). We compare the two approaches in simulation studies and evaluate them on completely recorded real data, where missing information due to death is artificially induced. For several scenarios, the bias is seen to be reduced compared to an ad-hoc analysis that right-censors the death cases at the last visit.

C40 Model performance evaluation

C40.1
A new measure of predictive ability in a survival model: the total gain statistic
B Choodari-Oskooei[1], P Royston[1], MKB Parmar[1]
[1]MRC Clinical Trials Unit at UCL, London, United Kingdom

The results of prognostic factor studies are usually summarized in the form of statistics resulting from statistical significance testing, i.e. estimated parameters, confidence intervals, and p-values. These statistics do not inform us about whether prognostic factor information will lead to any substantial improvement in the prognostic assessment. Predictive ability measures can be used for this purpose since they provide important information about the practical significance of prognostic factors. R^2-type indices are the most familiar forms of such measures in survival models, but they all have limitations and none is widely used.
In this talk, we extend the total gain (TG) measure, proposed for a logistic regression model, to survival models and explore its properties using simulations and real data. TG is based on the binary regression quantile plot, otherwise known as the predictiveness curve. Standardised TG ranges from 0 (no explanatory power) to 1 (`perfect´ explanatory power).
The results of our simulations show that unlike most of the other R^2-type predictive ability measures, TG is independent of random censoring. It increases as the effect of a covariate increases, and it remains largely unaffected by the categorisation of continuous prognostic factors. Furthermore, it can be applied to different types of survival models, including models with time-varying effects. Finally, we applied TG to quantify the predictive ability of prognostic models developed in several disease areas. On balance, TG performs well in our simulation studies and can be recommended as a measure to quantify the predictive ability in survival models.

C40.2
A note on the time-profile of time-dependent area under the ROC curve for survival data
J Lambert[1], R Porcher[1,2,3], S Chevret[1,4,5]
[1]Inserm U1153, Paris, France, [2]Hôpital Hôtel Dieu, Paris, France, [3]Université Paris Descartes, Paris, France, [4]Hôpital Saint Louis, Paris, France, [5]Université Paris Diderot, Paris, France

In the setting of survival analysis, the time-dependent area under the receiver operating characteristic curve (AUC) has been proposed as discrimination measure of interest. In contrast with the diagnostic setting, the definitions of time-dependent sensitivity and specificity are not unique and three time-dependent AUC are used in practice: cumulative/dynamic, incident/dynamic and incident /static. This work evaluates the time-dependent profile of these AUC(t).
We show that, even when the effect of a binary biomarker on the hazard rate is constant, the value of AUC(t) varies over time according to the prevalence of the marker. The time- profile of continuous biomarker is illustrated with a simulation study, and data on several prognostic factors in AML are examined.

C40.3
A unified approach for testing goodness of fit in binary, multinomial, and ordinal logistic regression models

MW Fagerland[1], DW Hosmer[2]
[1]Oslo University Hospital, Oslo, Norway, [2]University of Massachusetts, Amherst, United States

Evaluating goodness of fit is an important step in the assessment of the adequacy of a regression model. Logistic regression models are popular because of their availability in software packages and the fact that the exponential form of the regression coefficients can be interpreted as odds ratios. For binary logistic models, the Hosmer-Lemeshow (HL) test is in widespread use. The test is based on a strategy of sorting and grouping the observations according to their estimated probabilities of event. The test statistic is the Pearson chi-squared statistic on a contingency table where the groups form the rows and the observed and estimated frequencies form the columns.
The HL approach for constructing a goodness-of-fit test can also be used for the multinomial logistic model and several ordinal logistic models: the proportional odds, adjacent category, and constrained continuation-ratio models. The required modifications of the HL test from the binary case to the multinomial and ordinal cases involve finding a suitable function of the estimated probabilities to use for sorting the observations and determining the appropriate degrees of freedom for the chi-squared reference distribution.
Simulations pit the HL tests against other goodness-of-fit tests and show that the HL tests are capable of detecting several different types of poorly fit models and can be recommended with moderate and large sample sizes. The recommendation comes with a warning: no single test can provide a complete assessment of model fit. Ideally, a battery of tests and case-wise diagnostic tools should be used.

C40.4
Nonparametric estimation of covariate-specific summary indices of ROC curves through regression models

JC Pardo-Fernández[1], EM Molanes-López[2], E Leton[3]
[1]Universidade de Vigo, Vigo, Spain, [2]Universidad Carlos III de Madrid, Madrid, Spain, [3]UNED, Madrid, Spain

The receiver operating characteristic (ROC) curve is a statistical tool of extensive use in diagnostic studies. The ROC curve allows for the visualization of the effect of different thresholds of the diagnostic variable in terms of sensitivity (probability of classifying a diseased individual as diseased) and specificity (probability of classifying a healthy individual as healthy). Some summary indicators, such as the area under the curve (AUC) or the Youden index, are often employed to describe the main features of the ROC curve.
In many studies, a covariate is available along with the diagnostic variable. The behaviour of the ROC curve may depend on the values of the covariate, and therefore it is interesting to study the impact of the covariate on the covariate-specific ROC curve. This work will be devoted to the study of a nonparametric estimatior of the covariate-specific ROC curve and its associated summary indices, specifically, the covariate-specific AUC and the covariate-specific Youden index. The incorporation of the information of the covariate over the diagnostic variable is modelled through nonparametric location-scale regression models.

C40.5
On bias of measures of explained variation for survival data

J Stare[1], D Maucort Boulch[2], N Kejžar[1]
[1]Faculty of Medicine, Ljubljana, Slovenia, [2]University of Lyon, Lyon, France

Papers evaluating measures of explained variation, or similar indices, invariably use independence from censoring as the most important criterion. And they invariably end up suggesting that some measures meet this criterion, and some don't, leading to a conclusion that the first are better than the second. As a consequence, users are offered measures that cannot be used with time-dependant covariates and effects, not to mention extensions to repeated events or multi state models.
We explain in this paper that the above mentioned criterion is of no use in studying such measures, since it simply favours those that make an implicit assumption of a model being valid everywhere. Measures not making such an assumption are disqualified, even though they are better in every other respect. We show that if these, allegedly inferior, measures are allowed to make the same assumption, they are easily corrected to satisfy the `independent-from-censoring´ criterion. Even better, it is enough to make such an assumption only for the times greater than the last observed failure time $\tau$. Which, in contrast with the `preferred´ measures, makes it possible to use all the modelling flexibility up-to τ, and assume whatever one wants after τ.
As a consequence, we claim that measures being proffered as better in the existing reviews, are exactly those that are inferior.

C41 Survival analysis II

C41.1 *Conference Award for Scientists*
Kernel estimation of hazard function for orthopedic data

J Zelinka[1], I Horová[1], S Katina[1], I Selingerová[1]
[1]Masaryk University, Faculty of Science, Brno, Czech Republic

The hazard function is an important tool in survival analysis and reflects the instantaneous probability that an individual will die within the next time instant. The hazard function can depend on any covariates as age, gender, etc. In the present paper the kernel estimators of the hazard function and of the conditional hazard function are discussed.
These methods are applied to the real data from Slovak Arthroplasty Register about implants of an artificial hip joint replacement implemented in all 40 orthopaedic and traumatology departments in the Slovak Republic (coverage of 99.9%) with a maximum duration of follow-up of ten years from Jan 1 2003 to Dec 31 2013. The set of 35 182 operations with 665 implant failures is stratified based on types of fixation, diagnosis, and gender. The hazard function conditioned on age in years is calculated for pre-specified data-subsets and visualized as color-coded surfaces. These results will lead to an improvement of the quality of care for patients after artificial joint replacements.

C41.2

Quantile regression and prediction intervals for survival data

M Mayer[1], Q Li[2,3]
[1]*Consult AG Bern, Zurich, Switzerland,* [2]*Swiss Group for Clinical Cancer Research (SAKK), Bern, Switzerland,* [3]*University of Bern, IMSV, Bern, Switzerland*

Cox models are by far the most traditional statistical modelling technique in survival data analysis, e.g. because the effects of predictor variables have a simple interpretation as hazard ratios. However, when used for predicting survival times of individual patients (based on patient characteristics such as age, sex etc.), Cox models are unhandy and there is no simple way to quantify the precision of these individual predictions.
A very powerful but still quite unknown alternative to the Cox model is quantile regression, originally introduced into survival data analysis by J. Powell in 1986. It allows modeling any quantile of the (log) survival time distribution, for instance the median and the two other quartiles, as a linear function of the predictor variables. Quantile regression is almost as simple to use and to interprete as a multiple linear regression and is e.g. available in Roger Koenker's "quantreg" library in R.
We illustrate the method and its flavor using real survival data and show a trick how to use it to obtain "forecast" intervals for individual patients. Such an interval does not only quantify the precision of the corresponding point prediction but also answers the question "how long will I survive" in an honest and patient focused way.

C41.3

A special case of the reduced rank model for modelling time varying effects in survival analysis

A Perperoglou[1]
[1]*University of Essex, Colchester, United Kingdom*

Consider the case of modelling time to event data, where the effect of some covariates on the hazard function might change with time. Starting from a proportional hazards model, one can introduce interactions of fixed covariates with time functions to model the dynamic behaviour of the effects. In Perperoglou et al (2006) Reduced Rank Hazard Regression was introduced as an approach to achieve parsimonious models with few parameters. The approach was further extended to include both fixed and time varying effects of the covariates (Perperoglou 2013). However, a serious issue remain, which of the covariates in the model should be allowed to have time varying effects and which not.
In this work we will present our findings on the suggestion of van Houwelingen and Putter, to fit a modified rank one model with all covariates having both time varying and time fixed effects. The special case of the rank=1 model can be written as:
$h(t|X)=h_0(t)\exp(Xb_1+Xb_2\gamma F')$
where X is a matrix of covariates, b-s are the vectors of coefficients for the fixed effects and γ is a vector of coefficients for the interactions of fixed covariates with a matrix of time functions F.
We will illustrate how to fit the model using an alternating least squares algorithm, the properties of this approach and results from a series of applications in real and simulated data.

C41.4

Estimating probability of non-response to treatment with survival data

A Callegaro[1], B Spiessens[1]
[1]*GSK Vaccines, Rixensart, Belgium*

The treatment effect reported from clinical trials represents the average of the individual benefit from treatment. Classical statistical approaches in cancer clinical trials evaluate treatment-effect heterogeneity by modeling the interaction between the treatment and some known covariates.
However, often the underlying mechanism that causes variability is unknown and the relevant covariates are not observed. Further, if there is biological or empirical evidence that only a portion of treated patients respond to the treatment it is interesting to estimate the probability of patients to respond to the treatment. This probability could be used for personalized treatment selection i.e. to define a target population for a future comparative study. Different approaches to model the treatment effect heterogeneity for survival data (e.g. mixture models) will be presented and compared by clinical trial simulations.
Simulation results will give an idea of the amount of information (sample size, proportion of responders, treatment effect in responders) necessary to accurately estimate the probability of non-response to treatment on (oncology) trial data with survival outcome.
Funding source: GlaxoSmithKline Biologicals SA

C41.5

An application of frailty modeling for family level clustering of infant mortality in Empowered Action Group states in India

K Mani[1], RM Pandey[1]
[1]*All India Institute of Medical Sciences, New Delhi, India*

Objectives: In India, the focused intervention policies led to a decline in mortality among children younger than five years, yet some of the states in India are having very high mortality rates. We explored the effects proximate determinants on infant mortality by accounting for family level clustering using Cox frailty model in Empowered Action Group states (EAG) in India and compared the results with standard models.
Methods: Analysis included 20,126 live births that occurred five years preceding the National Family Health Survey-3 (2005-06). The Cox frailty model was used to account for the family level clustering.
Results: Of the 20126 live births, 1223 babies died before reaching their first birthday. The Cox frailty model showed that mother's age at birth, composite variable of birth order and birth interval, size of the baby at birth and breastfeeding among proximate determinants were significant determinants of infant mortality after adjusting for familial effect. The familial frailty effect was 2.52 in the EAG states. The inferences on the determinants for all the three models were similar except the death of a previous child and mother's age at birth in the Cox frailty model, which had the highest R2 and lowest log-likelihood.
Public Health Implications: While planning for the child survival program in EAG states, parental competence which explains the unobserved familial effect needs to be considered along with significant proximate/ programmable determinants. The frailty model that provide statistically valid estimates of the covariate effects are recommended, when observations are correlated.

C42 Poly-omics studies & Systems Biology

C42.1

A stratified boosting approach for combining gene expression measurements from different platforms to identify prognostic markers

J Mazur[1], I Zwiener[1,2], H Binder[1]
[1]University Medical Center Mainz, IMBEI, Mainz, Germany, [2]Merck KGaA, Darmstadt, Germany

Development of gene expression risk prediction signatures in a survival setting is typically severely constrained by the number of samples. A natural approach which analyzes several data sets simultaneously is a pooled analysis of samples. However, gene expression studies are often performed on different platforms, like RNA-Seq and microarrays, such that direct pooling of individual patient data is not possible anymore.
To still be able to combine gene expression studies, we propose a stratified boosting approach for regularized estimation of Cox regression models. For every study, i.e. every stratum, a componentwise likelihood-based boosting algorithm is performed where the variable that is updated in each step is the one where the score statistic is the largest across studies.
For evaluation, the prediction performance of our stratified boosting approach is compared to the prediction performance of the pooled analysis for simulated data. Additionally, for simulated data, we quantify the performance with respect to identifying important genes for our stratified boosting approach, the pooled analysis and a setting where only gene lists but not the data itself is available for the gene expression studies. Finally, we apply our approach to RNA-Seq and gene expression microarray data from kidney clear cell carcinoma patients.
The results indicate that our newly proposed stratified boosting approach performs close to the pooled analysis where the latter is feasible, and in addition makes it possible to combine gene expression studies from different molecular platforms.

C42.2

Weighted penalized canonical correlation analysis to integrate multiple omics-data

A Zwinderman[1]
[1]Academic Medical Center of the University of Amsterdam, Amsterdam, The Netherlands

To integrate omics-data from multiple platforms and to integrate this with phenotypic data, we suggested canonical correlation analysis.
Since omics-data are usually high dimensional, we suggested a penalized version (PCCA). We used the elastic net because this method is capable to perform variable selection but also groups correlated variables (which may represent biological pathways) (Waaijenborg & Zwinderman 2009,2010,2011).
To associate the multiple platforms and clinical data with each other we maximized the sum of the multiple correlations between the canonical variates of each platform. Optimal penalty parameters were estimated by k-fold crossvalidation using a grid search and we optimized the absolute mean difference between the canonical correlations between the training and test sets. We now extended PCCA with a weighting scheme to account for (causal) direction in the association analysis. Such causal pathway is useful, for instance, when integrating genomewide SNP/DNA sequence data with genomewide methylation and expression data or with proteome/metabolome data.
In addition the weighting schemes may also be used to search specifically for cis-regulatory elements, either located physically-close to a particular gene or located metaphysically-close to a particular protein in a biological/metabolic pathway. We illustrate the weighted PCCA approach by analyzing the associations between 700K SNPs, 200K CNVs, beta-methylation values of 450K CpG-sites, 20K gene expression values and 100 phenotypes measured in 237 patients with Marfan syndrome.
We used a weighting scheme to test for the expectation that the phenotypic variation is influenced by SNPs, CNVs and methylation data only through the gene-expression values.

C42.3

Prediction performance as a measure for optimal mapping of methylation and RNA-Seq data

A Gerhold-Ay[1], J Mazur[1], H Binder[1]
[1]IMBEI, Universitätsmedizin Mainz, Mainz, Germany

Next-generation sequencing and microarray data are becoming more and more important for medical research. They enable us to develop gene signatures for prediction of clinical endpoints like death, via the integration of the information present in RNA-Seq data on gene expression and methylation data on CpG sites. This still has the challenge which CpG sites should be considered as being related to one specific gene. Our aim is to investigate how the prediction performance measure can be used as a measure for optimality to find the mapping of CpG sites to their related genes.
To find the optimal mapping for methylation to gene information, we define a length of nucleotides around all genes, which we call a window around these genes. In a two-step approach, we first use a likelihood-based componentwise boosting approach to estimate a gene signature only with RNA-Seq data. In the following step, the methylation data of the CpG sites that are falling in this window are used to estimate a new signature. For finding prognostic signatures, RNA-Seq and methylation data of kidney tumor patients are used. We analyze different window sizes for the mapping and show that they have an effect on the prediction performance with respect to the clinical endpoint.
Prognostic gene signatures can be a powerful tool for the classification of cancer patients. To underpin this tool, we propose the prediction performance measure as a criterion to find the optimal mapping window for RNA-Seq and methylation data and show its usefulness.

C42.4

Integration of somatic mutation, gene expression and functional data in predicting human breast cancer survival

C Suo[1], D Lee[2], D Saputra[1], H Joshi[1], S Pramana[1,3], S Calza[1,4], Y Pawitan[1]
[1]Dept of Medical Epi and Biostatistics, Karolinska Institutet, Stockholm, Sweden, [2]Dept of Statistics, Ewha Womans University, Seoul, Korea, Republic of, [3]Institute of Statistics, Jakarta, Indonesia, [4]University of Brescia, Brescia, Italy

Whole-genome and transcriptome sequencing experiments can be used to explore the understanding of human cancers comprehensively. The Cancer Genome Atlas breast cancer consortium provides a unique data structure by sequencing from sixty matched tumor and normal sample of the same female patient diagnosed with breast invasive carcinoma, allowing us to accurately infer somatic mutations and isoform-level expression. However, it is not immediately obvious how to subsequently construct and integrate the complex network of the diverse signatures discovered, owing to a lack of mature statistical tools. The fundamental challenges also lie in identifying patient-specific mutational event contributing to the heterogeneity pattern between tumors and translating the findings into clinically relevant aspects.
We propose a novel method to integrate genomic and transcriptomic profiles based on network enrichment analyses, revealing statistical evidence of the functional implications of the biomarkers found between- and within-patients. We develop a weighted driver gene score summarizing the mutated driver genes that are common across patients and those that are patients-specific. To contribute to the driver gene score, a gene has to

be frequently mutated, with high or moderate mutational impact, exhibiting an extreme expression and functionally linked to a large number of differentially expressed neighbors in the gene network.
We show that breast cancer patients carrying more mutated driver genes with functional implications and extreme expression pattern have worse survival than those with less mutated driver genes. We propose the driver gene score as an informative tool to predict survival for guiding patient care and clinical management.

C42.5

Nonparametric mixture modelling of dynamic Bayesian networks derives the structure of protein-networks in adhesion sites

Y Fermin[1], K Ickstadt[1], MR Sheriff[2], S Imtiaz[2], E Zamir[2]
[1]TU Dortmund University, Faculty of Statistics, Dortmund, Germany,
[2]Max Planck Institute of Molecular Physiology, Dortmund, Germany

Cell-matrix adhesions play essential roles in important biological processes including cell morphogenesis, migration, proliferation, survival and differentiation (Gumbiner, 1996; Hynes and Lander, 1992). The attachment of cells to the extracellular matrix is mediated by dynamic sites along the plasma membrane, such as focal adhesions, at which receptors of the integrin family anchor the actin cytoskeleton to components of the extracellular matrix via a large number of different proteins (Zamir and Geiger, 2001). Focal adhesions can contain over 100 different proteins, including integrins, adapter proteins, and intracellular signaling proteins (Zaidel-Bar et al., 2007). Due to the large number of components and diversity of cell-matrix adhesion sites, a fundamental question is how these sites are assembled and function.
In systems biology graphical models and networks have been widely applied as a useful tool to model complex biochemical systems. In this work we propose a nonparametric mixture of dynamic Bayesian networks to study interactions among proteins in the presence of the temporal structure and heterogeneity among focal adhesions. Nonparametric mixture modelling of dynamic Bayesian networks is developed by a combination of dynamic Bayesian networks (Ghahramani, 1997) and of nonparametric Bayesian networks (Ickstadt et al., 2011). This approach provides further grouping of focal adhesions according to their network structures.
We apply and illustrate our approach using multicolor live cell imaging datasets, in which the levels of four different proteins are monitored in individual focal adhesions.
Keywords: Cell-matrix adhesions; Dynamic Bayesian networks; Nonparametric Bayesian networks

Wednesday, 27th August 2014 – 16:00-17:30

Invited session

I6 Statistical methods for poly-omics studies

Organizers: Axel Benner and Manuela Zucknick

I6.1

From associations to mechanical understanding - data integration and causal inference in genomics

R Spang[1]
[1]*University of Regensburg, Regensburg, Germany*

If we want to find out whether a drug is effective in a certain disease, we have only one working option: We must test it in cellular assays, in mice and ultimately in clinical studies. Genomic reasoning is no option yet, but could it become one in the future? The problem requires a functional understanding of cells and organisms. In more statistical terms, we need to infer causal relations between perturbations of cellular pathways and their downstream effects.
In this talk I will give a brief introduction into cellular signaling and will then address a couple of statistical problems associated with their analysis: The detection of pathway activation in expression profiles of tumors, the construction of signaling models from perturbation data, and the estimation of causal effects from observational data.

I6.2

Bayesian models for integrative genomics

M Vannucci[1]
[1]*Rice University, Houston, United States*

Novel methodological questions are being generated in the biological sciences, requiring the integration of different concepts, methods, tools and data types. Bayesian methods that employ variable selection have been particularly successful for genomic applications, as they allow to handle situations where the amount of measured variables can be much greater than the number of observations.
In this talk I will focus on models that integrate experimental data from different platforms together with prior knowledge. I will look in particular at hierarchical models that relate genotype data to mRNAs, for the selection of the markers that affect the gene expression. Specific sequence/structure information will be incorporated into the prior probability models. All modeling settings employ variable selection techniques and prior constructions that cleverly incorporate biological knowledge about structural dependencies among the variables. Applications will be to data from cancer studies.

I6.3

Do we gain by jointly analyzing multiple types of genomics data?

WN van Wieringen[1]
[1]*VU University Medical Center, Dept. of Biostatistics, Amsterdam, The Netherlands*

Through integration of genomics data from multiple sources, we obtain a more accurate and complete picture of the molecular mechanisms underlying tumorigenesis. Thus sounds the promise. What about practice? In this talk I will show that we may indeed gain from integrative analysis of the multiple genomics data. But adding clinical information to the mix proves valuable.
To show joint analysis may deliver I concentrate on the integration of DNA copy number and gene expression data from oncogenomics studies with a two-sample set-up. These molecular levels are linked through the central dogma of molecular biology. In this context the aim is to identify differential (between the two clinical groups) regulation among the genes of a pathway.
For starters the gene-centered (univariate) analysis of such data is discussed. This reveals no differential expression between the two groups. Alternatively, no significant association between the genomic and transcriptomic level is detected when ignoring group information.
However, incorporate both clinical and genomic information and differential associations abound.
The main course features pathways. The interactions among the pathway´s molecular constituents are described by a structural equation model (SEM). With this model I am able to show that inclusion of DNA copy number data benefits the discovery of gene-gene interactions. Extension of the SEM to accommodate group information reveals differential regulation between the groups. But this differential gene-gene interaction pattern is missed when DNA copy number is not accounted for!
Time for desert: is more thus better? Only when the data is well shaken and stirred.

Contributed sessions

C43 Causal inference from observational data II

C43.1

Using different propensity score matching methods to construct comparable control groups for disease management program evaluation

R Riedl[1], A Berghold[1]
[1]*Inst Med Info, Stat & Docu, Medical University of Graz, Graz, Austria*

In observational studies, confounders, defined as variables associated with both, treatment and disease outcome may induce a bias in the estimates of association. Different matching methods are frequently used to reduce systematic differences of baseline characteristics between treated and untreated individuals. However, if the number of confounding variables is large, matching on the variables itself becomes challenging. Propensity scores (PS), defined as conditional probability of treatment assignment given observed baseline covariates are used to overcome this dimensionality problem. In the literature, the performance of several matching methods, including optimal matching, nearest neighbour matching or matching within calipers, have recently been investigated by simulation studies for constructing matched pairs. It has been noted that for different matching methods and also for the order in which individuals are selected for matching can result in different qualities of the matches.
We investigate the influence of different propensity score matching methods in combination with exact matching methods on the ability to induce balance on baseline covariates between the treatment groups of the matched samples in praxis. We apply these methods to data of a disease management program in patients with type 2 diabetes. Furthermore, we investigate the impact on balance if more than one control per case is matched.

C43.2

Double propensity-score adjustment: a solution to incomplete matching

P Austin[1]
[1]*Institute for Clinical Evaluative Sciences, Toronto, Canada*

Propensity-score matching allows for estimation of the average treatment effect in the treated (ATT). However, popular matching methods such as nearest neighbour caliper matching, often result in some treated subjects being excluded from the final matched sample. This can lead to loss of generalizability of the estimated treatment effect, since the estimand only applies to the matched treated subjects, and not to the entire population of treated subjects. Alternative matching methods such as nearest neighbour matching (NNM) and optimal matching result in the inclusion of all treated subjects in the matched sample, at the cost of the elimination of a lesser degree of bias due to confounding variables.
We propose a method based on using covariate adjustment using the propensity score within a sample constructed using NNM or optimal matching to address these two limitations. Using a series of Monte Carlo simulations, we compared the performance of double propensity-score adjustment to caliper matching, NNM, and optimal matching.
The proposed method results in improved generalizability compared to caliper matching and greater bias reduction compared to NNM or optimal matching alone. We illustrate the application of this method using a sample of patients hospitalized with a heart attack.

C43.3

A structural equation modelling approach to explore the role of interferon-α on chronic immune activation in successfully treated HIV-infected patients

M-Q Picat[1], I Pellegrin[2], J Bitard[2], L Wittkop[1], C Proust-Lima[1], B Liquet[3], J-F Moreau[4], R Thiébaut[1]
[1]*Centre Inserm U897- Epidémiologie - Biostatistique, Bordeaux, France,* [2]*Laboratoire d'Immunologie-Immunogénétique, Bordeaux, France,* [3]*School of Mathematics and Physics, Saint Lucia, Australia,* [4]*CNRS, UMR 5164, Bordeaux, France*

Background: Chronic Immune Activation (CIA) is a predictor of Human Immunodeficiency Virus (HIV) progression. In successfully treated patients, the understanding of mechanisms by which CIA persists is still limited. We hypothesized that cytomegalovirus (CMV) could be an important factor of CIA in these patients through persistent production of interferon α (IFN-α).
Methods: Data from 191 HIV-1-infected patients were analyzed. Patients initiated antiretroviral therapy between 2005 and 2008, and were treated with sustained virological suppression for at least two years. CMV-induced immune response was measured by QuantiFeron-CMV test (positive if >0.2 IU/mL) and CIA was defined by HLA-DR+/CD38+CD8+T-cells. Structural equation modeling (SEM) was used to evaluate the mediating role of IFN-α related gene-transcription (defined as one or several latent variables using 21 IFN-α-induced genes) in the relationships between CMV and CIA. Several definitions of the latent IFN-α were explored, including IFN-stimulated-genes (n=5) and MYD88 dependent (n=6) or independent (n=10) pathways.
Results: The hypothesized SEM model revealed a strong association between IFN-α latent variable and CIA (p=0.00034). This association persisted in modeling IFN-α by IFN-stimulated-genes (p=0.00129). Modelling IFN-α through two types of latent variables: MYD88 dependent and independent, revealed a strong association between IFN-α MYD88 independent variable and CMV (p = 0.00501) and CIA (p = 0.00001).
Conclusion: SEM provides a flexible framework to explore complex relationships between variables and to understand mediation. In our application, a major role of IFN- α was demonstrated in the association between CMV and CIA.

C43.4

Independent censoring in survival analysis: a causal approach

K Røysland[1]
[1]*University of Oslo, Dep of Biostatistics, Oslo, Norway*

"Independent censoring" is a common assumption when using the Kaplan-Meier estimator.
This means that an individual who has not (by chance) experienced the event in question has the same risk of experiencing the event in an infinitesimal period, regardless of any previous censoring. The formal definition relies on martingale theory and yields a dynamic concept that is much weaker than assuming the censoring and event in question are independent in the usual probabilistic sense.
It is tempting to think that independent censoring would mean that Kaplan-Meier curves represent the survival as would be seen if the censoring had been prevented.
This, however, is a claim about causation, and can be treated formally using graphical models and techniques from causal inference.
"Local independence graphs" and "local characteristics" provide an analogy to causal Bayesian networks, where the nodes also may represent counting processes.
Independent censoring is actually a special case of local independence, so these graphical models provide a natural framework for our purpose.
Suppose a model is causal with respect to change of censoring regimes.

The previous interpretation of independent censoring would be valid if the parameter of interest would not change after replacing the observational intensity of censoring with 0. We will discuss identifiability of parameters when subject to hypothetical censoring regimes. Especially the ones corresponding to stabilized and non-stabilised censoring weights. Using local independence graphs and delta-separation, we derive an analogy to the back-door criterion that applies to censoring in survival analysis.

C44 Validation of prediction models

C44.1
The need for a third dimension in the external validation of clinical prediction rules

W Vach[1]
[1]*Clinical Epidemiologiy, IMBI, University of Freiburg, Freiburg, Germany*

When clincial prediction rules have to be validated in an external data set, the focus is often on two dimensions: calibration and discrimination. However, these two dimensions do not cover the whole information about the discrepany between the true event probabilities and the suggested probabilities according to the clinical prediction rule. We present some (theoretical) examples with varying degree of agreement between true and suggested event probabilities, which give identical calibration scope, AUC and Brier score.
To overcome the problem, we can consider to estimate directly some measures of the agreement between true and suggested event probabilities, like the euclidian distance. However, such measures may be hard to interpret. As an alternative, we suggest to estimate the inverse calibration slope, i.e. the slope of a regression of the suggested vs. the true event probabilities. The joint interpretation of the inverse calibration slope and the ordinary calibration slope is simple: If both are 1, then we have perfect agreement. We demonstrate that the inverse calibration slope can be estimated by a boostrap bias correction of the naive estimate based on a flexible estimate of the true event probabilities.

C44.2
Multiple validation of prediction models: a framework for summarizing and interpreting results

D Nieboer[1], Y Vergouwe[1], TPA Debray[2], H Koffijberg[2], KG Moons[2], EW Steyerberg[1]
[1]*Erasmus MC, Rotterdam, The Netherlands,* [2]*UMC Utrecht, Utrecht, The Netherlands*

Aim: A commonly found rationale is that multiple successful validation across different settings increase the likelihood that a model is valid for new settings. We aimed to develop a framework to critically assess the evidence of validation studies.
Methods: We developed a model predicting 6 month mortality in patients with traumatic brain injury from a single observational study. We validated the model on 14 other cohorts from the IMPACT database (3 observational studies and 11 RCTs). Overall calibration was assessed with calibration-in-the-large and average predictor strength with the calibration slope. We constructed forest plots to summarize validation results. We quantified heterogeneity using the I^2 statistic and calculated prediction intervals (PIs). Meta-regression was used to identify factors explaining the observed heterogeneity.
Results: The pooled calibration slope indicated that predictor effects were less strong at validation (pooled estimate 0.72, PI 0.37-1.06), with substantial heterogeneity (I^2 95%). Meta-regression showed that type of cohort (observational study/RCT) explained most of this heterogeneity. The pooled estimate of the calibration-in-the-large indicated that predicted probabilities were on average too high (-0.62 PI [-1.50 - 0.26]). The observed heterogeneity was again substantial (I^2 94%), but could not be explained with meta-regression.
Conclusion: We propose the use of meta-analytic methods to summarize the cumulating evidence of validation studies for prediction models. If limited heterogeneity is observed, the model is likely generalizable to the studied settings.
However, if heterogeneity is observed meta-regression may identify sources of heterogeneity to guide the interpretation of the validity and applicability of the prediction model.

C44.3
Summarising the performance of prognostic models developed and validated using multiple studies

KIE Snell[1], TPA Debray[2], J Ensor[3], MP Look[4], KG Moons[2], RD Riley[3]
[1]*MRC Midland Hub for Trials Methodology Research, Birmingham, United Kingdom,* [2]*University Medical Center Utrecht, Utrecht, The Netherlands,* [3]*University of Birmingham, Birmingham, United Kingdom,* [4]*Erasmus MC Cancer Institute, Rotterdam, The Netherlands*

Internal-external cross-validation (IECV) is an approach for developing and validating a prognostic model when data from multiple studies are available. The model is developed multiple times, each time excluding a different study for external validation of its performance (discrimination and calibration). This produces multiple values for every validation statistic of interest (e.g. C-statistic, calibration slope).
In this presentation we extend IECV by using random-effects meta-analysis to combine and summarise the validation statistics across the omitted studies. We show it provides two crucial summaries: (i) the average model performance in the different populations, and (ii) the heterogeneity of model performance across populations. A good prognostic model will have excellent average performance with little or no heterogeneity. We explain how the meta-analysis approach also allows model implementation strategies to be compared; for example regarding the choice of intercept or baseline hazard.
The presentation concludes with some novel extensions. First, we use the meta-analysis results to produce 95% prediction intervals for the validation performance in a new population. Narrow intervals are desirable if a model is likely to perform consistently in new populations. Then we propose multivariate meta-analysis to summarise correlated validation statistics (such as the C-statistic and calibration slope), to determine the probability that both discrimination and calibration performance will be acceptable in practice. Real examples in breast cancer and deep vein thrombosis are used throughout.

C44.4
Incorporating retrospective information to reduce the sample size of prospective diagnostic-biomarker-validation designs

L García Barrado[1], E Coart[2], T Burzykowski[1,2]
[1]*I-Biostat, Hasselt University, Diepenbeek, Belgium,* [2]*International Drug Development Institute (IDDI), Louvain-la-Neuve, Belgium*

Problem setting: The sample size of a prospective clinical study aimed at validation of a diagnostic biomarker may be prohibitively large. A Bayesian framework that would incorporate available retrospective data on the accuracy of the biomarker might allow reducing the sample size and rendering the study feasible.
Methods: A Bayesian design is presented for planning and analyzing a prospective clinical validation study that incorporates retrospective data.

In particular, validation is based on the Bayesian testing of a hypothesis regarding possible values of AUC. Toward this aim, first, available information is translated into a prior distribution. Next, this prior information is used in a Bayesian design to estimate biomarker's accuracy. This allows reducing the sample size as compared to the "classical", frequentist approach, in which the availability of any information on biomarker's diagnostic performance is ignored.
Results: A simulation study is performed to evaluate the power of the proposed design. For each scenario, 200 studies of sample size 100, 400, 600, and 800 are simulated. The power of the Bayesian design to confirm a satisfactory diagnostic performance of a biomarker is compared to a corresponding frequentist design.
Conclusion: For each study size, the proposed Bayesian design leads to a significantly higher power than the frequentist design. In some of the considered simulation settings, the Bayesian design required as little as ¼ of the frequentist-trial sample size to reach approximately the same power.

C45 Multistate models and competing risks II

C45.1
Regression models for expected length of stay

MK Grand[1], H Putter[1]
[1]*Leiden University Medical Center, Leiden, The Netherlands*

A multi-state model is a stochastic process with outcomes in a finite space that represents the states. The expected length of stay (ELoS) is defined as the time the process is expected to spend, in total, in a given state. ELoS is not a straightforward object to relate to covariates and the traditional approach has been to construct regression models for the transition intensities, and from these calculate ELoS. The disadvantage of this approach is that the effect of covariates on the intensities are not easily translated into the effect on the ELoS. Furthermore, it typically relies on the assumption that the process is Markov.
We propose using pseudo-observations (Andersen et al., Biometrika 2003) to make regression models for ELoS, thereby allowing a direct interpretation of covariate effects and evading the Markov assumption. For this approach, all we need is a non-parametric (asymptotically) unbiased estimator for ELoS. For every subject (and for every state of interest) a pseudo-observation is constructed and they are then used as outcome variables in the regression model.
We furthermore show how to construct longitudinal (pseudo-) data when combining the concept with landmarking. Covariates may then be time-varying and potential time-varying effects can be explored.
The models can be fitted using generalised estimating equations and by applying the sandwich estimator. The method is illustrated using data from the US Health and Retirement Study to explore the impact of socio-economic factors on ELoS in health and disability. The efficiency of our approach is investigated through simulations.

C45.2
A multistate model to assess the impact of menstrual status in premenopausal breast cancer patients

S Weber[1], M Schumacher[1]
[1]*Center for Medical Biometry and Medical Informatics, Freiburg, Germany*

Adjuvant treatment in premenopausal breast cancer patients may affect the menstruation, a cessation of menses is possible. Depending on treatment, recovery of menses may occur afterwards. How does the menstrual status impact disease-free survival?
The motivating question leads to a multistate model which allows the analysis of subsequent events concerning menstrual status and the respective transitions into the absorbing state defined by tumor recurrence and death. Problems like competing risk, right-censoring and left-truncation have to be considered.
The Zoladex Early Breast Cancer Research Association (ZEBRA, Jonat et al, Journal of Clinical Oncology 2002) study compares a hormone therapy with goserelin and a chemotherapy as adjuvant treatment in premenopausal patients with node-positive breast cancer. Since goserelin works via suppressing the ovarian estrogen production it induces cessation of menses but recovery is possible.
We investigate the effects of time-dependent menstrual status within a multistate model thereby using several Cox models including time-dependent covariates.
Furthermore, results of a similar clinical trial (trial VIII) of the International Breast Cancer Study Group are considered and compared with those derived from the ZEBRA study.

C45.3
Variable selection in the illness-death model

M Lauseker[1], J Schemenau[2], U Germing[2], VS Hoffmann[1]
[1]*Ludwig-Maximilians-Universität München, München, Germany,* [2]*Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany*

Preferably, variable selection should be done by content-related reflections. But this is often not possible. The illness-death model is used more and more frequently, but recommendations on variable selection are rare. Data were simulated according to an illness-death model without recurrent events. Using this simulated data a semi-Markov model was fitted. Simulated data sets included an interaction effect between the independent variables and a strong correlation between two variables.
Backward selection based on AIC and BIC was used on the one hand directly, on the other hand applying a bootstrap step. The performance of the selection procedures was measured via the inclusion fraction and the bias of the estimated coefficients.
In the simulations, both selection criteria yielded reasonable models. As expected, BIC led to more parsimonious models than AIC, regardless if bootstrapping was used. BIC performed slightly better with regard to both inclusion fraction and bias of the coefficients. Bootstrapping did not generally improve the results.
The results were illustrated with a real world data set on myelodysplastic syndromes (MDS) patients concerning an illness-death model with the states "MDS", "acute myeloid leukaemia" and "death". In this data a time-dependent effect was only detected by AIC.
Model selection via AIC and BIC works well for illness-death models, even though the true model was only found infrequently. Especially when strong correlation between the covariates is present, bootstrapping can lead to difficulties.

C45.4
Statistical models for improving prognosis of chronic cardiovascular diseases: hazard reconstruction and clustering of patients affected by heart failure

F Ieva[1], AM Paganoni[2], T Pietrabissa[2]
[1]*Department of Mathematics, Università degli Studi di Milano, Milano, Italy,* [2]*Department of Mathematics, Politecnico di Milano, Milano, Italy*

Heart Failure (HF) is nowadays among the leading causes of repeated hospitalisations in over 65 patients. The longitudinal dataset resulting from the discharge papers and its analysis are consequently becoming of a great interest for clinicians and statisticians worldwide in order to have insights of the burden of such an extensive disease.

We analysed HF data collected from the administrative databank of an Italian regional district (Lombardia), concentrating our study on the days elapsed from one admission to the next one for each patient in our dataset.
The aim behind this project is to identify groups of patients, conjecturing that the variables in our study, the time segments between two consecutive hospitalisations, are Weibull differently distributed within each hidden cluster. Therefore, the comprehensive distribution for each variable is modeled by a Weibull Mixture. From this assumption we developed a survival analysis in order to estimate, through a proportional hazards model, the corresponding hazard function for the proposed model and to obtain jointly the desired clusters.
We find that the selected dataset, a good representative of the complete population, can be categorized into three clusters, corresponding to "healthy", "sick" and "terminally ill" patients. Furthermore, we attempt a reconstruction of the patient-specific hazard function, adding a frailty parameter to the considered model.

C45.5
Smooth non-parametric estimation of the cumulative incidence functions for arbitrarily censored data

A Nguyen Duc[1], M Wolbers[1,2]
[1]Oxford University Clinical Research Unit, Ho Chi Minh City, Viet Nam, [2]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

The cumulative incidence function (CIF) describes the absolute risk of a specific event type over time and is a fundamental quantity to appropriately describe and analyze competing risks data. The most popular CIF estimator is the nonparametric Aalen-Johansen estimator which produces a step function. However, a smooth function might be a more realistic approximation to the truth for many applications and nonparametric approaches have nonstandard asymptotic properties under interval censoring even in the survival setting. In contrast, parametric models rely on restrictive distributional assumptions.
We introduce a novel flexible competing risks model which produces smooth CIF estimates for data with arbitrary censoring and truncation while relaxing the parametric assumptions. Our model is based on a mixture factorization of the joint distribution of the time (T) and type (D) of an event and the conditional distributions T|D are modeled using "smooth non-parametric densities" (SNPD), i.e. truncated (sieve) Hermite series expansions with an adaptive choice of the degree of truncation. Of note, SNPD have previously been successfully applied to econometrics and survival models.
An algorithm for fitting our models will be outlined and simulations presented which show that in many scenarios, our CIF estimator has lower integrated mean squared error compared to both nonparametric and parametric estimators. We will also present the application of our method to an interval-censored dataset of the time to fungal clearance (favorable event) or death (competing unfavorable event) for patients with cryptococcal meningitis. Finally, we will discuss extensions of our approach to regression modeling.

C46 Multiple imputation

C46.1
A multi-stage multiple imputation in a large-scale cohort study

K Furukawa[1], I Takahashi[1]
[1]Radiation Effects Research Foundation, Hiroshima, Japan

Multiple imputation (MI) has been recognized as a flexible and general approach to analysis involving missing data. Practically, however, it is unclear when and how data should be imputed in a project having data missing on variable(s) to be used in several analyses targeted for different subsets of the study subjects. To ensure MI to be valid, imputing for each individual analysis is ideal, but if the analysis is targeted on a small subset of the subjects, we may lose information that is potentially available on the rest of the subjects to increase precision of imputation.
This study proposes an alternative approach to impute data by estimating the imputation model in multiple stages to improve the efficiency of the main analysis while keeping the consistency. Suppose that our main interest is evaluating the effect of a covariate X on the outcome Y, where X is available for the entire cohort (S_1) but subject to missing and Y is measured only on a small subset (S_2) of the cohort. Also suppose that Z is an important predictor for X and available on S_1. We consider imputing data for missing Z from a distribution $f[X|Y,Z] \propto f_1[X|Z]f_2[Y|X,Z]$ where f_1 is estimated with S_1 and f_2 with S_2.
We apply this approach to analysis of cardiovascular disease incidence among a clinical subset of the Life Span Study cohort of more than100,000 Japanese atomic-bomb survivors, for whom radiation dosimetry and basic demographic factors are mostly available but lifestyle factors such as smoking habits are substantially missing.

C46.2
Sequential imputation for large epidemiological data sets

NS Erler[1], J van Rosmalen[1], ETM Leermakers[1], VWV Jaddoe[1], OH Franco[1], EMEH Lesaffre[1,2]
[1]Erasmus Medical Center, Rotterdam, The Netherlands, [2]L-Biostat, KU Leuven, Leuven, Belgium

Missing data are a challenge in cohort studies. An established procedure dealing with this is multiple imputation, as for instance with MICE (Van Buuren, 2012). After creating multiple imputed data sets, each can be analysed using standard software. Derived estimates are then pooled to obtain overall results. Multiple imputation for classical models is available in standard software, however, for more complex models the pooling procedure is not always implemented and difficult to compute manually.
For complex models, Bayesian methods may offer a solution. By specifying parametric distributions for the variables with missings, (e.g. a sequence of univariate conditional distributions as suggested by Ibrahim et al. (2002)), they can be imputed within the same MCMC-procedure used to estimate the model of interest, rendering pooling unnecessary. Furthermore, the Bayesian approach is theoretically justified and allows for a wide range of estimation models. However, Bayesian methods are often computationally intensive and may have convergence issues.
In our study, we evaluate how practical this sequential Bayesian imputation is in the context of epidemiologic questions that require analyses of large data sets with a high rate of missing values. We compare this procedure with multiple imputation with regards to 1) ease of implementation (using R and JAGS), 2) computational time, 3) robustness to modeling choices, and 4) the resulting estimates. To illustrate the method, we analyse the effect of sugar-sweetened beverage consumption on BMI trajectories in young children with a linear mixed model in data obtained from the Generation R Study at Erasmus MC, Rotterdam.

C46.3
Comparison of methods for imputing limited-range variables: a simulation study
L Rodwell[1,2], KJ Lee[1,2], H Romaniuk[1,2], JB Carlin[1,2]
[1]Murdoch Children's Research Institute, Melbourne, Australia, [2]The University of Melbourne, Melbourne, Australia

Multiple imputation (MI) was developed to enable valid inferences in the presence of missing data rather than to re-create the missing values. Within the applied setting, it remains unclear how important it is that imputed values should be plausible. One variable for which MI may lead to implausible values is a limited-range variable, where imputed values may fall outside the observable range. The aim of this work was to compare methods for imputing limited-range variables.
We consider three variables, based on different scoring methods of the General Health Questionnaire (GHQ). These variables resulted in three continuous distributions with mild, moderate and severe positive skewness. In an otherwise complete dataset, we set 33% of the GHQ observations to missing at random; creating 1000 datasets with incomplete data.
We imputed values on the raw scale and following transformation using: regression with no rounding; post-imputation rounding; truncated normal regression; and predictive mean matching. We estimated the marginal mean of the GHQ and the association between the GHQ and a fully observed binary outcome, comparing the results with complete data statistics.
Imputation with no rounding performed well when applied to the raw scale data. Post-imputation rounding and truncated normal regression produced higher marginal means for data with a moderate or severe skew. Predictive mean matching produced under-coverage of the complete data estimate. For the association, all methods produced similar estimates. For highly skewed limited-range data, MI techniques that restrict the range of imputed values can result in biased estimates for the marginal mean.

C46.4
Validation of prediction models based on lasso regression with multiply imputed data
JZ Musoro[1], AH Zwinderman[1], MA Puhan[2], G ter Riet[1], RB Geskus[1]
[1]Academic Medical Center of Amsterdam, Amsterdam, The Netherlands, [2]Institute for Social and Preventive Medicine, Zurich, Switzerland

Background: In prognostic studies, the lasso technique is attractive since it improves the quality of predictions by shrinking regression coefficients, compared to predictions based on a model fitted via unpenalized maximum likelihood. Since some coefficients are set to zero, parsimony is achieved as well. It is unclear whether the performance of a model fitted using the lasso still shows some optimism. Bootstrap methods have been advocated to quantify optimism and generalize model performance to new subjects. It is unclear how resampling should be performed in the presence of multiply imputed data.
Method: The study data were based on a cohort of Chronic Obstructive Pulmonary Disease (COPD) patients. We constructed models to predict Chronic Respiratory Questionnaire (CRQ) dyspnea 6 months ahead. We investigated optimism of the lasso model, and compared three approaches of handling multiply imputed data in the bootstrap procedure, using the study data and simulated data sets.
Results: The discriminative model performance of the lasso was optimistic. There was suboptimal calibration due to over-shrinkage. The estimate of optimism was sensitive to the choice of handling imputed data in the bootstrap resampling procedure.
Conclusion: Performance of prognostic models constructed using the lasso technique can be optimistic as well. Resampling in the presence of multiply imputed data should be performed such that a bootstrap sample selects the same subjects across the imputed data sets, which should differ solely by the imputed values, not by the individuals.

C46.5
Impact of incomplete follow-up when exploring associations between baseline characteristics and outcome in a longitudinal study
S Crichton[1], C Wolfe[1], J Peacock[1]
[1]King's College London, London, United Kingdom

Aim: To assess the impact of missing data when identifying predictors of poor outcome after stroke.
Methods: Data were extracted from South London Stroke Register (N=3617) which collects data at onset, 3 months and annually after stroke. Outcomes are assessed using the Barthel index (categorised as independent, mildly, moderately or severely disabled), Frenchay Activities Index (active, slightly active, inactive) and the Hospital Anxiety and Depression scale. Follow-up rates are typically 60-70%. Models, with varying missing data assumptions, were applied to explore relationships between baseline characteristics and outcomes up to 5 years after stroke. These were Generalised Estimating Equations (GEEs) (assuming missing completely at random data), weighted GEEs (WGEE), GEE combined with multiple imputation (MI-GEE), and multi-level mixed-effects models (all assume missing at random). GEE and mixed-effect estimates were compared to appropriate shared parameter and pattern mixture models, which allow for missing not at random data. All models for binary outcomes were logistic, proportional odds models used for activity level and multinomial for disability level (as proportional odds assumptions were violated).
Results: In univariable and multivariable models for anxiety and depression the same factors were consistently identified as significant. Population averaged effect sizes were comparable across models estimated using GEE's as were subject specific effect sizes from mixed-effects models. GEE, WGEE and MI-GEE models for disability and activity level produced similar results. Findings from other disability and activity models will also be compared.
Conclusions: Missing data appears to have limited impact when looking at associations between baseline and post-stroke outcomes.

C47 Special types of censored data

C47.1
Analysing disease recurrence with missing at risk information
M Pohar Perme[1], T Štupnik[2]
[1]University of Ljubljana, Ljubljana, Slovenia, [2]University Clinical Center, Ljubljana, Slovenia

When analysing time to disease recurrence, we sometimes stumble over data where we are certain that we have all information on recurrence, but do not know whether the studied patients are still alive. This may happen with diseases of benign nature where patients are only seen at recurrences or in poorly designed national registries with insufficient patient identifiers to obtain their dead/alive status. When the average time to disease recurrence is long enough in comparison to the expected survival of the patients, the statistical analysis of such data may be significantly biased.
Under the assumption that the expected survival of an individual is not influenced by the disease itself, we try to reduce this bias by using the general population mortality tables.
We show why the intuitive solution of simply censoring the patients with their expected survival time does not give unbiased estimates and provide an alternative framework that allows for unbiased estimation of the usual quantities of interest in survival analysis. Our results are supported by simulations and real data examples.

C47.2
Weibull regression for a right-censored endpoint with one censored and an arbitrary number of non-censored covariates

K Rufibach[1], S Hubeaux[2]
[1]*Biostatistics Oncology, F. Hoffmann-La Roche, Basel, Switzerland,* [2]*École polytechnique fédérale de Lausanne, Lausanne, Switzerland*

Biomarker data is often subject to limits of quantification or limits of detection. Statistically, this corresponds to left- or interval-censoring. In applications, e.g. when a biomarker is a covariate in a regression model, such data is often imputed in some way, e.g. by considering the limit of detection an actual measurement. In order to be able to correctly account for the nature of the data, we have implemented maximum likelihood estimation in Weibull regression for a right-censored endpoint, one interval-censored, and an arbitrary number of non-censored covariates.
We discuss the assumptions made in the model and how to set up the likelihood function and maximize it. Inference for estimated parameters can be received using standard maximum likelihood theory. We quantify the bias and mean-squared error for parameter estimates compared to commonly used imputation methods.
We illustrate the methodology by applying it to assess Prentice´s criteria for surrogacy in data simulated from a randomized clinical registration trial. The software is available on CRAN, as the package SurvRegCensCov.

C47.3
Accelerated failure time model with interval censored data and cure

S Scolas[1], C Legrand[1], A El Ghouch[1]
[1]*Université Catholique de Louvain-La-Neuve, Louvain-la-Neuve, Belgium*

Mild cognitive impairment (MCI) may be a precursor of Alzheimer disease or other dementia. Studying the time until conversion to MCI makes use of survival analysis theory. Generally, within this field, it is assumed that if the follow-up time is long enough, then the event of interest will be observed for each individual. In our case, not everybody will show signs of impairment. We then say that a proportion of the population is "cured", or "long-term survivor".
Also, patients come to scheduled interviews and thus we can only detect MCI to have appeared between two visits. That is, the database contains interval censored data. Thus, we propose to extend the existing survival models to the case where interval censored data and cure may be present. In this paper, we present the method we want to use: to model event times (i.e. the latency part), we utilize an accelerated failure time (AFT) regression model, adapted to interval censored data, together with an extended generalized gamma (EGG) distribution for the error term of the AFT. In addition, modeling the cure proportion (i.e. the incidence part) is made by a logistic regression.
Furthermore we show the good behavior of the method thanks to results of simulations. Then, we address some issues concerning variable selection in such a model and finally, we apply this method to our Alzheimer disease database, which consist in 241 at-risk patients followed-up between 1998 and 2008 with regular checks for the appearance of MCI.

C47.4
Semiparametric Bayesian frailty model for clustered interval-censored data

A Cetinyurek Yavuz[1], P Lambert[1,2]
[1]*Universite de Liege, Liege, Belgium,* [2]*Universite Catholique de Louvain, Louvain-la-Neuve, Belgium*

Recently, there has been an increasing interest in statistical analysis of interval-censored time-to-event data. This type of data is quite usual for clinical trials or longitudinal studies especially in practical settings of AIDS and cancer research where the individuals have pre-scheduled visits but the event of interest occurs between the visits. Then, the event times are not known exactly but rather to lie in an interval of time.
Moreover, in clinical trials, the units may be collected in clusters and they share some observed or unobserved characteristics, i.e. patients from multiple centres, teeth of multiple subjects; and hence they tend to be correlated. Interval-censored data is a natural generalization of right censored time-to-event data for which a large number of statistical techniques are developed. However, less well developed procedures are available for analysing interval-censored data.
Here, we propose a semiparametric Bayesian frailty model for analyzing correlated interval censored data. We discuss parametric specifications for frailty distribution in the analysis of such data. Afterwards we call particular attention to nonparametric specification of the frailty distribution.
The results of the simulation study suggest that the proposed approach is robust to misspecification of the frailty distribution. Moreover, the performance of the proposed methodology is quite good in practical situations where the frailty distribution is multimodal or skewed. The approach is applied to dental data arising from the Signal Tandmobiel Study.

C48 Drug development

C48.1
Bayesian response-adaptive design development: practical experiences from the DexFEM trial

CJ Weir[1,2], CH Hansen[3,4], P Warner[1], HOD Critchley[1]
[1]*University of Edinburgh, Edinburgh, United Kingdom,* [2]*Edinburgh Health Services Research Unit, Edinburgh, United Kingdom,* [3]*Mwanza Intervention Trials Unit, Mwanza, Tanzania, United Republic of,* [4]*London School of Hygiene and Tropical Medicine, London, United Kingdom*

Background: Heavy menstrual bleeding (HMB) is common but non-surgical treatments are often judged ineffective by women. We developed DexFEM, a UK MRC-funded Bayesian response-adaptive parallel group trial to investigate whether oral dexamethasone reduces HMB and to identify its optimal dose.
Methods: We sought a design comparing placebo and several dexamethasone doses, with randomisation probabilities adapting based on outcome data from patients already randomised to maximise learning about the dose-response. Bayesian Normal Dynamic Linear Modelling flexibly accommodated a range of potential shapes of dose-response curve.
Design options considered were: number of doses; proportion assigned to placebo; adaptation criterion; number and timing of adaptations. We assessed design performance across plausible scenarios for: dose-response curve shape; treatment effect magnitude; outcome variance; recruitment rate; interaction and heteroscedasticity. A fractional factorial simulation study used SAS for data handling, generating scripts and executing analysis in WinBUGS. 200 trials were simulated for each of 150 scenarios. Normal linear modelling estimated the effect of each design option on empirical type I error and statistical power.

Results: The design chosen was: 100 patients randomised to one of 6 dexamethasone doses or placebo; 28% allocated to placebo; 5 evenly spaced adaptations; adaptation criterion based on precision of estimated response at the ED95 (the minimum dose with near-maximal efficacy). Averaged across scenarios, this design gave statistical power of 93.8% (95% confidence interval 91.9%, 95.8%).
Conclusion: Adaptive designs offer flexibility and efficiency. Our integrated approach, using SAS to control simulations by executing analysis in WinBUGS is a practical tool for their development.

C48.2
Sample size optimization for phase II/III drug development programs

M Kirchner[1], M Kieser[1], H Götte[2], A Schüler[2]
[1]Institute of Medical Biometry and Informatics, Heidelberg, Germany, [2]Merck KGaA, Darmstadt, Germany

About 50% of development programs in phase III do not get regulatory approval (Arrowsmith, 2011). Usually, sample size of phase III trials is based on the treatment effect estimated from phase II data. As the true treatment effect is uncertain, a high intended statistical power for the phase III trial does not necessarily translate into a high success probability. Hence, the variability of the estimate has to be considered when sizing a study.
However, there is still a lack of methodology for sample size calculation across a phase II/III program including go/no-go decisions after phase II.
We investigate the impact of the uncertainty about the treatment effect estimate obtained from phase II trials on the sample size of subsequent phase III studies. Success probabilities of the complete phase II/III program are evaluated under consideration of the program-wise sample size. In order to optimize program-wise planning, utility as a function of phase II sample size is calculated for different scenarios. It is demonstrated that the go/no-go decision after phase II and the size of phase II trials strongly influence the distribution of the phase III sample size and the utility. Recommendations are given concerning an adequate choice of phase II sample size taking these aspects into account.
In summary, the presented methods for program-wise combined planning of phase II and III trials may help to improve the calculation of the sample size for phase II and phase III trials under the aim of reaching high success probabilities.

C48.3
Bayesian meta-analytical methods to incorporate multiple surrogate endpoints in drug development process

S Bujkiewicz[1], JR Thompson[1], KR Abrams[1]
[1]University of Leicester, Leicester, United Kingdom

Surrogate endpoints are increasingly being investigated as candidate endpoints in drug development process where measuring a primary outcome of interest may be too costly, too difficult to measure or require long follow-up time. A number of meta-analytical methods have been proposed that aim to evaluate surrogate endpoints as predictors of the target outcome. Bivariate meta-analytical methods can be used to predict the target outcome from the surrogate endpoint (while taking into account of the uncertainty around the surrogate outcome) as well as to combine evidence on both outcomes to "borrow strength" across outcomes when evaluating new health technologies.
Extensions to multivariate models will be discussed aiming to include multiple surrogate endpoints with a potential benefit of increasing precision of predictions. In our recent paper on Bayesian multivariate meta-analysis of mixed outcomes we model the between-study covariance in a formulation of a product of normal univariate distributions (Stat Med 2013; 32:3926-3943). This formulation is particularly convenient for including multiple surrogate outcomes. In this model however, two outcomes (which can be surrogate endpoints to the target outcome) are conditionally independent, conditional on the target outcome. Building on this model, we extend it to the case where this assumption is relaxed to allow for one of the surrogate endpoints to act as a surrogate to the other. The modelling techniques are investigated using example from multiple sclerosis (where the disability worsening is the target outcome, while relapse rate and MRI lesions have been shown to be good surrogates to the disability progression).

C48.4
Sequential meta-analyses of safety data

D Saure[1], K Jensen[1], M Kieser[1]
[1]Institute of Medical Biometry and Informatics Heidelberg, Heidelberg, Germany

While meta-analyses investigating the efficacy of therapies are mainly conducted retrospectively, there is a need for a prospective sequential approach for the assessment of safety data over several studies within a drug development program. Currently available methods for sequential meta-analyses, for example the procedure based on the combination of p-values (Jennison and Turnbull, JBiopharmStat 2005) or the repeated cumulative meta-analysis approach (Whitehead, StatMed 1997), are tailored to superiority trials. However, in the analysis of safety data including serious adverse events with low rates one is usually interested in demonstrating non-inferiority. We demonstrate the need for sequential methods in this situation and examine the applicability of the above mentioned approaches for different scenarios which are typical for drug development.
Our focus lies on fixed-effect meta-analyses with binary outcomes where we incorporate different effect measures and pooling methods. We calculate the exact type I error rate and the exact power for various situations occurring in a sequential approach of safety data. Various scenarios for event rates and non-inferiority margins are considered. The methods proposed by Jennison and Turnbull assume that the so called "p-clud" property of the p-values (Brannath et al., JAmStatAssoc 2002) holds true. We investigate whether this assumption is fulfilled in the current situation of sparse binary data and non-inferiority trials and, furthermore, in case of non-fulfillment of the "p-clud" condition, we examine the performance of those p-values within those methods.

C48.5
On the three-arm non-inferiority design including a placebo

T Tango[1,2], E Hida[1]
[1]Center for Medical Statistics, Tokyo, Japan, [2]Teikyo University Graduate School of Public Health, Tokyo, Japan

The design and the analysis of three-arm non-inferiority trials seems to have been focused on the fraction approach (e.g., Kock and Tangen 1999, Pigeot et al.,2003; Kock and Röhmel, 2004), which aim to show that the experimental treatment preserve a prescpecified fraction *f* of the active control treatment effect to placebo. The fraction approach has been modified and/or extended to several situations.
However, in many "common" two-arm non-inferiority trials conducted so far over the world, the non-inferiority margin Δ has been defined as a prespecified difference of treatments. So, we proposed a method with Δ for inference of the difference in means (Hida and Tango, 2011) and in proportions (Hida and Tango, 2013), in which we have to show the following inequality: $\theta_P < \theta_R - \Delta < \theta_E$ where θ_P, θ_R , θ_E denote the expected value of treatment outcome under the placebo, reference and experimental treatment, respectively. The first inequality implies the requirements for assay sensitivity that the superiority of the reference over the placebo should be more than Δ . To this substantial superiority condition, Röhmel and Pigeot(2011) and Stucke and Kieser(2012) expressed their concern. Kwong et al.(2012), on the other hand, stand against the fraction approach, but

revised our approach in an unrealisitc direction. Schlomer and Brannath (2013) proposed group sequential designs based on the fraction approach although they acknowledge some debate in the literature.
In this presentation, we shall discuss statistically sound procedures by clarifying the definition of assay sensitivity in three-arm non-inferiority trials.

Thursday 28 August 2014 – 9:00-12:30

Minisymposia

M1 Statistical challenges in the epidemiology of aging

Organizers: Carole Dufouil and Karen Leffondré

M1.1
Methodological challenges in the epidemiology of aging from a reproducible research perspective

SM Hofer[1]
[1]*University of Victoria, Victoria, Canada*

The analysis of longitudinal observational data can take many forms and requires many decisions, with research findings and conclusions often found to differ across independent longitudinal studies addressing the same question. Differences in measurements, sample composition (e.g., age, cohort, country/culture), and statistical models (e.g., change/time function, covariate set, centering, treatment of incomplete data) can affect the replicability of results. The central aim of the MELODEM Initiative and the Integrative Analysis of Longitudinal Studies of Aging (IALSA) research network (NIH/NIA P01AG043362) is to optimize opportunities for replication and cross-validation across heterogeneous sources of longitudinal data by evaluating comparable conceptual and statistical models at the construct-level. I will provide an overview of the methodological challenges associated with comparative longitudinal research, including the comparability of alternative models of change, measurement harmonization and construct-level comparison, retest effects, distinguishing and contrasting between-person and within-person effects across studies, and evaluation of alternative models for change over time. These methodological challenges will be discussed within the context of reproducible research on aging-related outcomes.

M1.2
Survival analysis aspects of the epidemiology of ageing

N Keiding[1]
[1]*University of Copenhagen, Copenhagen, Denmark*

Measurement of indicators of health and abilities is often restricted to observations at long intervals, and for the elderly necessarily further complicated by the likely possibility that the individuals die. There is a lively debate on the best ways to define targets of inference and associated methods of analysis under this *truncation by death* sampling pattern.
This talk will outline the positive body of concepts and tools from within survival and event history analysis which may be helpful in meeting these challenges, focusing on competing risks analysis with its descriptors cause-specific hazard rates and cumulative incidences, and on illness-death models, where an important issue is that of time origin: is survival measured from birth and/or first occurrence of disease? Disease *prevalence* is naturally formalized in illness-death models. Since disease status is often recorded intermittently, estimation may need to handle interval-censored data. A main theme is to be precise about the study base, that is, the population about which one wants to make statements.

M1.3
Modelling issues in the longitudinal study of cognitive aging

C Proust-Lima[1,2]
[1]*INSERM U897, Bordeaux, France,* [2]*University of Bordeaux, ISPED, Bordeaux, France*

In cognitive aging studies, there is a growing interest in the description of change over time of cognitive functions and the evaluation of risk factors of cognitive change. Indeed, as dementia is characterized by a progressive and continuous decline of cognitive functions, these longitudinal analyses better capture the dynamics of disease progression than survival analyses describing time-to-progression. However, the study of cognitive change entails several modeling issues.
First, markers of progression are psychometric tests with usually limited metric properties (ceiling/floor effects, curvilinearity) that translate in asymmetric distributions. To avoid large biases and misleading conclusions induced by these properties, mixed models adapted to psychometric data can be preferred to the more standard linear mixed model.
Second, multiple psychometric tests are usually collected and the interest is not directly in the change of one specific test but in the change of the underlying cognition that generated them. Latent process approaches that focus on the dynamics of a latent trait underlying a set of longitudinal outcomes are designed to address such multivariate and longitudinal data.
Finally, cognitive change may be associated with clinical events, mainly dementia and death. These informative events potentially bias the cognitive change estimates when not taken into account and bring essential information for understanding the natural history of cognitive aging. Joint models of longitudinal outcomes and times-to-events account for this informative dropout.
In this presentation, these statistical issues are gradually addressed and illustrated using data of a large prospective cohort study (PAQUID) on cerebral aging with a 22-year follow-up.

M1.4
Medical and conceptual challenges in conducting studies of the elderly

MD Koeller[1]
[1]*Medical University of Vienna, Vienna, Austria*

The objective of this talk is to identify conceptual challenges in medical research with older persons that should be addressed in the future work of biostatisticians cooperating and supporting geriatricians. Due to the diversity of medical problems and conditions in elderly patients clinical research studies require a specific design and analytic approach. These requirements are based on the multifactorial etiologies of geriatric syndromes and multimorbidity at the end of life. Moreover, manifold interventions, multiple outcomes, ceiling effects, missing data, or different methods are factors which have to be taken into account in planning studies in geriatric cohorts. Study design of multicomponent interventions in older persons is complicated. With respect to medical risk factors, participants often cannot be randomly assigned to all possible interventions. Under ideal study conditions, all the participants are retained with complete data. But in geriatric patients functional assessments, for instance, are frequently limited and inappropriate, causing semicontinuous data and consequently floor or ceiling effects may result. Coincidences, bias, or ambiguity are potential threats to successful interpretation of results with good and clear conclusions. Therefore, interdisciplinary thinking and the cooperation of researchers from gerontology and geriatrics, as well as biology, and the statistical discipline will have to be enhanced in the future.

M2 Genomics-based Personalized Medicine

Organizers: Andreas Ziegler and Georg Heinze

M2.1
Application of genomic tests in breast cancer management

M Filipits[1]
[1]*Medical University of Vienna, Institute of Cancer Research, Vienna, Austria*

Breast cancer is a heterogeneous disease at the clinical, biological and particularly at the molecular level. Gene expression profiling has improved the knowledge on the complex molecular background of this disease and allows a more accurate prognostication and patient stratification for therapy. Several genomic tests have been developed with the aim of improving prognostic information beyond that provided by classic clinicopathologic parameters. Some of these tests are currently available in the clinic and are used to determine prognosis and more importantly to assist in determining the optimal treatment in patients with hormone receptor-positive breast cancer.
Available data suggest that information generated from genomic tests has resulted in a change in decision making in approximately 25%-30% of cases. The clinical relevance of genomic tests and their ability to define prognosis and determine treatment benefit will be discussed.

M2.2
Risk prediction models using family and genomic data

JE Bailey-Wilson[1]
[1]*National Human Genome Research Institute, NIH, Baltimore, United States*

Advances in our ability to model personal risk of developing a disease have accelerated as large epidemiologic and genomic studies have increased our understanding of disease causation. Prediction of disease risk can be based on personal history of environmental exposures, family history of disease and personal genotypes at genetic susceptibility loci. Approaches to predicting risk of disease that utilize familial and genetic information will be discussed for a range of different causal models from simple Mendelian disorders that are caused by variants in a single gene to diseases caused by complex actions of multiple risk factors. The utility of adding family history and personal genotypes into disease risk models will be covered.
Accurate disease risk prediction can be important to individual health since it can encourage individuals to have more frequent screening procedures, to undertake environmental risk reduction, and to undergo preventive medical procedures and treatments.

M2.3
The importance of appropriate quality control in -omics studies as required for personalized and stratified medicine

B Müller-Myhsok[1,2,3]
[1]*Max Planck Institute of Psychiatry, Munich, Germany,* [2]*Munich Cluster for Systems Neurology (SyNergy), Munich, Germany,* [3]*Institute for Translational Medicine, University of Liverpool, Liverpool, United Kingdom*

Both personalized and stratified medicine are an important avenue for research at present and more likely so even in the future. The importance of deriving good predictors usually necessitates incorporating data from various -omics sources into the model, which has implications for the quality control of the corresponding data sets.
I will discuss some of these, including the need to very carefully take into account undesirable structure in the data and differing reliability of data from different data sets.
I will also show how some of the demonstrated procedures may as a consequence lead to better understanding and predictors.

M2.4
Study designs for predictive biomarkers

A Ziegler[1,2]
[1]*U Lübeck, Institute of Medical Biometry and Statistics, Lübeck, Germany,* [2]*U Lübeck, Center for Clinical Trials, Lübeck, Germany*

Biomarkers are of increasing importance for personalized medicine, including diagnosis, prognosis and targeted therapy of a patient. Examples are provided for current use of biomarkers in applications.
It is shown that their use is extremely diverse, and it varies from pharmacodynamics to treatment monitoring. The particular features of biomarkers are discussed. Before biomarkers are used in clinical routine, several phases of research need to be successfully passed, and important aspects of these phases are considered. Some biomarkers are intended to predict the likely response of a patient to a treatment in terms of efficacy and/or safety, and these biomarkers are termed predictive biomarkers or, more generally, companion diagnostic tests. Using examples from the literature, different clinical trial designs are introduced for these biomarkers, and their pros and cons are discussed in detail.

Conference course

Course 7

Designing adaptive clinical trials

Y Jemiai[1], R Mukherjee[2]
[1]Cytel Inc., Cambridge, United States, [2]Cytel Inc., Geneva, Switzerland

The objective of this course is to provide practical strategies and tools for efficient decision-making via interim analyses of ongoing clinical trials, using state-of-the-art methods for group sequential and adaptive designs.
Topics covered will include group sequential design and monitoring with sample size re-estimation, preserving type-1 error, computing power, obtaining point estimates, and computing confidence intervals in the adaptive setting. We will also cover modern methods for model-based dose-escalation in Phase 1 oncology studies.
Case studies in oncology and cardiology are used to reinforce the main points. The workshop includes a hands-on session with the East 6.3 software.
Firsthand experience with East will be used throughout as a running example to illustrate concepts.

Monday, 25th August 2014 - 15:30-16:00

Poster session P1

P1.1 Bayesian methods in biostatistics

P1.1.1

Coverage properties of Bayesian 95% probability intervals for odds ratio and relative risk

S Aghlmandi[1], M Zwahlen[1]
[1]Institute of Social and Preventive Medicine (ISPM), Bern, Switzerland

Fagerland & Newcombe[1] recently presented different methods to calculate 95% confidence intervals for the odds ratio (OR) and relative risk (RR) and advocated the use of the inverse hyperbolic sine interval transformation while adding different pseudo-frequencies. We assessed the coverage probability of Bayesian 95% probability intervals for the same situation using the 2.5% and 97.5% quantiles of the posterior distribution. We defined p_i as the probability of an event occuring in group i (i=1, 2), n_{i+} the total number of patients and n_{i1} the number of events in group i.
For the Bayesian analysis, we used beta priors for p_i which allow a beta-Binomial conjugate analysis for the p_i.
We used Monte-Carlo simulations to generate a 10^5 sample from the posterior distributions of p_1 and p_2 to then approximate the posterior distribution of the OR and RR. We used the same simulation scenarios as Fagerland & Newcombe[1]: $n_{1+}=n_{2+}=20$, true OR=2.41, RR=1.43, p_1 ranging from 0.01 to 0.99 in steps of 0.01 with 5000 data sets created for each p_1. Finally, we compared the coverage probability of Bayesian 95% probability intervals with the coverage Fagerland & Newcombe[1] obtained with their "best" methods, the inverse sinh formula with and without pseudo-frequencies. Results for the OR show that the Bayesian 95% probability intervals have a mean coverage near 95%, very similar to Fagerland's best results. Both approaches have a coverage above 95% when p_1 <10% or >90%. We obtained similar results for RR.
Reference: 1- Fagerland & Newcombe, Statistics in Medicine 2013;32:2823-2836.

P1.1.17

Comparison of two methods for futility analysis in vaccine efficacy trials

FP Bailleux[1], E Bassily[2], AJ Dunning[2]
[1]Sanofi Pasteur, Marcy l'Étoile, France, [2]Sanofi Pasteur, Swiftwater, United States

Aims: Vaccine efficacy (VE) trials involve a large number of subjects and considerable costs/resources. If the VE is not promising it may be beneficial to stop the trial.
Methods: The statistic to evaluate the VE in the interim or final analyses is derived from the number of cases in vaccine group, which follows a binomial distribution conditional on the total number of cases.
Two methods for futility analysis were evaluated.
The first method tests the null hypothesis that the VE is not greater than a predefined bound. This test is evaluated using the upper bound of confidence interval of VE at a fixed alpha level (independent of the alpha level used at the different interim efficacy analyses).
The second method is based on the Bayesian predictive power. At each futility analysis the Bayesian predictive probability to conclude at the end of trial is calculated, if this probability is lower than a cut-off then the trial stopped for futility.
Results: Methods are compared using different trials design with various theoretical vaccine efficacies, lower bounds and with a limited number of futility/interim analyses.
While the Bayesian predictive power permits stopping earlier in case of poor VE, the method based on the upper bound of CI of VE permits better control of the risk of stopping for futility in case of good VE.
Conclusion: Various methods exist to perform futility analyses in vaccine VE. Choice of the method is crucial to insure an optimization of the risk the sponsor will want to control.

P1.1.20

Choosing a gold standard: support of Bayesian inference methods for diagnostic accuracy of new biomarkers in pediatric urinary tract infection

S Bastide[1,2], P Landais[1,2], S Leroy[1,2]
[1]Department of Biostatistics and Epidemiology, Nîmes Hospital, Nîmes, France, [2]EA2415 Biostatistics Research Unit, Montpellier 1 University, Montpellier, France

Background: Acute pyelonephritis (APN - kidney infection) is a common pediatric bacterial infection. Biomarkers-based strategies (e.g. procalcitonin) aimed at promptly diagnosing APN, and were compared with DMSA scan, whose gold-standard quality raises concerns. We used for the first time Bayesian methods to estimate the diagnostic accuracy of procalcitonin and DMSA.
Methods: We used a Bayesian approach to explore disease prevalence and tests properties, using an independent model and both fixed and random effects models with conditional dependence between tests. Two levels of prior distribution were defined: one informative obtained from a published meta-analysis of individual patient data (1011 patients, 61% APN) and pediatrician beliefs for DMSA, and one non-informative. Standard procedures were used to achieve MCMC convergence, for model checking and for a sensitivity analysis.
Results: The fixed model yielded for procalcitonin: Se 74% [71-77], Sp 70% [66-74], and for DMSA: Se 94% [87-98], Sp 90% [80-97] with informative prior; whereas, with the non-informative prior, it achieved: Se 72% [59-90], Sp 75% [53-94] for procalcitonin, and for DMSA: Se 77% [64-92], Sp 74% [50-94] . Given the important amount of the additional information contained in prior samples, the non-informative prior seemed sounder. The same discordance between the priors was similarly observed with the independent model. A random effect model was completed to further explore this result.
Conclusion: A Bayesian approach allowed showing that the gold-standard test for APN, DMSA, was not perfect despite clinical beliefs. Support of Bayesian inference methods for diagnostic accuracy of new biomarkers should be fostered.

P1.1.21

Methodological review of Bayesian inference methods used in clinical decision rules and diagnosis studies

S Bastide[1,2], P Landais[1,2], S Leroy[1,2]
[1]Department of Biostatistics and Epidemiology, Nîmes Hospital, Nîmes, France, [2]EA2415 Biostatistics Research Unit, Montpellier 1 University, Montpellier, France

Background: In many clinical conditions, the existence of a gold standard is often disputable, or not available for all patients. The introduction of Bayesian methods appeared valuable in situations where other approaches failed. We aimed at studying the use of Bayesian inference methods dedicated to diagnosis decision making based on clinical decision rules (CDR) and diagnosis studies.
Methods: We searched for all articles using Bayesian methods for diagnosis and/or CDR studies in electronic databases using key words in MEDLINE and citations search of the cornerstone papers in ISI Web of Science. Eligibility was assessed on title/abstract, and finally inclusion on full-text.

We extracted the Bayesian inference methods characteristics (type, aim, justification, etc.), involvement of a statistician, clinical domain, and the study quality according to STARD and SAMPL guidelines.
Results: Among the 1666 identified articles, 281 were eligible, and 88 included. 44 (50%) papers were new methodological development articles, 29 (33%) clinical articles, 15 (17%) were intermediate between methodological development and clinical studies, and mostly re-analysis of clinical datasets using Bayesian methods. Articles were classified based on an arborescence assessing the evolution of the methods and their use over time. Among the intermediate and clinical groups, more than 68% dealt with biological and microbiological concerns; and only 10 papers concerned practical clinical issues.
Conclusion: Although there has been an increasing usage of Bayesian inference methods in the last few years with many methodological developments, their use remains limited in practice. The spread of Bayesian methods for those studies should be fostered.

P1.1.32
A Bayesian hybrid adaptive design for phase III survival trials

M Moatti[1], WF Rosenberger[2], S Chevret[1]
[1]INSERM UMR 1153, Université Paris 7, Paris, France, [2]Department of Statistics, George Mason University, Fairfax, United States

A few papers have described adaptive survival phase III randomised clinical trials in a frequentist context. We extend the method of Zhang and Rosenberger who proposed a response-adaptive randomization procedure that targets an optimal allocation for parametric survival data.
However, while they explored only fixed sample size designs, we incorporate an interim monitoring plan for estimating the log hazard ratio and propose stopping rules. Moreover, the main extension is in the Bayesian context, where we establish for the log hazard ratio a prior distribution based on either Spiegelhalter's skeptical or enthusiastic normal priors, or a normal mixture derived from experts´ opinions. Combining the prior with the normal likelihood, the mean posterior estimate of the log hazard ratio allows deriving the optimal target allocation.
We perform a simulation study to assess and compare the performances of this proposed Bayesian hybrid adaptive design to those of fixed, sequential or adaptive frequentist designs. When using stopping rules, there was a gain in reducing the proportion of observed deaths in adaptive vs. non adaptive designs; this gain was maximal using a Bayes mixture prior.
Such Bayesian hybrid adaptive survival trials may appear promising alternatives to reduce the duration and the costs of survival trials, as well as optimizing the ethical concerns for patients enrolled in the trial.

P1.1.36
Copula functions in the presence of cure fraction

EA Coelho-Barros[1], J Achcar[2], J Mazucheli[3]
[1]UTFPR, Cornélio Procópio, Brazil, [2]USP, Ribeirão Preto, Brazil, [3]UEM, Maringá, Brazil

We introduce bivariate Weibull distributions derived from copula functions in presence of cure fraction, censored data and covariates. Two copula functions are explored: the FGM (Farlie - Gumbel - Morgenstern) copula and the Gumbel copula. Inferences for the proposed models are obtained under the Bayesian approach, using standard MCMC (Markov Chain Monte Carlo) methods. An illustration of the proposed methodology is given considering a medical data set. The use of copula functions could be a good alternative to analyse bivariate lifetime data in presence of censored data, cure fraction and covariates.
Observe that in many applications of lifetime modelling we could have the presence of a cure fraction for individuals that are "long term survivors" or "cured individuals".

P1.1.43
Reliable confidence intervals for fractional polynomials: a simulation study

D Dunkler[1], M Gregorich[1]
[1]Medical University of Vienna, CeMSIIS, Vienna, Austria

Nonlinear associations of continuous risk factors with an outcome of interest can be modeled by fractional polynomials (FPs, Royston & Altman, 1994). Specifically, one or two elements of a polynomial transformation of the original variable are selected such that in subsequent regression analysis an optimal fit is obtained. This selection is usually ignored when constructing confidence intervals (CI) for the expected outcome at different values of the risk factor, or for contrasts between different risk factor values. This can lead to undercoverage if the selection is not stable, or if FPs are not flexible enough to capture a specific underlying shape of nonlinearity. Reliable CI can be obtained by the bootstrap, but this requires a large number of models to be evaluated.
Therefore, we consider some instant methods for CI estimation, such as model-based CIs with three degrees of freedom (MB3DF) to account for selection of two powers, and Bayesian model averaging (BMA) of several evaluated FP models.
Using the setting of logistic regression, we compared these approaches with simple model-based CIs and bootstrapped CIs in a simulation study, assuming various nonlinear associations between a continuous risk factor and a binary outcome.
Our simulations showed that with a true linear association BMA proves satisfactorily, and MB3DF may overcover. With nonlinear associations that can be modeled with FPs, both methods improve over simple model-based CIs without increasing the computational demand. In particular, we recommend the implementation of Bayesian model averaging CIs in software for FP estimation to stimulate their practical use.

P1.1.107
Determination of the minimum effective dose for correlated dose-response data using Bayesian variable selection (BVS) models

LK Muchene[1], M Otava[1], Z Shkedy[1], T Jacobs[2]
[1]Universiteit Hasselt, Center for Statistics, Diepenbeek, Belgium, [2]Janssen Pharmaceutical Companies of Johnson and Johnson, Beerse, Belgium

In drug development, the determination of a minimum effective dose for a compound is of primary interest. Classically this involves testing the difference in means of multiple doses against the mean in the first dose-level (typically, the control group) using analysis of variance with correction for multiple testing. The first dose level for which a significant difference is detected is defined as the minimum effective dose (MED).
Alternatively, Bayesian variable selection (BVS) models can be used for selecting the most probable model for the dose -response relationship given a set of known candidate models. The model with the highest posterior model probability is selected and the MED is determined based on the selected model. Hence, the BVS model allows to estimate the MED taking into account model uncertainty.
We apply Bayesian variable selection techniques to data from a differential reinforcement of low-rate 72 seconds (DRL-72) experiment. The DRL-72 experiment is commonly used in testing for clinically active anti-depressant compounds. In such an experiment, a rat pressed a lever and is expected to wait 72 seconds between two presses in order to receive a reward.
The number of times the rat presses a lever is Poisson distributed and the number of rewards obtained is binomial distributed. The Bayesian variable selection model is applied to the data using joint binomial/Poisson models while correcting for the design of the experiment, i.e. correlated measurements, cross-over drug administration design and over-dispersion in outcomes.
Keywords: DRL-72, Bayesian variable selection models, hierarchical Bayesian models, Minimum effective dose.

P1.1.110
On adapting the sample size in a Bayesian clinical trial in small populations
T Brakenhoff[1], S Nikolakopoulos[1], KCB Roes[1], I van der Tweel[1]
[1]UMC Utrecht, Utrecht, The Netherlands

In the design stage of a clinical trial in small populations there are several methodological challenges. An obvious fundamental obstacle is the limited available number of patients to be included in such a trial. As a consequence, limited information will be available, concerning design parameters, on which sample size calculations can be based.
Thus, framework of sample size re-estimation which is thoroughly studied in the frequentist paradigm, is an obvious alternative. We extend proposed Bayesian sample size estimation methodology to the situation where the sample size can be reevaluated at one or more interim stages. Working with normally distributed outcomes, such an approach handles the scenario where the variance observed in the trial is different from the one anticipated in the design stage.
By using a fully Bayesian predictive approach, our method can handle sample size re-estimation in combination with imposing a maximum sample size as a realistic constraint when conducting research in small populations. The approach is illustrated by reanalyzing data from a real randomized trial in the field of pediatrics.

P1.1.118
Bayesian analysis of parametric frailty models for repeated event data: estimating unreported event times using interval data
RK Owen[1], DG Tincello[1], PC Lambert[1,2], S Bujkiewicz[1], KR Abrams[1]
[1]University of Leicester, Leicester, United Kingdom, [2]Karolinska Institutet, Stockholm, Sweden

Background: We are often interested in analysing the time to recurrent events associated with chronic diseases following repeated treatments, but patient follow-up can be intermittent and event times are frequently unreported. Motivated by a clinical trial in overactive bladder syndrome, we want to analyse the potential diminishing effect of repeated injections of botulinum toxin, adjusting for severity status, on patient reported recurrence of symptoms.
Methods: We used a Bayesian framework to fit a Weibull proportional hazards model for repeated event data to obtain posterior predictive distributions from which to sample unreported event times. To further account for the correlation between repeated events within the same individual, we incorporate a shared frailty term. We applied this methodology to a clinical trial of patients receiving a maximum of 3 repeated injections of botulinum toxin for overactive bladder over a 5 year follow-up period.
Results: Bayesian approaches had an improved fit to the data compared to the frequentist alternative, and including posterior predictions of the missing severity covariate increased precision in the estimates. Patients with severe symptom severity had a considerably higher rate of symptom recurrence (HR: 3.07, 95%CrI: 1.56,7.15) compared to patients with normal severity status. Repeated injections appear to reduce the rate of symptom recurrence in severe patients.
Conclusions: With an increasing need to assess the time to symptom recurrence in chronic conditions, and the difficulties faced with intermittent follow-up, the use of a flexible Bayesian framework would appear to be advantageous.

P1.1.124
Phase II study to assess the safety of bevacizumab with neoadjuvant chemotherapy in ovarian cancer using a Bayesian approach
S Zohar[1], F Joly[2], R Rouzier[3], Y Ghazi[4], V Menguy[4], D Pau[4]
[1]INSERM UMRS 1138, Team 22, Paris, France, [2]INSERM U1086, Centre François Baclesse, Caen, France, [3]Institut Curie, Paris, France, [4]Roche, Boulogne-Billancourt, France

Background: Bayesian analysis is rarely used in randomized phase II clinical trials. In contrast to the frequentist approach, the Bayesian approach has no consideration of Type I error and no power calculation because the inference is based on the posterior or predictive distributions. Due to regulatory constraints, Bayesian analyses are typically only used in addition to the primary frequentist.
Objectives: This randomized phase II study will evaluate the feasibility of using two types of inferences: the safety is evaluated using Bayesian inference and the efficacy is evaluated using frequentist inference. The primary endpoint is to evaluate the benefit of neoadjuvant bevacizumab and chemotherapy assessed by the complete resection at surgery in patients with advanced ovarian cancer.
Methods: For both efficacy and safety endpoints sequential analyses are performed. For the efficacy endpoint a frequentist hypothesis testing is used and for the safety and point a Bayesian approach, based on a beta-binomial model with 3 prior distributions is performed. The parameters of each prior distribution were selected from expert's elicitation. Analyses are performed sequentially but more frequently for the safety endpoint as the investigators wished to stop early the trial if the treatment is estimated to be too toxic (for minimum 2 prior distributions out of three).
Conclusions: Bayesian approach provides flexibility in decision making process regarding the continuation or discontinuation of patient accrual regarding the safety endpoint. As the inference is not influenced by the number of interim analyses it allows stopping the trial as early as needed.

P1.1.130
Some inferential results in branching processes in random environments
M Molina[1], M Mota[1], A Ramos[2]
[1]University of Extremadura, Badajoz, Spain, [2]University of Extremadura, Cáceres, Spain

This work deals with mathematical modelling through branching processes. We are interested in developing stochastic models to describe the demographic dynamics of populations. We focus on the class of branching processes with progenitor couples in a random environment introduced in Molina *et al.* (2012). We provide several results concerning the extinction of the population and, under a parametric and nonparametric setting, we derive estimators for the offspring distribution and for its main moments. In order to determine the corresponding highest posterior density credibility sets, we also propose a computational method. By way of illustration, we include a simulated example in population dynamics.

P1.1.164
Bayesian methods in adaptive dose finding
C Tirodkar[1], S Solanki[1]
[1]Cytel Statistical Software and Services Pvt. Ltd., Pune, India

Biostatistics has seen a phenomenal growth in recent years and has been marked by considerable technical innovation in both methodology and computational practicality. The complexity and amount of data produced by biomedical studies is increasing at a staggering pace every year and is one of the most significant challenges experienced by clinical researchers. Hence, devising strategies to effectively design and obtain useful informa-

tion from such studies becomes crucial for current and future scientific endeavors.
Bayesian statistics provides an intuitive framework firmly grounded on probability theory to design and analyze complex data. Bayesian methods make it possible to incorporate prior information in the analysis and may be applied to problems whose structure is too complex for conventional methods to handle. With the aid of modern computing, the approach provides a flexible formulation to address applied problems realistically, and to incorporate the research goals into the analysis.
This paper/presentation will discuss the latest Bayesian developments in adaptive dose-finding studies. In particular, it will highlight Bayesian techniques used in the early phases of clinical drug development. In early-phase clinical trials, most often the data is scant (small sample sizes) and very little is known about the actual dose-toxicity relationship. This makes for a perfect setting for the use of Bayesian techniques and this paper will demonstrate some of the many practical benefits of adopting the Bayesian paradigm.

P1.2 Design and analysis of clinical trials

P1.2.5
Introducing continuity correction for the Laster-Johnson-Kotler non-inferiority asymptotic test

F Almendra-Arao[1], D Sotres-Ramos[2], Y Castillo-Tzec[2]
[1]UPIITA del Instituto Politécnico Nacional, México, D. F., Mexico, [2]Programa de Estadística, Colegio de Postgraduados, Texcoco, Mexico

Non-inferiority asymptotic statistical tests are frequently used in clinical trials. The "at least as good" criterion was introduced by Laster, Johnson and Kotler, for dichotomous data. In this approach (LJK), the margin of non-inferiority is taken as a percentage of the control response, rather than a fixed difference. The procedure is seen to be more efficient than the fixed margin approach yielding smaller sample sizes. Also, the procedure offers several advantages in the design, statistical efficiency and interpretability of non-inferiority trials.
However, the LJK procedure has the disadvantage that its size is much greater than the required nominal significance level (α). This latter issue is addressed in this work by using two continuity correction factors. The results show that the size of the modified LJK test yields a much better behavior than that of the original LJK test.

P1.2.7
An online calculator for futility interim monitoring rules in randomised clinical trials

A Alvarez-Iglesias[1], P Gunning[1], J Newell[1]
[1]HRB Clinical Research Facility, National University Ireland, Galway, Ireland

Multi-stage or Group sequential methods are common in many branches of scientific research. The aim of these methods is to pre-specify, prior to the start of data collection, the timing and manner in which a sequence of interim analyses will be conducted within the study.
Based on the results of these analyses, a decision about early stopping can be made, without any compromise in power.
For instance, in a Randomised Clinical Trial, one might want to investigate as soon as possible whether patients under a new treatment are being exposed to a harmful pharmaceutical drug, in which case the study should be stopped for futility.
On the contrary, if the new treatment is leading to positive results, early stopping means that patients will benefit earlier from the new treatment.
In this work we review a general approach proposed by Freidlin et al (2010) when there is a need for inefficacy monitoring rules in modelling time to event data. This approach has the advantage that the interim looks can be defined post design, meaning that there is no need for sample size modifications, even after some of the data have already been collected. We show some simulations for different survival distributions and, due to the usefulness of the method, an online calculator will be presented that can easily provide the stopping rules for any superiority design in which two populations are being compared using the log-rank test statistic.

P1.2.18
Best-after-breast design: challenges of nutrition intervention studies in infants

E Balder[1], S Swinkels[1]
[1]Danone Nutricia Research, Utrecht, The Netherlands

Breastfeeding is the preferred and recommended method of infant feeding. If a mother cannot or chooses not to (fully) breastfeed, a commercially prepared infant formula is the recommended alternative. This presentation will address the challenges of nutrition intervention studies in infants, due to practical, ethical and regulatory issues that are inherent to studying feeding regimens in infants. A Best-after-breast design was implemented in Danone Nutricia Research, which allows subjects to enter the study without interfering with the mother's or parents' choice of early nutrition for the infant. After inclusion (≤28 days of age), regardless of the feeding regimen, subject data is recorded. When the mother/parents autonomously decide to start formula feeding, the subject is randomized in a double-blind parallel design to one of the study products. After start with the study formula, the mother is free to continue breast feeding in combination with formula as long as she wants and/or she can switch to full formula feeding in her own pace at any time. The different feeding regimens in the Best-after-breast design may introduce time-varying confounding. To account for this, for each subject for each measurement the state and duration of full breast feeding, mixed feeding (combination of formula with breast milk and/or weaning foods) and full formula feeding is determined. However, the choice to breast or formula feed may introduce selection bias that cannot be eliminated in the statistical analysis. These and other challenges that are encountered in infant nutrition intervention studies and possible solutions will be presented.

P1.2.28
When does an interim analysis not jeopardise the type I error rate ?

P Broberg[1]
[1]Oncology & Cancer Epidemiology, LU/Skåne Hospital, Lund, Sweden

Interest in adaptive clinical trial designs has surged during the last few years. One particular kind of these called *sample-size adjustable designs* (sometimes sample size re-estimation designs) has come to use in a number of trials lately. Following a pre-planned interim analysis this design offers the options of

- closing the trial due to futility
- continuing as planned
- continuing with an increased sample size

Recent research has identified situations when raising the sample size does not lead to inflation of the type I error rate. Mehta and Pocock (StatMed 2011) identifies a set of promising outcomes where it is safe to raise the sample size in two-stage trials. Denote the observed test statistic at the interim by z, the originally planned sample size by N_0, the number of observations at the interim by n, and the raise considered by r. Call the final test statistic Z^*_2.
Then the reference finds that the modified rejection threshold $c(z, N_0+r-n)$ ensures protection of type I error: $P0(Z^*_2 \geq c(z, N_0+r-n))=\alpha$. In Broberg (BMC

Med Res Methodol 2013) a bound expressed in terms of the observed test statistic at the interim is exhibited : If $z \geq b(q, V)$, for a certain function b where $q = n/(N_0+r)$ and $V=(N0+r)/N0$, then one may use the traditional threshold z_α at the final analysis without compromising the type I error rate.
The current work proves these results to be equivalent. Also, further details regarding the asymptotics of the bound b are derived.

P1.2.33
Use of an adaptive approach to design and evaluation of multi-regional clinical trials

C Chiang[1], J-P Liu[1], C-F Hsiao[2]
[1]National Taiwan University, Taipei, Republic of China (Taiwan), [2]National Health Research Institutes, Zhunan, Republic of China (Taiwan)

In recent years, global collaboration has become a conventional strategy for new drug development. To accelerate the development process and shorten approval time, the design of multi-regional clinical trials (MRCTs) incorporates subjects from many countries/regions around the world under the same protocol. At the design stage of a multi-regional clinical trial, a common treatment effect and an equal variability of the primary endpoint across regions are usually assumed.
However, at any data monitoring point of a MRCT, we may discover regional differences. In this case, modification of the planned design will be inevitable. In this paper, we develop an adaptive selection design if we discover regional differences at an interim analysis in a MRCT. More specifically, we will stop recruiting subjects in regions without expected treatment differences. Statistical method on sample size modification will also be developed.

P1.2.35
Appropriate or inappropriate use of REMARK guidelines

C Ciria[1], S Love[1], C Roberts[1], M Shanyinde[1], D Altman[1]
[1]University of Oxford, Oxford, United Kingdom

Background: Appropriate reporting is critical for the interpretability and application of clinical research. REMARK is a guideline of reporting recommendations for tumour marker prognostic studies that encourages complete reporting of relevant information.
Objective: Evaluate the appropriateness of citations of REMARK in published articles.
Methods: We searched the Web of Science database in September 2013 and identified 878 articles that cited REMARK. A random sample of 100 articles was selected for review. Each article was read by two reviewers with professional experience in medical research. The citation of REMARK was classified as appropriate reporting or conduct, or inappropriate. Appropriate reporting was considered when REMARK was used as a reporting guideline, appropriate conduct when REMARK criteria were followed throughout the study, and inappropriate use when the indication was neither of these.
Results: In the random sample of 100 articles, most were oncology research (94%) of prognostic biomarkers (45%) reporting on a single dataset of patients (71%). The use of REMARK was considered appropriate in 67% of the articles (52% reporting; 15% conduct) and inappropriate in 33%. Appropriate uses included REMARK as a reporting guideline (38%), a tool to conduct the study (13%), and importance of reporting guidelines (7%). The most common inappropriate use was as a methodology guideline (28%).
Conclusion: REMARK is commonly used as a guideline of reporting recommendations. However, inappropriate use as a methodology guideline is also frequently observed with potential negative consequences over the design of the study.

P1.2.47
Using longitudinal toxicity score to detect time trend in dose-finding trials. Application to 19 phase I studies

M Ezzalfani[1], M Marous[1], L Collette[2], E Rizzo[2], X Paoletti[1]
[1]Institut Curie / U900, Paris, France, [2]EORTC Headquarters, Brussels, Belgium

The risk of late and cumulative toxicities is a concern with molecularly targeted agents in oncology dose-finding trials, because of their long-term administration. The power to detect time trend in toxicity has been shown low using binary or ordinal endpoints.
We propose a new approach to detect time trend using a quasi-continuous toxicity score endpoint which is defined as the Euclidean norm of the grades of toxicities experienced by a patient at each cycle (C). The association between the toxicity score and the binary endpoint dose-limiting toxicity (DLT) was studied using phase-I data. A mixed model is being developed to detect time trend with the toxicity score. This model will be compared with proportional odds mixed-effect models (POMM) for ordinal endpoint with longitudinal measurements.
Eight different molecules were investigated for the first 6 cycles. 536 patients were treated (1387 cycles, median cycle was 2). 5742 toxicities were observed during C1, 8% grade3-4 after C1.The mean toxicity score was 0.068 (min=0; max=0.31). The median and the interquartile range were 0.056 and 0.066, respectively. At cycle-1, the mean score (95%CI) of patients with DLT was significantly different from that of patients with no DLT: 0.170 (0.147, 0.194) *versus* 0.061 (0.057-0.065). A stratified analysis on cycle showed a significant association (p-value<0.0001) between the mean score and grades3-4 group: 0.126 (0.122-0.131) versus 0.050 (0.048-0.052). A significant time effect was found with POMM for two studies (p-value<0.05).
Using repeated toxicity score data in dose-finding trials is promising to investigate cumulative effects.

P1.2.48
Standardizing safety analyses for clinical trials: a story of success in the making

M Fayaz[1], M Nilsson[2]
[1]Tehran University of Medical Sciences, Tehran, Islamic Republic of Iran, [2]Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, United States

Industry standards have evolved over time for data collection (CDASH), observed data (SDTM), and analysis datasets (ADaM). Although substantial progress has been made, additional standardization can improve product development. Development of standard tables and figures with associated analyses will lead to improved product life-cycle evaluation by ensuring reviewers receive the desired analyses for the evaluation of patient safety.
More importantly, having an organized process for shared learning of improved methodologies can lead to earlier safety signal detection and better characterization of the safety profile of our products. A cross-industry working group (the PhUSE Computational Science Symposium Development of Standard Scripts for Analysis and Reporting Working Group) is providing recommendations for analyses, tables, and figures for data that is common across therapeutic areas (laboratory measurements, vital signs, electrocardiograms, adverse events, demographics, medications, disposition, hepatotoxicity, pharmacokinetics). A Script Repository has been created to contain associated code for the recommended analyses and displays.
This poster will provide an update of this effort, and instructions for how to participate in the development and review process. We encourage all medical researchers to adopt this collaborative culture change!

P1.2.57

Trial-situations where stratification in randomization is advantageous - results of a simulation study

A Glass[1], G Kundt[1]
[1]*University of Rostock, Institute for Biostatistics, Rostock, Germany*

Objective: Randomization of patients has been established to a gold standard in clinical trials. But pre-stratification of randomization has been discussed controversially in this context. Randomization of patients in strata might help to prevent type I and type II errors via reduction of variance. But, stratification forces administrative efforts, and an increasing number of strata bears some problem.
To support investigators decision concerning stratification we investigated the impact with respect to the risk of prognostic imbalance under H_0 by a simulation approach, where differently designed hypothetical trials were simulated (at least 1,000 times) for two therapy groups and two strata.
Results and conclusions: The risk of prognostic imbalance could be quantified with maximum 60% for complete randomization. In larger trials, and/or with a factor of less prevalence this risk decreases. Restricted randomization as blocking (alone) can reduce this risk marginally from 32% to 29.6% for a factor of high prevalence (50%) in small studies (N = 100). Stating this, stratification of randomization can be helpful to provide comparable groups with higher probability for certain trial constellations. Decreased type I error by maximum 16pps due to stratification were detected for prognostic factors with large differences (80pps) of success rates between strata concerning small trials (N = 100). For trials with less than 400 patients and differences in success rates ≥30pps, subgroup and interim analyses stratification is recommended to reduce the expected risk of prognostic imbalance.
For large trials with N≥400 patients relevant imbalance were not observed, independently of any factor prevalence.

P1.2.59

CompARE: web platform to choose the primary endpoint of a randomized clinical trial

M Gómez-Mateu[1], G Gómez Melis[1]
[1]*Universitat Politècnica de Catalunya, Barcelona, Spain*

CompARE is a free web-based interface. Its main aim is to help investigators to choose the primary endpoint at the design stage of a trial. This tool provides a quantitative measure to decide between a relevant endpoint (RE) and a composite endpoint (CE) defined as the union of the RE and an additional endpoint (AE).
Calculations in *CompARE* are based on the ARE method developed by Gómez and Lagakos. The ARE (Asymptotic Relative Efficiency) can be interpreted as the ratio of the required sample sizes to detect a specific treatment effect to attain the same power for a given significance level. The method is internally programmed and executed in *R* and run in the server, returning results in a dynamically generated webpage. *R* does not have to be installed in your computer, neither knowledge of *R* is required. *CompARE* is run through interactive HTML forms. Users are required to enter a list of candidate endpoints together with the probability of observing them in the control group and the relative treatment effect given by the hazard ratio. Results in *CompARE* are shown immediately with tables and intuitive plots in terms of ARE values for different scenarios.
We encourage investigators to try *CompARE*, which is currently accessible as a beta version, by freely registering into the website (http://composite.upc.edu/CompARE). Furthermore, you can refer to the user's guide to see how *CompARE* works.

P1.2.63

Effect of one-patient clusters on power in cluster-randomized trials

S Hayoz[1], D Klingbiel[1]
[1]*SAKK Coordinating Center, Statistics Unit, Bern, Switzerland*

Aims: The aim of this research was to assess the effect of one-patient clusters on the power in cluster-randomized trials with small cluster sizes.
Methods: The assumptions for the simulations were based on the trial SAKK 95/06 where 40 physicians (clusters) with equal patient numbers of 4 (160 patients in total) were planned. A linear mixed model with physician as random effect should yield a power of 81.9% with alpha 5%. In reality the cluster sizes varied between 1 and 10, with 22.6% one-patient clusters. Simulations were performed with 10000 repetitions per scenario.
Results: With 40 physicians, cluster sizes between 1 and 10 and 5.0%-25.0% one-patient clusters, the achieved power varied between 79.6% and 80.7%. If the number of patients per cluster was restricted to a maximum of 6 and the number of physicians increased to 41-55 with 4.9%-36.4% one-patient clusters, the achieved power varied between 81.8% and 82.6%. Resampling from the cluster sizes of SAKK 95/06 with 7%-43% one-patient clusters yielded a power between 78.1% and 79.3%. When the one-patient clusters were excluded from the analysis the achieved power was 0.3%-0.9% lower. The proportion of models with numerical problems was never higher than 0.2%.
Conclusion: In all considered scenarios the decrease in power in comparison to the theoretical model with equal cluster sizes was negligible. When the number of patients per cluster was restricted and instead the number of physicians was increased there was no relevant loss in power even with a high proportion of one-patient clusters.

P1.2.64

Comparison of design options for phase IB clinical trials in oncology: simulation results

M Bigler[1], S Hayoz[1], A Xyrafas[1], D Klingbiel[1]
[1]*Swiss Group for Clinical Cancer Research (SAKK), Bern, Switzerland*

Background: Phase I trials play an essential role in the development of new drugs in cancer research. Many designs and many comparisons of these designs with simulations exist. To the best of our knowledge, no simulation results exist when there are only few dose levels to compare, as it is often the case in phase IB trials, where for example toxicity data in other indications are available.
Methods: We compared the 3+3 design, the continual reassessment method (CRM), the modified toxicity probability interval method (mTPI) and the rolling-6 design in a dose-finding study with only 2-4 dose levels of a drug. We explored the properties of the designs in several scenarios.
Results: The 3+3 design generally performed poorly.
When the prior probabilities for toxicity were close to the true values, CRM was the best method in terms of selecting the correct dose as maximum tolerated dose (MTD) and number of patients treated at the MTD. However, it suffered from misspecification of the priors.
In all scenarios the rolling-6 design required the shortest time to complete the trial.
The mTPI method was—depending on the design parameters—slightly better than the 3+3 and rolling-6 designs as it selected the correct dose more often and treated more patients at the MTD.
Conclusion: CRM has the best performance in terms of different metrics for phase IB trials in oncology given that the priors are well specified. As investigators seem to feel uncomfortable with this complicated method, there may be alternatives like mTPI to consider.

P1.2.67
Adapted levels of evidence for small populations
G Hlavin[1], F König[1], M Posch[1], C Male[2], P Bauer[1]
[1]Medical University of Vienna, CeMSIIS, Vienna, Austria, [2]Medical University of Vienna, Paediatrics, Vienna, Austria

A full independent development programme to demonstrate efficacy may not be feasible in small populations such as paediatrics populations or orphan indications. We propose clinical trials designs, which make use of prior knowledge on efficacy and safety for inference in the small population. For example, we establish how studies conducted in related (larger) populations and/or studies conducted with similar drug substances can serve as basis to partly extrapolate efficacy and safety in the small population groups of interest. By using this strategy, the regulatory need for independent evidence in the small population groups of interest can be lowered. Depending on the degree of similarity and prior knowledge, e.g. between the paediatric and adult populations, the required level of evidence, testing strategy and sample sizes will be adopted. The less uncertainty is left, the higher the frequentist significance level for the validation study in the small population group may become.
We show how such an adapted significance level can be motivated in terms of prior information in the framework of the false positive reporting probability as well as in the Bayesian paradigm. We translate Bayesian decisions rules into classical frequentist decision making criteria.
Acknowledgement: This project has received funding from the European Union's 7th Framework Programme for research, technological development and demonstration under Grant Agreement no 602552.

P1.2.68
Retrospective evaluation of the futility analysis in LUME-Lung 2, a phase III trial of nintedanib plus pemetrexed for NSCLC patients
J Hocke[1], P Glomb[1], R Kaiser[1], J Barrueco[2], B Gaschler-Markefski[1]
[1]Boehringer Ingelheim Pharmaceuticals GmbH & Co. KG, Biberach, Germany, [2]Boehringer Ingelheim Pharmaceuticals Inc., Ridgefield, United States

Background: LUME-Lung 2 (NCT00806819; 1199.14) was a placebo-controlled phase III trial of nintedanib + pemetrexed for 2nd-line non-squamous NSCLC. The primary endpoint was progression-free survival (PFS) by central review. A preplanned futility analysis based on investigator-assessed PFS (secondary endpoint) was conducted by an independent data monitoring committee (DMC) after ~50% events. The futility criteria as defined in the DMC charter were met and the trial was halted with 713/1300 planned patients. Based on the DMC recommendation to stop the trial (without any safety concerns), ongoing patients were unblinded and follow-up was continued according to an amended protocol.
Methods: Final study data were used for a retrospective evaluation of the conditional and predictive power over time that led to the DMC recommendation. Both investigator and centrally assessed PFS data were used.
Results: The retrospective analysis confirmed that based on investigator-assessed PFS, the futility criteria were met at the time of the DMC analysis. However, when analyzed at other timepoints, the same statistical measures remained above the predefined futility criteria. Also, when centrally reviewed PFS was analyzed over the same time period, fluctuations in conditional and predictive powers were observed but both remained above the formal futility criteria. Furthermore, a significant prolongation of PFS by central review was shown based on final data.
Conclusions: Retrospective investigations suggest that had the DMC analysis been performed at another timepoint or had centrally reviewed PFS data been used, the futility outcome may have been different and the trial may have been continued.

P1.2.86 *Cancelled*
Comparison of two-stage versus two separate single-stage settings for bioequivalence studies with crossover design
S Knahl[1], A Seidel[1], F Fleischer[1]
[1]Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riß, Germany

Usually the aim of conducting a bioequivalence trial is to compare the bioavailability of two or more different formulations in a statistical sense. A gold standard design for this purpose is given by a two-period, two-sequence crossover. Historically, such a design has been performed in a single-stage setting where recently it is more and more common to implement two-stage designs which are also acknowledged by regulatory authorities [4][5].
Potvin et al. [1] proposed four possible methods of such a two-stage design. The properties of those methods were further evaluated by Montague et al. [2] as well as further characterized by Karalis and Macheras [3].
In this talk the different approaches are investigated and compared. The purpose is to identify scenarios and rules, which help to decide under which circumstances which specific setting is preferable. Various test cases, which vary in terms of sample size, power, geometric mean ratio, and intra-subject variability are considered and compared by means of simulations.
These investigations can support the decision process on the preferred setting for further bioequivalence trials.
References:
[1] PotvinD. et al. Sequential design approaches for bioequivalence studies with crossover design. Pharmaceutical Statistics 2008;7:245-262.
[2] MontagueTH. et al. Additional results for 'Sequential design approaches for bioequivalence studies with crossover designs'. Pharmaceutical Statistics 2012;11:8-13.
[3] KaralisV, MacherasP. An insight into the properties of a two stage design in bioequivalence studies. Pharmaceutical Research 2013;30(7):1824-35.
[4] FDA-Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products.March2003.
[5] EMA-Guideline on the Investigation of Bioequivalence.January2010

P1.2.88
Adaptive two-stage bioequivalence trials with early stopping and sample size re-estimation
F König[1], M Wolfsegger[2], T Jaki[3], H Schütz[4], G Wassmer[5]
[1]Medical University of Vienna, CeMSIIS, Vienna, Austria, [2]Baxter Innovations GmbH, Vienna, Austria, [3]Lancaster University, MPS Research Unit, Lancaster, United Kingdom, [4]BEBAC, Vienna, Austria, [5]Aptiv Solutions, Cologne, Germany

Bioequivalence between two products has to be demonstrated as an essential part of the generic approval process (new formulation vs. innovator product), bridging an innovator´s product from the formulation used in clinical phase III to the marketed formulation, or in case of major variations of an approved product. The most common design of bioequivalence studies is a two-sequence two-period two-treatment crossover design, where inclusion of 90% confidence intervals of pharmacokinetic metrics in a pre-defined acceptance range has to be shown.
Alternatively, bioequivalence can be established by using Two One-Sided Tests (TOST) each at an alpha level of 5%. %. However, a fixed sample approach offers no flexibility, e.g., if there is high uncertainty about the assumed ratio and/or CV.
We propose a two-stage adaptive design using combination tests to combine stagewise p-values. This will strictly control the type I error rate in case data-driven design modification have to be performed at an interim analysis. We derive 90%-repeated confidence intervals for the adaptive TOST approach. We investigate different sample size reassessment strategies using conditional power arguments. We discuss how futility stop-

ping can be sensible implemented. The operating characteristics will be assessed by clinical trial simulations.
Furthermore, the proposed adaptive design would allow to use different cross over designs for the first and second stage, e.g., switch from a classical two-period design to a more complex replicate design. Another application of the proposed method would be for establishing biosimilarity of two products.

P1.2.89

Adaptive designs for confirmatory model based decisions using MCP-Mod

S Krasnozhon[1], A Graf[1], B Bornkamp[2], F Bretz[2], G Wassmer[3], F König[1]
[1]Medical University of Vienna, CeMSIIS, Wien, Austria, [2]Novartis Pharma AG, Basel, Switzerland, [3]ADDPLAN Inc., an Aptiv Solutions Company, Cologne, Switzerland

Adaptive seamless designs for confirmatory clinical trials have attracted a lot of attention because they offer the possibility to combine different phases of drug development into a single trial. This is of paramount interest in small populations, e.g., when developing drugs for rare diseases. Though the sample size is limited still an appropriate dose has to be found and sufficient evidence for its efficacy to be demonstrated.
We propose adaptive clinical trial designs with multiple doses and use modelling approaches to (i) establish a positive dose-response profile, (ii) increase the power of declaring effective dose statistically significant, and (iii) support dose selection at an adaptive interim analysis. We extend MCP-Mod methodology to adaptive two-stage designs by using the closed-testing principle and applying an adaptive combination test to each intersection hypothesis. Combining the data from both stages in adaptive confirmatory designs allow for flexible interim decisions based on all (interim) data available of the ongoing trial while always ensuring strict type I error control. In particular, the MCP-Mod approach can be used to obtain model-based dose effect estimates at interim to guide early futility stopping and/or re-design the second stage (e.g. choice of doses, sample size, allocation ratio) and analysis (e.g., dropping of inadequate response models).
By the means of clinical trial simulations we show the operating characteristics (e.g., power for PoC or individual dose-control comparison, bias of effect estimates) for specific adaptations rules.

P1.2.93

Residual plots for censored data: a new approach

M Law[1], D Jackson[1], S Brill[2], W Wedzicha[2]
[1]MRC Biostatistics Unit, Cambridge, United Kingdom, [2]University College London, London, United Kingdom

All statistical models make assumptions. For the resulting statistical inference to be valid, the model must describe the data well. For regression models, and when data are uncensored, residual plots are a standard tool for assessing how well the model fits the data. However, when some data are censored, standard residual plots are less appropriate for checking modelling assumptions. Various approaches have been suggested, which we summarise and compare with our own proposed approach.
I will present a method for producing residual plots, analogous to those used for uncensored data, which take into account both the parameter uncertainty and the uncertainty in the location of the censored data. I will illustrate this method by examining the model fit for an analysis of bacterial load data from a trial for chronic obstructive pulmonary disease.
Here we use a standard linear regression model but some data are censored because they are only observed to be less than a threshold value. We conclude that the model fit is acceptable but is less good than might initially be thought before appropriately handling the censored observations.

P1.2.97

Longitudinal cluster analysis with application to identify mortality associated SOFAtrends in critical care medicine

K Liu[1,2], V Cornelius[1], M Shankar-Hari[1,2], C Dangoisse[2,3], M Terblanche[1,2]
[1]King's College London, London, United Kingdom, [2]Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom, [3]King's College London Hospital NHS Foundation Trust, London, United Kingdom

Background: In critically ill patients Sequential Organ Failure Assessment (SOFA) score is used as a surrogate for end organ function and is associated with mortality. In ICU literature, logistic regression analyses often used to predict patient mortality, but it does not acknowledge and are hampered by the heterogeneous nature of the patient population. We hypothesised that detecting homogenous patient groups in terms of organ dysfunction trajectory using longitudinal cluster analysis will identify mortality clusters in critically ill patients; thus improving risk prediction.
Method: A two-stage approach was conducted. In stage one k-means for longitudinal data (KML) was adopted to determine homogeneity of patient SOFA trajectories during their first 4 ICU days. In stage two a logistic regression model was fitted using the cluster allocation as predictor variable to examine the association with mortality. Model comparison was conducted using ANOVA with two conventional approaches.
Result: 4438 patients admitted between April 2010 and October 2013 were included. The KML analysis identified 6 distinct patterns of organ dysfunction. These six clusters had distinct outcome patterns: mortalities for clusters A through F were 2.2%, 6.0% 18.2%, 22.6%, 42.1% and 54.1%, respectively. ANOVA analysis showed that our approach is significantly better than the logistic models only considering baseline or day 1-3 change in SOFA score as the regressors of mortality.
Conclusion: Longitudinal cluster analysis using SOFA identified homogenous clusters which were associated with ICU mortality. Our approach outperformed the conventional approaches and thus would be a better approach for risk stratification in clinical trials.

P1.2.100

Using simulation to examine the effects of varying cluster size on the precision of stepped wedge cluster randomised trials

JT Martin[1], K Hemming[1], A Girling[1]
[1]University of Birmingham, Birmingham, United Kingdom

The use of a stepped wedge cluster randomised trial (SW-CRT) is increasing, especially in health service evaluations. The stepped wedge design incorporates a sequential rollout of the intervention to all clusters over multiple time periods. The SW-CRT typically includes equally spaced steps and an equal number of patients in each cluster. However, in practice, it is unlikely that an equal number of patients will be present across all clusters at each time point. Here we consider the impact of varying cluster sizes on the precision of a SW-CRT.
To this end, a formula devised by Hussey & Hughes to compute precision for SW-CRTs was extended to include varying cluster size. The degree of variation in cluster sizes can be described by the coefficient of variation (CV). By altering the CV, along with the intra cluster correlation coefficient (ICC), the number of clusters and the number of randomisation points, the effects of varying cluster sizes could be determined. Simulation methods were then used to form an average precision over varying designs.
For comparative purposes, the precision of a parallel CRT was also estimated. We found that standard methods of adjustment in a CRT may underestimate the precision.
Simulations show that the SW-CRT is less affected by varying cluster sizes than the parallel CRT. There are also situations in which for equal cluster sizes, the parallel CRT is of higher precision, but as the CV increases, the SW-CRT becomes the trial type with higher precision.

P1.2.106

Analysis of cluster randomised cross-over trials with binary outcomes

KE Morgan[1], BC Kahan[1], RH Keogh[2]
[1]*Queen Mary University of London, London, United Kingdom,* [2]*London School of Hygiene & Tropical Medicine, London, United Kingdom*

In cluster randomised cross-over (CRXO) trials clusters are randomised to receive treatments in a particular order. Data from a CRXO trial may have a complex dependency structure, with outcomes correlated within clusters and, potentially, additional correlation within treatment periods. This additional correlation arises when outcomes in one period are more similar to each other than they are to outcomes in another period in the same cluster. Assessing whether analysis methods adequately account for these correlations is important.

We used simulation to compare methods of analysing a binary outcome, including logistic mixed effects models, generalised estimating equations (GEEs) and cluster-level summary methods. Models that ignore additional within-period correlation led to increased Type I error rates. An unweighted linear cluster-level summary regression gave nominal error rates in all scenarios, but lost power if extra correlation was present especially for small numbers of clusters. A mixed effects model with random effects for cluster and cluster-by-period interaction gave nominal error rates only in scenarios with large numbers of clusters, if there was additional correlation. GEEs did not give correct error rates in any scenarios considered.

Assessing whether extra within-period correlation is possible is important when conducting a CRXO trial, and if it is this should be accounted for in both the sample size calculation and the analysis method. Unweighted cluster level summaries are the most robust method of analysis; however with a very large number of clusters a mixed effects model with random effects for cluster and cluster-by-period interaction could also be used.

P1.2.109

Optimal sampling times for pharmacokinetic modelling of a cocktail of phenotyping drugs

TT Nguyen[1], H Bénech[2], M Delaforge[2], A Pruvost[2], F Mentré[3], N Lenuzza[1]
[1]*CEA, LIST, Data Analysis and Systems Intelligence Laboratory, Gif-sur-Yvette, France,* [2]*CEA, DSV, iBiTecS, Gif-sur-Yvette, France,* [3]*IAME, UMR 1137, INSERM - University Paris Diderot, Paris, France*

"Cocktail" of drugs is of high interest to determine enzyme activity responsible for drug metabolism and pharmacokinetics. Phenotyping indexes can be derived from a few samples using nonlinear mixed effect models (NLMEM) for analyzing drug concentrations and maximum a posteriori estimation (MAP) of individual parameters. We proposed an informative design common to two molecules for a phenotyping study: midazolam (probe for CYP3A activity) and digoxin (P-glycoprotein).

Using data of a previous study, NLMEM for midazolam, its 1-OH-metabolite and digoxin were developed in software MONOLIX4.2. Based on these models, we proposed a common design using a compound optimality criterion which is a weighted sum of log determinants of population Fisher information matrix (FIM) for each compound. The resulting design was evaluated for MAP and predicted shrinkages were reported, based on Bayesian FIM, using R function PFIM4.0. Finally, sampling windows were computed around the optimal times, satisfying an expected joint loss of efficacy (evaluated by Monte-Carlo simulations) <5%.

The common design was composed of six samples (0.25, 1, 2.5, 5, 12, 48h post-administration) instead of ten samples if considering separately each molecule. Predicted relative standard errors of derived phenotyping indexes were <30%, with shrinkages <40%. The sampling windows provided more flexibility while maintaining 95%-efficacy, compared to the optimal design.

By combining NLMEM, compound design and sampling windows based on FIM, we were able to determine sparse samples allowing correct estimation of parameters for three compounds. This approach can be extended to efficiently design studies with cocktails including more drugs.

P1.2.112

Generalization of the big stick randomization rule to more than two treatment groups and unequal allocation rates

P Ofner-Kopeinig[1], M Errath[1], A Berghold[1]
[1]*Inst Med Info, Stat & Docu, Medical University of Graz, Graz, Austria*

The Big Stick randomization rule suggested by Soares and Wu provides good performance according to predictability and balance behavior. Treatments are allocated at random until a tolerance "a" is reached. In case "a" is reached, the under-represented treatment will be allocated next. The Big Stick procedure has so far only been described for two arm studies with equal allocation rates.

The idea of using absolute differences to identify treatment imbalances does not work anymore if there are more than two treatments or unequal allocation rates. To allow for a generalization of the Big Stick randomization rule we developed a generalized measure of treatment imbalance which is based on the differences of observed and expected frequencies. This measure of imbalance takes into account the expected frequencies of the treatments at each randomization step and also treatment weights. Using this generalized measure of imbalance one can expand the Big Stick randomization procedure for unequal allocation rates and more than two treatment groups.

To evaluate the generalized Big Stick randomization procedure a simulation study using the simulation tool of the "Randomizer for Clinical Trials" was performed. For an accuracy of 1 % and a confidence interval of 95 %, each simulation was performed 10000 times. Treatment imbalances were calculated to show the balance behaviour. To estimate the predictability the probability of correct guessing and the probability of deterministic allocation was calculated. The procedure was compared with complete randomization and permuted block randomization with different block lengths.

P1.2.114

The implications of differential clustering for the analysis of binary outcomes in cluster randomised trials

N O'Leary[1], SA Roberts[1], C Roberts[1]
[1]*University of Manchester, Manchester, United Kingdom*

Introduction: Cluster randomised trials usually assume homogeneity of the clustering effect. In cluster randomised trials of professional behaviour change interventions the clustering effect may differ between intervention arms.

We examine the robustness of a range of standard techniques for the analysis of cluster randomised trials where the outcome is binary in the presence of between arm differences in the intra-cluster correlation coefficient (ICC).

Methods: Binary data were simulated for a two-arm cluster randomised trial from a logistic-normal model with different values of the manifest ICC in each arm and a range of proportions.

Analysis methods assessed were (1) an adjusted test of proportions, (2) a summary measures analyses, (3) a logistic generalised estimating equation model with an exchangeable correlation structure, and logistic-normal models with (4) a random intercept or (5) random coefficients for each arm. We assessed consistency, small sample bias of estimates and test size.

Results: Treatment effect estimates for the random coefficient models were not consistent with estimates from other analyses; over-estimating the effect when proportions were low and under-estimating it when proportions were high.

In small sample settings, the two-sided test size of all methods was higher than the nominal significance level. For all methods, heterogeneous clustering led to small sample bias and asymmetry of the one-sided test-sizes in some scenarios.
Conclusion: The random coefficient model is not suitable for the analysis of cluster randomised trials with binary outcomes. Care is need interpreting other analysis methods where there is heterogeneous clustering for binary data.

P1.2.121
Treatment crossovers in time-to-event non-inferiority randomized trials of radiotherapy in subjects with breast cancer

S Parpia[1], JA Julian[1], L Thabane[2,3], C Gu[1], TJ Whelan[1,4], MN Levine[1,4]
[1]Ontario Clinical Oncology Group, McMaster University, Hamilton, Canada, [2]Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, Canada, [3]Centre for Evaluation of Medicine, Hamilton, Canada, [4]Juravinski Cancer Centre, Hamilton, Canada

The non-inferiority randomized trial design is commonly used to compare novel experimental breast radiation regimens to standard breast irradiation for the prevention of local recurrence in patients with breast cancer who have undergone breast conserving surgery. Prior to beginning radiation therapy (RT) the patient will undergo a planning process to establish the treatment fields to target the tumour and avoid normal tissues. Planning generally occurs after randomization. Sometimes it is not possible to deliver the experimental regimen and the patient will cross over to the standard RT. In addition, sometimes the patient decides to receive the usual RT. Although the intention-to-treat (ITT) analysis is the preferred approach for superiority trials, its role in non-inferiority trials is still under debate. The ITT is generally perceived to produce a diluted treatment effect and, therefore, is anti-conservative in demonstrating non-inferiority. This has led to the use of alternative approaches such as the per-protocol (PP) analysis or the as-treated (AS) analysis, despite the inherent biases of such approaches.
Using simulation, we investigate the effect of random and non-random crossovers, under various scenarios, on the ITT, PP, AS, and the combination of the ITT and PP analyses, with respect to type I error in trials with time-to-event outcomes. We also evaluate bias and standard error of the estimates from the ITT, PP and AS approaches. Results will be presented. Our research will guide methodologists in the analysis of non-inferiority trials with crossovers.

P1.2.122
An application of non-parametric factorial MANOVA in health research

O Pasin[1], H Ankarali[1], S Cangur[1], MA Sungur[1]
[1]Duzce University, Institute of Health Sciences, Biostatistics, Duzce, Turkey

In health research, factors can have effects on multiple correlated outcomes. Analyzing these related outcomes separately can cause an increase in type I error. However in practical applications, this error is made frequently by using univariate analyses, which are simple and easily interpretable. In addition a majority of variables used in studies do not exhibit normal distributions or homogeneity of variance which are assumptions of univariate parametric models. For these reasons, introduction and application of new approaches, which provide more reliable results, are becoming important.
Our aim is to introduce the theoretical characteristics of a nonparametric MANOVA (PERMANOVA) and to discuss its application to a data set from a clinical trial.
The PERMANOVA model uses distance or dissimilarity measures between pairs of subjects or variables. The null hypotheses regarding factor effects are tested with Pseudo F statistics and type I error of these statistics are calculated by a permutation approach or Monte Carlo simulation. In this study, we investigated the relationships among presence of the Hasimoto illness, gender and lipid profile by PERMANOVA. Data analysis was performed using the algorithm developed by Marti J. Anderson. We observed that there was no significant interaction between gender groups, but the lipid values in Hasimoto patients are significantly higher than in healthy individuals (P=0.0177).
As conclusion we can say that in our example the true biological structure is better reflected, if related variables are simultaneously evaluated.

P1.2.142
Comparison of different methods for controlling false positives in adverse event reports analysis

S Romio[1,2], L Scotti[1], A Zambon[1], L Hazell[3], N Schmedt[4], M Sturkenboom[2], G Corrao[1]
[1]University of Milano-Bicocca, Milan, Italy, [2]Erasmus University Medical Center, Rotterdam, The Netherlands, [3]Drug Safety Research Unit, Southampton, United Kingdom, [4]Leibniz Institute-BIPS GmbH, Bremen, Germany

Adverse event reports databases can be used to generate hypotheses of associations between drug use and adverse events. False positive (FP) associations need to be minimized when multiple association tests are performed simultaneously. Several methods to control for multiple comparisons in hypothesis testing have been developed but few methods are available for testing one-sided hypothesis.
I this context, we compared two modifications of the False Discovery Rate (FDR) approach: pFDR proposed by Storey and robust FDR proposed by Pounds and Cheng (rFDR) to control for FP.
Within the SAFEGUARD project, adverse events reports available in FDA's Adverse Event Report System and EudraVigilance databases in 2004-2012 were analyzed. The Proportional Reporting Ratio (PRR) was used to evaluate associations between non-insulin blood glucose lowering drugs and selected outcomes. Two identification criteria (broad/narrow) were used for each event. Data were analysed applying pFDR and rFDR approaches to the hypothesis tests on the PRR using proc multtest in SASv9.3 and a specific R script developed by Pounds and Cheng respectively. Statistical significance threshold was 0.05.
The pFDR showed a reduction of the number of statistically significant associations from 12% to 17% in both databases and event definitions systematically higher than that from rFDR approach (0% to 2%). Different performance in controlling FP could depend on the sensitivity of rFDR to non-monotone distribution of original p-values.
From these early findings the pFDR seems to be a robust tool for controlling FP in pharmacovigilance.

P1.2.146
Adaptive crossover designs for phase II dose-finding trials

S Simpson[1], LV Hampson[1], J Whitehead[1], B Jones[2]
[1]Lancaster University, Lancaster, United Kingdom, [2]Novartis Pharma AG, Basel, Switzerland

Finding the optimal dose of a new medicine is an important part of drug development. Adaptive dose-finding procedures use accumulating data to determine which doses should be allocated to each new cohort of patients recruited to the trial, with the aim of obtaining an accurate estimate of the target dose. In this presentation, we explore adaptive crossover designs for estimating the dose which provides a proportion, π, of the maximum effect of a drug, i.e., the ED100π. We restrict attention to designs where each patient receives placebo and three active doses of the drug in a sequence determined by a Williams square.

Bayesian optimal adaptive procedures are considered which recommend that each new cohort of patients receives the combination of doses that minimises the variance of the posterior modal estimate of ED100π. Prior opinion about the dose-response relationship is represented as pseudo-data. It is assumed that the dose-response relationship follows an E_{max} model. However, fitting E_{max} models can be challenging due to problems of non-convergence. With this in mind, when the E_{max} model fails to converge at an interim analysis we investigate Bayesian procedures which use a cubic approximation to the E_{max} model for the purposes of making dose recommendations.
An algorithmic procedure is also considered for dose-finding, which assumes only that the dose-response relationship is monotonic when making interim dose recommendations. Simulation is used to compare the algorithmic procedure and Bayesian optimal design for estimating the ED100π, using a non-adaptive incomplete block design as a benchmark for comparison.

P1.2.166
Statistical methods for centralised risk-based monitoring in clinical trials

JV Torres-Martin[1], M Rodriguez[1], K Haas[2], M Horneck[2]
[1]Syntax for Science S.L, Palma de Mallorca, Spain, [2]maxclinical GmbH, Freiburg, Germany

Due to a global crisis in combination with an increase in drug development costs, pharmaceutical companies are under enormous pressure to make their processes more cost-effective. In August 2013, the FDA released a guideline outlining its position on the current practice of clinical monitoring. This guideline opened a new perspective acknowledging that traditional costly practices based on on-site monitoring might not be the most efficacious. The FDA encourages centralised risk-based monitoring which allocates resources across centres based on their level of risk. Centralised risk-based monitoring stands out as a promising area wherein its use can not only reduce costs but also improve research quality and patient safety.
A number of statistical methods used in other areas that can be applied to risk-based monitoring are reviewed.
We explore the application of these methods under the following situations using real data:
(1) error and fraud detection (multivariate outlier detection in combination with missing data imputation, overdispersion and underdispersion);
(2) blinded monitoring of nuisance parameters that can affect the principal objectives of the clinical trial (variability of the primary variable for continuous endpoints, or event rates for survival and count endpoints); and,
(3) patient recruitment. Statistical modeling can also be used to assess trends of undesired operational trial events, and quantify the risk of these events happening in the near future.

P1.2.167
Re-sampling methods for internal model validation in diagnostic and prognostic studies: review of methods and current practice

JV Torres-Martin[1], H Chadha-Boreham[2]
[1]Syntax for Science S.L, Palma de Mallorca, Spain, [2]Actelion Pharmaceuticals Ltd., Basel, Switzerland

Multivariable logistic regression models are extensively used in diagnostic and prognostic studies. Examples can be found in cancer (e.g., prostate cancer test), cardiovascular diseases (e.g., Framingham Risk Score) or pulmonary arterial hypertension (e.g., DETECT PAH Risk Score). Variable selection is an important part of model building. Good models are those which show good performance characteristics (calibration and discrimination) not only with the data used to fit the model but also with new external data. As external data are not readily available in all disease areas, the data used to fit the model is commonly used to validate the model (i.e. internal model validation).
The objective of this work is to review methods for internal model validation, focusing on variable selection and discrimination by means of resampling methods. Methods for examining model performance and detection of overfitting will be reviewed.
Two systematic reviews are conducted: (1) in "Statistics in Medicine" to gather relevant research on re-sampling methods for internal model validation; and (2) in the "New England Journal of Medicine" to assess the extent to which these methods are used in medical research. The application of these methods in SAS and R is described. We finish our work by illustrating the methods applied to real data, and describing the challenges we faced in real life situations.

P1.2.168
Adaptive increase sample size with count endpoints: the path from statistical simulation to the development of an explicit formula

A Rodríguez[1], A Mir[2], M Rainisio[3], JV Torres-Martín[1]
[1]Syntax for Science S.L., Basel, Switzerland, [2]University of the Balearic Islands, Palma de Mallorca, Spain, [3]Abanovus, Sanremo, Italy

Adaptive designs have been a hot topic of discussion during the past recent years by biostatisticians from pharmaceutical companies, regulatory agencies and academia. Mehta and Pocock proposed an adaptive design where the trial starts with a small up-front sample size commitment compared to the traditional group sequential method. Additional sample size resources are committed to the trial only if *promising results* based on conditional power are obtained at interim analysis.
The authors proved that this design does not require any multiplicity adjustment being preserved the type I error if (1) the information obtained in the interim analysis is only used to decide whether the sample size is increased, and (2) the sample size is increased only when the interim conditional power falls within a promising zone. This design was developed for normal, binomial and survival endpoints, but not for count endpoints.
We describe how we use the previous design to plan a clinical trial with a count endpoint using statistical simulations. The required increase in sample size based on the conditional power at interim is obtained, and it is shown that the type I error is preserved. Statistical simulations are not considered by the regulatory agencies as a valid method to prove any statistical result. We describe our achievements and the challenges we face to prove the previous results by means of explicit mathematical formulation.

P1.2.172
Comparison of different allocation procedures in clinical trials in small population groups with respect to accidental and selection bias

D Schindler[1], D Uschner[1]
[1]RWTH Aachen University, Institute for Medical Statistics, Aachen, Germany

Each medical treatment available on the market has been tested by extensive clinical research. For statistically proving the effectiveness of a medical intervention, the randomized controlled clinical trial is considered the "goldstandard".
Usually, patients arrive sequentially to the clinical trial and have to be allocated to the treatments arm immediately. The allocation is realized using a randomization procedure. The accrual character of clinical trials is the source of different biases that may arise even though randomization and blinding have been employed effectively. This results particularly in a biased estimator for the treatment effect. Hence the effectiveness of a placebo might be alleged or, conversely, an effective treatment might be found inefficient and be banned from the market forever. The use of a suit-

able randomization procedure can minimize the influence of present bias. There are many people suffering from rare diseases, thus there is an urgent need for new therapies and hence for statistical methods that can be used in small clinical trials. As most current statistical methods are unsuitable for the use in small populations, the EU launched several programs investigation new methods for small population groups.
The scope of this presentation is to investigate and compare several randomization procedures that minimize selection and accidental bias such as introduced by trends and unobserved covariates in small clinical trials. The theoretical results are illustrated by simulations based on an R package the authors implemented for this purpose.
Acknowledgement: The research is embedded in the IDeAl EU-FP7 project, Grant-Agreement No. 602 552.

P1.2.178

Factors associated with fever control and identification of subgroups in sepsis trials: a regularization based approach

AN Vidyashankar[1], S Katsahian[1], Y Dong[2], F Schortgen[3]
[1]George Mason University, Fairfax, United States, [2]Temple University, Philadelphia, United States, [3]Réanimation Médicale Groupe Hospitalier Henri Mondor, Créteil, France

Aims: In a recent randomized controlled clinical trial, the effect of using external cooling in febrile ICU patients was evaluated and overall beneficial effects of external cooling was identified. However, the overall analysis did not take into account multiple sources of variation present in the data. The primary objectives of this study are: (a) to incorporate additional clinically relevant covariates and their interactions to identify subgroups of patients benefitting from external cooling and (b) evaluate the statistical and clinical gains of the proposed alternative data analysis approach.
Methods: When several clinically relevant covariates and their interactions with demographic variables are introduced into modeling, the number of parameters far exceeds the sample size leading to large p small n problem. Additionally, accounting for variability amongst the centers and the patients lead to statistical models involving both fixed and random effects. We use the adaptive LASSO method, to perform variable selection and inference accounting for uncertainty in variable selection.
Results: Incorporation of relevant clinical variables in the analysis of sepsis trials facilitates data based methods for identifying clinically benefitting subgroups within the patient population. Additionally, systematic accounting for various heterogeneities into statistical modeling improves power for identifying effects of cooling between the subgroups. Importantly, the methodology facilitates construction of alternative clinical scores for monitoring of sepsis severity.
Conclusions: Including several clinically relevant factors and their interactions into regression models lead to data-based methods for identifying clinically benefitting subgroups in sepsis trials. The methodology identifies variables for routine monitoring of sepsis progression.

P1.2.195

Benefit and cost of clinical trial operation quality monitoring - experiences and lessons from the ProTECT III trial

W Zhao[1], S Yeatts[1], Y Palesch[1]
[1]Medical University of South Carolina, Charleston, United States

Funded by NINDS, ProTECT III (NCT00822900) is a randomized controlled large multicenter phase III trial comparing the safety and efficacy of progesterone vs. placebo for the treatment of traumatic brain injury. Due to variability in neurological emergency treatment settings, the heterogeneous physiology of the disease, and the large number of clinical sites, this trial was vulnerable to the diversity of operation transgressions regarding the clinical standardization and guideline for patient care across sites - a serious concern for both sponsor and investigators.
To quantify this variability, a detailed monitoring effort was designed and implemented to track transgressions in study patient care. Thirteen transgression Forms were defined and included in the study book; data on these physiological parameters were collected 24 hours a day for each patient, up to 30 days after injury. At the end of the study, a total of 882 subjects were enrolled from 38 clinical centers with a median enrolment of 18. For these subjects, 28,927 transgression CRFs were submitted with a total of 694,285 hourly records. To effectively monitor the trial operation quality, 116 data validation rules where created within the study database, and captured 34,794 transgression rule violations.
The transgression monitoring protocol is a unique feature of the ProTECT III trial. It provides a rare opportunity for us to examine the benefit and cost of transgression monitoring, and the impact of trial operation transgressions on trial outcomes.

P1.3 Pharmacoepidemiology and drug development

P1.3.51

Pharmacogenetic study of delayed hyperbilirubinemia in a cohort of 4,000 infants

BL Fridley[1], A Wang[1], J Jackson[2], JS Leeder[2], SM Abdel-Rahman[2]
[1]University of Kansas Medical Center, Kansas City, United States, [2]Children's Mercy Hospital, Kansas City, United States

Despite recommendations by the American Academy of Pediatrics (AAP), screening for hyperbilirubinemia is not universally applied to newborns prior to their hospital discharge. As a result, hyperbilirubinemia remains the single most common reason for hospital readmission during the first 2 weeks of life.
The goal of this study was to identify genetic biomarkers, in conjunction with clinical risk factors, associated with risk for progression to hyperbilirubinemia. A cohort of 4,055 newborns (>34 weeks) were prospectively enrolled. Demographic data and a cord blood sample were collected, along with known risk factors for hyperbilirubinemia. The primary phenotype of interest was progression to hyperbilirubinemia after discharge. Seven candidate markers in genes related to conjugation, transport and clearance of bilirubin were genotyped (*SLCO1B1, RXRA, SLCO1B3, SULT1B1, SLCO1A2, APOA2, LST3).* Genetic association analysis of these markers found no statistically significant associations when analyses were conducted separately for white non-Hispanic (N = 2210) and African American babies (N=965). As breast feeding represents a risk factor for hyperbilirubinemia, we also investigated interaction of food source with genotype (gene-environment interaction) where we observed a modest association in African American newborns for rs7696239 (p = 0.045).
Additional research is on-going to expand the pharmacogenomic assessment of potential genetic risk factors and develop a prediction model that will support a clinical nomogram to detect hyperbilirubinemia risk.

P1.3.123

Choice-based conjoint (CBC) analysis to evaluate patient's perception regarding their erythropoiesis stimulating agent (ESA) treatment in chronic kidney disease (CKD)

D Pau[1], G Choukroun[2], O Moranne[3], C Vigneau[4], C Isnard-Bagnis[5], N Belamri[1], S Pibre[1]
[1]Roche, Boulogne-Billancourt, France, [2]CHU Amiens, Amiens, France, [3]Hôpital Pasteur, Nice, France, [4]Hôpital Pontchaillou, Rennes, France, [5]AP-HP Pitié-Salpêtrière, Paris, France

Background: In recent years, a growing interest emerged for patients' perception of treatment they received. In CKD patients receiving ESA, few data exist regarding their preference/satisfaction with medications.
Objectives: CBC analysis is a major market research methodology for studying how people value the characteristics of products/services. The objective of CBC is to determine what combination of a limited number of characteristics is most influential on respondent choice. This technique has been implemented in a clinical study to describe patient's perception and preference regarding their ESA treatment.
Methods: Percepolis is a 6-month multicenter prospective non-interventional study. The primary endpoint was the relative importance according by CKD patient to characteristics of their ESA treatments. CBC questionnaires were developed using the following multiple components:

- 7 ESA characteristics,
- 2 or 3 levels/characteristics,
- each possible answer includes 1 level for 2 characteristics,
- 2 choices/question,
- 7 questions/questionnaire.

Twenty questionnaires have been generated in order to mix all possible treatment characteristics/levels. Patients only had to answer one questionnaire and randomization was used to obtain equal number of respondents for each questionnaire.
Conclusions: CBC analysis results were made available and interpreted. CBC analysis can be considered as a useful technique to evaluate patient's perception and can be implemented in outcomes research when selective data collection is pre-planned.

P1.3.136

Fast track assessment of generics in Portugal

M Ribeirinho[1]
[1]INFARMED - National Authority on Medicines & Health Products, Lisboa, Portugal

Background: In 2013, Portugal implemented the fast track assessment of generics in terms of reimbursement and price decision, making mandatory to evaluate generics in 4 days while in the past the deadline was 75 days. This managed to speed up the entrance of generics in the Portuguese market reflecting an increase in their use potentiating several health gains.
Aim: To establish a relation between the fast track of Health Technology Assessment with the increase availability and use of generics. To determine the impact of this intervention in the National Health Service by modelling all events and agents. Methods Data was collected from generics evaluated since 2012 combining utilization data from prescriptions in NHS ambulatory care. Statistical methods were performed such as ANOVA procedures and Regression Intervention Models.
Results: With the fast track, generics are now fully assessed in 4 days while in 2012 the average conclusion time was 84 days. As a result, they now enter the reference price system much faster allowing more affordable prices. This procedure had an impact in the use of generics, increasing 18%, whose market share raised 5% now reaching 40%. Statistical relation was significant at $p<0.07$.Regression model revealed that savings achieved for both Patients and NHS can be address to the fast track ($p<0,08$).
Conclusions: One year after the implementation of this new assessment scheme, generics increase and expenditure savings due to it were statistical significant. Time and resources saved in this evaluation can now be used to assess innovative medicines.

Tuesday, 26th August 2014 - 10.30-11.00

Poster session P2

P2.1 Longitudinal data analysis

P2.1.54

Six- and 12-month follow-up of an interdisciplinary treatment of patients with fibromyalgia: results of a randomised trial

J Martín[1], S García[1,2], F Torre[3], U Aguirre[1,2], A Callejo[3], A Arizaga[3], F Mendoza[4], JM Quintana[1,2]
[1]Galdakao-Usansolo Hospital, Galdakao, Spain, [2]Health Services Research on Chronic Diseases Network-REDISSEC, Galdakao, Spain, [3]Pain Management Unit, Galdakao-Usansolo Hospital, Galdakao, Spain, [4]Department of Internal Medicine, Galdakao-Usansolo Hospital, Galdakao, Spain

Objective: To assess the efficacy of a 6-week interdisciplinary treatment that combines coordinated psychological, medical, educational, and physiotherapeutic components (PSYMEPHY) over time compared to standard pharmacologic care.
Methods: Randomized controlled trial with follow-up at 6 months for the PSYMEPHY and control groups and 12 months for the PSYMEPHY group. Participants were 153 outpatients with FM recruited from a hospital pain management unit. Patients randomly allocated to the control group (CG) received standard pharmacologic therapy. The experimental group (EG) received an interdisciplinary treatment (12 sessions). The main outcome was changes in quality of life, and secondary outcomes were pain, physical function, anxiety, depression, use of pain coping strategies, and satisfaction with treatment as measured by the Fibromyalgia Impact Questionnaire, the Hospital Anxiety and Depression Scale, the Coping with Chronic Pain Questionnaire, and a question regarding satisfaction with the treatment.
Results: Six months after the intervention, significant improvements in quality of life (p=0.04), physical function (p=0.01), and pain (p=0.03) were seen in the PSYMEPHY group (n=54) compared with controls (n=56). Patients receiving the intervention reported greater satisfaction with treatment. Twelve months after the intervention, patients in the PSYMEPHY group (n=58) maintained statistically significant improvements in quality of life, physical functioning, pain, and symptoms of anxiety and depression, and were less likely to use maladaptive passive coping strategies compared to baseline.
Conclusion: An interdisciplinary treatment for FM was associated with improvements in quality of life, pain, physical function, anxiety and depression, and pain coping strategies up to 12 months after the intervention.

P2.1.60

Assessment of the bias introduced by excluding patients from the analysis set due to missing post-randomization

S Hantel[1]
[1]Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany

The analysis set in randomized clinical trials is often restricted to patients for whom at least one measurement after start of randomized medication is observed. Several methods exist to handle situations that not all data after start of treatment are available like mixed model for repeated measurement (MMRM), pattern mixture models, multiple imputation (MI) or single imputation methods like last observation carried forward (LOCF), and the complete case analysis. Most of these methods are only valid under strong assumptions.
Since the exclusion of patients based on post randomization findings like early drop-outs can introduce a bias, this poster investigates the performance of methods using the restricted set of patients with post randomization data (full analysis set, FAS) compared to methods using the all randomized patients set (itt set). In order to apply the MMRM to the itt set, post randomization data are imputed by baseline data carried forward (BOCF) or MI for patients without any post-randomization data. The assessment of the bias is based on simulation studies for normally distributed longitudinal data.
The simulation show that the use of the restricted analysis set can introduce a substantial bias. In this context, baseline carried forward and multiple imputation using values more extreme than the observed values can play an important role in assessing the potential bias.

P2.1.74

Progress of bilateral monitoring - case study

K Hrach[1]
[1]J.E.Purkyne University in Usti nad Labem, Usti nad Labem, Czech Republic

In our study, when observing neurological patients, oxygen, lactate/pyruvate ratio, glycerol and glucose values are monitored bilaterally (from the healthy and the affected hemisphere of the brain), i.e. eight variables totally. Data collection is still ongoing. This contribution is dedicated to the exemplary editing data from one patient. It illustrates merging of variables (monitoring is not perfectly synchronized) and emphasizes the need for standardization (monitored variables have very different range).
An important aspect is the selection of appropriate data summarization (averaging in time). In the case of analyzed patient, the eight resulting time series show similar patterns in the graph. The closest course has always the same variable on healthy and the affected hemisphere, generally. At least similar to all the remaining variables is lactate/pyruvate ratio (from both hemispheres).
Confirmation of the visual analysis of this graph is an application of cluster analysis - dendrogram for gradual clustering of variables confirmed the aforementioned similarities (or dissimilarities) totally.
The study is realized under the support of grant project n. NT13883-4/2012 of the Czech Ministry of Health.

P2.1.77

Assessing the 'General Health Questionnaire' and 'Center of Epidemiological Studies Depression Scale' for depression screening: stroke and cancer patients

SCE Ifeagwu[1], S Ayis[1]
[1]King's College London, London, United Kingdom

Background: The English Longitudinal Study of Ageing (ELSA) is a health survey that entails information in waves on British people aged 50 and over from 1998 to 2011 (waves 0-5). Depression was assessed by the General Health Questionnaire (GHQ-12) and Center of Epidemiological Studies Depression Scale (CES-D8).
Aims: (1) To investigate the internal consistency, criterion validity and predictive power of the CES-D8 and GHQ-12 within waves 1 and 3 of the survey, where the two scales were used simultaneously for screening. (2) To assess the agreement of the two scales in case identification.
Methods: 12,099 respondents were screened within wave 1 and 9,771 in wave 3. Cronbach's alpha analysis was used to investigate the internal consistency between both instruments and the predictive power was calculated via STATA 12. ROC curve analyses were used to evaluate the optimal cut-off point of both scales and to assess the accuracy in identifying depression in patients with different comorbidities.
Results: Excellent item consistency and good criterion validity were shown by the two scales. Optimal cut-off scores were ≥3 for CES-D8 and

≥5 for GHQ-12. The predictive power for depression by comorbidities was relatively low, with an AUC of 0.69 and 0.63 for cancer and stroke respectively.
Conclusion: The study confirmed the reliability and validity of both screening instruments. Agreement of the two scales is between moderate to good, highlighting fairly high rates of false negatives and false positives if assessment relies on either. Caution is necessary when one scale is to be used.

P2.1.92
Investigating trend in the rate of suicide using regression methods in Hungary between 1963 and 2011

AM Laszlo[1], J Csicsman[2], F Bari[1], TA Nyari[1]
[1]University of Szeged, Szeged, Hungary, [2]Budapest University of Technology and Economics, Budapest, Hungary

Introduction: Rates of suicide have been analyzed from multiple points of views in Hungary in the last decades. However, descriptive rates have only been reported. An epidemiological study was carried out to characterize the pattern of annual Hungarian suicide rates during a long period.
Methods: Annual Hungarian suicide rates (per 100,000 population) were defined using gender, age group and suicide method based on tables published in Demographic Yearbooks and population data by gender and age group from Hungarian Central Statistical Office between 1963 and 2011.
Trends were determined using joinpoint regression analyses to segment the investigated period. Beyond, trends and relative risks for suicide rates were examined using negative binomial regression models overall and by gender, age group and suicide method.
Results: Significant peak in 1982 and nadir in 2006 were found in annual Hungarian suicide rates, which shifted to 1981 in females and 1983 in males using joinpoint regression in 1963-2011. Suicide rates remained constant for males after 1983. Different segmented patterns were observed for suicide rates in age groups.
Overall 178,323 suicides were committed in Hungary during the investigated period (50,265 female and 128,058 male). Suicide rates were significantly higher in males than in females based on relative risks calculated by negative binomial regression methods overall, in all age groups and most suicide methods.
Conclusion: Our findings support the results of joinpoint regression models for suicide frequencies first time in Hungary and give internationally comparable results.
Acknowledgement: This research was supported by grants TÁMOP-4.2.4.A/ 2-11/1-2012-0001, TÁMOP-4.2.2.A-11/1/KONV-2012-0052.

P2.1.119
Health-related mortality predictors among Krakow older citizens. 25-year follow-up study

A Pac[1], B Tobiasz-Adamczyk[1], M Florek[1], M Brzyska[1]
[1]Chair of Epidemiology and Preventive Medicine JU MC, Kraków, Poland

The aim of this study was to assess the predictive ability of self-reported health status on 25-year all-cause mortality in community-dwelling older people in Krakow. Poland.
The study population consisted of 2432 respondents, aged 65 and over, resident in Krakow city centre who were recruited in years 1986-1987. Mortality data were collected on regular basis from Vital Municipal Records up to September 31st, 2011.
During the 25-year observation period 2210 deaths were observed. To test the associations between health status and all-cause mortality six parametric survival models and flexible Royston-Parmar survival models were tested. For the final analysis the Royston-Parmar proportional hazard model (df=5) with time-dependent covariates was choosen. The health status was assessed using the self-reported questionnaire data. All results were adjusted for known mortality predictors (demographic, life-style, functional activity).
Among women, we have observed that higher mortality was related to diabetes (HR=1,82; 95%CI: 1,44-2,30), coronary heart disease (CHD) (HR=1,32; 95%CI: 1,11-1,58), hypertension (HR=1,31; 1,10-1,55) and asthma (HR=1,25; 95%CI: 1,00-1,57). No time-dependent effect of those predictors was observed. For men, the most important health-related predictors were diabetes mellitus (HR=1,59; 95%CI: 1,14-2,22) and asthma (HR=1,70; 95%CI: 1,24-2,34) as well as CHD (HR=1,36; 95%CI: 1,07-1,75) for this last predictor time-dependent effect was observed.
The impact of health-related mortality predictors among elderly Polish respondents was stable over 25-year follow-up. The only exeption was CHD among men - its impact was strongly time-dependent and significant only during the first few years of observation.

P2.1.129
An IRT longitudinal model for graded repeated responses: IADL and ADL hierarchy and functional dependency trajectories in the elderly

A Edjolo[1,2], C Proust-Lima[1,2], F Delva[1,2], J-F Dartigues[1,2], K Pérès[1,2]
[1]ISPED - University of Bordeaux, Bordeaux, France, [2]INSERM U897, Bordeaux, France

This study aimed at describing the hierarchy of a combined Lawton and Katz Instrumental (IADL) and basic (ADL) activities of daily living scale, and the trajectories of functional dependency before death in the elderly population by using a longitudinal Item Response Theory (IRT) model.
A 2-parameter probit IRT model combined with a mixed model for repeated graded responses was performed on a sub-sample of 3238 dead community dwellers aged 65 years and over at baseline in 1988 from the Paquid prospective cohort on brain and functional ageing. IADL and ADL were collected at home every 2 to 3 years over 22 years on a 3-point rating quotation. The model investigated the 11-scale items sequence and the functional nonlinear trajectories adjusted for education and gender.
A hierarchy of the combined scale was confirmed with early losses in half of IADL (shopping, partial transporting, finances and telephoning), last losses in ADL (total toileting, continence, eating and transferring), and an overlapping of concomitant IADL and ADL in the middle of the dependency continuum. The more discriminating items were basic activities in bathing, toileting, dressing and eating. The findings on functional trajectories showed a persistent postponement of functional dependency brought by education in men, but not in women.
IRT model shows that the ability to perform daily tasks is hierarchically affected from shopping to transferring disabilities. An in-depth understanding of this sequence provides an early warning of functional decline or a signal to continue further functional assessments.

P2.1.147
Factors associated with under five child mortality in mothers´ employed in agriculture, India

R Singh[1]
[1]HMC, Doha, Qatar

Objective: To examine factors associated with under five child mortality in mothers's employed in agriculture.
Design, Setting and Population: Data was retrieved from 2005-06 NFHS-3 in India (2008). The study population constituted a national representative cross-sectional sample of single children aged 0 to 59 months and born to mothers aged 15 to 49 years employed in agriculture from all 29 states of India. Demographic and socio-economic variables were considered as covariates in the Cox Proportional Hazard model.

Main Outcome Measure: Under five child mortality was the primary end point.
Results: Increase in Mother's age corresponded with an increase in child survival. Breastfeeding increases survival significantly by 71% (HR 0.29, 0.24-0.34, p=0.001). Standard of Living improved 22% child survival with medium (HR 0.78, 0.65-0.93, 0.005) and 43% with high (HR 0.57, 0.44-0.74, p=0.001) standard of living in comparison to low standard of living. Received prenatal care and breastfeeding health nutrition education were found significant protective factors for child mortality (HR 0.42, 0.31-0.67, p=0.001) and (HR 0.45, 0.31-0.67, p=0.001) respectively.
Conclusion: In a nationally representative sample of households in India, mother's age, breastfeeding, standard of living, prenatal care and breastfeeding Health Nutrition education were associated in reducing child mortality.

P2.1.161
Disease evolution of spinocerebellar ataxia type 2 patients: interruption of follow-up considerations

S Tezenas du Montcel[1,2], S Molho[2], H Jacobi[3], T Klockgether[3]
[1]UMPC - Paris 6, Paris, France, [2]Groupe Hospitalier Pitié-Salpétrière, AH-HP, Paris, France, [3]University Hospital of Bonn, Bonn, Germany

Aim: Spinocerebellar ataxia type 2 is a genetically heterogeneous ataxia disorder with a well-established clinical phenotype. A European cohort was formed to establish the disease evolution of some spinocerebellar ataxias. Follow-up was continued in order to establish long-term follow-up. However some patients were lost of follow-up due to center or patients reasons. The aim of the current study is to establish long term disease evolution taking in account the follow-up interruptions.
Method: Patients were included between May 2005 and September 2006 from 18 European centers and followed on a yearly basis for the first 3 years and then based on center practice. All data closed at 2013/08/15 were analyzed. Linear mixed models were used to analyze the longitudinal data.
Results: From the 146 patients followed at least one year (2 to 8 visits, mean: 2 years ± 2), 26% patients were followed less than 3 years: 11% patient withdrawn consent, 27% patients interrupted for disease reasons. The patients followed less than 3 years appeared more severe at baseline. The missing completely at random hypothesis was rejected. As a linear evolution could be evoked, we considered a missing at random process. Using pattern mixture models, we estimated that the disease evolution was of 1.46 ± 0.08 per year and 1.41± 0.08 after adjustment by the initial score. Age at onset and the genotype were associated with the disease evolution.
Conclusion: Taking in account the missing process we were able to estimate the unbiased disease evolution and the determinants of this evolution.
For the Eurosca Group.

P2.2 Methodological issues and case studies in epidemiology

P2.2.16
Use of propensity score in plastic and reconstructive research; rare complication events in elective surgeries

S Bagher[1], T Zhong[1,2], SWMC Maass[1], AC O'Neill[1,2], SOP Hofer[1,2]
[1]Plastic & Reconstructive Surgery, University Health Network, Toronto, Canada, [2]Plastic & Reconstructive Surgery, University of Toronto, Toronto, Canada

Aim: This study introduces propensity score (PS) methods and their application in studies with small sample size and rare events. A real-case dataset was evaluated, exploring two breast reconstruction (BR) methods (MS-Tram and DIEP) and their postoperative complication rates.
Method: Data was pulled from a prospectively maintained institutional database. Regression adjustment with and without PS method was studied. Complications were categorized as abdominal or breast. One PS was estimated for both breast and abdominal complications. The choice of variables to include in the PS model was examined (treatment, outcome or both); final variables used in the estimation of PS were associated with both treatment and outcome.
To capture the effect of other key covariates their association was examined in the final multivariate model and they were not included in the treatment effect model. Inclusion of interaction terms was examined. For missing data other than outcome, a multiple imputation procedure was performed. Complete cases analyses were performed as sensitivity test.
Results: There were 83 (28%) complications, (20% breast; 8% abdomen). Using PS, the adjusted odds of abdominal complications were still significantly higher in free MS-TRAM vs. DIEP flaps (OR=2.73). With prior chemotherapy, increase in BMI significantly increased abdominal complications (OR=1.16).There was no significant association between reconstruction method and breast complications; diabetics had significantly increased breast complications (OR=4.19).
Conclusion: To handle the risk of calculation bias, PS methods can be used in evaluating elective surgeries with rare events. PS analysis indicated significantly higher abdominal complications in free MS-TRAM compared to DIEP flaps.

P2.2.29
Pharmacoepidemiological characteristics of influenza-like illness in hospitalized children in Russia: a case-control study

VA Bulgakova[1], VF Uchaikin[2], VV Maleev[3], OV Shamsheva[2], EA Osipova[4], IV Prostyakov[4]
[1]Scientific Centre of Children's Health, Moscow, Russian Federation, [2]Russian National Research Medical University, Moscow, Russian Federation, [3]Central Research Institute of Epidemiology, Moscow, Russian Federation, [4]JSC Pharmstandard, Moscow, Russian Federation

Objectives: To evaluate pharmacoepidemiological picture of the course of influenza and other acute respiratory infections (ARI) in children during the first post-pandemic season.
Methods: A retrospective analysis of medical histories of patients hospitalized with influenza and other ARI was performed. Using the random sample method 2044 case histories of patients aged 1-18 years were selected. All the data were entered into a database for statistical analysis. Patients with laboratory confirmation of another inflectional disease were not included in the study. Continuously distributed variables were compared using criterion Mann-Whitney, categorical data or proportions were compared using $\chi 2$.

Results: Most of the patients (56%) came to the hospital in early terms and in a timely manner has received antiviral therapy. Flu-positive patients experienced more chills, myalgia, nasal congestion, dry cough, conjunctival injection, dyspnea than patients withother ARI ($\chi 2$, $p<0,01$). The most frequent complications of admission were bronchitis (13.7%), pneumonia (5.6%), sinusitis (1.9%).
For therapy of the most commonly prescribed Umifenovir (Arbidol): at the pre-hospital stage 10.2%, in the hospital from 65.2% in monotherapy, and at 14.3% in combination with Oseltamyvir (2.6%) and recombinant interferon (11.7%).
The early beginning therapy of influenza (Umifenovir, Oseltamyvir) to minimize the risk development of complications and showed the effective ness of the basic criteria.
Conclusion: Results complement the criteria for a rational and effective choice of the drug. An early beginning of antiviral therapy (Umifenovir or/ and Oseltamyvir) is one of the major factors ensuring the effectiveness of influenza therapy and reduction of the risk for complications.

P2.2.41
Sensitivity analysis for possible bias due to event-dependent observation periods in self-control case series analysis
MAJ de Ridder[1], GMC Masclee[1], CN Dodd[1], S Romio[1,2], MCJM Sturkenboom[1]
[1]Erasmus University Medical Center, Rotterdam, The Netherlands, [2]University of Milano-Bicocca, Milan, Italy

The self-controlled case series (SCCS) method analyses the association between a time-varying exposure and an event, using data on successive time episodes of cases only, including information on exposure status and occurrence of an event. A key assumption is independence between total length of observation and time of event. To allow event-dependent censoring, e.g. event-related death, a modified method was presented (Farrington 2011). After modelling the censoring process, weights are allocated to all episodes, depending on sequence. However, when using public health databases, this sequence may be unknown due to removal of dates for privacy regulations. We did sensitivity analyses to investigate possible bias caused by event-dependent observation times in a study on the effect of concomitant drug use on the risk of upper gastrointestinal bleeding.
Survival after the event was modelled with a mixture of an exponential (short-term) and a Weibull (long-term survival) distribution, with parameters depending on age at event. Random and extreme sequences of episodes were generated and estimates of unmodified and modified SCCS analyses using these different sequences studied.
Compared to unmodified, estimated RRs using both random and extreme sequences were moderately decreased and most were within the original 95% CI. As example, extreme RRs for concomitant use of non-selective NSAIDS and steroids were 2.81 and 4.42, random RRs were between 3.69 and 3.82, versus unmodified RR 4.60 (95% CI 2.43 - 8.68).
Using the modified SCCS analysis we were able to show in our data small bias due to ignoring event-dependency of the observation time.

P2.2.52
Smoking statistics in the mid-1920s birth cohorts
I Funatogawa[1]
[1]The Institute of Statistical Mathematics, Tokyo, Japan

Early smoking initiation may greatly increase mortality, but the subsequent deaths from smoking typically become evident in elderly individuals. This obscures the impact of early smoking on mortality in most nations. Although mortality in older ages has not yet been observed for younger birth cohorts, the mid-1920s birth cohorts are currently in their late 80s.
Therefore, we showed changes with age in smoking statistics (smoking initiation, smoking prevalence, and lung cancer mortality) of the 1925 birth cohorts of men and women in Japan and the United States (Funatogawa et al. Bull WHO 2013). In this study, we further compare changes in smoking statistics of the mid-1920s birth cohorts across several countries, including Japan, the United States, Great Britain, Spain, Italy and Norway based on WHO mortality database and literatures. Spanish women showed extremely low smoking prevalence and low lung cancer mortality which levels are similar to those in non-smokers. Japanese and Italian women showed modest smoking prevalence and modest lung cancer mortality. Although Norwegian women showed higher smoking prevalence, lung cancer mortality was similar to those in Japanese and Italian women. American, British and Italian men showed highest lung cancer mortality. American and British men showed high proportion of early smoking initiation and high smoking prevalence in young adulthood. Italian men showed lower smoking prevalence compared to the other two countries. Generally smoking initiation and prevalence are corresponded to lung cancer mortality.

P2.2.53
Non-monotonic trends in smoking statistics
I Funatogawa[1]
[1]The Institute of Statistical Mathematics, Tokyo, Japan

Many studies concerning smoking and health implicitly assume smoking habit spreads monotonically and then declines. However, Japanese women showed a long-term (i.e., six or seven decades) trend with a decrease and increase in smoking prevalence with birth year, instead of the increase and decrease (Funatogawa et al. Bull WHO 2013). Short-term decreases and increases with birth year (i.e., within two decades) were also seen in Japanese men and British men and women (Funatogawa et al. Bull WHO 2013, Funatogawa et al. BMJ Open 2012).
In this study, we further examine changes in age-specific smoking indicators (smoking initiation, smoking prevalence, lung cancer mortality) with birth year in other nations based on WHO mortality database and literatures. Chinese women showed a trend with an increase, decrease and increase again in lung cancer mortality with birth year.
Although a long-term decreasing trend in smoking prevalence with birth year was seen, the mean age of smoking initiation became younger with birth year. The effect of early smoking initiation on mortality has not yet been fully observed due to the long time lag. Smoking habit does not necessary spread monotonically.
Therefore, age-specific smoking measures should be paid more attention than summary smoking statistics which ignore non-monotonic trends with birth year.

P2.2.65
Exemplifying the usefulness of combining difference and equivalence tests in spatial maps
T Waldhör[1], H Heinzl[2]
[1]Medical Univiversity of Vienna, Center for Public Health, Epidemiology, Kledering, Austria, [2]Medical University of Vienna, CeMSIIS, Vienna, Austria

A choropleth map provides a widely accepted method for graphically presenting spatially distributed health data. Quite often local health authorities and the general public are concerned whether the health situation in certain areas is relevantly worse than a reference standard.
Highlighting statistically significant areas allows the conclusion that the health situation in these areas is better or worse than the reference value. However, this approach does not allow concluding that the situation in some areas is sufficiently close to the reference value. Since statements with regard to the relevance of observed extreme results are of interest for health policy making as well, a combined integration of statistical dif-

ference and equivalence tests into choropleth maps has been suggested. The approach will be graphically exemplified comparing mortality rates for the circulatory system, respiratory system and Alzheimer disease. Data provided by Statistics Austria covering all 121 administrative Austrian districts for the years 1998 to 2004 will be used and compared with traditional choropleth maps provided by Statistics Austria.
Choropleth maps, which only show the variable of interest and their corresponding difference test results, can easily misguide local health authorities and the general public. The adding of equivalence test results enables a better understanding of regional health care results as the issue of relevance is explicitly addressed through a pre-defined equivalence range.

P2.2.79
Medical use of allergic rhinitis under two healthcare system in South Korea

B-H Jang[1], I-H Choi[1], J-S Park[1], J-H Park[2], S-H Sun[3], Y Shin[1], S-G Ko[1]
[1]College of Korean Medicine, Kyung Hee University, Seoul, Republic of Korea, [2]College of Korean Medicine, Gachon University, Seoul, Republic of Korea, [3]College of Korean Medicine, Sangji University, Seoul, Republic of Korea

In South Korea, there are two healthcare system - Western Medicine (WM) and Traditional Korean (TKM). This study aims to investigate medical use of allergic rhinitis (AR), especially under two healthcare system.
We analyzed characteristics of AR patients from 2011 National Patients Sample (NPS) data of Health Insurance Review & Assessment Service. AR patients were defined as those who were diagnosed with AR as the primary disease (J30.x from ICD-10). We analyzed by dividing into three groups - visiting only the WM institute (A), and both WM and TKM (B), and only the TKM (C).
Among 1,375,751 patients in 2011 NPS data, AR patient was 12.5%. A female/male ratio was 1.20, and Average age was 31.8. More than half AR patients were diagnosed as J30.4 (Allergic rhinitis, unspecified), the proportion of primary clinics was 88.2%, and the most frequent in September and October.
Outpatient clinics were visited by 99.9% of AR patients. Group A was 97.7%, and B was 0.9%, and C was 1.4% of AR patients. In Group A, 36.2% was below 20 age, but 66.5% in B, and 59.1% in C. The annual average cost and the mean visit time per person were the highest in group B. 5.6% of group A had diagnostic test of AR at least one time, and 10.8% of group B had it.
According to this study, TKM was not widely used by AR patients in South Korea. In order to reflect the actual medical field, further study including uninsured items is needed.

P2.2.85
Multidimensional outcome. Does the interpretation change with the analysis method?

C Klersy[1], L Scudeller[1], P Maras[2], S Doimo[2], A De Silvestri[1], C Tinelli[1]
[1]IRCCS Fondazione Policlinico San Matteo, Pavia, Italy, [2]Ospedali Riuniti Trieste and University of Trieste, Trieste, Italy

Binary outcome measures that indicate the presence or absence of certain medical conditions are widely used in epidemiologic investigations. Moreover, these studies frequently measure an array of such indicators for different medical conditions to make an overall assessment, making it multidimensional. The situation where multiple binary outcomes are simultaneously assessed on the same individual presents some basic methodological problems in that proper statistical modeling here should account for the following features: A) Each individual contributes to multiple outcomes. Thus, the different outcomes are likely to be correlated. B) These multiple outcomes possibly measure the same underlying condition or construct. C) Outcome-specific exposure (i.e., which specific outcomes are associated with the exposure of interest) may still be of scientific interest in many situations.
To assess the association of any risk factor/exposure with these outcomes, several analysis methods are found in the literature: A) Each outcome is considered separately and independently (binary); B) Each outcome is considered in the framework of repeated measures (binary); C) All outcome components or a certain (acceptable) number of outcomes are satisfied (binary); D) The number of satisfied components is computed (count); E) The number of satisfied components is computed (few, ordinal); F) All outcome components are expression of an underlying (latent) concept (continuous).
Starting from a motivating example in outcome research for the secondary prevention of ischemic heart disease, with the primary analysis, based on structural equation modeling, following scenario F, we will discuss the implications of using alternative methods for the interpretation of results.

P2.2.105
Sensitivity analysis for the misclassification of competing outcomes in a cohort study in Japan

MN Mieno[1], N Tanaka[2], M Sawabe[3], T Arai[4], S Ishikawa[1]
[1]Jichi Medical University, Shimotsuke, Japan, [2]National Center for Global Health and Medicine, Tokyo, Japan, [3]Tokyo Medical and Dental University, Tokyo, Japan, [4]Tokyo Metropolitan Geriatric Hospital, Tokyo, Japan

Background: In epidemiological and clinical studies, causes of death are often measured as outcome. However, the risk estimates are susceptible to bias due to misclassification of causes of death, and the degree of misclassification actually remains unknown. In this study, we conducted a sensitivity analysis for quantifying the magnitude of bias to the risk estimates for three competing causes of death, such as cancer, cardiovascular disease and other causes of death.
Methods: The motivated data was from a multi-center population-based cohort study in Japan (The Jichi Medical School cohort study, n=10,692) and we analyzed whether the low lipoprotein(a) [Lp(a)] concentration was related to mortality using competing risks approach. Based on the dataset, we conducted simulations for the sensitivity analysis with several settings assuming differential or non-differential misclassification proportion of cause of death as 10%, 20%, 30% and 40%, including covariates which could relate to the misclassification mechanism.
Results: The observed risk ratio (low / not-low Lp(a)) for cancer as cause of death was 1.484 [95%CI: 1.147-1.919]. If the misclassification was non-differential and not with over-recording of a specific cause of death, the cumulative incidence rates and risk ratios from Fine & Gray proportional hazards model were not changed so much, whereas in the other differential cases, the estimates were changed depending on the degree of misclassification.
Conclusions: When the outcome misclassification are expected, careful consideration must be given to the interpretation of the results and sensitivity analysis like demonstrated here is useful for evaluating the robustness of the results.

P2.2.115

Timing for definitive cure in clinical trials for visceral leishmaniasis

R Omollo[1,2], N Alexander[3], N Omolo[2,4], P Oleche[2], M Wasunna[1], T Edwards[3]
[1]DNDi, Kenya Medical Research Institute, Nairobi, Kenya, [2]Maseno University, Private Bag, Maseno, Kenya, [3]London School of Hygiene & Tropical Medicine, London, United Kingdom, [4]Jaramogi Oginga Odinga University of Science & Technology, Bondo, Kenya

Background: Visceral Leishmaniasis (VL) is a parasitic neglected tropical disease, endemic in developing countries. Treatment outcomes in VL clinical trials are measured at two time points: a) initial cure at end of treatment (EOT) and b) definitive cure (DC) measured commonly at six months post EOT. Additional assessment is commonly conducted at either one or three months post EOT for patient monitoring. This paper investigates justification for shortening of the six month follow-up time point in the assessment of DC.
Methods: A three state Markov model has been applied to estimate the transition probabilities for DC using data from a phase III clinical trial conducted in East Africa to compare the safety and efficacy of three different treatment regimens. At baseline all patients were untreated but their status changed to either treatment failure (TF) or success (TS) at EOT, month 3 or 6 follow-up with TF considered as an absorbing state.
Results: Conditional on a patient ever having a TS, 98.5% of observations have a TS outcome at the next visit while that of ever having a TF is 61.4%. From the transition probabilities, approximately 1% of patients change cure status between month 3 and 6. Overall, there was high stability for the treatment outcomes (>97%).
Conclusion: Advantages of shortening follow-up time include a possible reduction in loss to follow-up and time to availability of effective new treatments. Limitations to be addressed in simulation studies include the possibility of false negative results and differences in relapse rates between treatment regimens.

P2.2.127

Etiological age-components of cervical cancer in Finland in 1953-2011

K Seppä[1], J Pitkäniemi[1], N Malila[1], M Hakama[1]
[1]Finnish Cancer Registry, Helsinki, Finland

The objective of the study was to assess whether age incidence of cervical cancer in Finland is consistent with two etiological components and how the changes in the incidence is concordant with the organized screening programme.
The study utilized a population-based register of cervical cancer cases from the Finnish Cancer Registry diagnosed at 20-84 years of age in 1953-2011 and female population at risk. Age-specific incidence of cervical cancer was estimated by using a Poisson regression model with the assumption of two normally distributed latent components. A hierarchical Bayesian model was applied and life time net risks and crude numbers of cancer cases of the two components were estimated from the joint posterior. Before the screening started the life time net risks were 0.5% for younger and 1.3% for older (RR=0.43, 95% CI 0.29-0.61) with actual cases of 154 and 206 per year, respectively. The component of incidence occurring in younger women disappeared in 1970s after the organized mass screening program was initiated (in 1973-1977, RR=0.02, 95% CI 0.00-0.08).
Since that, the risk for younger increased to 0.2% whereas the risk for older decreased to 0.5% in 2008-2011 (RR=0.58, 95% CI 0.25-0.94) with actual cases of 80 and 61 per year, respectively. Existence of the two components is likely to be due to different etiological exposures.
The trend in risk of the both components followed closely both the effects of organized screening and changes in sexual mores. The increase in the first is consistent with increase in HPV exposure.

P2.2.132

Location-scale tests for non-negative data with skewed distribution, with focus on parasitology research

J Reiczigel[1], S Yehia[2]
[1]Szent Istvan University, Faculty of Veterinary Science, Budapest, Hungary, [2]Tanta University, Faculty of Science, Tanta, Egypt

Two-sample comparison of parasite infection data is usually made by location tests. As more infected samples have both higher mean and higher SD, location-scale tests might be more powerful. We investigated this by simulation. Our results may also be relevant in other fields where skewed distributions are common, e.g. in the analysis of health care cost data.
Methods: We compared Cucconi's location-scale test (CU) with 3 commonly used location tests: Welch-t-test (WT), Mann-Whitney-test (MW), and bootstrap-t-test (BT), for 5 right-skewed theoretical distributions and 5 empirical parasite distributions. Sample sizes varied from 10 to 100. Alpha was compared assuming that the two samples came from identical distributions. Power was compared assuming that differences would appear at higher infection levels, reflecting to the observation that even in heavily infected populations many hosts (those with good defence) remain free or almost free.
Results: Alpha error rate of MW, CU, and BT was acceptable (<6% at nominal 5%) if ratio of sample sizes (SSR) was below 2, and slightly elevated (<7.3%) for SSR=3. WT showed alpha=8.9% for SSR=3, and 7.2% for SSR=2, therefore we excluded WT from the power comparison. Power depended strongly on the distribution. For theoretical distributions CU performed best. For empirical parasite distributions there was no clear "winner" but both CU and BT had considerably higher power than MW.
Conclusion: Location-scale tests may be useful in parasitology. CU was more powerful than the location tests, among which BT was best. All tests had elevated alpha levels for unbalanced designs.

P2.2.155

The challenges of conducting a multidisciplinary trilogy of studies: diagnostic accuracy & use of the SeHCAT test in England

JA Summers[1,2], A Pascoal[2], S Keevil[2], C Lewis[2], G Vivian[3], R Logan[3], V Cornelius[1], J Peacock[1,2]
[1]King's College London, London, United Kingdom, [2]King's Health Partners AHSC, London, United Kingdom, [3]King's College London Hospital NHS Foundation Trust, London, United Kingdom

Bile Acid Malabsorption (BAM) is common in many diarrhoea conditions, however, robust data on the prevalence of BAM and care pathways for diagnosis and treatment of the condition do not exist.
Confirmation of BAM can be made using SeHCAT (tauroselcholic [^{75}selenium] acid), a radiolabelled synthetic bile acid. The SeHCAT test measures the retention of radioactivity in the patient following administration of a capsule containing SeHCAT. Bile acid sequestrants (BAS) are a treatment option for BAM.
The aim of this research is to gather and assess evidence regarding the use of SeHCAT in the diagnosis of BAM in NHS centres.
Several challenges have been identified in conducting this ongoing research:
- Patchy adoption of SeHCAT testing across NHS centres;
- Limited information on outcome of patients diagnosed as BAM negative;
- Poor patient adherence to BAS treatment;
- Current lack of evidence on clinical value of the SeHCAT test;
- Expertise is required in various areas: nuclear medicine, gastroenterology, study design, statistics and database development and management;
- Statisticians working and coordinating successfully with clinicians in an area with little background knowledge.

Based on the identified challenges, a trilogy of sequential studies is being

undertaken to address the research aim. This consists of a) Retrospective Audit based on existing data, b) Prospective Survey (to inform continuum of results, prevalence and severity of BAM) and c) Multi-Centre Prospective Observational Study (comparing diagnostic accuracy of SeHCAT, help to develop diagnostic threshold of SeHCAT and assess efficacy of BAS treatment).

P2.2.156

The value of using modern epidemiological approaches in studying past influenza pandemics: combining history, war and statistical methods

JA Summers[1]
[1]*King's College London, London, United Kingdom*

Knowledge of pandemic influenza risk factors, the likely speed and pattern of spread, and the expected impact on a healthcare system, is based on the study of past pandemics. Therefore the ability to transform historical records into a quantifiable form for epidemiological/statistical analysis is imperative for such research.
As the 1918-19 H1N1 influenza pandemic occurred during World War One (WW1), many of the best-documented aspects from this period are held within military files.
Several recent studies have used military records from diverse settings (such as New Zealand (NZ), Australian, British, Japanese, US, and Canadian sources).
These records are relevant to understanding influenza epidemiology, enabling exploration of mortality risk for sub-populations.
In particular, a range of statistical methods (i.e. multivariate logistic regression) has been used to assess variables extracted from enlistment data, using both cohort and case-control designs to assess the following: occupational class, geographic variables, health status, military variables, ethnicity, body mass index, marital status and even complexion.
The use of archival military sources for epidemiological research is growing; both because of the increasing availability of records, but also because of the increased scientific need to better understand the nature of influenza pandemics.
Further research into past influenza pandemics will improve our understanding of influenza, especially regarding the control of its impact and the optimal targeting of limited health care resources during a pandemic. Additionally, given the severity of the 1918-19 pandemic, this event could arguably be described as a worse-case scenario for guiding future population-based pandemic planning.

P2.2.163

Assessment of neighbourhood effect on neonatal mortality: translation of area level variance in odds ratio scale in multilevel logistic regression

B Thakur[1], V Sreenivas[1], S Dwivedi[1], A Pandey[2,3]
[1]*All India Institute of Medical Science, New Delhi, India,* [2]*Indian Council of Medical Research, New Delhi, India,* [3]*National Institute of Medical Statistics, New Delhi, India*

Introduction: Logistic regression is frequently used in epidemiological and public health research to measure the binary outcome. The variability at different levels is not directly comparable in multilevel model. Quantifying area-level variance in a meaningful way is a challenge in multilevel logistic regression.
Method: We obtained individual and district level information on the binary outcome neo-natal mortality from District Level Household Survey-3. The exploration of data structure confesses the consideration of only two-level structure in analysis, conceptualized as children nested within districts.
Estimations of Variance Component Model (empty model) and Random Intercept Model in multilevel logistic analysis were carried out. The median odds ratio translates the area level variance on the odds ratio scale.
Result: The median odds ratio was equal to 1.60, in the empty model which shows if a person moves from one district to another district with a higher probability of neonatal mortality, their risk of mortality will increase by 1.6 times, when randomly picking out two persons in different districts. Adjusting the individual effect in random intercept model, this ratio reduced to 1.54. Area level variance and Intra-class correlation were 0.246 and 0.067 in the empty model as well as 0.210 and 0.059 in the subsequent model respectively.
Conclusion: The usual odds ratio are not proper interpretable for district-level covariates because it is impossible to make comparison within district. As MOR quantifies cluster variance in terms of odds ratios, it is comparable to the fixed effects odds ratio and can be useful in epidemiological studies.

P2.2.182

Regression models for rare events – stroke mortality rates over the last 30 years in Hungary

K Virág[1], T Nyári[1], K Boda[1]
[1]*University of Szeged, Szeged, Hungary*

Modeling the number of occurrences of a disease is a common task in epidemiological investigations. When the dependent variable describes the counts of rare events then its distribution is skewed to the right, hence the use of the ordinary linear regression is inappropriate.
The natural model for count data is a Poisson regression, which fits Poisson distribution to the number of occurrences (or rates) of the event. Poisson model assumes that the mean is the same as the variance. In epidemiological studies we often find that the variance is greater than the mean, therefore the data is not well modeled by Poisson regression. Negative Binomial regression can deal with overdispersion.
The aim of this study is to compare different count regression models for rare events using SAS 9.2, SPSS 22.0, STATA 9 and R statistical softwares. We perform simulations to compare the coverage probabilities of the confidence intervals given by different methods under different conditions.
As an empirical application, we analyze stroke mortality rates (ICD: 1981-1995: 430-438; 1996-2010: I60-I69) in Hungary between 1981 and 2010.
Acknowledgement: Katalin Virág was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP 4.2.4.A/2-11/1-2012-0001 'National Excellence Program'.

P2.2.183

Dependence of the effect of altitude on infant as well as maternal related variables on birth weight

T Waldhoer[1], K Klebermass-Schrehof[2]
[1]*Medical Univiversity of Vienna, Center for Public Health, Epidemiology, Vienna, Austria,* [2]*Medical University of Vienna, Dept. of Pediatrics, Vienna, Austria*

Decrease of birth weight with increasing altitude has been described numerously and shown for different populations as well as for different types of studies. Only few studies have attempted to test whether the effect of altitude depends on other risk factors. Corresponding results showed rather clear independence on other variables.
In this study we use about 2 million Austrian birth certificates in order to test for interaction terms with infant and maternal related variables as e.g. gestational age, sex, education, age of mother, year of birth, parity, time to previous birth.
Results show that there obviously exist significant as well as relevant interactions which can not be detected in small sized studies because of lack of power.

P2.2.185
Time series analysis of *Campylobacter* incidence in Switzerland
W Wei[1], G Schüpbach[2], L Held[1]
[1]University of Zurich, Zurich, Switzerland, [2]Veterinary Public Health Institute, University of Bern, Bern, Switzerland

Campylobacteriosis is the most common food-associated infectious disease in Switzerland since 1995. Contact with and ingestion of raw or undercooked broiler are considered the dominant risk factor for infection.
In this study, we investigate the temporal relationship between disease incidence among humans and prevalence of *Campylobacter* in broiler in Switzerland from 2008 to 2012. We use a time series approach to describe the pattern of the disease by incorporating seasonal effects and autocorrelation.
Our analysis shows that prevalence of *Campylobacter* in broiler, lagged by two weeks, has a significant impact on disease incidence in humans. Therefore *Campylobacter* cases in human can be partly explained by contagion through broiler meat. We also found a strong autoregressive effect among human infections, and a significant increase of infections during Christmas and new year holiday.
In a final sensitivity analysis, we corrected for the sampling error of broiler prevalence estimates which gave similar conclusions.

P2.3 Methods for handling missing data

P2.3.2
Multiple imputation is not necessary for performing analyses in pre-post studies
U Aguirre[1,2], I Arostegui[2,3], JM Quintana[1,2]
[1]Hospital Galdakao-Usansolo, Galdakao-Usansolo, Spain, [2]REDISSEC Health Services Research on Chronic Patients Network, Bilbao, Spain, [3]Applied Mathematics, Statistics Operational Research, UPV/EHU, Bilbao, Spain

Aims: Pre-post studies based on health related quality of life (HRQoL) variables are motivated to determine the potential predictors of the mean change of the outcome of interest. It is very common in such studies for data to be missing, which can bias the results. The use of Multiple Imputation (via Markov Chain Monte Carlo, MCMC) has been increased when handling missing data. However, it has been discussed whether only Complete Case (CC) with mixed models are also effective for this performance.
Methods: We compared CC analysis and MCMC methods to assess their performance for handling missing data under different situations (rate: 10% and 30%; mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR)). Moreover, in both cases mixed-models techniques were used. These strategies were applied to a pre-post study of 400 patients with chronic obstructive pulmonary disease (COPD). We analyzed the relationship of the changes in subjects' HRQoL over one year with clinical and sociodemographic characteristics. A simulation study was performed (500 and 1000 runs), where the standardized bias of the regression coefficient of the interaction between the Time effect and the covariate was computed.
Results: In both 500 and 1000 simulation-runs, CC with mixed models showed the lowest standardized bias coefficients for MCAR and MAR scenarios. However, in MNAR setting, both approaches provided biased coefficients.
Conclusions: MCMC has not additional benefit over CC when handling missing data for MCAR and MAR settings. There is no consensus in MNAR scenario.

P2.3.44
A new method for significance testing of categorical covariates after multiple imputation
I Eekhout[1,2], MA van de Wiel[1,3], MW Heymans[1,2,3]
[1]VU University Medical Center, Amsterdam, The Netherlands, [2]EMGO Institute for Health and Care Research, Amsterdam, The Netherlands, [3]VU University, Amsterdam, The Netherlands

In medical prognostic research, logistic regression analysis is frequently used. Unfortunately, missing data is common in these studies. As a solution multiple imputation is recommended, which generates multiple imputed datasets. Subsequently, logistic regression models are applied in each imputed dataset and finally parameter estimates are pooled using Rubin's Rules (RR). For significance testing of dichotomous and continuous covariates in these models, RR can easily be applied.
However, to consider whether a categorical covariate as a whole significantly contributes to the model, RR cannot be used. Instead, to obtain an overall p-value, Meng and Rubin (MR) proposed to pool the log likelihood ratio test statistics for each parameter and obtain the significance level from that pooled statistic. This procedure is complicated and not available in standard statistical software.
We propose a new method which is much easier to use with power at least equal to that of the MR method. Our method uses the median of the p-values of all separate likelihood ratio tests in each imputed dataset: the Median P-Rule (MPR). In a large simulation study, it was shown that for non-significant categorical covariates the type I error is controlled and the statistical power of the MPR was at least equal to that of the MR method for significant ones. An illustrative empirical data example showed similar results.
We recommend using the median of the p-values from the imputed data analyses (MPR). This method performs at least equally well as the MR method, but is much easier to apply.

P2.3.137
Missing categorical data: the influence of imputation technique on regression analysis in an opioid maintenance treatment setting
M Riksheim[1], J Røislien[1,2]
[1]Norwegian Centre for Addiction Research, University of Oslo, Oslo, Norway, [2]Department of Biostatistics, University of Oslo, Oslo, Norway

Missing data is a recurring topic in observational studies and can be decisive in some settings. In Opioid Maintenance Treatment (OMT) mainly one of two substitution medications is used to treat opioid dependence and best practice is discussed. Studies on the OMT population are important to optimize treatment, but with OMT patients being hard to reach, the issue of missing data is non-negligible.
For missing continuous data, several imputation methods have been proposed and extensively researched. For categorical data, however, no clear recommendations exist. In this methodological study we applied four different imputation techniques on missing categorical data to explore the influence of method choice on results in a subsequent regression analysis on data from the Norwegian OMT programme.
In the Norwegian OMT programme questionnaire data regarding patients' treatment status are collected annually. For the present study, we used data from the eastern region of Norway collected 2005-2010. The data comprised of 9039 questionnaires with 12 questions from 2886 patients. Missing ranged from 0% to 10% per question.
Four missing data techniques were tested: Expectation Maximization with Bootstrapping; Multiple Imputations by Chained Equations; Hot Deck Imputation, and Multiple Imputation using Latent Class.
The imputed data sets were tested in logistic regression analyses with type of medication as outcome and 11 covariates, including measures of social situation, age and sex. The imputation methods gave differences in both

parameter estimates and statistical significance, indicating that the manner in which missing data is handled is essential to obtain correct results in regression analysis.

P2.3.140

Imputation of an ordinal exposure derived from a semi-continuous variable with missing data: a simulation study

L Rodwell[1,2], H Romaniuk[1,2], KJ Lee[1,2], JB Carlin[1,2]
[1]Murdoch Children's Research Institute, Melbourne, Australia, [2]The University of Melbourne, Melbourne, Australia

Multiple imputation (MI) requires careful specification of the imputation model, with which there are often a number of possible methods. We focus on a specific, albeit relatively common, scenario where the analysis model includes an ordinal exposure variable with categories derived from a semi-continuous variable with missing data.
We based our simulations on weekly alcohol consumption data from a longitudinal study of adolescents. We varied the proportion of zeros in three different semi-continuous distributions (25, 50 or 75 per cent). The semi-continuous data were then categorised to represent levels of alcohol consumption: none, 1-10 units, 11-20 units and 21+ units. Finally, we created a binary outcome, with which higher alcohol consumption was associated with higher odds. We generated 2000 sets of 1000 observations. Within each of these we set a random 33% of observations for the alcohol variable(s) to missing.
We examined five imputation methods which involve either deriving the ordinal variable first: (1) projected distance based rounding , (2) ordinal logistic regression; or imputing the semi-continuous variable: (3) two-part model, (4) the 'just another variable' method, (5) predictive mean matching.
We assessed the performance of the imputation approaches by comparing the average estimates across 2000 simulations with those from a pseudo-population of 1 million observations.
All of the imputation methods performed reasonably well when used to estimate the association with the binary outcome, with over-coverage more common than under-coverage. We intend to investigate these approaches in more realistic conditions where the data are missing at random.

P2.3.169

Regularized approach for missing data problem

C-H Tseng[1], Y-H Chen[2]
[1]UCLA, Los Angeles, United States, [2]Academia Sinica, Taipei, Republic of China (Taiwan)

We present a framework to analyze missing data when the missing data mechanism is unknown. The regularized approach is used to accommodate both ignorable and non-ignorable missing data.
We investigate the impact of missing data mechanism uncertainty based on simulation and show that proposed method can provide stable and reliable estimates for both ignorable and nonignorable missing data. We apply our method to a longitudinal clinical trial of hypertension where nonignorably missing data were a concern.

P2.3.188

What if my doctor would be as receptive to innovations in therapies as to innovations in statistical methods?

A Güttner[1], Y Gong[2], M Akacha[1], F Hornig[1], P Mesenbrink[3], S Witte[1]
[1]Novartis, Basel, Switzerland, [2]Novartis, Shanghai, China, [3]Novartis, East Hanover, United States

For chronic auto-immune diseases such as psoriasis, the evaluation of long-term effects is important for physicians and patients with respect to treatment decisions. In long-term clinical trials, missing data due to treatment interruptions or discontinuation of studies have to be handled within the statistical analysis of treatment comparison or estimation of response rates or disease activity.
Statistical methods for imputation of missing data vary and improve over time and health authorities may change the acceptance of imputation methods. Within the same indication, this leads to publication of long-term data where different imputation methods were used.
Hence, comparison of different compounds, e.g. for the education of treating physicians and patients, although not studied in the same study, are not always straight-forward.
Non-statisticians may not understand different assumptions underlying the imputation methods or associated bias. For effective and safe treatments, the amount of missing data is moderate, but not negligible. For response variables in psoriasis, non-responder imputation is often used for short-term comparisons to controls, but leads to decreasing response rates over time. If the handling of drop-outs is not considered in the interpretation, a diminishing effect may be, incorrectly, attributed to the treatment. Next to non-responder imputation, last-observation-carried-forward, less stringent non-responder imputation, observed data only and multiple imputations are other options for dealing with missing data. Comparing these methods on the same dataset shows that decreases in response rates over time are driven by non-responder imputation whereas other methods are relatively indistinguishable, in particular show constant response rates over time.

P2.3.196

Anemia is a risk factor for poor cognitive outcome after ischemic stroke

VD Zietemann[1], FA Wollenweber[1], M Dichgans[1,2,3]
[1]Institute for Stroke and Dementia Research, Munich, Germany, [2]Interdisciplinary Stroke Centre, Munich, Germany, [3]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

Background: Anemia is common in patients with stroke and has recently been shown to be a risk factor for poor functional outcome and mortality after stroke. However, the impact of anemia on cognitive outcome after stroke remains unexplored.
Methods: 252 consecutively recruited patients with acute ischemic stroke and without pre-stroke dementia were included in this prospective observational study. Anemia was defined by the WHO criteria (hemoglobin concentration, 13 g/dl for men and 12 g/dl for women). Blood samples were taken on the first morning after admission between 6 and 9 AM. Cognitive outcome was assessed by the Telephone Interview for Cognitive Status (TICS) 3 months post-stroke. Ordinal logistic regression was used to adjust for confounders. Complete case (CC) analyses as well as sensitivity analyses (inverse probability weighting (IPW), multiple imputation (MI)) were performed to control for loss to follow up.
Results: 20% of the patients had anemia. 36% of patients with anemia and 14% of patients without anemia had missing TICS data. The proportion of subjects with missing TICS data was higher in patients with poor functional status and poor functional status was associated with poor cognitive status. Anemia was associated with poor cognitive outcome in CC analysis (n=206; OR=4.19, 95%CI: 1.40-12.55) and results were confirmed using

sensitivity analyses (n=252; IPW: OR=5.92, 95%CI: 1.80-19.42; MI: OR=3.12, 95%CI: 1.22-8.02).
Conclusions: Anemia is a risk factor for poor cognitive outcome 3 months after ischemic stroke. Two different sensitivity analyses accounting for missing outcome data confirmed the negative impact of anemia.

P2.4 Penalized methods in high- and in low-dimensional regression analyses

P2.4.23
Dimensional reduction in the flexible B-spline Cox model using functional principal components analysis

MA Benadjaoud[1,2,3], H Cardot[4], F de Vathaire[1,2,3]
[1]Radiation Epidemiology Group CESP - INSERM 1018, Villejuif, France, [2]Université Paris Sud, Le Kremlin-Bicêtre, France, [3]Institut Gustave Roussy, Villejuif, France, [4]Institut de Mathématiques de Bourgogne, Dijon, France

Radiation epidemiology is a rare events context, which restricts the number of estimated parameters in flexible survival models. This requires the use of a low-dimensional spline regression constructed from a reduced number of interior knots or through Penalized splines (P-Splines) where the choosing of optimal degree of smoothness is crucial and still an issue even if various selection method were proposed such as Akaike information criterion (AIC), Bayesian information criterion (BIC) or generalized cross-validation (GCV).
We propose a new dimension reduction technique based on a functional principal component analysis (FPCA) where the spline basis is replaced by a smaller number of score functions that summarize the effects of the initial interior knots sequence. We perform a simulation study in a flexible Cox regression context based on a dataset intended to reflect realistic radiation-epidemiological data.
We vary the shape of the dose-response function, the number of interior knots, the number of principal components incorporated in the model. For assessing performance, we consider the integrated mean squared error (IMSE). Finally, we apply the proposed method to investigate the exposure-response relation between the radiation dose to the thyroid and the radio-induced tumor risk.
The FPCA estimator minimizes the IMSE better than the P-splines estimator and our rule of thumb is that 6 or 7 interior knots and 3 or 4 functional principal components are adequate for several practical situations.
In conclusion, this study suggests the FPCA as a form of regularized estimation which could represent an alternative to the classical penalized approach.

P2.4.56
Firth's bias reduction method revisited: software implementation boosts application

A Geroldinger[1]
[1]Medical University of Vienna, CeMSIIS, Vienna, Austria

This presentation reviews David Firth's bias reduction method for maximum likelihood estimates of regression coefficients (Firth, Biometrika, 1993). Instead of correcting bias after estimation, Firth's approach prevents bias by introducing a penalty to the likelihood function. This allows one to compute reliable finite regression coefficients even in the situation of separation. Separation, sometimes also termed 'monotone likelihood', is frequently observed in models used in clinical biostatistics, including the logistic, the multinomial or the Cox model.
Firth's seminal publication was also followed by a series of papers discussing different aspects of the penalization; for instance issues of inference, connections with Bayesian methods, interpretability of the estimates, or the application to high-dimensional predictor space. Some of these developments are highlighted in our presentation.
After the Firth correction was made available in add-on packages for standard software, it has been widely used by many researchers to solve their small sample regression problems. Firth's correction is now also available in the standard distribution of SAS, further enhancing its accessibility. We will provide an overview of implementations of Firth's correction in statistical software, and will correlate the number of medical and non-medical citations of Firth's method with the release dates of these software implementations. This analysis reveals that besides methodological excellence, software availability is a very likely causative factor for getting many citations.

P2.5 Statistical methods for systems biology and genetics

P2.5.40 *Cancelled*
A two-stage approach to test for gene-gene interactions in family data based on within-family and between-family information

L De Lobel[1], K Van Steen[2]
[1]StatGent CRESCENDO - Ghent University, Ghent, Belgium, [2]Montefiore, Liege, Belgium

The search for susceptibility loci in gene-gene interactions imposes a methodological and computational challenge for statisticians due to the large dimensionality inherent to the modelling of gene-gene interactions or epistasis. In an era where genome-wide scans have become relatively common, new powerful methods are required to handle the huge amount of feasible gene-gene interactions and to weed out the false positives and negatives from these results.
One solution to the dimensionality problem is to reduce the data by preliminary screening of markers to select the best candidates for further analysis. Ideally, this screening step is statistically independent of the testing phase. To obtain two independent steps to test for associations in family data, we can split up the genotypic information in a between-family and within-family component as is done in the QTDT. Those two components are orthogonal so that one of the components can be used for screening and the other can be used for testing. The QTDT proposes a definition of these components for one locus.
In our research, we define analogous components for gene-gene interactions and investigate the proporties of this screening technique in different types of simulations.

P2.5.78
Topology-based pathway analysis of microarray and RNA-Seq data: an evaluation of existing methods

I Ihnatova[1], E Budinska[1,2]
[1]Institute of Biostatistics and Analyses, Brno, Czech Republic, [2]Masaryk Memorial Cancer Institute, Brno, Czech Republic

Pathway analysis methods for transcriptional microarray data analysis have reached their third generation, currently incorporating pathway topology information. However, high throughput parallel sequencing of transcriptome (RNA-Seq) has recently emerged as an appealing alternative to microarrays and becomes widely available.

As to our knowledge, there is no specific topology-based pathway analysis method for RNA-Seq data. Here, we present simple adaptations of the topology-based methods for RNA-Seq data and compare their ability to identify differentially expressed pathways on the example of real data. As a model we chose colorectal cancer (CRC), where we compare microsatellite instable (MSI) and microsatellite stable (MSS) tumors, which have distinct prognosis and specific transcriptional activity. We also compare the performance of our adapted methods applied on RNA-Seq data with the original methods applied on microarray data.
To this end we use four publicly available datasets from both The Cancer Genome Atlas (TCGA) database and The Gene Expression Omnibus (GEO) and discuss the number of identified pathways, the ranks of the pathways and the overlaps between individual methods.

P2.5.116
Contribution of alternative splicing variants to gene expression variation

W Ouwerkerk[1], AH Zwinderman[1]
[1]Amsterdam Medical Center, Amsterdam, The Netherlands

Alternative splicing of messenger RNAs provides cells with the opportunity to create protein isoforms of a multitude of functions from a single gene by excluding or including exons during post-transcriptional processing. Reconstructing the contribution of these splicing variants to the total amount of gene expression remains difficult.
We introduced a probabilistic model of the alternative splicing reconstruction problem using a finite mixture model, and provide a solution based on the maximum likelihood principle. Our model is based on the assumption that the expected expression level of exons in a particular splicing variant is the same for all exons in that variant but allows for measurement error.
In this model the expression (Y) of patient i can be written as a weighted sum of the number of splicing variants, K, mixture multivariate Gaussian densities:
$f(Y_i)=\Sigma P_k \times g_k(Y_i|\theta_k)$ {k=1,…,K}. The k^{th} variant can be described by the factor Z_{jk}. Where Z_{jk}=1 or 0 depending on if exon j is included or excluded. We estimated parameters θ_k of the mixture Gaussian densities by maximizing the total likelihood using a Nelder and Mead optimization algorithm in R.
We applied this model to three genes (SLC2A10, TGFβR2 and FBN1) associated with marfan's syndrome in gene/exon expression data of 63 patients with Marfan's syndrome. We compared the likelihood, AIC and BIC of 5 scenario's: Normal Mixture Modeling estimated by Mclust, known splicing variants, no splicing variation, all possible variants, which existed of 2^5, 2^9, 2^{65} possible splicing variants, for SLC2A10, TGFβR2 and FBN1.

P2.5.120
Multi-purpose SNP selection method in genetic association study

M Park[1]
[1]Eulji University, Daejeon, Republic of Korea

Recent development of high-throughput technologies in biology has resulted in the production of huge amount of data. In genetic association study, those are characterized by thousands of SNPs with small number of samples, which could cause the "large p, small n" problem. For this reason, single marker-based analysis is commonly adopted in many studies despite of various merits of the joint analysis of multiple markers. Existence of the redundant SNPs may also bring about many problems in further analysis.
Therefore, it is necessary to eliminate the near-redundant SNPs and hence to determine the subset of SNPs that should be included in the joint analysis.
In this study, we propose an unsupervised SNP selection algorithm based on the principal variable method. Minimum trace of partial variances of the unselected SNPs unexplained by selected SNPs is used as criterion. The resulting subset of SNPs could be used for further analysis on multiple purposes. This method is illustrated with real genotype datasets.

P2.5.138
Measurement Error in GWAS: what have we missed?

RCA Rippe[1]
[1]Institute of Education and Child Studies, Leiden University, Leiden, The Netherlands

Genetic associations with any behavior or disease are commonly found using using Genome Wide Association Analysis, among other methods. Biology predicts that single as well as multi-locus effects do exist but are generally very small (Davis et al, 2010, Vinkhuyzen et al, 2012).
While GWAS was introduced as promising methodology, the amount of significant empirical results seems to be less impressive. Underlying this observation is the possible influence of measurement error in both the outcome as well as (co)variates. Imperfections can be due to the use of (self report) questionnaires (Hofstee, 1994; Spain, Eaton & Funder, 2000) for outcome and determinants like age, weight and height, as well as determination of the genotype (Rabbee & Speed (2006); Rippe, Meulman & Eilers, 2012; Ziegler, König & Thompson, 2008) and reported ethnicity (Price et al, 2006).
Measurement error can distort results either through error in the determinants, diluting estimates of the association toward zero, and through error in the outcome, inflating standard errors (Hutcheon et al, 2010).
The current study illustrates to what extent the expected effects have remained undetected due to these errors. A large scale simulation study was set up using the highly efficient GWAS implementation of Sikorska et al. (2013) in order to evaluate genetic effect detection for different error levels in the variables involved. We observe that up to 20% more and 10% stronger genetic associations could be detected under smaller measurement error, showing possibly stronger biological effects than those currently reported.

P2.5.144
Classification in high-dimensional feature spaces

F Schroeder[1]
[1]Austrian Institute of Technology, Vienna, Austria

The characteristic property of many data sets in modern scientific fields, such as genomics, is the high-dimensionality of its feature space. It poses a significant challenge for statistical methods for classification and has thus been the object of intensive research.
This work studies the different approaches, with which standard classification methods, such as Discriminant Analysis, Support Vector Machines and Logistic Regression, have been modified to account for high-dimensionality, and compares their performance in different simulation experiments. Both prediction as well as model selection performance are examined under different parameters, including sample size, signal-to-noise ratios, and different structures of dependence.
The results are supposed to guide the applied researcher in one of the most tricky questions: Choosing the most suitable method for a given research question and data set.

P2.5.194

A non-homogeneous hidden Markov-model for gene mapping based on whole-genome sequencing data

F Zamanzad Ghavidel[1], J Claesen[1], T Burzykowski[1]
[1]*I-Biostat, Hasselt University, Diepenbeek, Belgium*

The analysis of polygenetic characteristics for mapping quantitative trait loci (QTL) remains an important challenge. QTL analysis requires two or more strains of organisms that differ substantially in the (poly-)genetic trait of interest, resulting in a heterozygous offspring. The offspring with the trait of interest is selected and subsequently screened for genetic markers such as single nucleotide polymorphisms (SNPs) with next generation sequencing (NGS). Gene mapping relies on the principle of co-segregation, the tendency for closely linked genes and genetic markers to be inherited together.
For each marker, observed mismatch frequencies between the reads of the offspring and the parental reference strains can be modeled by a multinomial distribution with the probabilities depending on the state of an unobserved (hidden) Markov process (Claesen and Burzykowski, 2014). After fitting the model to data, the Viterbi algorithm can be used to predict the most probable state for each of the SNPs. The predicted states can be used to infer whether a SNP is located in a (vicinity of a) QTL or not. Consequently, genomic loci associated with the QTL can be discovered by analyzing hidden states along the genome.
The aforementioned hidden Markov-model does not take into account the variation in the location of SNPs across the genome. To address this issue, we develop a non-homogenous hidden Markov-model with a transition matrix that depends on a set of distance-varying observed covariates. The application of the model is illustrated on the data from a study of ethanol tolerance in yeast.

P2.6 Software aspects of efficient statistical analyses

P2.6.26

Easy-to-use R-application to evaluate bioequivalence studies

M Borsos[1], M Müller[1]
[1]*AdWare Research Ltd, Balatonfüred, Hungary*

Aim: The aim of our work was to develop an easy-to-use application for researchers to be able to evaluate their bioequivalency study data from the calculation of pharmacokinetic parameters to obtain a raw statistical report with easily understandable tables and graphs.
Methods: Our goal was to apply only free softwares and still provide a user-friendly solution that doesn't require high programming or statistical knowledge from the researcher. We developed two R macros: one for calculating the pharmacokinetic parameters using non-compartmental methods, and one for performing the standard statistical analysis required by the FDA for 2x2 crossover bioequivalence studies. Then we validated our programs with the help of previously programmed and validated SAS codes by comparing the results on numerous simulated databases. Finally, we integrated our R programs with Sweave to provide the required tables and graphs automatically in a pdf report.
Results: We obtained two LaTeX codes with R programs that only needs the study data given in easy pre-defined formats and after running the applications, it results in dynamically changing pdf reports including tables, graphs and standard texts fitted to the actual results.

P2.6.30

Boosting diagnosis performance of biomarkers with nonparametric logistic type classification functions

Y-CI Chang[1]
[1]*Institute of Statistical Science, Academia Sinica, Taipei, Republic of China (Taiwan)*

The binary classification task is very common in medical diagnosis, where subjects are classified into one of two groups based on observed values of variables. As there are usually many variables of interest in a study, optimizing the combination of these biomarkers is an important problem. A linear combination is usually preferred because of its ease of interpretation. In the literature, there are already numerous published reports on achieving the best linear combination of biomarkers that maximizes the area under a receiver operating characteristic (ROC) curve, which is a popular tool for measuring classification performance of a classifier. However, there is often a lack of information about the relationship between disease status and the value of each biomarker. Hence, an improved method is always in demand. Here we propose a nonparametric classification function based on a general additive logistic model. It is proved that the proposed method gives a greater area under ROC curve than that of a linear combination. Moreover, because of the property of the method of additive functions, the proposed method retains the information of the relationship between biomarker and the disease status in the sense of a general additive model. Numerical results based on both synthesized and real data are reported.

P2.6.73

Repeated observations design analysed with ANOVA tools in MS-Excel

K Hrach[1]
[1]*J.E.Purkyne University, Faculty of Health Studies, Usti nad Labem, Czech Republic*

Repeated measures design (i.e. the situation, when each subject in the study is exposed to each level of the factor or factors) was discussed in a similar contribution last year in Munchen. Now, repeated observations design (i.e. the situation, when each subject in the study is measured several times under the same conditions) is in focus. Such designs are often used in practical trials.
The aim of this contribution is to show, how to perform the analysis of such a data, with the help of the common Excel analytical tools (ANOVA 1-factor and ANOVA 2-factors with replications, namely). This contribution proves the validity of the approach and performs a practical manual, how to do it. The situation concerning balanced single-factor repeated observations data, is discussed. Concrete example is solved as an illustration.
These findings were formed as a by-product of the grant project n. NT 14448-3/2013 of the Czech Ministry of Health.

Wednesday, 27th August 2014 - 11.00-11.30

Poster session P3

P3.1 Survival analysis, multistate models and competing risks

P3.1.19

Competing-risk regression model to explore risk factors associated with lost to follow-up prior to antiretroviral therapy: a multicentric observational cohort

M Bastard[1], J-F Etard[1,2]
[1]Epicentre, Paris, France, [2]UMI 233 TransVIHMI, IRD, Université Montpellier 1, Montpellier, France

In clinical research, death and lost to follow-up (LTFU) are from part of the most commonly reported outcomes. The objective here is to present the competing-risk regression model, an alternative method to Cox proportional-hazards model in presence of multiple outcomes that are not independent, and to apply it to explore factors associated with LTFU prior to antiretroviral therapy (ART).
Standard techniques assume that the distribution of the censoring time and the time-to-event distribution are independent. In case of administrative censoring, this hypothesis seems reasonable. However, if patients died, censoring is related to the time-to-event of interest (LTFU), and the independence assumption is violated leading to biased estimates of LTFU rates and hazard ratios. Fine and Gray specify a semi-parametric model which uses the hazard of the subdistribution of the event of interest to model the cumulative incidence function. With this model, patients with the competing event are kept at risk and continue to contribute person-time with the remaining time at risk weighted by the inverse probability of censoring.
To illustrate the method, we apply it to assess factors associated with LTFU before ART initiation in a large multicentric cohort of HIV-infected patients, treating death before ART as a competing event. We also compare it to standard Cox proportional-hazard model.
Competing-risk models should be considered with interest in clinical research when conducting survival analysis with competing events as standard methodology could provide biased estimates of both the cumulative incidence of the event of interest and hazard ratios of associated factors.

P3.1.22

Description of disease progression and relevant predictors in diabetic foot ulcer patients using a Markov chain model

A Begun[1], S Morbach[1,2], G Rümenapf[3], A Icks[1,4]
[1]German Diabetes Center at the Heinrich-Heine-University, Düsseldorf, Germany, [2]Department of Diabetes and Angiology, Marienkrankenhaus, Soest, Germany, [3]Diakonissen-Stiftungs-Krankenhaus Speyer, Mannheim, Germany, [4]Department of Public Health at the Heinrich-Heine-University, Düsseldorf, Germany

Macro- and microvascular complications among diabetic patients can lead to foot ulceration and increased risks of (minor or major) amputation and death. We have used an eight-state Markov chain model to study the course of the diabetic foot syndrome. The diabetic patients were staged in accordance with their amputation status (no, minor or major), present or previous foot ulceration (yes, no), and death as absorbing state. Potential risk factors - such as gender, age at arriving in the state, smoking habits, diabetes duration, neuropathy, peripheral arterial disease (PAD), chronic renal failure (CRF) and others - were included in the model in the form of the Cox-regression covariates. In addition, the impact of revascularization procedures was studied.
We used long-term data of a patient cohort from one single specialized diabetic foot center in North Rhine-Westphalia (Germany). The estimates of unknown baseline transition intensities and Cox-regression coefficients were derived from stepwise regression with backward elimination based on the likelihood ratio test at level 5%. Amongst others, we found that established risk factors as gender, PAD or CRF were predictive for the transition between some stages, while not others. For instance, male patients with diabetic foot syndrome but without previous amputations showed an increased risk of foot ulcer recurrence compared to females, while there was no gender difference regarding the risk for transition to neither minor nor major amputation. This model can help us to quantify the disease progression and its predictors.

P3.1.27

Reconstructing individual patient level data: a simulation approach

RH Boucher[1], KR Abrams[1], PC Lambert[1,2]
[1]University of Leicester, Leicester, United Kingdom, [2]Karolinska Institutet, Stockholm, Sweden

There are times when the reported analysis for a time-to-event outcome may be considered to be inappropriate. For example, if a proportional hazards (PH) model has been used, despite the PH assumption being clearly violated. It is, therefore, desirable to re-analyse the dataset using more appropriate methods.
However, this is only realistic if the original individual patient data (IPD) is accessible. Nevertheless, if only summary data are available, the method outlined here simulates multiple datasets that are representative of the original IPD. These can then be analysed using the desired method and the results averaged over in order to produce a more appropriate result. This approach relies on routinely reported summary information and the Kaplan-Meier curve. Coordinates are extracted from the Kaplan-Meier curve. A model is fitted to these coordinates, and then used to simulate survival times for individual patients. The censoring distribution is formed using information published on recruitment times, or from the 'numbers at risk' table. The minimum of the survival and censoring time is then taken as the patient's observed survival time. The last three steps are repeated to generate multiple datasets.
An application of this method highlights its ability to replicate the reported statistics, and hence its success in representing the original IPD. In addition, this example is one for which the IPD was available and, thus the comparison between the appropriate analysis on the original IPD and the averaged result over the simulated datasets can also be made.

P3.1.69

Joint modelling of multiple longitudinal markers and recurrent events of multiple types

MH Hof[1], JZ Musoro[1], RB Geskus[1], AH Zwinderman[1]
[1]Academic Medical Center of the University of Amsterdam, Amsterdam, The Netherlands

Our study was motivated by post-kidney transplantation data, where we observed four longitudinal markers and nine different recurred infection types. Moreover, as a consequence of low marker values and multiple infections, individuals dropped-out of the study. We used a joint modelling approach to correct for informative drop-out. Our main interests were the relations between the markers and the infection rates.
We discretized the time-scale into small intervals such that individuals could experience at most one event per interval. We parameterized the sub-model for the nine competing infection risks and the drop-out risk with a multinomial regression model with subject-specific random effects. The sub-model for the marker trajectories was parameterized by a multi-

variate linear mixed effect model. By including the marker values as latent terms in the event sub-model, both sub-models shared the random effects of the markers.
Because we had random effects for all four biomarkers and for all nine event types, our joint model involved a high dimensional integral. In this case, quadrature approximations are too computationally expensive. As a solution, a Monte-Carlo approach was used to evaluate the integral. To improve the accuracy of the approximation, we used a Quasi-Monte Carlo (QMC) approach with a deterministic point set based on scrambled Sobol sequences.
With simulations, we showed that the joint model with QMC integration gave accurate estimates. Moreover, the joint model successfully captured informative drop-out. We fitted the joint model on our post-kidney transplantation data.

P3.1.75
Analyzing clinical pathways in observational studies: pitfalls and approaches
M Huebner[1], DW Larson[2]
[1]Michigan State University, East Lansing, United States, [2]Mayo Clinic, Rochester, United States

An enhanced recovery pathway in colorectal surgery consists of a sequence of treatments in preoperative, intraoperative, and postoperative phases before discharge from the hospital. It is of interest to determine key pathway elements that are associated with short-term outcomes such as length of hospital stay.
Studying the individual association of elemental compliance with the outcome is not sufficient to establish their importance, since other factors impact the compliance or have an effect on the outcome. For example, occurrence of complications may lead to a longer length of stay and may modify the postoperative pathway. Preoperative and intraoperative elements may be associated with a decreased need for opiates and thus lead to a faster recovery after surgery. Comparisons of research reports between institutions are difficult due to heterogeneity of patient populations and ignoring confounding factors in the analyses.
We will present some approaches including multistate models that have been helpful in shedding some light on these issues.

P3.1.111
Remedy for 'IntCox' in partly interval-censored survival data
M Nishikawa[1], H Mizukami[2], T Morikawa[3], T Yokoyama[1]
[1]National Institute of Public Health, Wako,Saitama, Japan, [2]Sanofi K.K., Shinnjyuku, Japan, [3]Kurume University (formerly), Kurume, Japan

In survival analysis, common data type are right-censored. However, we often encounter "partly interval-censored data" (PICD) in oncology studies where observed data include both exact times of event and interval-censored, for example progression free survival (PFS). PFS is defined as the time from randomization to the date of disease progression or death, whichever is earlier.
The most common regression model in survival analysis is proportional hazard model. It is also called Cox model for right-censored data. Several interval-censored data regression analyses have been proposed (e.g. Finkelstein(1986), Pan(1999)). Among them, Pan(1999)'s method is implemented in the R package 'IntCox' as an extension of Cox model. Chen et al. (2012) report performance of 'IntCox' for interval-censored data.
However, its performance for PICD is unknown. 'IntCox' cannot work for PICD. We found that some data manipulation (jittering) is necessary before calling 'IntCox', and incomplete result of 'IntCox'(no improvement of likelihood possible) is frequently observed.
We consider some remedies for 'IntCox' in the analysis of PICD; these are deterministic imputations and bootstrap methods.
In this talk, we compare the performances and properties of the three deterministic imputation methods (the right-point, the mid-point and the left-point of the censoring interval), and 'IntCox' with/without bootstrap for PICD by simulation study of which design was based on actual clinical trials.
Our simulation suggests that the left-point imputation may be generally better in MSE than the right-point or the mid-point imputations if there are right-censored data before the planned end-of-the-study.
The details will be shown in the presentation.

P3.1.125
Comparison of survival between allogeneic haematopoietic stem cell transplantation and continued drug treatment when differentiating between risk groups at diagnosis
M Pfirrmann[1], R Hehlmann[2], M Lauseker[1], J Hasford[1]
[1]IBE, Ludwig-Maximilians-Universität München, München, Germany, [2]III. Medizinische Klinik Mannheim, Universität Heidelberg, Mannheim, Germany

Aims: In the German study IIIA, patients with chronic myeloid leukaemia providing consent and eligibility for allogeneic haematopoietic stem cell transplantation (HSCT) were randomized in accordance with the availability of a (matched) related donor to receive either HSCT or continued drug treatment. Primary endpoint was overall survival (OS) from diagnosis. Subgroups with different survival risks after transplantation were of particular interest.
Methods: At transplantation, survival probabilities can be differentiated through subgroups defined by the prognostic factors age, phase of disease, time to transplantation, donor matching, and recipient-donor sex combination. At diagnosis, patients were in chronic phase of disease and age known. It was assumed that for all patients, transplantation was planned in the first year. For non-transplanted patients, values for the factors "donor matching" and "recipient-donor sex combination" were randomly drawn from their actual distribution among the patients transplanted. This was repeated to gain 999 results of a log-rank statistic for each subgroup. Treatment comparisons were in accordance with randomisation.
Results: Of 166 patients randomized to HSCT, 151 were transplanted. Here, donor matching and recipient-donor sex combination were assumed to have been known at diagnosis but were randomly drawn for the 15 remaining patients and all 261 patients randomized for drug treatment. Median p-values for the 999 OS comparisons of HSCT versus drug treatment were not significant for any of the subgroups.
Conclusions: This innovative approach prevents time-to-transplantation and selection bias. To decide for the option "HSCT" at diagnosis depicts reality. The issue of statistical power needs to be discussed.

P3.1.148
Modeling cause-specific survival in cancer patients compared to the general population, a large population based study
MC Småstuen[1], DA Barzenje[2], A Kolstad[3], H Holte[3]
[1]University of Oslo, Department of Biostatistics, Oslo, Norway, [2]Department of Oncology, Ostfold Hospital Trust, Oslo, Norway, [3]Department of Oncology, Oslo University Hospital, Oslo, Norway

Aims: Competing risk is a recognised methodology however not often used and correctly interpreted by clinicians. Kaplan Meier approach is still the method of choice even though it might result in wrong and misleading results. Cause-specific survival is often cited and compared in the medical literature however in many studies calculated using a wrong methodology.

Methods: We analysed 404 Follicular lymphomas cancer patients (FL) diagnosed between January 1, 1980 and December 31, 2005. Each FL was matched on age and gender with 4 individuals randomly selected from the general population. Causes of death were divided into 3 groups: cardiovascular disease (CVD), cancer lymphoma related and other. For the FL patients we have investigated overall survival (OS), progression free survival (PFS) and time to next treatment (TNT). When comparing with the general population, we have computed overall survival and cumulative incidences stratified by the three causes of deaths using competing risk methodology.
Results: This population-based study revealed larger overall mortality risk but not due to CVD in FL patients than in the general population. The cumulative incidence of TNT was still elevated for stage II patients compared to Stage I, but both estimates were much lower than when modelled with Kaplan Meier methodology.
Conclusions: Many cancer patients live longer and respond well to treatment so they might experience a relapse/need for new treatment long time after the initial diagnosis when also other competing events might interact with the main event of interest. Therefore, the choice of the correct methodology is crucial.

P3.1.149

Cox model with multiple events: an application to mammography screening intervals in the Portuguese primary health care system

A Sousa[1], DP Tavares[2], H Mouriño[1], P Nicola[3]
[1]Faculdade de Ciências - Universidade de Lisboa, Lisboa, Portugal, [2]Universidade de Lisboa, Lisboa, Portugal, [3]Faculdade de Medicina - Universidade de Lisboa, Lisboa, Portugal

According to the World Health Organization, breast cancer is the top cancer in women both in the developed and the developing world. In 2012, breast accounted for 522 000 deaths worldwide. Although these numbers, breast cancer mortality has been falling in many European countries due to the combined effects of breast screening and better treatments.
So far the only breast cancer screening method that has been proved effective is mammography screening. In 2003, the European Council has suggested to all the Member States to undertake screening for women aged 50-69 years and the Portuguese Directorate-General of Health has adopted this recommendation and women must be screened every two years. Nowadays the screening covers 60% of the territory.
This study aims at identifying variables associated with an increase of the time interval between screenings events. We focus on all women enrolled at Family Health Units (FHU) from Lisbon. It covers the period from 2000 to 2013. The variables used are age at study entry, body mass index, age at menarche, alcohol consumption, smoking, menopausal status, contraceptive use and the number of primary care visits.
The Cox model with multiple events has been estimated. This model allows for multiple mammographies per woman. We found out that high body mass index, hormonal contraceptive use, menopausal status and number of primary care visits are related to the time between screening examinations.

P3.1.150

Statistical modelling of biomarkers incorporating non-proportional effects for survival data

J Stephen[1], G Murray[1], J Bartlett[2,3], D Cameron[3]
[1]University of Edinburgh, Edinburgh, United Kingdom, [2]Ontario Institute for Cancer Research, Toronto, Canada, [3]Edinburgh Cancer Research Centre, Edinburgh, United Kingdom

Personalised medicine is replacing the one-drug-fits-all approach with many prognostic models incorporating biomarkers available for risk stratifying patients. Evidence has been emerging that the effects of biomarkers change over time and therefore violate the assumption of proportional hazards when performing Cox regression. Analysis using the Cox model when the assumptions are invalid can result in misleading conclusions.
We report the results of a review of existing approaches for the analysis of non-proportional effects with respect to survival data which identified a number of well-developed approaches but a lack of application of these approaches in practice. The review indicated there is a need for more widespread use of flexible modelling to move away from standard analysis using a Cox model when the assumption of proportional hazards is violated.
We further illustrate the use of two key approaches; the multivariable fractional polynomial time (MFPT) approach by Sauerbrei *et al.* and flexible parametric models proposed by Royston & Parmer, to develop a model for predicting survival of patients with early breast cancer. We illustrate their respective advantages and disadvantages in the development and evaluation of a prediction model.

P3.1.151

Impact of length of follow-up on the evaluation of prognostic scores with an example using two breast cancer studies

J Stephen[1], G Murray[1], J Bartlett[2,3], D Cameron[3]
[1]University of Edinburgh, Edinburgh, United Kingdom, [2]Ontario Institute for Cancer Research, Toronto, Canada, [3]Edinburgh Cancer Research Centre, Edinburgh, United Kingdom

Background: We investigate the impact of follow-up duration on two residual risk models, IHC4 and Mammostrat, for predicting risk in early breast cancers using two studies with different lengths of follow up; the Edinburgh Breast Conservation Series (BCS) and the Tamoxifen versus Exemestane Adjuvant Multinational (TEAM) trial.
Methods: The multivariable fractional polynomial time (MFPT) algorithm was used to determine which variables had possible non-proportional effects and the best fitting fractional polynomial to model these effects. The performance of the scores was assessed at various lengths of follow-up using measures of discrimination and calibration.
Results: We observed a strong time-dependence of both the IHC4 and Mammostrat scores. Both scores were significant independent predictors of outcome restricted to the first five years of follow-up, after which the scores were not associated with distant recurrence free survival. The models performed statistically better with shorter follow-up compared to full follow-up with differences in D statistic between 0.4 and 0.5 and R^2 between 7 and 13%.
Conclusion: Our analyses confirm that it is important to consider the length of follow-up and violations of the Cox proportional hazards assumption when evaluating prognostic models. Longer follow-up resulted in strong degradation of the performance of the scores.

P3.1.157

On the estimation of survival of HIV/AIDS patients on anti-retroviral therapy: an application to interval censored data

PK Swain[1], G Grover[1]
[1]*University of Delhi, Delhi, India*

The main objective of this paper is to estimate the survival of HIV/AIDS patients who are undergoing Antiretroviral Therapy treatment in an ART centre, Delhi, India. Non Parametric Maximum Likelihood Estimation NPMLE (E-M) for interval censoring and KM survival plot for left, right and mid-point imputation have been used to estimate the survival of these patients. It has been observed that that the mid-point imputed survival plot has a very similar and consistent pattern as obtained by NPMLE (E-M) method. Considering these mid-point imputed values as right censored data, Cox PH model and Accelerated Failure time Model (AFTM) have been applied to study the effects of prognostic factors like age, sex, mode of transmission, baseline CD4 cell count, hemoglobin, baseline weight and smoking habits on the survival of the patients. The Akaike Information Criterion (AIC) has been employed to compare the efficiency of the models and Cox-Snell residual to test proportionality assumption.

P3.1.160

The relevance of joint modelling of longitudinal and competing risks data in the analysis of a peritoneal dialysis program

L Teixeira[1], I Sousa[2], A Rodrigues[3,4], D Mendonça[1,5]
[1]*ICBAS-UP, Porto, Portugal,* [2]*Department of Mathematics and Applications, UMinho, Guimarães, Portugal,* [3]*CHP-HGSA, Porto, Portugal,* [4]*UMIB/ICBAS-UP, Porto, Portugal,* [5]*ISPUP, Porto, Portugal*

In many clinical studies such as on peritoneal dialysis program, the presence of a longitudinal outcome repeatedly registered along the follow-up time and the occurrence of a specific event is common. The many well-established models proposed to analyse longitudinal and time-to-event outcomes separately are not suitable to analyse data when the longitudinal and survival outcomes are associated. Then, a joint modelling approach is required.

In the last years, joint modelling of longitudinal and survival data has received much attention and an increase in the use in clinical studies was verified. Although, some joint models were adapted in order to allow for competing endpoints, this methodology has not been widely disseminated.

The present study has as main objectives to compare different joint modelling approaches of longitudinal and survival data in a competing risks setting and to illustrate their relevance in the analysis of a peritoneal dialysis program. With these models it was possible to evaluate the association between a longitudinal clinical parameter (such as albumin) and the events of interest (death, transfer to haemodialysis and renal transplantation), besides the identification of predictors of each of these outcomes.

Results obtained with this methodology, which could not have been obtained with standard survival models, produced new information about peritoneal dialysis and contributed for a better knowledge and management of peritoneal dialysis program.

P3.1.162

An illness-death model of chronic kidney disease progression

A Thakkinstian[1], P Vejakama[1], A Ingsathit[1]
[1]*Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand*

Objectives: To estimate probability of ESRD, death, and death after ESRD (ESRD-death) in CKD patients using illness-death model.

Methods: Retrospective CKD cohort data were retrieved from one province (20 districts) in Thailand years 1997-2011. Illness-death models consisted of 3 transitions: death (transition 1), ESRD (transition 2), and ESRD-death (transition 3). A flexible-parametric survival with restricted-cubic spline was applied, and probability of each state was estimated.

Results: Among 32106, 30634 CKDs I-IV were initial state, 2573/30634 (8.4%) subjects developed ESRD, but 55 subjects were lost since ESRD diagnosis, 6175/30634 (20.2%) subjects died without ESRD, and 1472/32106(4.6%) subjects initially enrolled with ESRD. Of 3990 ESRDs, 2457 (61.6%) subjects died. Probability of ESRD at 2-years, 5-years, and 10-years were 6.33 (95% CI: 6.13%, 6.55%), 5.30 (5.07%, 5.53%), and 2.47% (2.27%, 2.66%), respectively. These corresponding probabilities were respectively 4.98% (4.78%, 5.18%), 16.50 (16.12%, 16.89%), 36.37% (35.57%, 37.17%) for death; and 2.58% (2.43%, 2.73%), 7.27 (7.00, 7.54), and 13.04 (12.62%, 13.45%) for ESRD-death.

Finally, probabilities of survival with ESRD-free were 86.03 (85.70%, 86.38%), 70.85% (70.37%, 71.33%), and 48.12% (47.31%, 48.93%). Risk of death was higher in diabetes than non-diabetes with hazard ratios of 1.21 (1.15, 1.28), 1.83 (1.15, 1.28), 1.65 (1.47, 1.86) for death, ESRD, and ESRD-death, respectively. In addition, ESRD-diabetes was about 1.37 (1.20, 1.56) times significantly higher risk of death than non-ESRD-diabetes.

Conclusions: This study provided progression of CKD in Thai setting. Probabilities of ESRD, ESRD-death, and death with ESRD-free were estimated. Diabetes was higher risk for both ESRD and death than non-diabetes.

P3.1.170

An illness-death model of HIV infection

U Udomsubpayakul[1], P Subhaluksuksakorn[1], P Chancharas[2], B-O Thepthien[3], S Sriwanichakorn[2]
[1]*Ramathibodi Hospital, Mahidol University, Bangkok, Thailand,* [2]*Ministry of Public Health, Nonthaburi, Thailand,* [3]*ASEAN Institute for Health Development, Mahidol University, Nakhon Pathom, Thailand*

Objectives: To estimate probability of lost to follow-up (lossFU), death, and death after lossFU in HIV infected patients using illness-death model.

Methods: Data were retrieved from ThaiHIV-registry of 21 provinces, the National Health Security Office (NHSO) years 2008 to 2012. Illness-death models were constructed with 3 transitions: death (transition 1), lossFU (transition 2), and lossFU-death (transition 3). State's probability was estimated using a restricted-cubic spline regression. Prognostic factors (i.e., sex, age, opportunistic infection(OI), anti-retroviral treatment (ART), health-coverage, and hospital-change) were then assessed.

Results: Among 8692 HIV patients entered to the initial state, 2453 (28.2%) patients were lossFU, 6,239 adhered with clinics but 769 died. Of 2453, 783 patients died after lossFU. Probability of lossFU at 2-years and 5-years were 17.48% (95% CI: 16.72%, 18.23%) and 21.77 (20.64%, 22.91%), respectively. Corresponding probabilities were respectively 8.15% (7.58%, 8.72%) and 10.93 (10.17%, 11.70%) for death; 8.96% (8.36%, 9.56%) and 12.32 (11.25%, 13.02%) for lossFU-death. Age ≥ 55 years and males were 1.28 (95%CI: 1.124, 1.47) and 1.21 (1.14, 1.29) times higher risk than age <55 years and females. ART, social, and government health-coverage were respectively 85% (84%, 86%), 17%(8%, 25%) and 41% (28%, 50%) lower risk than non-ART and universal scheme. Conversely, OI and hospital-change were 55% (40%, 71%) and 26% (17%, 36%) higher risk than non-OI and non-hospital-change.

Conclusions: Our study provided probabilities of lossFU and death in HIV patients. ARTs and health coverage were preventive-factors whereas males, old-age, OI, and shopping-around hospital were risk factors of loss-FU and death.

P3.1.189

New insights on therapy choices in non-small cell lung cancer using a flexible extension of the standard Cox's model

W Wynant[1], M Abrahamowicz[1], I Gioulbasanis[2], G Kasymjanova[3], JS Agulnik[3], B Gagnon[4]
[1]McGill University, Montreal, Canada, [2]University Hospital of Larissa, Larissa, Greece, [3]Jewish General Hospital, Montreal, Canada, [4]Laval University, Quebec, Canada

Non-small cell lung (NSCL) cancer accounts for about 80% of all cases of lung cancer. Most of the patients present with advanced-stage disease at diagnosis and have a poor prognosis. Moreover, chemotherapy confers a modest survival improvement, compared with supportive care alone.
Therefore, probability of survival conditional on patient characteristics is a meaningful metric for prognosis and therapy choices. In NSCL cancer, this could guide treatment by recognizing patients at high risk for poor survival who might be considered to earlier intervention and could help for optimal clinical management.
We estimated survival probabilities from data on 269 patients with NSCL cancer. As misspecification of covariates' effects can have a huge impact, we extended the standard Cox´s model to allow (i) non-linear effects of the covariates on the logarithm of the hazard and (ii) covariates effects to change over time. We also considered a flexible modeling of the baseline hazard to avoid step functions, biologically implausible.
Our results emphasize the importance of taking into account the potential time-dependent and non-linear effects of biomarkers and new insights obtained from survival curves. For example, our survival probability estimate at 6 months after chemotherapy, for patients who smoke, had a double-agent chemotherapy but with low blood levels of albumin and C-reactive protein and average level of various other biomarkers (including neutrophil counts and alkaline phosphatise) was only 40%, whereas the estimate from the widespread standard Cox´s model was 85%.
This highlights a profile of patients who may be targeted to earlier intervention.

P3.1.191

Heterogeneous M/M/1 type queuing models

S Yiu[1]
[1]MRC Biostatistics Unit, Cambridge, United Kingdom

Motivated by the need to account for unobserved heterogeneity from a missing important covariate in a longitudinal reversible count data setting, this work presents three M/M/1 type queuing models with random effects.
The standard M/M/1 queuing model is considered with and without an absorbing state, followed by the M/M/1 queuing model with one modified transition intensity. By convenient choice of mixing distributions, closed form expressions for the marginal likelihoods are available, thereby providing tractability.
The methodology is illustrated with an application to a psoriatic arthritis data set where modelling the number of active joints is of interest.

P3.2 Diagnostic studies

P3.2.91

Sample size calculations for confidence limits of prevalence of disease adjusted for estimated sensitivity and specificity

Z Lang[1], J Reiczigel[1]
[1]Szent István University, Faculty of Veterinary Science, Budapest, Hungary

Prevalence of a disease or other characteristic of a target population is frequently estimated by diagnostic tests. Lang and Reiczigel (2014) constructed approximate confidence intervals for prevalence when apparent prevalence, sensitivity and specificity were estimated from independent binomial samples. When the sample sizes are small, the confidence intervals obtained may prove to be too wide to cope with a required precision. The solution for this problem is to calculate suitable sample sizes based on preliminary diagnostic parameter estimates stemming from earlier studies or estimated from the actual data and draw new samples from both the target population and the populations to re-estimate sensitivity, specificity and prevalence. In this presentation we provide sample size formulas e.g. when the planned length of the confidence interval of prevalence is prescribed and the estimated values of the diagnostic parameters remain unchanged.
When the variances of the estimates of sensitivity or specificity are poor it is advisable to re-estimate them from larger new independent samples. If prevalence is small then the variance of specificity has to be reduced first of all.
Analogously, when prevalence is close to 1 the variance of sensitivity is advised to be controlled.

P3.2.95

Systematic review and meta-analysis of diagnostic accuracy of FDG-PET in dementia and Alzheimer's disease

YE Lee[1], JE Choi[1], J-Y Kim[1], YH Yoo[2], DY Lee[3], S-H Park[1], S-K Son[1], Y-K Lee[1], E Shin[1], CH Ryu[2], C-H Sohn[3], J-Y Lee[4], Y Kim[4]
[1]National Evidence-based Healthcare Collaborating Agency, Seoul, Republic of Korea, [2]Department of Nuclear Medicine, Gangnam Severance Hospital, Seoul, Republic of Korea, [3]Seoul National University Hospital, Seoul, Republic of Korea, [4]SMG-SNU Boramae Medical Center, Seoul, Republic of Korea

As aging population is rapidly increasing, it is estimated that prevalence of dementia among older adults would be doubled every twenty years and number of patients would increase by one million until 2027. Dementia would lead to a burden of care on family members, care-givers, and even societies due to social and economic cost. Therefore, early diagnosis of dementia is important step to prevent further worsening of disease and improve quality of life of dementia patients and their family. In this study, we assessed diagnostic accuracy of FDG-PET in evaluation of dementia, which known as a tool for detecting reduced glucose metabolism in patient's brain even before the development of dementia symptoms.
To evaluate diagnostic accuracy in early detection of dementia and Alzheimer's disease, we conducted systematic reviews of published articles, and identified 9 cross-sectional studies and 13 delayed cross-sectional studies. Bivariate Meta-analysis of 9 cross-sectional studies resulted in a pooled sensitivity(SN) of 0.61(95% CI: 0.42-0.79), a pooled specificity(SP) of 0.81(95% CI: 0.55-1.07). In 13 delayed cross-section studies, it resulted in a pooled SN of 0.81(95% CI: 0.72-0.91), and a pooled SP of 0.78(95% CI: 0.65-0.92). With subgroup analyses in amnestic mild cognitive impairment(MCI) patients, the result suggested a pooled SN of 0.92(95% CI: 0.75-1.00), a pooled SP of 0.88(95% CI: 0.77-0.98). These results indicate that FDG-PET among amnestic MCI patients was most accurate in the aspects of pooled SN and SP in delayed cross-sectional studies.

P3.2.101
Smooth time-dependent ROC curve estimators
P Martínez-Camblor[1], JC Pardo-Fernández[2]
[1]OIB-FICYT, Oviedo, Spain, [2]Universidad de Vigo, Vigo, Spain

The ROC curve is a popular graphical method frequently used in order to study the diagnostic capacity of continuous (bio)markers. When the considered outcome is a time-dependent variable, two main extensions have been proposed: the cumulative/dynamic and the incidence/dynamic ROC curves. In both cases, the principal problem for developing appropriate estimators is the estimation of the joint, time-to-event and marker, distribution.
As usual, different approximations lead to different estimators. In this work, the authors explore the use of a bivariate kernel density estimator with this goal. The performance of the cumulative/dynamic and the incidence/dynamic versions of the time-dependent ROC curve is studied from Monte Carlo simulations. In addition, the influence of the bandwidth selection on the obtained results and the use of different indices to measure the global diagnostic capacity are also studied. Finally, some real-world applications are reported.
Results suggest that the smooth estimators provide good approximations, in special, when the area under the ROC curve is not too large. As usual, the main handicap of this technique is the impact of the used bandwidth on the obtained estimations. A reasonable empirical rule to choose this parameter is also proposed.

P3.2.103
Seasonality in testing for systemic lupus erythematosis
EJ McKinnon[1], M John[2]
[1]Murdoch University, Perth, Australia, [2]Royal Perth Hospital, Perth, Australia

Systemic lupus erythematosis (SLE) is a chronic autoimmune disease that can affect most of the organ systems of the body. The disease generally follows a relapsing and remitting course and is characterized by a range of symptoms including lethargy, fever, rashes and muscle and joint aches. There is no gold standard diagnostic test for SLE, but in practice a positive diagnosis typically comes after clinical assessment combined with a series of laboratory tests. These begin with an initial screen to detect the antinuclear antibodies (ANA) that mark disease activity of SLE and other rheumatic conditions. Choice of follow-up testing to confirm specificity of the antibodies is guided by observed fluorescence patterns, and include those based on detecting antibodies to extractable nuclear antigens (ENA) or anti-double-stranded DNA (anti-dsDNA).
Here we explore patterns of seasonality in ANA screening/monitoring, and investigate how they translate to follow-up confirmatory testing. Analysis is based on sequential test results from a large state-wide laboratory database. For each test type, numbers of results per individual are quite variable (range ANA: 1-19; ENA: 1-14; anti-dsDNA: 1-38) and heavily skewed, with only a minority of patients having multiple measures (ANA: 16%, ENA: 10%, anti-dsDNA: 9%).
In this presentation we will contrast inferences obtained from several methodological approaches that differ in how they take account of the between-individual variability in testing frequency.

P3.2.117
Clinical factors affecting bias between different eGFR measurements based on the weighted Deming regression
AJ Owczarek[1], K Wieczorkowska-Tobis[2], A Skalska[3], A Więcek[4], J Chudek[5]
[1]Division of Statistics, Medical University of Silesia, Sosnowiec, Poland, [2]3 Dept. of Geriatric Med. & Gerontology, Univ. of Med. Sci., Poznan, Poland, [3]6 Dept. of Int. Med. and Geront., Jagiel. Univ. Med. Col., Cracow, Poland, [4]Dept. of Nephr., Endocr. & Met. Dis., Med. Univ. of Silesia, Katowice, Poland, [5]Dept. of Pathophysiology, Medical University of Silesia, Sosnowiec, Poland

Six different eGFR calculation methods were done in 3503 subjects elder than 65. eGFR based on full MDRD formula was chosen as a gold standard. According to intraclass correlation coefficient, through the chronic kidney diseases classes, the most compatible formulas were as follow: CKD-EPI (ICC=0.88), short MDRD (0.76), Cystatine-Creatynine (0.72), Cockroft-Gault (0.57), Hoek (0.52). The highest overestimation occurred in C-G (44.3%), while the lowest in MDRD short formula (0.1%). The highest underestimation occurred in Hoek formula (45.2%), while the lowest in the CKD-EPI formula (13.3%).
Results of agreement were presented using the agreement chart. Next the bias in corresponding five weighted Deming regression models were calculated, for patients with eGFR lower than 60. It has been shown, that diabetes, hypertension, heart failure, stroke, obesity and immobility were factors that affect the bias effect.
Conclusions: It has been shown, that different formulas to calculate estimated glomerular filtration rate have different level of bias in the presence of important clinical factors.

P3.2.153
When do latent class models outperform an imperfect gold standard? A problem revisited
MR Oliveira[1], A Subtil[2]
[1]Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal, [2]Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

The dynamical biomedical sciences and pharmaceutical industry steadily seek to produce new diagnostic tools. Yet, new tests should only be introduced into medical practice after its clinical value is thoroughly evaluated, including the test's ability to correctly identify diseased and nondiseased patients. Standard performance measures, can be estimated by comparison with a gold standard test. Since such perfect reference test is frequently unavailable, alternative approaches are needed.
An available test perceived as the best one can be used as an imperfect reference test, against which the new test is compared. However, it is known that the imperfect reference, in general, leads to biased estimates. Latent class models (LCM) provide an alternative approach for this problem. A widely used LCM admits a binary latent variable, that indicates the disease status, and manifest binary variables, that express the tests results. This LCM assumes that the test results are independent conditional on the disease state, which may fail in practice and can result in substantial bias.
In this work, for the special case of 3 tests, we compare the LCM's estimators of performance measures with alternative estimators. In contrast with simulated comparisons, we take the theoretical viewpoint, based on the estimators analytical forms.
In the absence of a gold standard, LCM create a consensual "gold standard", based on the multiple test results, which can be used to classify patients as diseased or nondiseased. We discuss, from the theoretical perspective, the validity and potential usefulness of this classification as a clinical diagnostic tool.

P3.2.154

Bayesian latent class models for the evaluation of diagnostic tests in multiple populations

A Subtil[1,2], PZ Bermudez[1,2], L Gonçalves[2,3]
[1]Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal, [2]CEAUL, Lisboa, Portugal, [3]IHMT, Universidade Nova de Lisboa, Lisboa, Portugal

The evaluation of a diagnostic test's ability to correctly discern between diseased and non-diseased individuals is crucial to establish the test's clinical relevance and practical utility. In many situations, the true disease state of the individuals is unknown, because it is not possible to apply a perfect reference test (gold standard). In such cases, latent class models are often used to estimate diagnostic tests performance measures, such as sensitivity and specificity, as well as the disease's prevalence.
In this work, we look into situations where, under the absence of a gold standard, multiple diagnostic tests are applied to multiple subpopulations, admitting that dependencies between these subpopulations may exist. Plausible dependencies between prevalences may arise, for instance, between parents and offspring.
Distinguishing population subgroups allows differentiating prevalences and the tests' performance measures, and thereby further detailing the case under study. While stratified sampling naturally defines subpopulations, it may also be acceptable to artificially construct populations with a practical meaning, when an appropriate sampling scheme is missing.
A bayesian approach can be particularly relevant in complex settings, with multiple populations and covariates. Furthermore, it allows for the introduction of prior information, such as experts opinions or findings from previous studies, which may improve the inferences and avoid non-identifiability.
We explore and compare alternative bayesian latent class models with different strategies to model dependencies between subpopulations.

P3.3 Analysis of electronic health records

P3.3.49

A validation algorithm for detecting dose increase from longitudinal data of psychotropic drug users, using Monte Carlo simulation

F Feuillet[1,2], C Victorri-Vigneau[1,3], J-B Hardouin[1,2], V Sébille[1,2]
[1]EA4275, University of Nantes, Nantes, France, [2]Nantes University Hospital - Biometric Department, Nantes, France, [3]CEIP, Nantes University Hospital, Nantes, France

Introduction: Several methods have been recently developed from French National Insurance Health System (IHS) database concerning problems in drug use (misuse, addiction). No valid indicators are available to characterize dose increase which could reveal drug inefficiency or compulsive use.
Objective: To validate properties of an algorithm for detecting dose increase from longitudinal data (sensitivity, specificity).
Methods: Moving average method was applied for detecting dose increase. Several steps were defined for the algorithm: 1) defining a reference dose 2) calculating average doses for each drug delivery 3) comparing each dose with the reference dose, according to a defined detection threshold depending on dosage of the study drug.
Monte Carlo simulations were used to vary different parameters. Population parameters: study duration, increase rate and increase duration. Algorithm parameters: moving average method (one to four-period) and detection threshold.
Results: 1 million patients per dataset were simulated. For a dataset with duration of 12 months and an increase rate of 50%, algorithm with two-period moving average method achieves a sensitivity of 76% and a specificity of 94%. When study duration increases, specificity decreases and sensitivity increases. The one-period moving average method is sensitive (90%) but less specific (64%). Conversely, the four-period moving average method is not very sensitive (10%) but highly specific (100%). A high detection threshold results in a poor sensitivity of the algorithm.
Conclusion: This algorithm for detecting dose increase has good properties. The two-period moving average method optimizes properties of the algorithm. Usage recommendations may be proposed based on study objectives (population, pharmacological class, potential drug dependence...).

P3.3.96

How much of socioeconomic differences in breast cancer patient survival can be explained by stage at diagnosis and treatment?

R Li[1], R Daniel[1], B Rachet[1]
[1]London School of Hygiene & Tropical Medicine, London, United Kingdom

Socioeconomic inequalities in breast cancer survival persist in England. The main contributing factors could be presentation at different stages and variation in access to treatment.
Information on 36,793 women diagnosed with breast cancer during 2000-2007 was routinely collected by an English population-based cancer registry. Surgical treatment information from Hospital Episode Statistics was dichotomised into "major" versus "minor or no procedures". A deprivation category was allocated according to each patient's area of residence at diagnosis.
We estimated the proportion of the effect of deprivation on short-term survival mediated by stage and by treatment using G-computation procedures. Single stochastic imputation was incorporated to handle missing stage (8%).
Net survival differed between the most affluent and most deprived patients at one year (97% vs 94%), and at five years (86% vs 76%) after diagnosis.
Adverse stage distribution was associated with more deprived patients ($p<0.01$). The more advanced the stage at diagnosis, the less likely the patient was to receive major surgical treatment ($p<0.01$).
The most deprived patients were almost three times more likely to die within six months after diagnosis than the most affluent (OR: 2.77 [2.17-3.53]). One third of this excess mortality was mediated by adverse stage distribution whilst none was mediated through differential surgical treatment.
Our results showed that the effort to advance the diagnoses is important, but would reduce the socio-economic inequalities in cancer survival only by a third. We did not have reliable information on comorbidity, which could be another mediator on the causal pathway.

P3.3.152

Adjustment for hidden confounding in the analysis of pneumococcal vaccination effectiveness using electronic health records

AJ Streeter[1], A Ble[1], J Foster[1], D Melzer[1], WE Henley[1]
[1]University of Exeter Medical School, Exeter, United Kingdom

Vaccination against pneumococcal infection is currently recommended for adults aged over 65y in the UK. However the practical and ethical difficulties in conducting trials in this age group limit the evidence for this policy. Observational studies based on analysis of routinely collected patient records provide an alternative source of information for evaluating effectiveness of the vaccine in the population, away from the ideal environment of the clinical trial. A major challenge in adopting this approach is addressing the potential for bias due to hidden confounding.
We used a quasi-experimental approach to estimate vaccine effectiveness

in the UK elderly population using data from a large national Primary Care database. This method is not contingent on the identification of the confounders, and can be applied, where outcomes are recorded in the data before and after intervention. Extensions to the method were explored to address the potential for bias from informative censoring due to subsequent vaccination of the controls. This prompted the authors to further develop the method for wider applicability. This has important consequences for interventions, such as vaccination, where an entire population is targeted, leaving few untreated controls for comparison, or where outcomes are not observed before an intervention.

P3.4 Comparative effectiveness and outcomes research

P3.4.25a

Short- versus long-term outcomes after treatment for tuberculosis

LJ Bonnett[1], G Davies[1]
[1]*University of Liverpool, Liverpool, United Kingdom*

Background: TB remains a major killer amongst infectious diseases and current treatment involves a four-drug regimen for at least six months. Clinical development of a single novel TB drug is expected to take at least six years. A completely novel combination regimen would require twenty years or more. New drugs and regimens are required to shorten treatment duration, reduce toxicity and combat drug resistance.
We reviewed the ability of short-term outcomes from phase II trials to predict longer-term outcomes from phase III trials and hence improve selection of optimum combinations of new and existing drugs for development in pivotal trials.
Methods: Phase II or phase III trials of combinations of eight agents for drug sensitive individuals with tuberculosis were included in our systematic review. Definitive clinical endpoints included treatment failure and treatment relapse. Early clinical endpoints incorporated positive or negative culture at various time points, time to sputum culture conversion and serial viable colony counts.
For categorical data, the odds ratio will be calculated using the Mantel-Haenszel method, and for continuous data, such as colony counts, the mean difference will be calculated. Time to event outcomes will be summarised via the generic inverse-variance method. Additionally, early endpoints will be evaluated as surrogate outcomes for poor outcome via the generalised R2 statistic.
Results: 2865 trials were identified for potential inclusion in the review. Of these, 49 phase II and 478 phase III trials were included. Data is currently being extracted and results will be presented.

P3.4.34

The association of inhaled bronchodilators with the risk of acute myocardial infarction

S Choi[1], EJ Jang[1,2], Y Kim[1], JM Kim[1], C-H Lee[1,3], J-J Yim[3], HI Yoon[4], DK Kim[5]
[1]*National Evidence-based Healthcare Collaborating Agency, Seoul, Republic of Korea,* [2]*Andong National University, Andong, Republic of Korea,* [3]*Seoul National University Hospital, Seoul, Republic of Korea,* [4]*Seoul National University Bundang Hospital, Seongnam, Republic of Korea,* [5]*SMG-SNU Boramae Medical Center, Seoul, Republic of Korea*

Background: Among cardiovascular adverse events, acute myocardial infarction (AMI) has been regarded as one of the most important issues on drug safety. The objective of this study is to investigate whether inhaled bronchodilators affect the risk of AMI.
Methods: A nested case-control study based on the Korean national claims database included new adult users of inhaled medications between January 1, 2009 and December 31, 2011. Patients diagnosed with AMI after enrollment were identified as cases and up to five control individuals matched for age, sex, initiation date, diagnosis of hypertension, diabetes mellitus, chronic obstructive pulmonary disease(COPD), ischemic heart disease, other heart disease, and Charlson Comorbidity Index were selected. The association between the use of inhaled bronchodilators and AMI were investigated by conditional logistic regression.
Results: From the eligible cohort, 11,054 patients with AMI and matched 47,815 controls were selected. Mean age was 67 years old and the proportion of males was 53.6%. In unadjusted analysis, short acting beta agonists(SABA) [OR=1.2, 95% CI=(1.1, 1.3), p-value<0.001] and long acting beta agonists(LABA) [OR=1.3, 95% CI=(1.1, 1.6), p-value=0.013] significantly increased the risk of AMI. After adjusting with other inhaled medicines, age, respiratory disease, comorbidities, concomitant medication, and health care utilization, SABA [OR=1.2, 95% CI=(1.1, 1.3), p-value<0.001] and LABA [OR=1.3, 95% CI=(1.1, 1.6), p-value=0.011] increased the risk of AMI.
Conclusions: Our population based nested case-control showed that the use of SABA and LABA increased risk of clinically significant AMI.

P3.4.83

Economic evaluation of cervical cancer screening strategy

J Kim[1], Y Kim[1], M-J Ko[1], S Choi[1], J Shim[1], Y Lee[1]
[1]*National Evidence-based Healthcare Collaborating Agency, Seoul, Republic of Korea*

National Cervical Screening Programme in Korea recommends that a biannual cervical pap smear test interval is appropriate for women over 30 years. Conventional cytology is relatively simple and cheap method . However, several studies in abroad have shown the relatively high false negative rate and reported that HPV testing is more sensitive than cytology.
We assessed the cost-effectiveness of incorporate human papillomavirus (HPV) DNA testing into existing cervical cancer screening program in South Korea. The model compared two management of screening methods: (1) Pap smear, (2) Triage with HPV DNA testing(HPV DNA screening test after atypical or abnormal pap results at routine cervical cancer screening). To compared current screening policy with new strategy: (1) screening interval, (2) screening start age, (3) screening period . We considered these kind of strategies combination.We conducted cost-utility analysis applying QALYs to which takes into account life span expansion and the quality of life. Markov model was used with one year cycle and life time analysis period.Sensitivity analysis was conducted to reflect the uncertainty of variables.
As a result, pap smear test with 5 year interval was most inexpensive strategy, pap smear test with 1 year interval was most effectiveness strategy.

In the South Korea, most of the interventions with an ICER below KRW 30,000,000won/QALY are recommended routinely. Below on threshold 30,000,000won/QALY, pap smear test with 2 year interval (aged 20 to 79 years) was most cost-effectiveness strategy.

P3.4.113

Dichotomising highly skewed outcome data using a distributional method: a simulation study

M Ofuya[1], O Sauzet[2], JL Peacock[1]
[1]King's College London, London, United Kingdom, [2]Universität Bielefeld, Bielefeld, Germany

Researchers commonly dichotomise outcome data for ease of interpretation. The distributional method provides a way to dichotomise a continuous outcome without losing power by considering the proportion below a given cut-off as a function of the parameters of the Normal distribution (Peacock et al., Stat Med, 2012). This method works when data are Normal or can be transformed to Normal.
However, if the data is highly skewed, the commonly used log transformation may not completely remove the asymmetry and in such situations, the Gamma distribution may fit better than the Log-normal. In this study, simulations are used to compare the results provided by the distributional method using log transformed data to the true gamma distributional values, with standard error obtained through resampling methods.
Simulations were performed using published data parameters based on a study investigating the effect of Hepatitis A vaccine on antibody titre levels and 10 mIU/ml was considered to be a clinically relevant cut-point. Random gamma variables for a two sample design were generated and the simulated data were log transformed for varied distribution parameters. Distributional estimates of differences in proportions, risk and odds ratios and their standard errors were obtained for both distributions and compared.
Provisional results indicate the distributional method on log-transformed data provides acceptable estimates of standard error for both distributions but for differences in proportions, the effect sizes have bias 25%, risk ratios bias 8.5%.

P3.4.143

Comparison of classification models for sex determination of Polish skulls

K Salapa[1], KA Tomaszewski[1], R Chrzan[1], M Pliczko[1], M Gomulska[1], P Fraczek[1], A Sliwinska[1], S Baczkowski[1]
[1]Jagiellonian University Medical College, Cracow, Poland

Sexual dimorphism reveals in the whole skeleton. Many studies shown that sex can be determined by measurements of single bones. One of the most reliable bone structure for this is the skull. Various studies were performed to determine the sex by employing different measurements of the skull and usually discriminant analysis or logistic regression were applied. The purpose of this study is to determine whether there are significant differences in correct sex identification based on those two algorithms.
The study consisted of 500 archived Polish adult head CT scans (237 [48.18%] males and 255 [51.82%] females), age >21, without any malformation. The measurements of both right and left sides of the palatal bones and skull base, in millimeters, were considered in analysis: the depth of the greater palatine canal, distances between greater palatine foramen and incisive foramen, median palatine suture and posterior nasal spine.
A model was created using SPSS Modeler v15.0. Its tasks involved calculating both descriptive characteristics, correlations coefficients, mean comparison in males and females and stepwise logistic and discriminant (with a leave-one-out cross-validation) functions. The input data was randomly divided into training and testing samples (using a ratio of 70%:30%).
Final assessment of model quality was based on percentages of correct sex identification obtained using testing sets (N=152). Percentages of correct sex identification for all model were very similar and slightly exceeded 69%. The best classification function was derived from discriminant analysis, which used as predictors the measurements only of the left side of skulls.

P3.4.180

Quality of life in Portuguese cancer patients. A structural equation modeling application

E Vilhena[1,2,3], JL Pais Ribeiro[4,5], I Silva[6], H Cardoso[7], RF Meneses[6], L Pedro[8], AM da Silva[7], D Mendonça[2,3]
[1]Polytechnic Institute of Cavado and Ave, Barcelos, Portugal, [2]ICBAS, University of Porto, Porto, Portugal, [3]ISPUP, Institute of Public Health, University of Porto, Porto, Portugal, [4]Faculty of Psychology and Educational Sciences, Porto, Portugal, [5]UIPES, Lisboa, Portugal, [6]University of Fernando Pessoa, Porto, Portugal, [7]UMIB/ICBAS and Hospital Santo António/CHP, Porto, Portugal, [8]ESTeSL Polytechnic Institute of Lisbon (IPL), Lisboa, Portugal

Living with a chronic disease is a demanding experience that may affect multiple aspects of an individual's life. In general, chronically ill patients are responsible for the management of a wide range of psychosocial factors which contribute to their quality of life (QoL). QoL has become an important concept for health care. Cancer can produce many different symptoms. An increasingly important issue in oncology is to evaluate QoL in these patients.
The aim of the present study was to test the hypothetical model to evaluate the simultaneous impact of optimism, treatment adherence and social support on QoL (general well-being, physical and mental health), controlling for socio-demographic and clinical variables.
This study included a sample of 210 in Portuguese cancer patients approached by their physicians, in outpatient departments of the main hospitals in Portugal. All patients completed self-report questionnaires to assess socio-demographic and clinical, psychosocial and QoL variables. Structural Equation Modeling (SEM) was used to test the quality of the hypothesized model.
Results (performed using EQS 6.1) showed that the hypothesized model fitted the data reasonably, CFI=0.85, RMSEA=0.06, x^2/df=1.77. All factors had a simultaneous independent statistically significant impact in QoL, demonstrating that an attitude more optimistic, a better treatment adherence and more social support contribute to a better general well-being, a better physical health and a better mental health. Structural Equation Modeling techniques are considered a major component of applied multivariate statistical analysis for addressing complex scientific questions.

P3.4.181

A structural equation modeling application to test mediation of optimism between stigma and quality of life in Portuguese obese patients

E Vilhena[1,2,3], JL Pais Ribeiro[4,5], I Silva[6], H Cardoso[7], D Mendonça[2]
[1]Polytechnic Institute of Cavado and Ave, Barcelos, Portugal, [2]ICBAS, University of Porto, Porto, Portugal, [3]ISPUP, Institute of Public Health, University of Porto, Porto, Portugal, [4]Faculty of Psychology and Educational Sciences, Porto, Portugal, [5]UIPES, Lisboa, Portugal, [6]University of Fernando Pessoa, Porto, Portugal, [7]UMIB/ICBAS and Hospital Santo António/CHP, Porto, Portugal

Quality of life (QoL) has become an important concept for health care. It is a construct composed of a number of factors that contribute to individual's well-being and adjustment to chronic diseases. Obesity is considered one of the more relevant problems of public health in modern societies, as it is a factor predominant risk for the development of various diseases. They are patients are forced to live with the limitations imposed by their conditions.

The aim of the present study was to test a hypothetical model to evaluate: 1) optimism, stigma perception and social support, have an impact on QoL, controlling for socio-demographic and clinical variables; 2) optimism exerts a mediator effect between stigma perception and QoL.
Study comprises a sequential sample of 215 volunteer obese patients, approached by their physicians, in outpatient departments of principal hospitals in Portugal and completed self-report questionnaires to assess socio-demographic and clinical, psychosocial and QoL variables. Structural Equation Modeling (SEM) was used to test the quality of the hypothesized model.
Results showed that the hypothesized model fitted the data reasonably well, CFI=0.9, RMSEA=0.06, X(276)2 =525.75, p<0.001 (sensible to sample size). Controlling for socio-demographic and clinical variables, all factors had a simultaneous independent statistically significant impact in QoL, demonstrating that a more optimistic attitude, a lower stigma perception and more social support contribute to a better general well-being, a better physical health and a better mental health. Results also showed a partial mediation effect of optimism between stigma perception and general well-being/mental health.

P3.5 Development and validation of clinical prediction models

P3.5.3

Logistic regression and linear discrimant analysis for assessing factors related to genetic anemia: a comparison of both approaches

U Aguirre[1,2], E Urrechaga[3]
[1]Research Unit, Hospital Galdakao-Usansolo, Galdakao-Usansolo, Spain, [2]REDISSEC Health Services Research on Chronic PatientsNetwork, Bilbao, Spain, [3]Laboratory, Hospital Galdakao-Usansolo, Galdakao-Usansolo, Spain

Aims: Logistic regression (LR) and linear discriminant analyses (LDA) are statistical methods which can be used for the evaluation of the associations between various covariates and a categorical outcome. Both methodologies have been extensively applied in research, especially in medical and sociological sciences. Although the theoretical properties have been studied extensively throughout the literature, the choice of the proper method in data analysis is still a question for the researcher. The aim of this work is to explore the performance of the two analytical methods to the detection of genetic anemia.
Methods: A set of 1108 blood samples was divided into the training (60%) and test groups (40%). Red Blood Cells (RBC), hemoglobin (Hb), mean cell volume (MCV), mean cell hemoglobin (MCH) and RBC distribution width (RDW) were considered as independent variables whereas the genetic anemia as outcome. LR and LDA methods were applied to both data sets. Sensitivity, Specificity, Negative and Positive Predictive Values and Accuracy have been evaluated.
Results: When trying to classify genetic anemia, 82.5% and 78.21 % of the genetic anemias were correctly classified by LDA and LR, respectively. As for the sensitivity, LR showed higher value than LDA approach. Overall accuracy was higher (81.37%) using LDA than LR (79.30%).
Conclusions: LDA presents the advantage over classical analysis (LR) that it can be applied to discriminate two groups: acquired anemia and genetic anemia, independently of the clinical state of the carrier at the moment of the analysis.

P3.5.8

An investigation of performance measures developed to validate risk models for survival data

G Ambler[1], S Rahman[2], B Choodari-Oskooei[1], RZ Omar[1]
[1]UCL, London, United Kingdom, [2]Institute of Statistical Research and Training, Dhaka, Bangladesh

When developing a risk prediction model for survival data it is essential that the performance of the model is evaluated in validation data using appropriate performance measures. Although a number of measures have been proposed, there is only limited guidance regarding their use in practice.
A simulation study based on two clinical datasets was conducted to investigate a wide range of performance measures. Measures were selected from categories that assess overall performance (Graf's Brier score and IBS, Schemper's V and measures from Kent and O'Quigley, and Schmid), discrimination (Harrell, Uno and Gonen's concordance indices and Royston's D) and calibration (calibration slope) of a model and were evaluated with respect to their robustness to censoring and ease of interpretation. Some of the measures needed to be modified for use in validation data.
The overall performance measures were all reasonably robust to moderate levels of censoring. The most commonly used discrimination measure, Harrell's C, was considerably affected by censoring and tended to increase as censoring increased. In contrast, Uno and Gonen's C indices were reasonably stable in the presence of censoring, as was Royston's D. The calibration slope was not affected by censoring.
We recommend that Uno's C is used in practice to quantify concordance and that D is reported alongside since it has an appealing interpretation. Any of the overall performance measures could be recommended but we prefer Graf's measures as they are robust to high levels of censoring. The calibration slope can also be recommended.

P3.5.10

PREVEXEPOC: a computer tool for risk stratification of patients with exacerbated COPD based on a predictive severity scoring system

I Arostegui[1,2], MJ Legarreta[1], I Barrio[1], A Unzurrunzaga[3], JM Quintana[3]
[1]University of the Basque Country UPV/EHU, Leioa, Spain, [2]REDISSEC, Galdakao, Spain, [3]Hospital Galdakao-Usansolo, REDISSEC, Galdakao, Spain

Limited information is available about predictors of short-term mortality in patients with exacerbated COPD (eCOPD). The goal of the study was to propose a method for the development of prognostic severity scores for risk stratification of patients with eCOPD and to make them available as easy to use tools for clinical decision-making process.
The method we propose started with the development of a prediction model for short-term mortality in patients with eCOPD internally and externally validated. The next step consisted on creating a prognostic severity score that predicted the risk of mortality based on the previous statistical model. The predictive accuracy of the severity score was internally and externally validated and comparison with the original model was performed. A final step consisted on categorizing the severity score into 4 levels that defined risk groups of mortality using a novel approach to categorizing continuous variables in prediction models. Discrimination ability was tested using the area under the receiver operating characteristic (ROC) curve (AUC) and comparison of continuous and categorical versions of the score was performed with the integrated discrimination improvement (IDI).
The methodology was applied to a cohort study of patients with an eCOPD. A severity score was created without significant loss of predictive accuracy compared to the predictive model. The score was categorized into risk categories, without significant loss in discrimination ability, IDI

= 0.004. An easy to use computer tool named PREVEXEPOC was created, whereby the clinicians could stratify patients with eCOPD by their risk of short-term mortality.

P3.5.11

Assessing the prediction accuracy of cure in the Cox proportional hazards cure model

J Asano[1], A Hirakawa[2], C Hamada[3]
[1]Pharmaceuticals and Medical Devices Agency (PMDA), Tokyo, Japan, [2]Nagoya University Graduate School of Medicine, Nagoya, Japan, [3]Tokyo University of Science, Tokyo, Japan

Aims: A cure rate model is a survival model incorporating the cure rate with the assumption that the population contains both uncured and cured individuals. It is a powerful statistical tool for prognostic studies, especially in cancer. The cure rate is important for making treatment decisions in clinical practice. The proportional hazards (PH) cure model can predict the cure rate for each patient. This contains a logistic regression component for the cure rate and a Cox regression component to estimate the hazard for uncured patients. A measure for quantifying the predictive accuracy of the cure rate estimated by the Cox PH cure model is required, as there has been a lack of previous research in this area. Actually, we used the Cox PH cure model for the breast cancer data; however, the area under the receiver operating characteristic curve (AUC) could not be estimated due to the fact that many patients were censored.
Methods: In this study, we propose to use imputation-based AUCs to assess the predictive accuracy of the cure rate from the PH cure model. We examined the precision of these AUCs using simulation studies.
Results: The results demonstrated that the imputation-based AUCs were estimable and their biases were negligibly small in many cases, although ordinary AUC could not be estimated.
Conclusion: The proposed imputation-based AUCs are useful for assessing the predictive accuracy of cure rates from the Cox PH cure model.

P3.5.25

External validation of a prognostic model

LJ Bonnett[1], AG Marson[1], A Johnson[2,3], L Kim[4], JW Sander[5,6,7], N Lawn[8], E Beghi[9], M Leone[10], C Tudur Smith[1]
[1]University of Liverpool, Liverpool, United Kingdom, [2]MRC Biostatistics Unit, Cambridge, United Kingdom, [3]MRC Clinical Trials Unit, London, United Kingdom, [4]London School of Hygiene & Tropical Medicine, London, United Kingdom, [5]NIHR Biomedical Research Centre, King's College London, London, United Kingdom, [6]University College London, London, United Kingdom, [7]Epilepsy Society, Buckinghamshire, United Kingdom, [8]Royal Perth and Fremantle Hospitals, Perth, Australia, [9]IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy, [10]Ospedale Maggiore della Carita, Novara, Italy

Background: Prognostic models can be used to stratify patients. However, before a model can be used for this purpose, it needs to be validated externally in independent data. We demonstrate methods of external validation, including methods for handling a covariate missing from the validation dataset, via a prognostic model for risk of seizure recurrence following a first ever seizure.
Methods: Three independent datasets were obtained. External validation was evaluated for each dataset via discrimination using Harrell's c-index. Calibration plots were also considered as was a measure of prognostic accuracy, R^2 Brier. Five imputation methods were examined to handle a covariate missing from one validation dataset. These included hot deck imputation and multiple imputation.
Results: Trial data for 620 people with epilepsy was used to develop the original model; the validation datasets consisted of 274, 307 and 847 trial participants respectively. The model generalised relatively well to the other datasets. All five methods of imputation performed fairly similarly.
Conclusions: Prognostic models can be validated by considering calibration and discrimination methods and predictive accuracy measures. Although there are limitations to external validation, it is still a necessary part of modelling. Our prognostic model, used to inform driving regulations, has been validated and consequently has been proven as a valuable tool for predicting risk of seizure recurrence following a first seizure in people with various combinations of risk factors. Additionally, there is evidence to support one worldwide overall prognostic model for risk of second seizure following a first.

P3.5.37

Partial least square discriminant analysis of neurological outcome after cardiac arrest using bispectral index

O Collignon[1], P Stammet[2], Y Devaux[1]
[1]CRP Santé, Strassen, Luxembourg, [2]Centre Hospitalier de Luxembourg, Strassen, Luxembourg

Early prediction of neurological outcome after cardiac arrest would represent a major breakthrough towards personalized medicine. This would indeed allow physicians to adapt patients' healthcare in the first hours of intensive care unit stay.
To achieve this goal, the bispectral index (BIS), which has initially been developed as a tool to measure the depth of anesthesia, was monitored minute-by-minute during the first 24h after cardiac arrest by electroencephalography (EEG) in 96 patients in order to predict their cerebral performance category (CPC) at 6 months of follow-up. Since the predictive BIS time points were highly correlated and more numerous than the number of patients, Partial Least Square Discriminant Analysis was performed to model these functional data. Mean BIS, area under BIS curve and BIS slope over time were also evaluated.
These methods were compared at each time point by plotting ROC curves and detecting the cut-offs maximizing Youden index. Sensitivity and specificity were corrected for optimism using bootstrap internal validation.
Finally, PLS-DA, mean BIS and area under BIS curve outperformed the slope method and all reached the same optimal prediction around 12h, with a ROC AUC of 0.89, a specificity of 89% and a sensitivity of 86% (resp. 87% and 85% after correction). Added value of mean BIS to a multivariate clinical model was evaluated using continuous Net Reclassification Index (1.44%,p=4E-12) and Integrated Discrimination Improvement (0.29, p=2E-9).

P3.5.45

Internal-external cross-validation (IECV) for prognostic model research using data from multiple studies: potential & pitfalls

J Ensor[1], K Snell[2], D Moore[1], D Fitzmaurice[1], RD Riley[1,2]
[1]University of Birmingham, Birmingham, United Kingdom, [2]MRC Midland Hub for Trials Methodology Research, Birmingham, United Kingdom

Background: Validation is a crucial step in the acceptance of any prognostic model within clinical practice. While the majority of models are internally validated, very few are externally tested. However, the availability of individual participant data (IPD) from multiple studies can address this through IECV. We highlight strengths and weaknesses of the approach with application to a venous thromboembolism (VTE) recurrence prediction model.
Methods: The IECV approach iteratively selects N-1 studies from the N total studies available, and the prognostic model is developed within this subset of studies, leaving the remaining study for external validation of the model. A total of N different models are derived (one for each set of included studies) and each is validated in the omitted study. Model perfor-

mance can then be summarised across the omitted studies.
Application: Utilising IPD from seven studies to develop a prognostic model for VTE recurrence in unprovoked patients, we describe strengths and complications of the IECV approach, and subsequent adaptations. Challenges include dealing with studies with small numbers of events, missing data from one or many studies and heterogeneity across study populations.
Conclusions: There are strong benefits to using the IECV approach for prognostic model building, and using IPD from several studies can produce more robust models. However challenges remain to be addressed.

P3.5.55
Creation and validation of a predictive model to assess poor outcomes in acute decompensated heart failure

AG Unzurrunzaga[1], S Garcia-Gutierrez[1], I Arostegui[2], I Barrio-Beraza[2], U Aguirre[3], A Antón[1], JM Quintana[1]
[1]Hospital Galdakao-Usansolo, Osakidetza-REDISSEC, Usansolo, Spain, [2]Universidad del País Vasco-REDISSEC, Leioa, Spain, [3]Hospital Galdakao-Usansolo, REDISSEC, Usansolo, Spain

Objective: To create a predictive model to assess severity in acute decompensated heart failure at acute setting.
Methodology: Prospective cohort study. Patients over 18 years who were seen between April 2011 and April 2013 for acute decompensated heart failure were included. Sociodemographic, cardiovascular risk factors, comorbidities, functional status, and general medical history of heart disease and analytical and echocardiographic data were collected. Dependent variable was "poor evolution", a composed variable defined as:
1) mortality in acute period (one week);
2) admission to an intensive care unit (ICU);
3) need for invasive mechanical ventilation (IMV);
4) cardiac arrest;
5) use of non-invasive mechanical ventilation (NIMV).
Statistical analysis: Categorical variables are expressed as frequencies and percentages and continuous variables as means and standard deviations. Model was created by means of multilevel logistic regression model adjusted by hospital . The model as well as its AUC was validated by split validation techniques. Bootstrapping was used to internally validate the models' AUC.
Results: The 51.35% of the 1856 patients of our sample were women, mean age was 79.6(9.7) with more than two comorbidities in the 61.69% of the cases. Background of acute myocardial infarction, previous visits to ED in the two previous years, glycemia and BUN entered into the model (AUC=0.78, 95%CI 0.73-0.83). Split validation techniques showed similar AUC (0.79, 95%CI 0.74-0.84) as bootstrapping techniques (AUC=0.97, 95%CI 0.75-0.83).
Conclusions: We achieve a simple to use predictive model to assess poor evolution in acute heart failure.

P3.5.62
Model building using learning methods to identify SNPs related to the pharmacokinetics of high-dose methotrexate in pediatric acute lymphoblastic leukemia

A Harnos[1], K Csordas[2], O Lautner-Csorba[3], AF Semsei[3], M Hegyi[2], DJ Erdélyi[2], OT Eipel[2], C Szalai[3], GT Kovács[2]
[1]Dept. of Biomathematics, Szent István University, Budapest, Hungary, [2]2nd Dept. of Pediatrics, Semmelweis University, Budapest, Hungary, [3]Dept. of Genetics, Semmelweis University, Budapest, Hungary

High-dose methotrexat plays an important role in the consolidation therapy of acute lymphoblastic leukemia. We investigated the influence of single nucleotide polymorphisms (SNPs) in genes of the folate metabolic pathway, transporter molecules and transcription proteins on the pharmacokinetics of methotrexate (MTX) and 7-hydroxy-methotrexate (7-OH-MTX). 63 SNPs of 14 genes were genotyped. Totally 463 treatment courses were analyzed (4 measurements of each patient).
As a first step, random forest method (RF) was applied to calculate variable importance measures because of the large number of explanatory variables and the relatively small sample size. With the RF, we selected the important ones from all explanatory variables for further analyses.
In the second step, we built classification and regression trees (CART) with the preselected explanatory variables to generate clinical decision rules and to explore the relationship between the response and the explanatory variables.
In the last step general linear mixed model (testing the relationship between SNPs and the logarithmic transformed serum levels) was applied to prove the significance of the selected variables and their interactions and to estimate effect sizes. SNPs (rs4948502, rs4948496, rs4948487) of the ARID5B gene were associated with the serum levels of MTX ($P<0.02$), serum levels and AUC of 7-OH-MTX ($P<0.02$). The rs4149056 of the SLCO1B1 gene showed also a significant association with the serum levels of MTX ($P<0.001$).
Our findings confirm the association of novel genetic variations in folate-related and ARID5B genes with the serum MTX levels.

P3.5.82
Prediction of outcome after severe and moderate head injury by classification and regression tree technique

VK Kamal[1], RM Pandey[1], D Agrawal[1]
[1]All India Institute of Medical Science, New Delhi, India

Traumatic brain injury is the leading cause of disability and death all over the Globe. Our aim is to develop and validate a prognostic model, which is simple and easy to use for In-hospital mortality and unfavourable outcome at 6-months in patients with moderate and severe head injury involving rapidly and easily available variables in daily routine practice.
For this, a classification and regression tree (CART) technique was employed in the analysis by using trauma database (n=1466 patients) of consecutive patients. A total of 24 prognostic indicators were examined to predict In-hospital mortality and outcome at 6 months after head injury. For In-hospital mortality, there were 7 terminal nodes and the area under curve was 0.83 and 0.82 for learning and test data sample respectively.
The overall classification predictive accuracy was 82% for learning data sample and 79% for test data sample, with a relative cost 0.37 for learning data sample. For 6-months outcome, there were 4 terminal nodes and the area under curve was 0.82 and 0.79 for learning and test data sample respectively. The overall classification predictive accuracy was 79% for learning data sample and 76% for test data sample, with a relative cost 0.40 for learning data sample.
Methodologically, CART is quite different from the more commonly used statistical methods with the primary benefit of illustrating the important

prognostic variables as related to outcome. This seems less expensive, less time consuming, and less specialized measurements and may prove useful in developing new therapeutic strategies and approaches.

P3.5.145

A clinical diagnostic model using biomarkers and clinical characteristics for the identification of MODY patients

BM Shields[1], T Mcdonald[2], K Owen[3], M Malecki[4], R Besser[1], A Jones[1], S Ellard[2], W Henley[1], A Hattersley[1]
[1]University of Exeter Medical School, Exeter, United Kingdom, [2]Royal Devon and Exeter NHS Foundation Trust, Exeter, United Kingdom, [3]University of Oxford, Oxford, United Kingdom, [4]Jagiellonian University Medical College, Krakow, Poland

Aims/Objectives: Maturity-onset diabetes of the young (MODY) is a genetic form of diabetes, often misdiagnosed as Type 1 or Type 2 diabetes, resulting in inappropriate treatment. A diagnostic model to determine a patient's probability of MODY based on clinical features (www.diabetesgenes.org) has been useful in determining which patients should receive expensive genetic testing. Biomarkers can also help identify MODY. We aimed to determine whether combining clinical features with biomarkers provides the best diagnostic model.
Methods: Type 1 biomarkers (C-peptide and islet autoantibodies) and Type 2 biomarkers (hsCRP and HDL-cholesterol) were measured in 144 Type 1, 209 Type 2, and 172 MODY patients. Logistic regression models were developed using the clinical features from the original model and biomarker results, with adjustment for the prevalence of MODY. Discriminative ability and classification improvement was assessed using ROC curves and net reclassification improvement (NRI).
Results: In patients insulin treated from diagnosis, the addition of Type 1 biomarkers improved discrimination over clinical features alone (c-statistic=0.977 v 0.914, p=0.0004); NRI=1.57 (95% CI 1.29,1.85), p<0.0001). In patients not insulin treated from diagnosis, Type 2 biomarkers made a small improvement in discrimination (c=0.986 v 0.981, p=0.06; NRI=0.80 (95% CI 0.56,1.04), p<0.0001). Both models had low prediction errors on jackknife crossvalidation (<6%) and good model fit.
Conclusion/Summary: Excellent discrimination between MODY and Type 1/Type 2 diabetes is achieved when biomarkers are used in combination with clinical characteristics. The addition of these markers to the online model will improve referrals sent for genetic testing for MODY.

P3.5.173

Comparing different methods to develop prediction models for polytomous outcomes

K Van Hoorde[1,2], K De Raedt[1], D Timmerman[3,4], S Van Huffel[1,2], B Van Calster[3]
[1]KU Leuven Dept of Electrical Engineering / ESAT-STADIUS, Leuven, Belgium, [2]iMinds Medical Information Technologies, Leuven, Belgium, [3]KU Leuven Dept of Development & Regeneration, Leuven, Belgium, [4]University Hospitals Leuven Dept of Obstetrics & Gynecology, Leuven, Belgium

When diagnostic problems have a polytomous endpoint (i.e. more than two outcome categories) polytomous risk estimation can be useful although not common in clinical practice. For example, available models for ovarian tumor diagnosis simply predict the risk of malignancy, yet borderline malignancies often require less invasive fertility-sparing procedures than invasive cancers.
The standard approach for polytomous outcomes is the multinomial logistic regression (MLR) model. We compared MLR to four approaches that involve a combination of dichotomous logistic regression (LR) models to predict the risk that an ovarian tumor is benign, borderline, or invasive: (1) sequential dichotomous models, (2) combining models for each pair of categories, (3) individualized LR, and (4) combining models for each outcome category vs the rest.
From a dataset including 5912 women and eleven predictor variables we drew a development and a validation set. The size of the development set was varied to have 5, 10, or 20 events per variable. Models were validated in terms of discrimination using the Polytomous Discrimination Index. Results for 300 development/validation set draws were averaged.
Combining one-vs-rest models yielded the lowest discrimination between categories. The reason could be that the pooling of categories in a rest group obscures differences between individual categories. For this reason, the performance of sequential dichotomous models depended on the chosen sequence: borderline vs rest followed by benign vs invasive was the worst sequence. As expected apparent performance was overoptimistic when EPV=5. With EPV=20 optimism was notably smaller yet not negligible.

P3.5.174

Assessing the influence of case-mix heterogeneity on the discriminative ability of a risk model: the model-based concordance-index

D van Klaveren[1], M Gönen[2], EW Steyerberg[1], Y Vergouwe[1]
[1]Department of Public Health, Erasmus MC, Rotterdam, The Netherlands, [2]Epidemiology & Biostatistics, Memorial Sloan Kettering, New York, United States

The discriminative ability of a risk model in a new population depends on the validity of the regression coefficients, but also on the heterogeneity of the case-mix. We propose a model-based concordance-index (mbc-index) that assumes correct regression coefficients, to assess the influence of case-mix heterogeneity. We aimed to study the behaviour of the mbc-index in external validation data.
We compared three concordance measures in a simulation study: Harrell's c-index; the mbc-index, i.e. an extension to binary outcomes of Gönen and Heller's censoring-robust concordance measure for time-to-event outcomes; and a previously proposed case-mix corrected c-index based on resampled outcomes. We first generated hypothetical samples of patients (n=400 or n=1,600) for fitting regression models. To mimic different external validation settings, we simulated new patient data with different case-mix heterogeneity and different true regression coefficients. We performed the simulations (10.000 replications per setting) for binary and for time-to-event outcomes. As expected, the mbc-index was identical to the case-mix corrected c-index for binary outcomes. The mbc-index remained stable at the true value with increasing proportions of censoring of time-to-event outcomes, while the case-mix corrected c-index increased unfavourably. The mbc-index was further illustrated in two clinical examples.
Comparison of the mbc-index in a new population with concordance measures in the model development population straightforwardly quantifies the loss (or gain) in a model's discriminative ability due to less (or more) case-mix heterogeneity. The mbc-index is robust to censoring of time-to-event outcomes, in contrast to the previously proposed case-mix corrected c-index.

P3.5.175

A censoring-robust concordance measure for proportional hazards regression models in external validation data: the calibrated Gönen and Heller estimator

D van Klaveren[1], M Gönen[2], EW Steyerberg[1], Y Vergouwe[1]
[1]Department of Public Health, Erasmus MC, Rotterdam, The Netherlands, [2]Epidemiology & Biostatistics, Memorial Sloan Kettering, New York, United States

For validation of proportional hazards regression models within the model development data, the Gönen and Heller (GH) concordance probability estimator is a censoring-robust alternative to Harrell's concordance-index (c-index). In an external validation population it merely assesses the influence of case-mix heterogeneity since it uses the regression coefficients of the development population. To estimate the concordance probability in external validation data we propose to apply the GH estimator to predictions with the regression slope calibrated to the external validation data, i.e. the calibrated GH estimator. We aimed to study the behaviour of the calibrated GH estimator in external validation settings with a focus on its sensitivity to censoring.

In a simulation study we compared the calibrated GH estimator with Harrell's c-index. We first generated hypothetical samples of patients (n=400 or n=1,600) for fitting Cox regression models. To mimic different external validation settings, we simulated new patient data with different true regression coefficients and different case-mix heterogeneity (10.000 replications per setting). In each setting we varied the amount of censoring from 0% to 70%. The calibrated GH estimates remained stable at the true value with increasing proportions of censoring, while the c-index increased unfavourably. The calibrated GH estimator was further illustrated in a clinical example.

We conclude that the calibrated GH-estimator is an attractive, censoring-robust alternative for the c-index when assessing the discriminative ability of a proportional hazards regression model in external validation data.

Wednesday, 27th August 2014 - 15.30-16.00

Poster session P4

P4.1 Meta-analysis and network meta-analysis

P4.1.4

Developing and validating risk prediction models in an individual participant data meta-analysis

I Ahmed[1], TPA Debray[2], KGM Moons[2], RD Riley[1]
[1]University of Birmingham, Birmingham, United Kingdom, [2]University Medical Center Utrecht, Utrecht, The Netherlands

Background: Risk prediction models estimate the risk of developing future outcomes for individuals based on one or more underlying characteristics (predictors). We review how researchers develop and validate risk prediction models within an individual participant data (IPD) meta-analysis, in order to assess the feasibility and conduct of the approach.
Methods: A qualitative review of the aims, methodology, and reporting in 15 articles that developed a risk prediction model using IPD from multiple studies.
Results: The IPD approach offers many opportunities but methodological challenges exist, including: unavailability of requested IPD, missing patient data and predictors, and between-study heterogeneity in methods of measurement, outcome definitions and predictor effects. Most articles develop their model using IPD from all available studies and perform only an internal validation (on the same set of data). Ten of the 15 articles did not allow for any study differences in baseline risk (intercepts), potentially limiting their model's applicability and performance in some populations. Only two articles used external validation (on different data), including a novel method which develops the model on all but one of the IPD studies, tests performance in the excluded study, and repeats by rotating the omitted study.
Conclusions: An IPD meta-analysis offers unique opportunities for risk prediction research. Researchers can make more of this by allowing separate model intercept terms for each study (population) to improve generalisability, and by using 'internal-external cross-validation' to simultaneously develop and validate their model. Methodological challenges can be reduced by prospectively planned collaborations that share IPD for risk prediction.

P4.1.38

Meta-analysis of single-arm survival studies: a distribution-free approach for estimating summary survival curves with random effects

C Combescure[1], Y Foucher[2], D Jackson[3]
[1]University of Geneva & University Hospitals of Geneva, Geneva, Switzerland, [2]University of Nantes, Nantes, France, [3]Institute of Public Health, Cambridge, United Kingdom

Regression models have been proposed to combine survival curves from various studies in a single summary survival curve. This approach is useful for meta-analysis of survival studies. We propose an alternative approach, which is distribution-free, for meta-analysis of single-arm survival studies. It is an extension of the product-limit estimator of survival for aggregate data accounting for between-study variability.
The survival estimates are read off from survival curves at a set of time-points. The conditional survival estimates, derived from the collected survival estimates, go through an arc-sine transformation and then are pooled. The between-study variability is accounted by using the extension of the DerSimonian and Laird's methodology for multiple outcomes. A back transformation is applied to obtain the pooled conditional survival estimates. The pooled survival at each time point is obtained by a product of pooled conditional survival estimates. Statistics I^2 and H^2 are used to quantify the impact of the heterogeneity in the published survival curves and we propose a statistical test for the between-strata comparison. The performance of the proposed approach was evaluated in a simulation study with various Weibull survival models, sample sizes and censoring rates.
This simulation study showed that our approach is biased when events are rare and sample sizes are small. It performs well in other situations. An application of our approach on aggregate data of 27 studies will be given, with the aim to synthesize the survival of untreated patients with hepatocellular carcinoma.

P4.1.42

Publication bias tests for survival data: a simulation study

TPA Debray[1], KGM Moons[1], H Koffijberg[1], RD Riley[2]
[1]UMC Utrecht, Utrecht, The Netherlands, [2]University of Birmingham, Birmingham, United Kingdom

The presence of publication bias is often verified in meta-analyses by visual inspection of the funnel plot. Statistical tests may estimate the association between the reported effect size and their standard error (Egger´s test), total sample size (Macaskill´s test) or inverse of the total sample size (Peter´s test). Although these tests have been evaluated for pooling odds ratios, their application may be less appropriate for survival data where censoring influences statistical significance (and thus selective reporting) of the hazard ratio.
Here, we propose and evaluate two new publication bias tests for survival data that are based on the total number of events (Test-1) and the total survival time (Test-2). We compare their performance to existing tests in an extensive simulation study where more than 20,000,000 meta-analyses were generated. Here, we varied the true hazard ratio (HR=0.20-1.00), the number of available studies (N=10-100), the censoring proportion (cp=0.10-0.50) and the scale of the hazard distribution. Furthermore, we used a set of predefined reflecting meta-analyses of randomised clinical trials in the medical literature.
When treatment groups are balanced, simulation results demonstrate that type-I errors are problematic for Egger´s test (averaging from 0.110 for N=10 to 0.195 for N=100), but consistently good (around 0.10) for Peter´s test and Test-1. The power of all tests was low; for example Test-1 yielded power from 0.112 (for N=10) to 0.208 (for N=100). Finally, we compare and discuss the performance of Peter´s test and Test-1 in imbalanced treatment groups, and make recommendations for practice.

P4.1.80

Meta-analysis of mobile phone-based interventions for smoking cessation trials in different countries

Y Jiang[1], Y Huang[1], M Ybarra[2], L Abroms[3], C Free[4], R Whittaker[5]
[1]The University of Auckland, Auckland, New Zealand, [2]Center for Innovative Public Health, San Clemente, United States, [3]George Washington University, Washington, United States, [4]London School of Hygiene & Tropical Medicine, London, United Kingdom, [5]National Institute for Health Innovation, Auckland, New Zealand

Tobacco use contributes to 12% of all deaths among adults ages 30 years and older. Smoking cessation programs delivered via mobile phone text messaging are used worldwide to help people quit smoking. Although results suggest increases in self-reported quitting, at least in the short-term, little is known how this relates to different cultures.
Individual participants' data (N=8,315) collected in five randomised control trials in New Zealand, UK, USA and Turkey were included for meta-

analysis. Frequency and intensity of intervention strategies and outcome assessments were compared. The primary outcome was self-reported seven-day point prevalence (PP) at 4-weeks since the quit day. Secondary outcome was self-reported PP at 3 months.
Treatment evaluations were performed on the principle of intention-to-treat. Participants who were lost to follow-up were treated as smokers. Generalized linear mixed models were fit to estimate the overall treatment effect, while accounting for clustering within individual studies. Estimates were adjusted for age, biological sex, socioeconomic status, previous quit attempts, and the baseline Fagerstrom score.
There was no evidence of heterogeneity across individual studies. Overall, 28% of the smokers in the intervention group (1188/4202) quit smoking at 4-weeks, compared to 12% in the control (499/4113). The adjusted odds ratio (OR) was 2.9 (95% CI 2.58-3.28, p-value<.0001). A strong but reduced treatment effect was observed at 3-months (OR 1.88, 95% CI 1.53-2.32, p-value<.0001).
Data from across four countries suggest that mobile phone-based smoking cessation programs have a consistent effect at increasing the chance of quitting across a diversity of cultures.

P4.1.87
Diagnostic accuracy of Pap testing and human papillomavirus DNA testing in cervical cancer screening in Korea

MJ Ko[1], Y Lee[2], J Shim[1], S Choi[1], J Kim[1]
[1]National Evidence-based Healthcare Collaborating Agency, Seoul, Republic of Korea, [2]CHA Bundang Medical Center, Gyeonggi-do, Republic of Korea

Although screening program based on Pap testing (PAP) has substantially reduced cervical cancer in Korea, there is a paucity of study to compare the diagnostic accuracy among screening methods. This study performed a systematic review and meta-analysis of accuracy for PAP and Human papillomavirus DNA Testing (HPV) among asymptomatic women.
Studies were included if PAP and/or HPV were conducted as index test among Koreans and published after 1995. We extracted sensitivity and specificity for the detection of cervical intraepithelial neoplasia 1(CIN 1) or worse based on colposcopy and histology of colposcopy-targeted biopsies. Two independent reviewers appraised and rated the quality of studies using the Quality Assessment of Diagnostic Accuracy (QUADAS-II). Summary ROC curve and bivariate random effects model were applied to synthesize diagnostic accuracy with 95% confidence interval. The ratings of risk of bias among 44 papers were as follows; most papers were categorized into unclear in the selection and application criteria. Bivariate random effects meta-analysis estimated the pooled sensitivity and specificity of 21 studies for PAP as 0.85(95% CI: 0.77-0.90) and 0.74(95% CI: 0.58-0.86), respectively for detecting ascus or worse. In HPV test, based on 24 studies, the pooled sensitivity and specificity was 0.76(95% CI: 0.68-0.82), and 0.80(95% CI: 0.73-0.86), respectively. PAP appeared to have a higher sensitivity but lower specificity than that of HPV.
However, included studies had poor or unclear quality with respect to interpretation of the test results. More rigorous studies are needed.

P4.1.108
Mean platelet volume and coronary artery disease: a systematic review and meta-analysis

N Sansanayudh[1], D Muntham[2], T Anothaisintawee[3], P Numthavaj[3], S Rattanasiri[3], MM Evoy[4], J Attia[4], A Thakkinstian[3]
[1]Phramongkutklao Hospital, Bangkok, Thailand, [2]Mahidol University, Bangkok, Thailand, [3]Ramathibodi Hospital, Mahidol University, Bangkok, Thailand, [4]Faculty of Health and Medicine, University of New Castle, Callaghan, Australia

Background: Platelets play an important role in the pathophysiology of coronary artery disease (CAD). Mean platelet volume (MPV) has been proposed as platelet reactivity, and might emerge as a potential marker of CAD risk.
Objective: To conduct a meta-analysis aiming at comparing mean difference in MPV between CAD and controls, and pooling the odds ratio (OR) of CAD in those with high versus low MPV.
Methods: We searched MEDLINE since initiations to 12 March 2013. All observational studies that reported MPV in CAD and control groups were included**.** Two reviewers independently extracted data. An unstandardised mean difference (USMD) with a random-effects model was applied for pooling the mean difference, and DerSimonian & Laird method was used for pooling ORs.
Results: Of the 1092 identified studies, 40 were included in this meta-analysis. The MPV was significantly larger in patient with CAD than controls with the USMD of 0.70 fL (95% CI: 0.55, 0.85). The USMD of MPV in acute coronary event and stable CAD were 0.84 fL (95% CI: 0.63, 1.04) and 0.55 fL (95% CI: 0.32, 0.78), respectively. Patients with higher MPV (≥ 7.3 fL) also had higher odds of having CAD than patients with lower MPV with a pooled OR of 2.28 (95% CI: 1.46, 3.58).
Conclusion: Larger MPV was associated with CAD. Patients with MPV ≥ 7.3 fL was higher odds of having CAD. These findings suggest that MPV may be an useful marker in patients with coronary artery disease.

P4.1.131
The association between the IRF6 genes and non-syndromic cleft lip with or without cleft palate: Systematic review and meta-analysis

S Rattanasiri[1], K Wattanawong[1], A Thakkinstian[1]
[1]Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

Objective: A systematic review and meta-analysis was performed aiming at determining association between IRF6 genes and non-syndromic cleft lip with or without cleft palate (NSCLP).
Methods: Studies published from 2004-2013 were retrieved for 3 polymorphisms: rs2235371 (n=13), rs2013162 (n=7) and rs642961 (n=8). Data were independently extracted by two reviewers. Minor allele frequencies (MAF) in control groups were pooled. Odds ratio (OR) of genotypic effects were estimated using a mixed-effect logit regression.
Results: For rs2235371 polymorphism, the pooled MAFs in 7 and 5 Asian- and Caucasian-studies were 0.36 (95%CI: 0.28, 0.43) and 0.08 (95% CI: 0.05, 0.12), respectively. The pooled OR_1 (AA vs GG) and OR_2 (GA vs GG) were respectively 0.49 (95% CI: 0.37, 0.63) and 0.60 (95% CI:0.51,0.71) in Asians, and 0.38 (95% CI: 0.18, 0.80) and 0.67 (95% CI: 0.53, 0.86) in Caucasians. For rs2013162, the pooled MAF in 4 Caucasian-studies was 0.42 (95%CI: 0.32, 0.51). The pooled OR_1(AA vs CC) and OR_2 (CA vs CC) were 0.87(95%CI: 0.71, 1.07) and 0.93(95%CI: 0.81, 1.08), respectively. Among 4 Caucasian studies for rs642961, the MAF was 0.20 (95% CI: 0.18, 0.23), and the pooled OR_1 (AA Vs GG) and OR_2(GA Vs GG) were1.938 (95% CI: 1.443, 2.604) and 1.568 (95% CI:1.353, 1.819), respectively.
Conclusion: This meta-analysis suggested association between IRF6 and NSCLP. Robust estimate suggested protective effect of rs2235371 in Asians and Caucasians and indicated risk effect of rs642961 in Caucasians.

P4.1.134

Bayesian meta-analysis without MCMC

KM Rhodes[1], RM Turner[1], D Jackson[1], JPT Higgins[2,3]
[1]MRC Biostatistics Unit, Cambridge, United Kingdom, [2]University of Bristol, Bristol, United Kingdom, [3]University of York, York, United Kingdom

Background: Many meta-analyses combine results from only a small number of studies, a situation in which between-study variance is imprecisely estimated when standard methods are applied. Bayesian meta-analysis allows incorporation of external evidence on heterogeneity, providing the potential for more robust inference on the effect size of interest.
Methods: We propose two methods for performing Bayesian meta-analysis, using data augmentation and importance sampling techniques. Both methods are implemented in standard statistical software and provide much less complex alternatives to Markov chain Monte Carlo (MCMC) approaches. In a simulation study, we compare the performance of the proposed methods.
Results: An importance sampling approach produces almost identical results to standard MCMC approaches, and results obtained through data augmentation are very similar. We compare the methods formally and also apply them to real datasets.
For example, a meta-analysis combining four studies evaluating the effectiveness of fluoride for lower limb pain is considered. In a conventional random-effects meta-analysis, the between-study variance τ^2 is high at 1.78, but very imprecisely estimated (95% CI 0.39 to 52.2).
The estimated summary odds ratio is 4.14 (95% CI 0.92 to 18.4). When incorporating an informative inverse-gamma prior for τ^2 using importance sampling, the heterogeneity estimate reduces to 0.54, with 95% credible interval 0.04 to 5.33. The summary odds ratio changes to 3.46 (95% CI 1.17 to 14.3).
Conclusion: The proposed methods facilitate Bayesian meta-analysis in a way that is accessible to the applied researchers who are commonly carrying out meta-analyses.

P4.1.135

Individual patient data network meta-analysis with a time-to-event endpoint in head and neck cancer

L Ribassin-Majed[1], G Le Teuff[1], J-P Pignon[1], S Michiels[1]
[1]Institut Gustave Roussy, Service de Biostatistique et d'Épidémiologie, Villejuif, France

Background: While the use of individual patient data (IPD) is the gold standard for meta-analysis (MA), network meta-analysis typically only uses summary data. We investigated how to extend network meta-analysis to IPD survival data in head and neck cancer.
Materials and methods: From a previous IPD-MA comparing radiotherapy (RT) + neoadjuvant chemotherapy (A) or RT + concomitant chemotherapy (B) to RT (reference group) or directly A to B, a network meta-analysis combining direct (6 trials, 861 patients) and indirect (84 trials, 14 317 patients) evidence was constructed, which included two-arm and three-arm trials.
The aim was to compare B versus A. We constructed stratified Cox model using both fixed and random effects with and without adjustment for patient covariates. The two random effects for the treatment contrasts are both assumed to follow a $\sim N(0, \sigma^2)$ distribution.
We propose to study consistency using an interaction test.
Results: Adjusted hazard ratios (HR) comparing B to A from direct, indirect and mixed treatment comparisons are respectively HR = 0.84 95%CI = [0.67-1.06]; 0.85 [0.79-0.92] and 0.84 [0.78-0.91] with a significant effect for the last two. The corresponding HRs using the approach with 2 random effects gave similar results with small variance components (σ^2 between 6.82×10^{-5} and 0.0414). The consistency was respected with an interaction p-value =0.65.
Discussion: Our framework allows combining direct and indirect evidence in a single Cox model, either with fixed effects or with two random effects. Mixed treatment comparison using IPD survival data allows a better control of potential confounding bias than summary data.

P4.1.177

Bayesian evidence synthesis and combining randomized and nonrandomized results: a case study in diabetes

PE Verde[1], C Ohmann[1]
[1]University of Duesseldorf, Düsseldorf, Germany

There is an increasing interest in combining results from randomized control trials (RCTs) and non-randomized studies in evidence synthesis. One motivation is the generalization of results of RCTs into clinical practice, in particular to a group of patients which may not be included in RCTs for ethical reasons.
In this work we present a Bayesian hierarchical meta-regression model for combining results from different study types (randomized and non-randomized) and different data types (aggregated and individual data). Under this model, experimental and observational data are viewed as complementary sources of evidence. The model explicitly includes two types of parameters: those which are the focus of inference (e.g. treatment effect) and those which are used to describe the data collection processes. The data collection processes parameters are used to directly model potential sources of bias or inconsistencies between sources of evidence.
We illustrate this approach by combining results from recent RCTs, which investigated treatment efficacy of diabetic foot problems and we extrapola these results to patients enrolled in a prospective cohort study.

P4.1.186

Meta-analysis of continuous outcomes: systematic review of methods available for dealing with missing mean and standard deviation values

I Butcher[1], M Brady[2], SC Lewis[1], GD Murray[1], P Langhorne[3], CJ Weir[1,4]
[1]Centre for Population Health Sci., University of Edinburgh, Edinburgh, United Kingdom, [2]NMAHP Research Unit, Glasgow Caledonian University, Glasgow, United Kingdom, [3]Inst. of Cardiovascular & Medical Sci. University of Glasgow, Glasgow, United Kingdom, [4]Edinburgh Health Services Research Unit, Edinburgh, United Kingdom

Background: Omission of mean or standard deviation (SD) data from clinical trial reports, perhaps due to the skewed distribution of the outcome, prevents the inclusion of the trial in a meta-analysis, potentially causing bias.
Aims: To identify and develop improved methods for handling continuous outcomes within meta-analysis, enabling best use to be made of clinical trial evidence available.
Methods: We investigate how best to impute the mean or SD where either of these has not been reported. In certain circumstances these are suitable summary statistics to analyse in a meta-analysis, regardless of the underlying continuous distribution. We report on a comprehensive review of the literature to identify all methods of deriving missing means and standard deviations using electronic resources (including MEDLINE, EMBASE, Web of Science, BioMed Central and The Cochrane Library), relevant journals and grey literature from inception up to March 2014.
This updates a previous review (Wiebe 2006) of methods used to determine the SD and extends it to include methods for imputing the mean. Our search focuses on trial-level imputation from non-parametric summaries, but also considers algebraic recalculation from parametric summaries, trial-level imputation from external data sources or another treatment

group internally, trial-level imputation of correlation and solutions at the meta-analysis level. For each method, we note whether it calculates the mean or the SD, list which summary statistics are needed to implement it, and summarise any assumptions.

P4.1.190
Effect of lifestyle and metformin for delaying or preventing type 2 diabetes: a network meta-analysis

K Yamaoka[1], T Tango[1,2]
[1]Teikyo University Graduate School of Public Health, Tokyo, Japan, [2]Center for Medical Statistics, Tokyo, Japan

Background: A sedentary lifestyle accompanied by dietary issues is a central theme in the development of type 2 diabetes (T2DM), and amelioration of related factors is a central theme in the prevention of T2DM. Lifestyle modifica-tions are considered effective means of delaying or preventing T2DM. However, the effect of different types of therapies on resolution of this is controversial.
Objective: To assess the effects of different types of therapies on delaying or preventing T2DM in high risk for T2DM.
Methods: A systematic review and network meta-analysis of randomized controlled trials. Electronic literature search of PubMed, Medline, and the Cochrane Library for studies published up to January 2014. Randomized controlled trials of lifestyle education and other therapies (metformin) in high risk of T2DM with a follow-up of at least 6 months, reporting onset of T2DM, levels of 2hPD, and HbA1c. Pairwise meta-analyses and Bayesian network meta-analysis combined direct and indirect evidence to estimate the relative effects between treatments were used.
Results and discussion: Overall estimates by the traditional meta-analysis were calculated using a random-effects model. Under the several assumptions, the network meta-analysis was conducted. Those estimates were examined by several models.
In the presentation, results of further analyses will be also presented.

P4.2 Modeling infectious diseases

P4.2.24
A predictive model for HIV spreading in hyper-endemic settings

S Blaizot[1,2,3], B Riche[1,2,3], D Maman[4], J-F Etard[4,5], R Ecochard[1,2,3]
[1]Hospices Civils de Lyon, Service de Biostatistique, Lyon, France, [2]Université de Lyon, Université Lyon 1, Lyon, France, [3]CNRS UMR5558, LBBE, Equipe Biostatistique-Santé, Lyon, France, [4]Epicentre, Paris, France, [5]Institut de Recherche pour le Développement/UMI 233, Montpellier, France

We present a modelling approach of HIV spreading in hyper-endemic settings taking into account the sex and age structures of the population, the natural progression of HIV through different CD4-cell-count stages, and the use of antiretrovirals. The specificity of this approach is the use of the same data for both estimation and prediction.
The approach involves estimating the model parameters with a statistical method that uses likelihood decomposition. In short, the complete model is split into several simple sub-models. The predictive model was formulated as a system of sex- and age-specific differential equations and was adapted to include prospective medical interventions such as increasing the coverages of antiretroviral therapy or male circumcision.
The use of the same data for estimation and prediction, by avoiding transportability bias, allows obtaining, to some extent, more reliable results. The model is able to take into account the heterogeneity in volume and parameter values by sex and age, which is of great interest when interventions such as circumcision or pre-exposure prophylaxis are modelled. The approach can be applied to more complex models that would like to capture additional information and to data from Demographic and Health Surveys, available in many Sub-Saharan African countries.
In the near future, the model will be applied to a recently designed household survey: the "Ndhiwa HIV Impact in Population Survey" (NHIPS). This survey is a district-representative cross-sectional population survey conducted in the district of Ndhiwa (Nyanza Province, Kenya).

P4.2.58
A dynamic regression analysis of pulmonary tuberculosis incidence

D Gomes[1], AL Brás[2], PA Filipe[1], B de Sousa[3], C Nunes[4]
[1]University of Évora, School of Sciences and Technology, Évora, Portugal, [2]Faculty of Veterinary Medicine, University of Calgary, Calgary, Canada, [3]University of Coimbra, Fac. of Psych. and Educ. Sciences, Coimbra, Portugal, [4]Nova University of Lisbon, National School of Public Health, Lisbon, Portugal

Tuberculosis is a global health concern, being the second cause of death from an infectious disease worldwide. Tuberculosis infection is highly associated with airborne transmission, hence, Pulmonary Tuberculosis (PTB) being the most common form is of special importance in Public Health. Understanding what characterizes patients who suffer from Pulmonary Tuberculosis is crucial when establishing screening strategies to better control of the disease.
Monthly PTB incidence rates for Portugal mainland were analyzed, from 2000 to 2010. Two official data sources were used: the National Program for Tuberculosis Control (PNT) provided the number of PTB cases by date of diagnose, and the Statistics Portugal provided estimates of the annual population at risk for the period at study. Dynamic Regression analysis was applied to model Pulmonary Tuberculosis incidence. Variables such as sex, age, being alcoholic, diagnosed with HIV, being a smoker, an inmate and/or a homeless are some of the factors that can contribute for the diagnosis of Pulmonary Tuberculosis and thus were used as covariates in our models. The key goal of all Tuberculosis control programs is reducing the transmission of the infectious agent within the community. The transmission of Tuberculosis from one individual to another occurs mainly due to sputum positive Pulmonary Tuberculosis form. Therefore, understanding Pulmonary Tuberculosis cases and comprehend the patterns of the disease is especially important for the control of a Tuberculosis endemic.

P4.2.66
Modelling infectious disease parameters using serological data

SA Herzog[1], A Berghold[1]
[1]Inst Med Info, Stat & Docu, Medical University of Graz, Graz, Austria

A challenge in modelling and investigating infectious diseases is to use the "correct" force of infection. A special issue arises for infectious agents which are transmitted only indirectly through environmental sources like contaminated food, animal, or ground. Infected individuals can possibly be identified but these individuals are not infectious. However, serological data provide information about whether or not a person has been infected before the time point at which the serological sample was taken.
Age-specific prevalence and force of infection can be estimated using serological data under the assumption of lifelong immunity and that the epidemic is at equilibrium. If the epidemic is not at equilibrium age- and time-specific prevalence and force of infection should be estimated. We will apply and compare these methods to serological data on toxoplasma infection which are routinely collected by screening pregnant women in Austria (1991-2012). Furthermore, we investigate how repeated measurements and information of seroconversion can influence the results.

P4.2.102

Determining the risk of *Neisseria gonorrhoeae* infection by meeting location among men who have sex with men in Amsterdam

A Matser[1,2], R Geskus[1,3], M Schim van der Loeff[1,3]
[1]*Public Health Service of Amsterdam, Amsterdam, The Netherlands,* [2]*University Medical Center Utrecht, Utrecht, The Netherlands,* [3]*Academic Medical Center of Amsterdam, Amsterdam, The Netherlands*

Introduction: In individuals with concurrent sexual partnerships, the source of *Neisseria gonorrhoeae* (NG) infection is usually unknown. For directed prevention measures, knowing the infection risk per meeting location is useful.
Methods: In 2008-2009, we collected information from 3034 men who have sex with men at the Public Health Service of Amsterdam. For up to four partners per participant (*n*=8032), meeting location, partner and partnership characteristics were asked. For NG-infected participants, we assumed that one of the reported partners was the source. We used logistic regression to relate infection risk per partnership to participant and partner characteristics and meeting location. In the likelihood we accounted for partially unobserved transmission status per partnership. We predicted the partners' probability of having NG based on this model, replacing index characteristics by partner characteristics. Each partnership was assumed to be the source based on the relative size of the probabilities. Each source was linked to a meeting location. We used a Bayesian approach.
Results: We considered 16 meeting locations; two (*A* and *B*) were streets in Amster-dam with gay venues. When the infection risk was equally allocated over the partnerships, the NG probabilities were 0.085 at *A* and 0.028 at *B*. The multivariate model predicted 0.019 (95% CI 0.017-0.021) for *A* and 0.056 (95% CI 0.052-0.060) for *B*.
Discussion: Modelled NG probabilities differed from equal allocation results. NG typing data can be included, which may improve accuracy of estimates. The model can be applied to other data with partially unknown source of infection.

P4.3 Novel designs and methods for simulations

P4.3.13

Bi-cross-validation for the choice of optimal number of non-zero loadings in sparse PCA methods

N Assi[1,2], P Ferrari[1], V Viallon[2,3]
[1]*International Agency for Research on Cancer, Lyon, France,* [2]*Université Claude Bernard Lyon I, Lyon, France,* [3]*UMRESTTE, Bron, France*

Recent developments in data analysis and dimension reduction introduced a variety of sparse methods to overcome problems related to high dimensionality. While PCA is a tool of choice for such analyses, factor loadings related to its components may be difficult to interpret. Alternate techniques allow yielding eigenvectors/components with sparse loadings i.e. with a few non-zero elements. These methods require the user to give the number of non-zero elements to appear on each eigenvector.
However the cardinality is not known a priori, and while some methods have a simple cross-validation routine included, the usual approach to obtain this parameter is prone to a level of subjectivity often found by trial and error.
Inspired by the work of Owen and Perry on unsupervised cross-validation for singular value decomposition, non-negative matrix factorization and k-means, we devised a bi-cross-validation algorithm to help in deciding the optimal cardinality to use in these methods. A bi-cross-validation for 3 different sparse eigenvalue problems was developed, including: Sparse PCA using the elastic net penalty (Zou, Hastie, Tibshirani, 2006), SPC, that uses a penalised matrix decomposition (Witten, Tibshirani, Hastie, 2009) and Truncated Power Method (Yuan and Zhang, 2011). The performance of this algorithm in terms of successfully recovering the underlying sparse covariance structure of the data, together with the reproducibility of results, was evaluated with a simulation study. The bi-cross-validation was applied to nutritional data and application to metabolomics is planned. Developments to make the procedure help identify the optimal number of components to retain are currently on-going.

P4.3.70

Calculating sample size for cluster-randomised trials with mid-point sample size re-assessment

R Hooper[1]
[1]*Queen Mary University of London, London, United Kingdom*

Sample size re-assessment in a clinical trial involves estimating nuisance parameters at an interim analysis, and using them to re-calculate sample size. This can rescue an initially underpowered trial. Though interim estimation biases the significance level of a naive statistical analysis, an unbiased test can be achieved by combining p-values from the two stages of the trial using methods for adaptive trials.
In individually randomised trials it is known that little bias arises from a naive analysis if the interim analysis includes 20 individuals or more, but adaptive methods could be particularly useful for cluster-randomised trials, where two nuisance parameters must be estimated at interim from (perhaps) a small number of clusters. In the latter case it may make sense to optimise the information available at the interim analysis by planning it to fall half-way through recruitment.
I describe how to choose a sample size re-assessment design with given power under initial assumptions about nuisance parameters, where the interim analysis is planned to occur after recruitment reaches its expected mid-point. Sample size in this case cannot be calculated analytically, but may be determined by simulation (I will demonstrate using Stata).
Important considerations when deciding whether to use a sample size re-assessment design are the variability in actual sample size, and the extent to which power is successfully controlled at the desired level under departures from initial assumptions. These aspects of performance should also be simulated before deciding against the use of a simpler, fixed sample size trial design.

P4.4 Observational studies and causal inference methods

P4.4.14

Prevalence of internet use and Internet addiction disorder among medical students: a case from low income country

S Awan[1]
[1]*Aga Khan University, Karachi, Pakistan*

Background: Prolonged use of internet caused a series of problems such as internet dependence, problematic internet use, compulsive internet use, internet abuse and Internet addiction disorder (IAD), This has aroused attentions of researchers all over the world, with IAD being recognized as a mental disorder.
The aim of study is to investigate the prevalence and risk of Internet addiction disorder (IAD) and the associated factors amongst medical students

Methodology: A cross-sectional study was conducted among 412 students at different medical colleges in Karachi.
For measuring internet addiction the Internet addiction test tool was applied. This test was developed by Yung (1998). Yung defined 20-39 points as an average user, 40-69 points as a possible or intermittent user, and more than 70 points as an addicted user. The subjects' psychological well-being were assessed by self-reporting questionaire-20 (SRQ20).
Results: Internet addiction test showed that 74% of the sample were average internet users, 24% as problem over-users and 2% as addicted to the internet. The score of psychological morbidity was higher in internet addicted or problem over-users as compare to average user (p value<0.001). over-users and addicts spending increasingly more time in online activities, being more socially anxious, and gaining greater support from internet social networks (p value=0.003) and online chatting (p value=0.01) than average internet users.
Conclusion: Internet addiction is prevalent among students and Key health messages and interventions to reduce stress and anxiety among students may help in curtailing the burden of this disease which has serious adverse consequences.

P4.4.15
Quality of life (QoL) after stroke in five European populations assessed by health survey form (SF-12) and EuroQoL (EQ-5D)
S Ayis[1], C Mckevitt[1], A Rudd[2], D Parkin[1], C Wolfe[1]
[1]King's College, London, United Kingdom, [2]St Thomas' Hospital, London, United Kingdom

Background: Health-related quality of life (HRQoL) of stroke patients is likely to be impaired, due to functional and psycho-social limitations. Little is known about differences in HRQoL in different countries. The generic measure (SF-12) and a preference based measure (EQ-5D) were used to understand the distribution of HRQoL in five European countries.
Methods: Data on 1,848 ischemic or intra-cerebral haemorrhage (ICH) stroke patients were obtained from population-based registers in five European countries: Dijon (France); Kaunas (Lithuania); London (UK); Warsaw (Poland), and Sesto-Fiorentino (Italy), between 2004 and 2006. Quality of life was measured using the physical component summary (PCS) and the mental component summary (MCS) of SF-12. EQ-5D scores were derived using a standardized response mapping algorithm, based that predicts probabilities of response for EQ-5D questions, using SF-12 responses. Mixed effect regression models were used to investigate variations in HRQol, between populations, adjusting for case-mix, and socio-demographic factors.
Results: The PCS and MCS scores differed significantly across countries. Lower HRQoL estimates were associated with old age, incontinence, and physical dependency in the acute phase. Dijon, has the lowest utility score with a difference of 0.11 (0.06-0.16) significantly lower than London (reference) and Sesto-Fiorentino, and Warsaw both have a utility of 0.07 (0.2-0.11) significantly higher than London.
Conclusion: HRQoL of stroke patients varied widely across populations. There were significant associations between quality of life and patient characteristics. Extracting EQ-5D provides important information that may be used to estimate quality adjusted life years (QALY), to interpret QALY and compare costs across populations.

P4.4.46
Discrepancies in exercise therapy prescriptions after hip replacement: a multicenter survey of surgeons, rehabilitation physicians and physiotherapists
C Eulenburg[1], A-L Rahlf[2], A Kutasow[2], A Zech[2]
[1]University Medical Center, Medical Biometry and Epidemiology, Hamburg, Germany, [2]Institute for Human Movement Science, Hamburg Univerity, Hamburg, Germany

Total hip replacement is one of the most performed surgeries across the world in the past decade. With ageing of the population, the number of patients requiring hip surgery may even increase. Rehabilitation after surgery is expected to reduce pain and optimize the functional recovery. By now, evidence-based therapy guidelines on rehabilitation after hip replacement are not reported. The objective of our study was the evaluation of the present rehabilitation practice in Germany and the analysis of discrepancies among rehabilitation facilities.
In a multicentre cross-sectional study, aims, interventions, exercise dosage prescriptions and individual beliefs for postoperative treatment were assessed among rehabilitation professionals using a standardized questionnaire. Furthermore, we collected information on profession and experience of subjects and on the type of facility. Statistical analysis contained linear mixed models, accounting for random effects of the centres. To evaluate the accordance among subjects, an agreement measure was constructed.
313 questionnaires from 28 rehabilitation centres were available for statistical analysis. Results show a high variability of the current practice. Subjects disagreed on more than half of the requested items. Recommendations concerning the time interval between surgery and start of rehabilitation differed significantly between professions.
Findings reveal that the current rehabilitation practice after hip replacement in Germany is very inconsistent. The measured variability among professions suggests that the therapies may follow personal experiences and convictions more than standardized guidelines.

P4.4.94
Application of propensity-score matching: comparison of CT and ultrasonography in the effect on negative appendectomy
J Lee[1], S Ahn[1], JH Park[1], HJ Kim[2], B Kim[1], SH Park[2], KH Lee[1]
[1]Seoul National University Bundang Hospital, Gyeonggi-do, Republic of Korea, [2]University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea

Purpose: The aim of this study was to compare computed tomography (CT) and ultrasonography (US) as the preoperative first-line imaging using propensity score matching, measured in terms of negative appendectomy rate (NAR) retrospectively.
Materials and methods: From the original database which was conducted by the Low-dOse CT Appendicitis Trial (LOCAT) group, we included 2312 adolescents and adults patients who visited the site emergency departments and underwent appendectomy at 11 hospitals in metropolitan Seoul in 2011. We carefully selected 11 independent variables regarding patients and hospital characteristics and evaluated any imbalance between the CT and the US group. The propensity score was calculated using a logistic regression with the confounders and their pair-wise interaction terms. Matching was performed with greedy algorithm and 1:1 ratio. Finally, the effect of imaging modalities on NAR was assessed using McNemar's test and conditional logistic regression analysis in the matched subgroups.
Results: Before matching, NAR in the US group was significantly higher than in the CT groups (Chi-square test: $\chi 2=7.429$; P=.006; odds ratio [OR], 2.429; 95% confidence interval [CI]: 1.319, 4.473; P=.004).
The results of propensity score matching revealed no significant differ-

ence in NAR between two groups (McNemar´s test: $\chi 2=0.235$; P=.628; OR=1.429; 95% CI: 0.544, 3.753; P=.469).

Acknowledgement: This study funded by the Korea Healthcare Technology R&D Project, Ministry of Health and Welfare, Republic of Korea (HI13C00040200), the Seoul National University Bundang Hospital Research Fund (04-2011-004), and Dasol Life Science Inc (06-2013-107), all of which have no role other than funding.

P4.4.99

The mediating role of mental health in the relations between dietary behaviors and general health

MR Maracy[1], S Nematollahi[1]
[1]Isfahan University of Medical Sciences, Isfahan, Islamic Republic of Iran

Aim: The aim of this study is to find a pathway to explaining the effect of dietary behaviors on mental health and psychological functioning.
Material and Methods: This cross-sectional study contains 4763 participants from the employees of Isfahan University of Medical Sciences (IUMS). In this study 4 questionnaires were administered which were provided as followed: 21-item dietary behaviors questionnaire, Hospital Anxiety and Depression Scale questionnaire (HADS), and 12-item General Health Questionnaire (GHQ-12). We started the analysis with the SEM model with 4 unobserved latent and 16 observed variables.
Results: The results show that the regression coefficient (SE) of diet behavior on mental health and mental health on GHQ were -1 (0.37) and 0.02 (0.01) with their p-values as 0.007 and 0.01, respectively. For this model the estimated RMSEA is 0.062 with 90% confidence interval of (0.060-0.065). In addition, the comparative fit indices were as the following: 0.860 for NFI, 0.810 for RFI, 0.866 for IFI, 0.818 for TLI, and 0.866 for CFI. All these indices represent fairly good fit of the model to the data.
Conclusion: We can conclude that "lower scores on diet behavior - higher scores on mental health problems" and " higher mean scores in depression and/or anxiety - higher scores in GHQ domains". The SEM results showed that dietary behaviors have significant impact on depression and/or anxiety and general health status.

P4.4.104

The distribution of apolipoprotein E genotype in relation to age and origin of birth

K Mehlig[1], S Kern[2], J Kern[2], H Zetterberg[2,3], DS Thelle[1,4], I Skoog[2], L Lissner[1], K Blennow[2], A Börjesson-Hanson[2]
[1]Public Health Epidemiology, University of Gothenburg, Gothenburg, Sweden, [2]Dept of Psychiatry and Neurochemistry, Gothenburg University, Gothenburg, Sweden, [3]UCL Inst of Neurology, London, United Kingdom, [4]Dept of Biostatistics, University of Oslo, Oslo, Norway

The e4 allele of the apolipoprotein E gene (*APOE)* has been associated with increased risk for age-dependent diseases, e.g. atherosclerosis and Alzheimer's disease. In contrast, the same allele has been suggested to protect against age-related macular degeneration. When examining the age-dependent risk for a disease, it is important to consider the differential risk of genotype-related mortality.
Several studies examining the proportion of *APOE* alleles over lifespan observed a stable prevalence of e4 in midlife and a decrease from a certain age, however, the value of this turning point differed between studies. In addition, geographical variation of e4 prevalence has been found, e.g. along the European-African axis, with a minimum at Mediterranean latitudes and maxima near equator and North Pole.
We examined *APOE* allele prevalences in 4570 subjects from four Swedish population studies in relation to age (25-99 years) and origin of birth. E4 prevalence varied between 16% in Scandinavia, 5% in Southern Europe, and 6% in the Middle East, with corresponding changes in e3 prevalence. When combined in two groups, e4 prevalence was age independent and constantly higher in subjects from Nordic compared to non-Nordic countries up to 80 years. After age 92, e4 prevalence decreased from 15% to 9%, observed in subjects from Nordic countries only. The high turning point is explained by presence of participants with APOE-related diseases such as dementia in our gerontologic population studies.
We show that age-dependent variation of ethnic composition and selection bias could explain conflicting results observed previously.

P4.4.126

Estimating the variance of a propensity score matching estimator: another look at right heart catheterization data

R Pingel[1]
[1]Uppsala University, Uppsala, Sweden

This study considers the implementation of a variance estimator when estimating the asymptotic variance of a propensity score matching estimator for the average treatment effect. We investigate the role of smoothing parameters in the variance estimator and also propose using local linear estimation. Simulations demonstrate that large gains can be made in terms of mean squared error by properly selecting smoothing parameters and that local linear estimation may lead to a more efficient estimator of the asymptotic variance. The choice of smoothing parameters in the variance estimator is shown to be crucial when evaluating the effect of right heart catheterization, i.e., we either show a negative effect on survival or no significant effect depending on the choice.

P4.4.171

Sensitivity analysis of dissociative principal strata effect: application to a bone fracture prevention trial

Y Uemura[1]
[1]Tokyo University, Tokyo, Japan

In medical studies, there are situations in which the interest is to estimate the treatment effect on outcome not due to change in the variable caused by the treatment. For example, in randomized trial for osteoporosis subjects, researchers may be interested in the preventive effect of the treatment on the fracture not mediated through bone mineral density. For such research question, we may use a framework of principal stratification, and estimate the casual effects within principal strata where the posttreatment variable (BMD) is unaffected by the treatment. This dissociative principal strata effects (DPSE) may be estimated using the methods proposed by Hyden et al (Biometrics, 2005), Chiba (Biom J,2011), and Diag et al (JASA, 2012).
However, the DPSE cannot be identified without making any assumptions. Different assumptions are required for each estimation method proposed, although some of which are quite difficult to interpret and may not hold in actual studies. Using the above example of osteoporosis patients, I interpret each assumption. Also, I propose a sensitivity analysis for the estimation method proposed by Diang.
This method uses a pretreatment covariate that assumes to be independent of outcome variable given principal strata and the treatment randomized. The above methods of estimating DPSE and the proposed sensitivity analysis will be applied to the data of randomized trial of osteoporosis study.

P4.4.192

Does a predisposition to kidney disease originate during prenatal development? A cohort study from the Born in Bradford Project

HM Yuen[1], P Roderick[1], R Jeffrey[2], KM Godfrey[3], J West[4], J Wright[4]
[1]PCPS, University of Southampton, Southampton, United Kingdom, [2]Bradford Teaching Hospitals NHS Foundation Trust, Bradford, United Kingdom, [3]MRC Lifecourse Epidemiology Unit, Southampton, United Kingdom, [4]Bradford Institute for Health Research, Bradford, United Kingdom

Background: Rates of advanced chronic kidney disease (CKD) and renal replacement therapy are higher in South Asian compared to White British populations. Low birthweight (LBW) is also more frequent in South Asian populations; LBW has been associated with increased risk of chronic kidney disease which maybe due to reduced number of nephrons at birth.
Aim: To measure foetal kidney size measured from ultrasound scans in South Asian and White British pregnant women at 34 weeks gestation.
Methods: Kidney dimensions (transverse diameter, antero-posterior diameter, length, and circumference) and derived volume were assessed in a random sample of 872 White British and 715 South Asian women who were participants in the Born in Bradford cohort study. Kidney measures were compared between ethnic groups, unadjusted and adjusted for maternal age, socio-economic factors, marital status, body mass index, smoking and alcohol use in pregnancy, parity, baby's gender and birthweight for gestational age.
Results: Birthweight for gestational age at 40 weeks was 200g less in South Asian compared to White British babies. Mean kidney volume for gestational age was 8.79 cm^3 in South Asian and 10.45cm^3 in White British babies, a difference of 1.66cm^3 (95%CI -1.93 to -1.40, p<0.001). The difference was robust to adjustment for the potential confounders mentioned above (adjusted difference 1.38cm^3, 95%CI -1.84 to -0.97, p<0.001). There were smaller reductions in other foetal measures.
Conclusions: South Asians babies have smaller kidneys compared to White British babies, even after adjusting for potential confounders including birthweight. This finding may partly explain the increased risk of adult CKD.

P4.5 Other

P4.5.6

Meta-analysis of genome-wide gene-environment interactions on colorectal cancer

H Alonso[1], U Peters[2], J Gauderman[3], T Harrison[2], L Hsu[2], PT Campbell[4], D Thomas[3], SB Gruber[3], V Moreno[1]
[1]Catalan Institute of Oncology, Hospitalet del Llobregat, Spain, [2]Fred Hutchinson Cancer Research Center, Seattle, United States, [3]University of Southern California, Los Angeles, United States, [4]American Cancer Society, Atlanta, United States

Susceptibility loci identified so far for colorectal cancer (CRC) only account for a small fraction of the estimated inheritance. We have undertaken an exhaustive analysis of gene-environment (GxE) interactions aiming to identify novel susceptibility loci for CRC in 6 GWAS within the ColoRectal Transdisciplinary Study (CORECT) Consortium. Here we explore diverse statistical methods to identify GxE and meta-analyze the results. The GWAS studies contribute with 3927 cases and 4203 controls that have been genotyped with a custom array (1.3M SNPs). Whole genome variation has been imputed using the 1000 Genomes Project panel. After a QC and filtering protocol (MAF>0.01, LD<0.99), more than 4.5 million SNPs have been analyzed for GxE interactions.
Here we analyze BMI and height, variables that have been centrally harmonized through a standardized process and individual level data are available for all analysis. Diverse statistical models have been used to scan for GxE interactions, including logistic regression for case-control (CC), case-only (CO) analysis, empirical Bayes (EB), as well as more powerful 2-step procedures, such as the cocktail (Hsu et al., 2012) and EDGxE (Gauderman et al., 2013) methods. All of these procedures were implemented in the GxEScan software program (http://biostats.usc.edu/software).
A weighted testing approach has been used for two-step methods to improve the likelihood of capturing significant interactions while multiple testing is controlled. To combine the results of each study, meta-analytical techniques have been used: combination of p-values, z-values, odds-ratios. These are compared with a pool-analysis of all studies combined with fixed and random effects models.
On behalf of CORECT, ColoRectal Transdisciplinary Study

P4.5.9

Dutch national cohort for environmental health issues. The DUELS Study

C Ameling[1], M Marra[1]
[1]National Institute for Public Health and the Environment, Bilthoven, The Netherlands

Background: Databases concerning population exposure to e.g. noise, air pollution and available green space in the Netherlands are available at increasing geographical resolution. These data can be combined with data on an individual level for all Dutch residents. A retrospective cohort combining exposure and mortality was built.
Aims: To allow assessment of the association of several types of individual exposure with cause specific mortality in the Dutch population.
Methods: Local environmental exposures such as air pollution, noise and available green space where linked to 98% of all addresses in the Netherlands. Based on this linkage, demographic and mortality data were generated for the ~15 million Dutch residents on 1 January 2004. Records from the municipal registration were available from 1 January 1995 through 31 December 2010. For an initial analysis ~9 million persons of all ages who had resided at the same address for at least five years before 2004 were selected. Survival analysis was carried out on ~7 million persons over 30 years of age using Cox' proportional hazards regression model. Hazard rate estimates for several causes of death for five-year residential exposure to various environmental risk factors were determined, adjusted for confounders at the individual level.
Results: Several of the available environmental risk factors showed a statistically significant linear association with increased mortality. Detailed results will be presented.
Conclusion: The cohort provides a valuable data source for analyzing and monitoring the association between environmental risk factors at a relatively high geographical resolution and mortality in the Dutch population.

P4.5.12

Correlation between the degree of grade of astrocytomas and the current value of motor-evoked potential during brain surgery

S Asano[1], T Nakagomi[1], K Furuya[1], A Morita[2], N Kawahara[3], T Toyoda[4], T Haisa[5], M Murakami[6], K Hoya[6], A Matsuno[1]
[1]Dep. of Neurosurgery, Teikyo University School of Medicine, Tokyo, Japan, [2]Dep. of Neurosurgery, Nippon Medical School, Tokyo, Japan, [3]Dep. of Neurosurgery, Yokohama City Univ. Hospital, Yokohama, Japan, [4]Dep. of Neurosurgery, JCHO Tokyo-Shinjuku Med. Center, Tokyo, Japan, [5]Dep. of Neurosurgery, JR Tokyo General Hospital, Tokyo, Japan, [6]Dep. of Neurosurgery, Teikyo Univ. Chiba Med. Center, Ichihara, Japan

Background: During brain tumor surgery, we have sometimes received unconfirmed grading of the astrocytomas in the intraoperative rapid histological diagnoses. However, we have had some correlation between the grading of the astrocytomas and the current value of the motor-evoked potential (MEP) monitoring.
Methods: From 1995 to 2013, we performed MEP monitoring with anodal five-train stimulation of the brain surface of the primary motor cortex and the lead of contralateral muscle of the ball of the thumb during tumor removal. We retrospectively reviewed the final histological results and the current values of MEP, nitrous oxide, muscle relaxant. In all cases, we used intravenous anesthesia with porpofol.
Results: Twenty-six cases were extracted. They were 15 females and 11 males, and mean age was 52±17 years. During their anesthesia, for 12 cases nitrous oxide was used, and for five cases muscle relaxant was administered. Mean current value of the beginning MEP was 13.7±5.9 mA, and that of the last MEP was 17.4±7.1 mA. In their histological final diagnoses, the number of grade II astrocytomas was 10 (B-group), and that of grade III and IV astrocytomas were 4 and 12, respectively (total 16 cases: M-group). In the multivariate analysis with logistic regression analysis with step-wise method, we yielded two significant variables: age (odds ratio[B/M] 0.936, p=0.021) and current of the last MEP (odds ratio[B/M] 1.16, p=0.034).
Conclusion: Our logistic results suggested that the current value of last MEP had negative correlation to the degree of malignancy of astrocytomas.

P4.5.61

Calculation of target range values on a continuous scale in immunosuppression following solid organ transplantation

M Harden[1], T Friede[1]
[1]Universitätsmedizin Göttingen, Göttingen, Germany

Solid organ transplantation is followed by lifelong immunosuppression to avoid rejection of the donor's graft. Since most immunosuppressive drugs might harm the recipient's body in the long run, drug doses need to be monitored and changed over time. In practice, one tries to keep the drug concentrations within a target range to avoid both, rejection of the organ and harmful effects due to too large doses. Since existing target ranges are mainly based on expert opinions, we present different approaches to construct target ranges based on statistical models.
We assume two binary endpoints with an event on one endpoint precluding the observation of an event on the other endpoint, and model the risks of these events simultaneously, related to a continuous covariate (e.g. drug concentration). Based on these risk models, we calculate a lower and an upper cutoff value for the dose.
We propose several methods to construct confidence intervals for both cutoff points, like inverse regression and normal approximations. Since repeated measurements play a major role in these kind of analyses, we discuss approaches to combine data/ account for time-varying effects. Performance of all methods will be compared based on simulated data, covering coverage probability and sample size. The investigations are motivated by clinical trials in liver transplantation.

P4.5.71

Population-based metabolic syndrome risk score and its determinants: the Isfahan Healthy Heart Program

SM Hosseini[1]
[1]Department of Biostatistics & Epidemiology, Isfahan, Islamic Republic of Iran

Background: Metabolic Syndrome (MetSy), a most important predisposing factor for non-communicable diseases. Given different definitions used for the MetSy, recently using a score termed "Continuous MetSy Risk Score (CMetSyS)" is recommended. The Aim of this study was to provide a CMetSyS in a population-based sample of Iranian adults, and to assess its determinants.
Methods: The data of the baseline survey of a community trial entitled "the Isfahan Health Heart Program "has been used. The MetSy was defined according to the Revised National Cholesterol Education Program Third Adult Treatment Panel. All probable predictive models and their predictive performance were provided using leave- one- out cross validated logistic regression and the Receiver Operation Characteristic curve methods. Multiple Linear regression was performed to assess factors associated with the CMetSyS.
Results: The study population consisted of 8313 persons. The MetSy was documented in 1539 persons (21.86%). Triglycerides and waist circumference were the best predictive components and fasting plasma glucose had lowest area under curve (AUC). The AUC for our best model was 95.36(94.83- 95.83%). The best predictive cut off for this risk score was -1.151 with 89% sensitivity and 87.93% specificity.
Conclusion: We provided four population based leave- one- out cross validated risk score models, with moderate to perfect predictive performance to identify the MetSy in Iranian adults. The CMetSyS had significant associations with high sensitive C-reactive protein, body mass index, leisure time and work place physical activity as well as age and gender.

P4.5.72

Comparison of reference curves of anthropometric indices in two national studies conducted among Iranian children in 2003-2004 and 2009-2010

SM Hosseini[1]
[1]Isfahan University of Medical Sciences, Isfahan, Islamic Republic of Iran

Objectives: This study aims to compare the curves of anthropometric measures obtained in two national studies conducted among Iranian children and adolescents in 2003-2004 and 2009-2010.
Methods: Anthropometric measures obtained in two nationwide surveys conducted in 10-18-year-old Iranian students (n=15883 in 2003-2004 and n=5312 in 2009-2010)were compared.LMS Chart Maker Pro program was used to develop age- and gender- specific percentiles and to smooth and fit the model. Outlier data were detected with Z Scores graph and excluded from the analysis. Normality of the data and goodness-of-fit were examined via Detrended Q-Q plot and Q-Tests, respectively. Then the percentile curve graphs were plotted by using SAS software.
Results: Body mass index was higher in girls than in boys in most percentiles, especially in the higher ones. Height curves did not show considerable changes in two studies. Boys´ weight was lower than girls´ up to 15 years, and thereafter it became higher. In approximately all ages, boys´ waist circumference was higher in boys than in girls. Waist-to-height-ratio, especially in higher percentiles, was greater in boys than in girls approximately up to 13 years.
Conclusion: The growth charts of Iranian children and adolescents aged 10-18 years have changed over 5 years. Given the changes of reference curves of anthropometric measures over time, repeated surveys should be conducted to provide age- and gender-specific reference curves in different ethnic groups.

P4.5.76

IL-1A gene promoter region polymorphism and the risk of familial CAD in a Pakistani population

S Hussain[1]
[1]COMSATS Institute of Information Technology, Islamabad, Pakistan

Coronary artery disease has complex etiology with acquired and inherited risk factors. Several factors including family history have been reported in the pathogenesis of CAD. The potential role of IL-1A gene variations in pathogenesis of CAD with familial history has not been investigated so far. Serum IL-1A levels were determined from 670 sporadic samples including 335 CAD patients and 335 healthy controls by using enzyme-linked immunosorbent assay. C-reactive protein levels were measured by Tinaquant C-reactive protein (latex) high sensitive assay. Genotyping for IL-1A-889C>T polymorphism was investigated by using PCR-RFLP method. Heritability of susceptible allele was investigated from 130 trio-families with CAD affected offspring in this study.

We observed a significant increase in IL-1A and hs-CRP levels in patients than in controls (P <0.0001, respectively). The IL-1A-889C>T polymorphism was significantly associated with CAD in patients compared with healthy controls (P <0.0001). The minor allele T at -889 was more prevalent in cases vs. controls (P <0.0001). The mutant allele T is more transmitted to affected offspring from heterozygous parents (χ^2_{TDT} = 17.88 with 1 df, OR = 2.6, 95%CI = 19.04-48.68, P <0.0001).

Further, we observed a significant increased in hs-CRP concentration in sporadic CAD patients after the comparison with healthy controls (P <0.0001). The mutant genotype CT+TT was significantly associated with high levels of hs-CRP from CAD patients (P <0.0001).

For the first time we demonstrate a significant association of IL-1A-889 functional polymorphism with CAD.

P4.5.81

Selecting a classification function for class prediction in gene expression data

VL Jong[1,2], PW Novianti[1], KCB Roes[1,3], MJC Eijkemans[1]
[1]Julius Center, Department of Biostatistics, UMC, Utrecht, The Netherlands, [2]Viroscience Lab, Erasmus Medical Center, Rotterdam, The Netherlands, [3]Directorate of Quality Control and Patient Safety, UMC, Utrecht, The Netherlands

Classifiers have been shown to perform differently across datasets. Thus, a key question is how to select a classifier for a particular data. In this study, we devised a guideline based on data characteristics in choosing an optimal classifier for such data.

We simulated data using correlation structures observed in a variety of real-life datasets. We varied sample size, number of genes, proportion of informative genes, absolute log2 fold changes and pair-wise correlations.

For each simulated scenario, nine classifiers chosen from discriminant analyses, tree based, regularization/shrinkage, nearest neighbors, neural networks and partial least squares methods were constructed. To arrive at an optimal choice, the resulting error rates were clustered by hierarchical clustering. Well performing classifiers were clustered together and were summarized in a table specifying recommended classifiers for each scenario. These results were applied to eight real-life datasets.

We hypothesized group 0 (optimal) and group 1 (non-optimal) classifier(s) for each data and constructed all classifiers. The median error of each classifier was compared to those in the opposite group to determine its observed value. Using predicted and observed values, sensitivity and false positive rate (FPR) of each data were computed and a bivariate random effects model was fitted to this pair. As expected, the optimality of classifiers depended on specific scenarios. On the datasets evaluated this predicted an optimal classifier with aggregated sensitivity of 80.8% and FPR of 13.7%.

Hence, our proposed guidance allows selection of an optimal classification function based on data characteristics with a sensitivity of 80.8%.

P4.5.84

Risk of pneumonia in patients with respiratory disease according to type of inhaler devices

Y Kim[1], EJ Jang[1,2], S Choi[1], J Kim[1], CH Lee[1,3], JJ Yim[3], DK Kim[4], HI Yoon[5]
[1]National Evidence-based Healthcare Collaborating Agency, Seoul, Republic of Korea, [2]Department of Information Statistics, Andong National, Andong, Republic of Korea, [3]Seoul National University Hospital, Seoul, Republic of Korea, [4]SMG-Seoul National University Boramae Medical Center, Seoul, Republic of Korea, [5]Seoul National University Bundang Hospital, Seongnam, Republic of Korea

Objective: The objective of this was to investigate the risk of a hospital admission or an emergency room visit for pneumonia in patients with respiratory disease between MDI and DPI of ICS with/without LABA.

Methods: A retrospective Cohort study was conducted using the Korean national claims database between January 1, 2009 and December 31, 2011. We performed the individual matching to minimize the selection bias. Cox proportional hazard regression model was applied to compare the pneumonia risk of MDI vs. DPI and all data manipulation.

Results: For ICS users, eligible cohort were 63,635 patients (18,780 in the ICS DPI group; 44,855 in the ICS MDI group), and 18,759 DPI users were one-to-one matched with 18,759 MDI users. The risk of pneumonia was higher in MDI users compared with DPI users after adjustment in total cohort(HR 1.6; 95% confidence interval (CI) 1.2 to 2) and in matched cohort(HR 1.7; 95% CI 1.3 to 2.2).

For ICS/LABA users, eligible cohort were 244,699 patients (236,724 in the ICS/LABA DPI group; 7,945 in the ICS/LABA MDI group) and 7,942 MDI users were one-to-five matched with 36,690 DPI users. The risk of pneumonia was higher in MDI users compared with DPI users after adjustment both total cohort and matched cohort(HR 1.6; 95% CI 1.3 to 1.9 in the eligible cohort; HR 1.6; 95% CI 1.3 to 2.0 in the matched cohort).

Conclusions: Use of MDI seems to increase risk of pneumonia compared to use of DPI in ICS users and ICS/LABA users.

P4.5.90

Finding variant X - a framework for identification of causative variants in WGS-data

R Kreuzhuber[1], RV Pandey[1], A Weinhäusel[1], R Kallmeyer[1], I Visne[1], E Dilaveroglu[1], A Yildiz[1], A Kriegner[1]
[1]AIT - Austrian Institute of Technology, Vienna, Austria

The number of rare monogenic diseases is estimated to be >5000. For half of these the underlying genes are unknown (McKusik V.A., 2011). An increasing proportion of common diseases, such as schizophrenia or autism, previously thought to be due to complex multifactorial inheritance, are now thought to represent a heterogeneous collection of rare monogenic disorders (Mitchell KJ, Porteous DJ, 2011), the large majority of which is still unknown.

For the efficient investigation of genetic mutations next generation sequencing (NGS) technology has revolutionized molecular diagnostics. Its big advantage is fast and cost-effective sequencing of enormous amounts of nucleic acids. Whole exome and whole genome NGS reveals a never before seen amount of variants for which filtering strategies a field of intensive research.

MutAid, which is implemented on the XworX-platform, is a comprehensive variant calling pipeline for Sanger and most NGS platforms using state of the art open source mappers and variant callers. MutAid provides solid data management and quality control combined with the various filtering approaches: Integrative use of existing data bases and information contained within phenotypic ontologies. Co-analysis of results produced by those sequencing platforms is one of its key-features.

This combination should enable the genetic researcher to identify causative mutations more easily, especially in the context of rare diseases.

Thanks to the XworX platform further analysis can be appended to the workflow by anyone with minimal programming skills, because the platform allows for easy and direct integration of R-, Python- and Java-code.

P4.5.98
Use of factor analysis in assessments of clinical teaching evaluations

J Mandrekar[1]
[1]*Mayo Clinic, Rochester, MN, United States*

Factor analysis is a generic term for a family of statistical techniques concerned with the reduction of a set of observable variables in terms of a small number of latent factors. It has been developed primarily for analyzing relationships among a number of measurable entities (such as survey items or test scores).
The underlying assumption of factor analysis is that there exists a number of unobserved latent variables (or "factors") that account for the correlations among observed variables, such that if the latent variables are partialled out or held constant, the partial correlations among observed variables all become zero. In other words, the latent factors determine the values of the observed variables. Factor analysis has been widely used, especially in behavioral sciences, to assess the construct validity of a test or a scale.
The focus of this talk is to provide an introduction to factor analysis in the context of research projects from Medical Education that involve clinical teaching evaluations, for example, resident-teacher evaluations, resident´s reflection on quality improvement etc.

P4.5.128
How to choose a two-sample test for continuous variables: a new solution to an old problem

A Poncet[1], DS Courvoisier[1], C Combescure[1], TV Perneger[1]
[1]*University Hospital of Geneva, Geneva, Switzerland*

Objectives: To explore the importance of normality and sample size when choosing the best two-sample test for continuous data.
Study design: Simulation study that compared tests (T, Mann-Whitney, robust T, permutation) applied to samples of various sizes (10 to 500) drawn from 4 distributions (normal, uniform, log-normal, bimodal) under the null hypothesis and under the alternate (difference between means of 0.25 standard deviation), with equal unit variance in all distributions.
Results: Type 1 errors were well controlled in all situations. The T test was most powerful for data drawn from the normal and the uniform distributions, but only by a narrow margin. The Mann-Whitney test was the most powerful option for data drawn from asymmetric distributions; compared to the T test the gain of power was often large, especially for the highly skewed log-normal distribution. Of note, even the T test was more powerful under asymmetric distributions than under the normal distribution. In presence of outliers (bimodal distribution), the robust T test was most powerful.
Conclusions: All tests performed well under the four distributions, at all sample sizes: type 1 errors were on target, and assumptions violations did not reduce power. This justifies opting for the test that fits the scientific hypothesis the best, regardless of normality or sample size. To select the most powerful test, the symmetry of the distribution is the key criterion: for asymmetric distributions the Mann-Whitney test is the most powerful, for symmetric distributions it is the T test.

P4.5.133
Properties of ANOVA and its alternatives under violation of their assumptions: a simulation study

Z Reiczigel[1], M Ladanyi[1]
[1]*Corvinus University of Budapest, Budapest, Hungary*

One-way ANOVA is applied in almost all research areas but many users are uncertain what to do if its applicability conditions, normality and homoscedasticity seem to be violated.
There are alternative methods for unequal group variances (Welch, Brown-Forsythe, WLS), and nonparametric methods like the Kruskal-Wallis test which are applicable under non-normality. There is a widespread misbelief that the Kruskal-Wallis test works fine under heteroscedasticity, or that ANOVA is robust against violations of the assumptions if the group sizes are equal.
Our aim is to carry out a systematic simulation study to explore the properties of some available methods and make recommendations for the non-statistician users based on the results.
We compared one-way ANOVA, Kruskal-Wallis test, Welch's method, Brown-Forsythe method, bootstrap ANOVA, as well as the method of pre-testing heteroscedasticity by Levene's test and using ANOVA or Welch depending on its result.
We used normal as well as non-normal distributions (uniform, chi-squared, exponential, symmetric bimodal and skewed bimodal), combined with heteroscedasticity (2-fold or a 3-fold SD). Nominal alpha was set to 5%, and we simulated the actual alpha.
Under non-normality combined with heteroscedascity, classical ANOVA performs rather poor, even in case of large samples and balanced design. Heteroscedasticity totally invalidates the Kruskal-Wallis test. Alternative methods, such as Welch, Brown-Forsythe, and bootstrap ANOVA perform better. Pre-testing the equality of variances by Levene's test and choosing between the Welch's test or classical ANOVA based on its result was found almost always worse than using the Welch's test without any pre-testing.

P4.5.139
Estimation of the ROC curve with a time-dependent disease variable in the presence of covariates

MX Rodríguez-Álvarez[1], L Meira-Machado[2]
[1]*University of Vigo, Vigo, Spain,* [2]*University of Minho, Guimarães, Portugal*

The receiver operating characteristic (ROC) curve is the most widely used measure for evaluating the performance of a diagnostic biomarker when predicting a binary outcome. The ROC curve displays the sensitivity and specificity for different cut-offs values used to classify an individual as healthy or diseased.
In many studies, however, the target of a biomarker may involve prognosis instead of diagnosis. In such cases, when evaluating the performance of the biomarker, several issues should be taken into account: first, the time-dependent nature of the outcome (i.e., the disease status of an individual varies with time); and second, the presence of incomplete data (e.g., censored data typically present in survival studies). Accordingly, to assess the discrimination power of continuous prognostic biomarkers for time-dependent disease outcomes, time-dependent extensions of sensitivity, specificity and ROC curve have been recently proposed.
In this work we present a new nonparametric estimator for the cumulative-dynamic time-dependent ROC curve that allows accounting for the possible modifying effect of current or past covariate measures on the discriminatory power of the biomarker. The proposed estimators can accommodate right-censored data, as well as covariate-dependent censoring. The behavior of the estimator proposed in this study will be explored through simulations and illustrated using real data.

P4.5.141
"Numberstories": Teaching statistics to kids - and clinicians

J Røislien[1,2], KF Frøslie[1,3]
[1]*Department of Biostatistics, University of Oslo, Oslo, Norway,* [2]*Department of Health Studies, University of Stavanger, Stavanger, Norway,* [3]*Norwegian Resource Centre Women's Health, Oslo Uni. Hospital, Oslo, Norway*

The use of statistics has revolutionized medicine, and statistical analysis is at the core of evidence based medicine and the communication of medical research to the general public. However, most non-statisticians struggle with the mathematical language of statistics. Indeed, people tend to use statistics like they use their cars: They're happy if the user interface is fine, if it works like it is supposed to, and takes them where they want to go. People generally care just as little about what goes on under the hood of their cars, as do clinicians about formulas and algorithms. Still, they get their driver's license. Still, they drive cars.
"Tall forteller" (Eng: *"Numberstories"*) takes a similar approach to teaching statistics. Awash with pop-art photos and anecdotal introductions, focus is on statistics as a vehicle for analyzing quantitative information, rather than mathematical formulae. Through nineteen simple research projects, which the reader is encouraged to perform herself, she is - unknowingly - guided through a basic course in descriptive statistics and bivariate analysis. The text is simple and can be understood by anyone having their first introduction to statistics; from teenagers to medical doctors. Common statistical terms are explained in plain language.
Formulas are replaced by a free app performing all calculations necessary to answer the nineteen research questions, presenting graphs and statistical tests at the push of a button, given that the reader knows the type and number of variables. *"Tall forteller"* teaches how to drive the statistics car.

P4.5.158
A model of population dynamics based on fuzzy cellular automata theory

A Szymański[1], A Rossa[1]
[1]*Institute of Stat. and Demography, University of Lodz, Lodz, Poland*

In the paper a hybrid model based on cellular automata and fuzzy logic is proposed to simulate the dynamics of a population, where the changes in fertility and mortality rates are induced by a stochastic variation of the environment.
The concept of the Cellular Automata (CA) is based on the original von Neumann's idea (1966), modified by Betel et al. (2011) in a form of the so-called fuzzy cellular automata (FCA). CA is a formal model composed of a rectangular region of MxN cells in which the evolution of each cell depends on its present state and a state of its neighboring cells. All the cells pass through the following generation at the same time, according to the transition function being common for all the cells.
In the proposed model the environment variables are treated as fuzzy sets. Their combination is treated as a result of factors determining population growth. The operations on fuzzy sets lead to fertility and mortality rates. The obtained results have been verified by means of the mortality data for Poland.
This research was supported for both authors by a grant from the National Science Center under contract DEC-2011/01/B/HS4/02882, for which the authors are indebted.

P4.5.159
Studying latent hydration level through multivariate longitudinal biological indicators: application of a latent process mixed model

J Tanguy[1], C Proust-Lima[2,3]
[1]*Danone R.D., Palaiseau Cedex, France,* [2]*INSERM U897, Epidemiology and Biostatistics Research Center, Bordeaux, France,* [3]*University of Bordeaux, ISPED, Bordeaux, France*

Clinicians are often faced with the problem of not being able to measure directly what they want to study. In this case, multivariate longitudinal variables measuring the latent construct of interest are collected, and the aim is to establish causal-effect relationship between explanatory covariates and this latent concept. In addition, understanding how outcomes are related to each other and with respect to their underlying concept can also be relevant.
To address these challenges, an appealing solution is to use latent process mixed models assuming that each longitudinal indicator (whatever its nature) is a noisy measure of the same underlying function.
In short, measurement models link the observed outcomes (quantitative, ordinal,...) to the latent concept, and a structural model describes the latent concept according to explanatory covariates. This model borrows strength across outcomes while taking into account the correlation within each outcome over time and between outcomes.
This work aims at illustrating the possibilities offered by this approach through a study case on the issue of measuring hydration in real-life. Hydration can be measured by multiple biological measures from urines or plasma and by total fluid and water intake. However, it is not clear which markers are good proxies of the absolute hydration level. Yet, identifying proxies of hydration becomes essential to quantify hydration in populations and evaluate optimal levels of hydration.
Goodness-of-fit, interpretation, limits and advantages of the method, but also perspectives with respect to regulatory requirements and specific challenges in clinical studies in nutrition are discussed.

P4.5.165
Item analysis and evaluation of the examinations in faculty of medicine

L Tomak[1], Y Bek[1]
[1]*Ondokuz Mayis, Samsun, Turkey*

Item analysis is an effective method in the evaluation of examinations. This study aims the comparative evaluation of the classical and latent methods used in Item Analysis, and the efficacy of these methods in the evaluation of the examinations of Faculty of Medicine.
The student exams in the Faculty of Medicine were evaluated using different methods. The methods to be used can be divided in two main classes; classical methods and latent methods. Among the classical methods, Cronbach´s alpha, split half methods, item discrimination and item difficulty were investigated. On the other hand, various models of item response theory and their statistics were compared in the group of latent methods.
Different statistics and measurements obtained by the application of classical and latent methods were evaluated, and we tried to determine "which method is more reliable under which evaluation criteria". The most ideal evaluation criteria, evaluation method and the software were determined by evaluating the implementation results of the computer softwares developed for this purpose.

P4.5.176

An empirical comparison of methods for a conjoint analysis survey of knowledge translation in women's addiction agencies

T Vanniyasingam[1], C Cunningham[1], G Foster[1], A Niccols[1], L Thabane[1]
[1]McMaster University, Hamilton, Canada

Research has identified effective models of addiction services for the prevention and treatment of substance abuse among women, including programs that incorporate pregnancy-, and child-related services. Despite these effective evidence-based methods, there is a lack of implementation of such findings in the delivery of addiction services.
We conducted a discrete choice experiment with 1379 service providers and administrators from 333 addiction agencies. The participants were presented with a set of scenarios comprised of combinations of different attributes and corresponding levels, using a fractional factorial design. Our objective is to determine professional development preferences by addiction service providers and administrators for the enhancement of addiction services.
We will empirically analyze discrete choice experiment data using various statistical methods to account for within-participant correlation—specifically, multinomial logit and probit models. We will report the relative importance (or ranking) of each attribute and level of attribute in determining participants' preferences.

P4.5.179

Outlier detection in functional time series: an application to mortality rates

J Vilar[1], G Aneiros[1], P Raña[1]
[1]University of A Coruña, A Coruña, Spain

A new procedure to detect outliers in functional data, which takes into account the dependence, is proposed. An application to mortality rates is presented.
There are some contexts in which the data should be treated as functions, rather than a string of values. This kind of data, known as functional data, appears in many areas, such as bio-medical. In this work we study a functional time series, which is a time series made of functional data. It is important then to consider the dependence found in this kind of data.
Our approach to detect outliers follows the idea of Febrero et al. (2008) which assumes the independence between the data, modifying some steps and adapting it to consider the dependence of the functional data. Briefly, the procedure obtains the functional depth of the data and looks for a cutoff using bootstrap techniques. The data with depth lower than the cutoff are classified as outliers.
A comparison between our proposal and other different methods to detect outliers in functional data, using simulated data, has been carried out showing the good behavior of our approach.
In this paper, the method is applied to mortality rates. Specifically, we use data of the French male age-specific log mortality rates, between 1899 and 2005. The effect of historical events, as the World War I or the Spanish flu pandemic, is reflected in the mortality rates and some of these years will be classified as outliers.

P4.5.184

Split and merge techniques for Gaussian mixture learning

Y Wang[1], M Titterington[2]
[1]King's College London, London, United Kingdom, [2]University of Glasgow, Glasgow, United Kingdom

EM algorithm has been the first choice for mixture density estimation. However, the EM, essentially a local search algorithm, has a number of limitations such as slow to converge, sensitive to initialization, and may get stuck in one of many local maxima of the likelihood function. Inspired by the idea of splitting and/or merging components sequentially or simultaneously based on certain criteria, a number of algorithms were introduced to overcome these limitations.
We tested 3 algorithms of this kind, namely, the IPRA (Iterative Pairwise Replacement algorithm) (merge) by Scott and Szewczyk (2001), the Greedy EM (split) by Vlassis and Likas (2002), and the SMEM (split and merge) by Ueda et al (2000), on both simulated and real-world genomics data. We commented on their performances with comparison to the standard EM+BIC approach (MCLUST package in R).
Further, we extended the IPRA into multidimensional problems by proposing a multivariate IPRA (mIPRA), which uses the minimal spanning tree (MST) to limit searches, thereby reduces the number of comparisons from $O(n^2)$ to O(nlogn). We found the mIPRA was efficient and effective when fitting mixtures with a large number of components.

P4.5.187

Preanalytical variation: computing variance estimates from systematic differences and models for clinical practice

MS Sylte[1], T Wentzel-Larsen[2,3,4], BJ Bolann[1,5]
[1]Haukeland University Hospital/Lab. of Clinical Biochemistry, Bergen, Norway, [2]Centre for Child and Adolescent Mental Health, Oslo, Norway, [3]Norwegian Centre for Violence and Traumatic Stress Studies, Oslo, Norway, [4]Haukeland University Hospital/ Centre for Clinical Research, Bergen, Norway, [5]University of Bergen, Department of Clinical Science, Bergen, Norway

Aims: A framework for estimating preanalytical uncertainty using linear mixed-effects models is previously developed, with fixed factors for systematic differences between preanalytical settings and random effects for structural levels. Settings for the systematic factors vary within clinical practice, and our aim is to combine explicit modelling of this variation for each systematic difference to be added to the appropriate random effect.
Methods: For a discrete fixed effects covariate, the additional variance is based on fixed effects coefficients and corresponding probabilities from clinical practice. For a continuous covariate, it is based on the fixed effect and an assumed rectangular distribution from clinical practice.
Results: The systematic and random effects created from different handling of blood samples before analysis e.g. using different needles, mixing of the blood samples and different transportation is combined with frequencies and calculated into a total preanalytical uncertainty. By this approach the interpretation of patient results can be improved.
Conclusions: We have developed a unified procedure for combining fixed effects with estimates from clinical practice to produce variance estimates to combine with random effects. The procedure is quite general and will be investigated further in our laboratory.

Author Index

A

Aalen O. C37.5
Abdel-Rahman S.M. P1.3.51
Abellan-Andres J. C02.2
Abrahamowicz M. I4.3, C28.1, S1.3, C33.1, P3.1.189
Abrahamsson L. C11.5
Abrams K.R. C48.3, P1.1.118, P3.1.27
Abroms L. P4.1.80
Achcar J. P1.1.36
Aghlmandi S. P1.1.1
Agrawal D. P3.5.82
Aguirre U. P2.1.54, P2.3.2, P3.5.3, P3.5.55
Agulnik J.S. P3.1.189
Ahlin C. C06.2
Ahmed I. P4.1.4
Ahn S. P4.4.94
Akacha M. P2.3.188
Akram M. C23.4
Albert A. C11.4
Alexander N. P2.2.115
Alkerwi A. C11.4
Almendra-Arao F. P1.2.5
Alonso H. P4.5.6
Alpar C.R. C16.1
Altman D. P1.2.35
Altman D.G. C19.1, C19.3
Alvarez-Iglesias A. C06.1, P1.2.7
Ambler G. C19.5, C36.5, P3.5.8
Ameling C. P4.5.9
Andersen P.K. C18.3, C28.4
Anderson P.J. C05.4
Andersson T.M.-L. C15.2
Andrews N. C12.5
Andrinopoulou E.-R. C15.4
Aneiros G. P4.5.179
Ankarali H. P1.2.122
Anothaisintawee T. P4.1.108
Antolini L. C27.1
Antón A. P3.5.55
Arai T. P2.2.105
Arizaga A. P2.1.54
Arostegui I. P2.3.2, P3.5.10, P3.5.55
Asakura K. C24.5
Asano J. P3.5.11
Asano S. P4.5.12
Assi N. P4.3.13
Atassi N. C20.2
Atherton J. C28.2
Attia J. P4.1.108
Austin P. C43.2
Awan S. P4.4.14
Ayis S. P2.1.77, P4.4.15
Ayles H.M. C02.5

B

Baart M. C32.2
Baayen C. C34.5
Bacon S.L. C28.2
Baczkowski S. P3.4.143
Bagher S. P2.2.16
Bagos P.G. C10.5
Bailey-Wilson J.E. M2.2
Bailleux F.P. P1.1.17
Balder E. P1.2.18
Banerjee M. C08.4
Barbieri A. C38.5
Bari F. P2.1.92
Barrio I. P3.5.10
Barrio-Beraza I. P3.5.55
Barrueco J. P1.2.68
Bartlett J. P3.1.150, P3.1.151
Barzenje D.A. P3.1.148
Bassily E. P1.1.17
Bastard M. C09.3, P3.1.19
Bastide S. P1.1.20, P1.1.21
Bastien P. C31.3
Batistatou E. C07.5
Bauer P. P1.2.67
Bayar M.-A. C01.5
Beals J. C09.4
Becher H. C03.2
Beghi E. P3.5.25
Begun A. P3.1.22
Bek Y. P4.5.165
Belamri N. P1.3.123
Bellomo R. C23.4
Belot A. C33.3, C33.5
Benadjaoud M.A. C11.3, P2.4.23
Bender R. C18.2
Bénech H. P1.2.109
Benedetti A. C14.4
Benoit A. C12.1
Berger M.P.F. C35.4
Berghold A. C43.1, P1.2.112, P4.2.66
Bermudez P.Z. P3.2.154
Bernasconi D.P. C27.1
Berry J. C20.2
Bertrand F. C31.3
Besser R. P3.5.145
Beyer U. C17.2
Bigler M. P1.2.64
Billingham L. C25.4
Binder H. C08.5, C16.4, C21.1, C26.2, C27.3, C27.5, C42.1, C42.3
Binder N. C39.5
Binquet C. C33.1
Bitard J. C43.3
Blagus R. C16.3, C34.2
Blaizot S. P4.2.24
Blanche P. C19.2
Ble A. P3.3.152
Blennow K. P4.4.104
Blettner M. S1.1
Blizzard L. C34.3
Blows F.M. C31.4
Boda K. P2.2.182
Bodicoat D. S1.1
Bogaerts J. C37.4
Böhning D. C35.6
Bolann B.J. P4.5.187
Bond S. C17.3
Bond S.J. C05.5
Bonnett L.J. P3.4.25a, P3.5.25
Borgan Ø. C37.5
Börjesson-Hanson A. P4.4.104
Bornkamp B. C36.2, C36.4, P1.2.89
Borsos M. P2.6.26
Borthwick N.J. C05.2
Bossard N. C33.5
Boucher R.H. P3.1.27
Bourke L. C29.5
Bouwmeester W. C19.4
Bowden J. C35.5
Brady M. P4.1.186
Brakenhoff T. P1.1.110
Brannath W. C03.1, C03.3, C03.4, C36.1
Brás A.L. P4.2.58
Breinegaard N. C32.3
Bretz F. C13.2, C36.2, C36.3, P1.2.89
Brill S. P1.2.93
Broberg P. P1.2.28
Brock K. C25.4
Brzyska M. P2.1.119
Budinska E. P2.5.78
Bujkiewicz S. C48.3, P1.1.118
Bulgakova V.A. P2.2.29
Burger S. C03.4
Burne R. I4.3
Burzykowski T. C44.4, P2.5.194
Butcher I. P4.1.186

C

Cabilio P. C35.2
Cadarso-Suarez C. C21.3
Cai T. I3.3
Caldas C. C31.4
Callegaro A. C41.4
Callejo A. P2.1.54
Calza S. C42.4
Cameron D. P3.1.150, P3.1.151
Campbell M.J. I1.1
Campbell P.T. P4.5.6
Canary J.D. C34.3
Candel M.J.J.M. C35.4
Cangur S. P1.2.122
Cardoso H. P3.4.180, P3.4.181
Cardot H. C11.3, P2.4.23
Carlin J.B. C05.4, C23.2, C46.3, P2.3.140
Carpenter J. C20.3
Castell F. C33.3
Castillo-Tzec Y. P1.2.5
Cederkvist L. C21.5
Cetinyurek Yavuz A. C47.4
Chadha-Boreham H. P1.2.167
Chancharas P. P3.1.170
Chang Y.-C.I. P2.6.30
Charlett A. C12.5
Chen O. C34.6
Chen S. C09.4
Chen Y.-H. P2.3.169
Cheong J.L. C05.4
Cheung S.H. C18.1
Cheung Y.B. C05.3
Chevret S. C02.4, C21.2, C40.2, P1.1.32
Chiang C. P1.2.33
Choi I.-H. P2.2.79
Choi J.E. P3.2.95
Choi S. P3.4.34, P3.4.83, P4.1.87, P4.5.84
Choodari-Oskooei B. C40.1, P3.5.8
Choukroun G. P1.3.123
Chrzan R. P3.4.143
Chudek J. P3.2.117
Cipriani A. C14.1
Ciria C. P1.2.35
Clack G. C04.5
Claesen J. P2.5.194
Clark A.E. C08.1
Coart E. C44.4
Coelho-Barros E.A. P1.1.36
Collette L. C04.1, P1.2.47
Collignon O. C38.2, P3.5.37
Collins G.S. C19.1, C19.3
Colzani E. C37.2
Combescure C. P4.1.38, P4.5.128
Copas A.J. C23.1
Corkum A. C35.2
Cornelius V. P1.2.97, P2.2.155
Corrao G. P1.2.142
Cortese G. C21.4
Cotterill A. C04.4
Courvoisier D.S. C32.1, P4.5.128
Crack L. C25.4
Crichton S. C46.5
Critchley H.O.D. C48.1
Cronin A. C29.2
Crowther M.J. C15.2
Csicsman J. P2.1.92
Csordas K. P3.5.62
Cudkowicz M. C20.2
Cunningham C. C38.3, P4.5.176
Curt A. C20.5
Czene K. C37.2

D

Da Silva A.M. P3.4.180
Dangoisse C. P1.2.97
Daniel R. P3.3.96
Danieli C. C33.5
Daniel-Spiegel E. C34.6
Dantan E. C33.4
Dartigues J.-F. P2.1.129
Davies G. P3.4.25a
De Hoop E. C23.5
De Lobel L. P2.5.40
De Raedt K. P3.5.173
De Ridder M.A.J. P2.2.41
De Silvestri A. P2.2.85
De Sousa B. P4.2.58
De Vathaire F. C11.3, P2.4.23
De Wreede L.C. C39.4
Dean J.M. C08.1
Debray T. C20.1
Debray T.P.A. C20.4, C44.2, C44.3, P4.1.4, P4.1.42
Dejardin D. C07.3
Delaforge M. P1.2.109
Delcoigne B. C37.2
Delva F. P2.1.129
Den Hurk K.V. C32.2
Deshpande N. C30.5
Devaux Y. P3.5.37
Dewé W. C12.1
Diack C. C17.2
Diao G. C10.2, C27.4
DiazOrdaz K. C02.1
Dichgans M. P2.3.196
Didelez V. C28.5
Dilaveroglu E. P4.5.90
Dimou N.L. C10.5
Dodd C.N. P2.2.41
Doimo S. P2.2.85
Dong G. C22.4
Dong T. C05.2
Dong Y. P1.2.178
Douiri A. C08.3
Doyle L.W. C05.4
Dunkler D. P1.1.43
Dunning A.J. P1.1.17
Durkalski-Mauldin V.L. C29.3
Dwivedi S. P2.2.163

E

Eberg M. C37.3
Ecochard R. P4.2.24
Edjolo A. P2.1.129
Edwards T. P2.2.115
Eekhout I. P2.3.44
Efendi A. C37.4
Efthimiou O. C26.1
Eijkemans M.J.C. P4.5.81, C16.2
Ein-Mor E. C34.6
Eipel O.T. P3.5.62
El Ghouch A. C47.3
Ellard S. P3.5.145
Emsley R.A. I1.3
Enki D. C12.5
Ensor J. C44.3, P3.5.45
Erdélyi D.J. P3.5.62
Erler N.S. C46.2
Errath M. P1.2.112
Etain B. C10.2
Etard J.-F. C09.3, P3.1.19, P4.2.24
Eulenburg C. P4.4.46
Evans S.R. C24.5
Evoy M.M. P4.1.108
Ezzalfani M. P1.2.47

F

Færch K. C11.1
Fagerland M.W. C40.3
Farrington P. C03.5, C12.5
Fayaz M. P1.2.48
Fenske N. C16.5
Fermin Y. C42.5
Ferrari P. P4.3.13
Ferrario P.G. C10.1
Feuillet F. C35.1, P3.3.49
Filion K.B. C37.3

Filipe P.A. P4.2.58
Filipits M. M2.1
Fisch R. C24.4
Fitzmaurice D. P3.5.45
Fleischer F. C22.1, P1.2.86
Florek M. P2.1.119
Floyd S. C02.5
Forbes A.B. C23.4
Forman J.R. C05.5
Foster G. C38.3, P4.5.176
Foster J. P3.3.152
Foucher Y. C33.4, C37.1, P4.1.38
Fraczek P. P3.4.143
Franco O.H. C46.2
Free C. P4.1.80
Fridley B.L. P1.3.51
Friede T. P4.5.61
Friger M. C15.3
Fröhlich H. Course 4
Frommlet F. C31.6, C36.6
Frøslie K.F. C11.1, P4.5.141
Funai T. C08.1
Funatogawa I. P2.2.52, P2.2.53
Furukawa K. C46.1
Furuya K. P4.5.12

G

Gagnon B. P3.1.189
Galimard J.-E. C02.4
Ganna A. C10.4
García S. P2.1.54
García Barrado L. C44.4
Garcia-Gutierrez S. P3.5.55
Garthwaite P. C12.5
Gaschler-Markefski B. P1.2.68
Gaudart J. C33.6
Gauderman J. P4.5.6
Gerds T. C21.1
Gerds T.A. I3.2, C18.3, C19.2
Gerhold-Ay A. C42.3
Gerke O. C13.2
Germing U. C45.3
Geroldinger A. P2.4.56
Geskus R. Course 2, I3.1, P4.2.102
Geskus R.B. C46.4, P3.1.69
Geva D. C15.3
Ghazi Y. P1.1.124
Ghebremichael-Weldeselassie Y. C03.5
Gichuru P. C38.4
Gieger C. C16.5
Gillaizeau F. C33.4
Giorgi R. C33.1, C33.2, C33.3, C33.6
Gioulbasanis I. P3.1.189
Giral M. C33.4, C37.1
Girling A. P1.2.100
Glass A. P1.2.57
Glimm E. C36.2
Glomb P. P1.2.68
Godfrey K.M. P4.4.192
Godfrey-Faussett P. C02.5
Goetghebeur E. S1.4
Gomes D. P4.2.58
Gomes M. C02.1
Gómez Melis G. P1.2.59
Gómez-Mateu M. P1.2.59
Gomulska M. P3.4.143
Gonçalves L. P3.2.154
Gönen M. P3.5.174, P3.5.175
Gong Y. P2.3.188
Götte H. C48.2
Grace J.J. C08.3
Graf A. P1.2.89
Grafféo N. C33.2, C33.3, C33.6
Grallert H. C16.5
Grand M.K. C45.1
Grayling M.J. C24.2
Gregorich M. P1.1.43
Grieve R. C02.1
Grobler A.C. C02.3
Groenwold R.H.H. I4.1
Grøn R. C18.3
Grover G. P3.1.157
Gruber S.B. P4.5.6
Gu C. P1.2.121
Gueorguieva R. C06.5
Guillaume M. C11.4
Gunning P. P1.2.7
Gustafson P. I4.2
Gutjahr G. C36.4
Güttner A. P2.3.188

H

Haas K. P1.2.166
Haisa T. P4.5.12
Hakama M. P2.2.127
Halkes C.J.M. C39.4
Hamada C. C25.5, P3.5.11
Hamasaki T. C24.5
Hampson L. Course 5, C24.4
Hampson L.V. C29.4, P1.2.146
Hanke T. C05.2
Hansen C.H. C48.1
Hantel S. P2.1.60
Harden M. P4.5.61
Hardouin J.-B. C35.1, P3.3.49
Harnos A. P3.5.62
Harrington J. C29.2
Harris T.B. C15.3
Harrison T. P4.5.6
Hasford J. P3.1.125
Hattersley A. P3.5.145
Haymart M. C08.4
Hayoz S. P1.2.63, P1.2.64
Hazell L. P1.2.142
Hedeker D. C32.1
Hegyi M. P3.5.62
Hehlmann R. P3.1.125
Heinzl H. C27.2, P2.2.65
Hejblum B. C12.3
Held L. C31.5, P2.2.185
Hemming K. P1.2.100
Henderson R. I5.3
Henley W. C12.4, P3.5.145
Henley W.E. P3.3.152
Hernandez-Fuentes M. C29.2
Herzog S.A. P4.2.66
Heßler N. C25.3
Heymans M.W. P2.3.44
Hida E. C48.5
Higgins J. C14.2
Higgins J.P. C14.1
Higgins J.P.T. P4.1.134
Hilton R. C29.2
Hinde J. C06.1
Hirakawa A. P3.5.11
Hjort N.L. I5.1
Hlavin G. P1.2.67
Hocke J. P1.2.68
Hof M.H. P3.1.69
Hofer S.M. M1.1
Hofer S.O.P. P2.2.16
Hoffmann I. C16.4
Hoffmann V.S. C45.3
Hofner B. C31.2
Holst K.K. C21.5
Holte H. P3.1.148
Holubkov R. C08.1
Hooper R. C29.5, P4.3.70
Horneck M. P1.2.166
Hornig F. P2.3.188
Horová I. C41.1
Hosmer D.W. C34.3, C40.3
Hosseini S.M. P4.5.71, P4.5.72
Hothorn T. C20.5
Hougaard P. C34.5
Hoya K. P4.5.12
Hrach K. P2.1.74, P2.6.73
Hsiao C.-F. P1.2.33
Hsu L. P4.5.6
Huang Y. P4.1.80
Hubeaux S. C47.2
Huebner M. S1.1, P3.1.75

Hughes R. C32.5
Huh I.-S. C10.3
Humphreys K. C11.5, C15.2
Hussain S. P4.5.76

I

Iacobelli S. C27.1
Icks A. P3.1.22
Ickstadt K. C42.5
Ieva F. C11.2, C45.4
Ifeagwu S.C.E. P2.1.77
Ihnatova I. P2.5.78
Imtiaz S. C42.5
Ingel K. C27.5
Ingsathit A. P3.1.162
Ishikawa S. P2.2.105
Isnard-Bagnis C. P1.3.123

J

Jackson C. C18.4, C35.3
Jackson D. P1.2.93, P4.1.38, P4.1.134
Jackson J. P1.3.51
Jacobi H. P2.1.161
Jacobs T. P1.1.107
Jacqmin-Gadda H. C09.5, C15.5
Jaddoe V.W.V. C46.2
Jahn-Eimermacher A. C27.5
Jaki T. Course 5, C04.4, C17.2, C24.4, C29.4, C30.3, P1.2.88
Jalali A. C06.1
James G. C04.5
James I. C05.1
Jang B.-H. P2.2.79
Jang E.J. P3.4.34, P4.5.84
Janssens M. C02.2
Jeffrey R. P4.4.192
Jemiai Y. Course 7, I2.2
Jenkner C. C03.2
Jennison C. C30.2
Jensen A.K.G. C28.4
Jensen K. C26.4, C48.4
Jiang L. C09.4
Jiang Y. P4.1.80
John M. P3.2.103
Johnson A. P3.5.25
Jolani S. C20.4
Joly F. P1.1.124
Jones A. P3.5.145
Jones B. P1.2.146
Jong V.L. P4.5.81, C16.2
Joshi H. C42.4
Julian J.A. P1.2.121

K

Kahan B.C. P1.2.106
Kaiser R. P1.2.68
Kallberg H. C13.5
Kallmeyer R. P4.5.90
Kamal V.K. P3.5.82
Karabulut E. C16.1
Kasymjanova G. P3.1.189
Kasza J. C28.3
Katina S. C41.1
Katsahian S. C10.2, C27.4, P1.2.178
Katsovskiy I. C20.2
Kawahara N. P4.5.12
Kechel M. C13.3
Keevil S. P2.2.155
Keiding N. M1.2
Kejžar N. C40.5
Kenward M.G. C02.1
Keogh R.H. C12.2, P1.2.106
Kern J. P4.4.104
Kern S. P4.4.104
Kieser M. C26.4, C30.4, C48.2, C48.4
Kifley A. C09.1
Kim B. P4.4.94
Kim D.K. P3.4.34, P4.5.84
Kim H.J. P4.4.94
Kim J.-Y. P3.2.95
Kim J.M. P3.4.34
Kim J. P3.4.83, P4.1.87, P4.5.84
Kim L. P3.5.25
Kim Y. P3.2.95, P3.4.34, P3.4.83, P4.5.84
Kirchner M. C48.2
Klebermass-Schrehof K. P2.2.183
Klersy C. P2.2.85
Klingbiel D. P1.2.63, P1.2.64
Klinglmueller F. C30.1, C36.3
Klockgether T. P2.1.161
Knahl S. P1.2.86
Ko M.-J. P3.4.83
Ko M.J. P4.1.87
Ko S.-G. P2.2.79
Koeller M.D. M1.4
Koffijberg H. C20.4, C23.5, C44.2, P4.1.42
Kolstad A. P3.1.148
König F. C30.1, C30.3, S2.1, C36.2, C36.3, P1.2.67, P1.2.88, P1.2.89
König J. C26.2
Kontopantelis E. C14.3
Kort W.D. C32.2
Kovács G.T. P3.5.62
Krahn U. C26.2
Krasnozhon S. P1.2.89
Kreuzhuber R. P4.5.90
Kriegner A. P4.5.90
Kueffner R. C20.2
Kundt G. P1.2.57
Kurum E. C09.2
Kutasow A. P4.4.46
Kwon M.-S. C10.3
Kyle R.P. C28.1, S1.3

L

Ladanyi M. P4.5.133
Lambert J. C40.2
Lambert P. C47.4
Lambert P.C. C15.2, P1.1.118, P3.1.27
Lancaster G. C38.4
Landais P. P1.1.20, P1.1.21
Lang Z. P3.2.91
Langan D. C14.2
Lange M.R. C04.3
Lange T. C22.2, C37.5
Langhorne P. P4.1.186
Larson D.W. P3.1.75
Laszlo A.M. P2.1.92
Lauseker M. C45.3, P3.1.125
Lautner-Csorba O. P3.5.62
Lavergne C. C38.5
Law L.M. C17.5
Law M. P1.2.93
Lawn N. P3.5.25
Le Borgne F. C37.1
Le Cessie S. S1.1, S1.4
Le Deley M.-C. C01.5
Le Teuff G. C01.5, P4.1.135
Leacy F.P. C02.5
Lebrun P. C12.1
Lee C.-H. P3.4.34
Lee C.H. P4.5.84
Lee D. C10.4, C42.4
Lee D.Y. P3.2.95
Lee J.-Y. P3.2.95
Lee J. P4.4.94
Lee K.H. P4.4.94
Lee K.J. C05.4, C46.3, P2.3.140
Lee W. C10.4, C13.5
Lee Y.-K. P3.2.95
Lee Y.E. P3.2.95
Lee Y. P3.4.83, P4.1.87

Leeder J.S. P1.3.51
Leermakers E.T.M. C46.2
Lefebvre G. C28.2
Legarreta M.J. P3.5.10
Legrand C. C12.1, C47.3
Leitner M. C20.2
Lenuzza N. P1.2.109
Leone M. P3.5.25
Leroy S. P1.1.20, P1.1.21
Lesaffre E. C07.3, C15.4, C32.2
Lesaffre E.M.E.H. C46.2
Leton E. C40.4
Leuchs A.-K. C07.2
Leucht S. C26.1
Levine M.N. P1.2.121
Levy Y. C12.3
Lewis C. P2.2.155
Lewis S.C. P4.1.186
Li Q. C41.2
Li R. C09.2, P3.3.96
Lin N. C12.4
Lingsma H.F. C01.4
Link J. C22.1
Liquet B. C43.3
Lissner L. P4.4.104
Liu H.Y. C24.1
Liu J.-P. P1.2.33
Liu K. P1.2.97
Lo J. C34.4
Logan R. P2.2.155
Look M.P. C44.3
Lorenz E. C03.2
Louis T.A. IP.1
Love S. P1.2.35
Lu T.-Y. C18.1
Lusa L. C06.2, C16.3

M

Maass S.W.M.C. P2.2.16
Macaskill P. S1.2
Magirr D. C30.3
Mahboubi A. C33.1
Majdič N. C34.2
Mäki-Petäjä K.M. C05.5
Male C. P1.2.67
Malecki M. P3.5.145
Maleev V.V. P2.2.29
Malila N. P2.2.127
Malina M. C31.6
Maman D. P4.2.24
Mandel M. C34.6, C39.3
Mander A. C17.3, C24.2
Mandrekar J. P4.5.98
Mangtani P. C12.2
Mani K. C41.5
Manju A. C35.4
Manson S.M. C09.4
Maracy M.R. P4.4.99
Maras P. P2.2.85
Marous M. P1.2.47
Marra M. P4.5.9
Marshall J. C04.5
Marshall R.J. C06.4
Marson A.G. P3.5.25
Martin J.T. P1.2.100
Martín J. P2.1.54
Martínez-Camblor P. P3.2.101
Masclee G.M.C. P2.2.41
Matser A. P4.2.102
Matsuno A. P4.5.12
Matthews G. C02.3
Maucort Boulch D. C40.5
Maumy-Bertrand M. C31.3
Mavridis D. C14.1, C26.5
Mayer M. C41.2
Mazucheli J. P1.1.36
Mazur J. C42.1, C42.3
McCandless L.C. I4.2
Mcdonald T. P3.5.145
McKevitt C. C08.3
Mckevitt C. P4.4.15
McKinnon E.J. P3.2.103
Mehlig K. P4.4.104
Meira-Machado L. P4.5.139
Meiser K. C07.4
Melzer D. P3.3.152
Mendonça D. C21.3, P3.1.160, P3.4.180, P3.4.181
Mendoza F. P2.1.54
Meneses R.F. P3.4.180
Menguy V. P1.1.124
Mentré F. C32.4, P1.2.109
Mesenbrink P. P2.3.188
Meyer N. C31.3
Michiels S. C01.5, C31.1, P4.1.135
Middleton G. C25.4
Mieno M.N. P2.2.105
Miller F. C30.1
Mir A. P1.2.168
Mitchell C.M. C09.4
Mittlböck M. C27.2
Mizukami H. P3.1.111
Moatti M. P1.1.32
Molanes-López E.M. C40.4
Molenberghs G. C02.3, C15.3
Molho S. P2.1.161
Molina M. P1.1.130
Mollevi C. C38.5
Moodie E. S1.4
Moodie E.E.M. I1.2, C28.1
Moons K.G. C19.1, C44.2, C44.3
Moons K.G.M. C20.1, C20.4, P4.1.4, P4.1.42
Moore D. P3.5.45
Moranne O. P1.3.123
Morbach S. P3.1.22
Moreau J.-F. C43.3
Moreno V. P4.5.6
Morgan K.E. P1.2.106
Morgan M. C29.2
Morgane R. C35.1
Morikawa T. P3.1.111
Morita A. P4.5.12
Mota M. P1.1.130
Mouriño H. P3.1.149
Muchene L.K. P1.1.107
Muenz D. C08.4
Mukherjee R. Course 7
Müller M. P2.6.26
Müller-Myhsok B. M2.3
Muntham D. P4.1.108
Murakami M. P4.5.12
Murray G. P3.1.150, P3.1.151
Murray G.D. P4.1.186
Musoro J.Z. C46.4, P3.1.69

N

Nakagomi T. P4.5.12
Nasserinejad K. C32.2
Ndiaye K. C33.6
Nematollahi S. P4.4.99
Neuenschwander B. C04.2
Nevo D. C34.6
Newcombe P.J. C31.4
Newcombe R.G. C18.2
Newell J. C06.1, P1.2.7
Nguipdop Djomo P. C12.2
Nguyen T.T. P1.2.109
Nguyen Duc A. C22.5, C45.5
Niccols A. C38.3, P4.5.176
Nicola P. P3.1.149
Nieboer D. C44.2
Nikolakopoulos S. C24.3, C30.2, P1.1.110
Nikolakopoulou A. C26.5
Nilsson M. P1.2.48
Nishikawa M. P3.1.111
Nomura S. C25.5
Norel R. C20.2
Noufaily A. C12.5
Novianti P.W. C16.2, P4.5.81
Numthavaj P. P4.1.108
Nunes C. P4.2.58
Nyari T.A. P2.1.92
Nyári T. P2.2.182

O

O'Neill A.C. P2.2.16
O'Donnell M. C06.1
Ofner-Kopeinig P. P1.2.112
Ofuya M. P3.4.113
Ogundimu E.O. C19.3
Ohmann C. P4.1.177
O'Leary N. P1.2.114
Oleche P. P2.2.115
Oliveira M.R. P3.2.153
O'Malley S.S. C06.5
Omar R. C19.5, C36.5
Omar R.Z. P3.5.8
Omollo R. P2.2.115
Omolo N. P2.2.115
Osipova E.A. P2.2.29
Otava M. P1.1.107
Ouwerkerk W. P2.5.116
Owczarek A.J. P3.2.117
Owen K. P3.5.145
Owen R.K. P1.1.118
Ozga A.-K. C27.5

P

Pac A. P2.1.119
Paganoni A.M. C11.2, C45.4
Pais Ribeiro J.L. P3.4.180, P3.4.181
Palermo G. C17.1
Palesch Y. P1.2.195
Palesch Y.Y. C39.2
Pandey A. P2.2.163
Pandey R.V. P4.5.90
Pandey R.M. C41.5, P3.5.82
Pang M. C37.3
Paoletti X. C04.1, P1.2.47
Pardo-Fernández J.C. C40.4, P3.2.101
Park J.-S. P2.2.79
Park J.-H. P2.2.79
Park J.H. P4.4.94
Park M. P2.5.120
Park S.-H. P3.2.95
Park S.H. P4.4.94
Park T. C10.3
Parkin D. P4.4.15
Parmar M.K.B. C40.1
Parpia S. P1.2.121
Pascoal A. P2.2.155
Pasin O. P1.2.122
Pau D. P1.1.124, P1.3.123
Pavlou M. C19.5, C23.1
Pawitan Y. C10.4, C13.5, C42.4
Peacock J. C34.4, C46.5, P2.2.155
Peacock J.L. P3.4.113
Pedersen T. C37.5
Pedro L. P3.4.180
Pellegrin I. C43.3
Pérès K. P2.1.129
Perneger T.V. P4.5.128
Peron E. C35.1
Perperoglou A. C41.3
Peters A. C16.5
Peters U. P4.5.6
Pfirrmann M. P3.1.125
Pharoah P.D. C31.4
Phung K.L. C08.2
Pibre S. P1.3.123
Picat M.-Q. C43.3
Pietrabissa T. C45.4
Pignon J.-P. P4.1.135
Pingel R. P4.4.126
Pitkäniemi J. P2.2.127
Platt R. C14.4
Platt R.W. C37.3
Pliczko M. P3.4.143
Pohar Perme M. C47.1
Polkinghorne K. C28.3
Pollack M.M. C08.1
Poncet A. P4.5.128
Poon W.-Y. C18.1
Popat S. C25.4
Porcher R. C40.2
Posch M. C30.1, C30.3, S2.1, C36.3, C36.6, P1.2.67
Pötschger U. C27.2
Pramana S. C42.4
Prevost A.T. C35.5
Prostyakov I.V. P2.2.29
Proust-Lima C. C15.5, C43.3, M1.3, P2.1.129, P4.5.159
Provenzano E. C31.4
Pruvost A. P1.2.109
Puhan M.A. C46.4
Putter H. Course 2, C39.4, C45.1

Q

Quantin C. C33.1
Quartagno M. C20.3
Querard A.-H. C37.1
Quinn S. C34.3
Quintana J.M. P3.5.55, P2.1.54, P2.3.2, P3.5.10

R

Rabe-Hesketh S. C32.3
Rachet B. P3.3.96
Rahlf A.-L. P4.4.46
Rahman S. P3.5.8
Rainisio M. P1.2.168
Ramos A. P1.1.130
Raña P. P4.5.179
Rattanasiri S. P4.1.108, P4.1.131
Rauch G. C30.4
Ravn H. C28.4
Raza Ali H. C31.4
Rebollo-Mesa I. C29.2
Reeves D. C14.3
Reiczigel J. P2.2.132, P3.2.91
Reiczigel Z. P4.5.133
Reigner B. C17.2
Reilly M. C05.2, C37.2
Reitsma J.B. C19.1, C20.1, C23.5
Remontet L. C33.1, C33.5
Resche-Rigon M. C02.4, C21.2
Rhodes K.M. P4.1.134
Ribassin-Majed L. P4.1.135
Ribeirinho M. P1.3.136
Richardson S. C31.4
Riche B. P4.2.24
Riedl R. C43.1
Riksheim M. P2.3.137
Riley R.D. C20.1, C44.3, P3.5.45, P4.1.4, P4.1.42
Rippe R.C.A. P2.5.138
Ristl R. C36.6
Rizopoulos D. C15.4, C32.2
Rizzo E. C04.1, P1.2.47
Roberts C. C07.5, P1.2.35, P1.2.114
Roberts S. C07.5
Roberts S.A. P1.2.114
Robertson D.S. C35.5
Roche L. C33.5
Roderick P. P4.4.192
Rodgers L.R. C12.4
Rodrigues A. C21.3, P3.1.160
Rodrigues L. C12.2
Rodriguez M. P1.2.166
Rodriguez R.N. I2.3
Rodríguez A. P1.2.168
Rodríguez-Álvarez M.X. P4.5.139
Rodwell L. C46.3, P2.3.140
Roes K.C.B. C16.2, C24.3, C25.2, C30.2, P1.1.110, P4.5.81
Røislien J. C11.1, P2.3.137, P4.5.141
Romaniuk H. C46.3, P2.3.140
Romio S. P1.2.142, P2.2.41

Rondeau V. Course 3
Rosenberger W.F. P1.1.32
Rosenkranz G.K. C01.1
Rosmalen J.V. C32.2
Rossa A. P4.5.158
Rossi A.M. C28.2
Rotolo F. C31.1
Rouanet A. C09.5
Roubideaux Y. C09.4
Rouzier R. P1.1.124
Rovers M. C20.1
Røysland K. C37.5, C43.4
Royston P. C13.1, C40.1
Rücker G. C14.5, C26.3
Rudd A. P4.4.15
Rudd A.G. C08.3
Rufibach K. C47.2
Rümenapf G. P3.1.22
Ryu C.H. P3.2.95

S

Sabanes Bove D. C17.1, C31.5
Sacks S. C29.2
Salanti G. C14.1, C26.1, C26.5
Salapa K. P3.4.143
Sander J.W. P3.5.25
Sansanayudh N. P4.1.108
Saputra D. C42.4
Sargent D. C01.5
Sariyar M. C16.4
Sauerbrei W. C03.2, C13.1
Saure D. C48.4
Sauvageot N. C11.4
Sauzet O. C34.4, P3.4.113
Sawabe M. P2.2.105
Scharpenberg M. C03.1, C03.3
Scheike T. C21.5
Schemenau J. C45.3
Schetelig J. C39.4
Schim van der Loeff M. P4.2.102
Schindler D. P1.2.172
Schmedt N. P1.2.142
Schmid M. C16.5
Schmidli H. C04.3, C07.4
Schmidt S. C36.1
Schmidtmann I. C21.1, C27.3
Schoenfeld D. C20.2
Schortgen F. P1.2.178
Schramm C. C27.4
Schritz A. C27.3
Schroeder F. P2.5.144
Schüler A. C48.2
Schulz C. C29.1
Schumacher M. C14.5, C39.5, C45.2
Schüpbach G. P2.2.185
Schürmann C. C22.3
Schuster T. C37.3
Schütz H. P1.2.88
Schwarzer G. C14.5
Scolas S. C47.3
Scotti L. P1.2.142
Scudeller L. P2.2.85
Seaman S. Course 1, C19.5
Seaman S.R. C23.1
Sébille V. P3.3.49
Seidel A. P1.2.86
Selingerová I. C41.1
Semsei A.F. P3.5.62
Sène M. C15.5
Senn S. C25.1
Seppä K. P2.2.127
Shahar D.R. C15.3
Shamsheva O.V. P2.2.29
Shankar-Hari M. P1.2.97
Shanyinde M. P1.2.35
Sharples L. C18.4, C35.3
Shen H. C04.2
Sheriff M.R. C42.5
Sherman A. C20.2
Shields B.M. P3.5.145
Shiffman S. C09.2
Shim J. P3.4.83, P4.1.87
Shin E. P3.2.95
Shin Y. P2.2.79
Shinozaki T. C25.5
Shkedy Z. P1.1.107
Shui A. C20.2
Sieben W. C22.3
Silva I. P3.4.180, P3.4.181
Simmonds M. C14.2
Simpson S. P1.2.146
Sinani E. C20.2
Singh R. P2.1.147
Sjolander A. C13.5
Skalska A. P3.2.117
Skoog I. P4.4.104
Skrondal A. C32.3
Slaets L. C37.4
Sliwinska A. P3.4.143
Småstuen M.C. P3.1.148
Smith S. C04.5
Snell K. P3.5.45
Snell K.I.E. C44.3
Sohn C.-H. P3.2.95
Solanki S. C30.5, P1.1.164
Somers J.M. I4.2
Sommer H. C13.4
Son S.-K. P3.2.95
Sotres-Ramos D. P1.2.5
Sousa A. P3.1.149
Sousa I. C15.1, P3.1.160
Spang R. I6.1
Spiessens B. C41.4
Springate D. C14.3
Sreenivas V. P2.2.163
Sriwanichakorn S. P3.1.170
Stammet P. P3.5.37
Stare J. C40.5
Steeves J. C20.5
Stephen J. P3.1.150, P3.1.151
Stephens D.A. I1.2
Sterne J. C32.5
Steyerberg E.W. C01.4, S1.2, C44.2, P3.5.174, P3.5.175
Stoehlker A.-S. C13.4
Stolovitzky G. C20.2
Streeter A.J. P3.3.152
Strohmaier S. C37.5
Ström P. C05.2
Štupnik T. C47.1
Sturkenboom M. P1.2.142, P2.2.41
Subhaluksuksakorn P. P3.1.170
Subtil A. P3.2.153, P3.2.154
Sugitani T. C36.3
Suhre K. C16.5
Sullivan T.R. C23.2
Summers J.A. P2.2.155, P2.2.156
Sun H. C13.2
Sun S.-H. P2.2.79
Sungur M.A. P1.2.122
Suo C. C42.4
Swain P.K. P3.1.157
Swinkels S. C18.5, P1.2.18
Sylte M.S. P4.5.187
Symeonides S. C04.5
Szalai C. P3.5.62
Szymański A. P4.5.158

T

Takahashi I. C46.1
Takkenberg J.J.M. C15.4
Talbot D. C28.2
Tamm M. C01.2
Tanadini L. C20.5
Tanaka N. P2.2.105
Tango T. C48.5, P4.1.190
Tanguy J. P4.5.159
Tanniou J. C25.2
Tarabelloni N. C11.2
Tavares D.P. P3.1.149
Taylor J.M. C15.5, I5.2
Teerenstra S. C25.2
Teixeira L. C21.3, P3.1.160
Tepper S. C15.3

Ter Riet G. C46.4
Terblanche M. P1.2.97
Ternès N. C31.1
Tezenas du Montcel S. P2.1.161
Thabane L. C07.1, C38.3, P1.2.121, P4.5.176
Thakkinstian A. P3.1.162, P4.1.108, P4.1.131
Thakur B. P2.2.163
Thas O. C34.1
Thelle D.S. P4.4.104
Thepthien B.-O. P3.1.170
Thiébaut R. C12.3, C43.3
Thoemmes G. C02.2
Thom H. C35.3
Thomas D. C14.4, P4.5.6
Thompson D.D. C01.4
Thompson D.K. C05.4
Thompson J.R. C48.3
Tilling K. C32.5
Timm J. C29.1
Timmerman D. P3.5.173
Tincello D.G. P1.1.118
Tinelli C. P2.2.85
Tirodkar C. P1.1.164
Titman A. C38.4
Titterington M. P4.5.184
Tobiasz-Adamczyk B. P2.1.119
Todd J. C17.3
Tomak L. P4.5.165
Tomaszewski K.A. P3.4.143
Torre F. P2.1.54
Torres-Martin J.V. P1.2.167, P1.2.166
Torres-Martín J.V. P1.2.168
Touraine C. C33.2
Toyoda T. P4.5.12
Tsai W.-M. C06.5
Tseng C.-H. P2.3.169
Tudur Smith C. P3.5.25
Turner R.M. P4.1.134

U

Uchaikin V.F. P2.2.29
Udomsubpayakul U. P3.1.170
Uemura Y. P4.4.171
Uhlmann L. C26.4
Uhry Z. C33.5
Unzurrunzaga A. P3.5.10
Unzurrunzaga A.G. P3.5.55
Urrechaga E. P3.5.3
Uschner D. P1.2.172

V

Vach W. C13.2, C13.3, S1.1, C38.1, C44.1
Vakulenko-Lagun B. C39.3
Valsecchi M.G. C27.1, C27.2
Van Buuren S. C20.4
Van Calster B. C19.4, P3.5.173
Van de Wiel M. C19.2
Van de Wiel M.A. P2.3.44
Van der Graaf R. C23.5
Van der Tweel I. C16.2, C23.5, C24.3, C25.2, C30.2, P1.1.110
Van Hoorde K. P3.5.173
Van Huffel S. C19.4, P3.5.173
Van Klaveren D. P3.5.174, P3.5.175
Van Rosmalen J. C07.3, C46.2
Van Steen K. P2.5.40
Van Wieringen W.N. I6.3
Vandenberghe S. C37.4
VanderWeele T. Course 6
Vanniyasingam T. C38.3, P4.5.176
Vannucci M. I6.2
Vansteelandt S. C28.5
Vejakama P. P3.1.162
Verde P.E. P4.1.177
Vergouwe Y. C19.4, C44.2, P3.5.174, P3.5.175
Viallon V. P4.3.13
Vickers A.V. S1.2
Vickerstaff V. C36.5
Victorri-Vigneau C. C35.1, P3.3.49
Vidmar G. C34.2
Vidyashankar A.N. C10.2, P1.2.178
Vigan M. C32.4
Vigneau C. P1.3.123
Vilar J. P4.5.179
Vilhena E. P3.4.180, P3.4.181
Virág K. P2.2.182
Visne I. P4.5.90
Vivian G. P2.2.155
Von Wangenheim U. C22.1
Vonthein R. C06.3

W

Waernbaum I. S1.4
Wahl S. C16.5
Waldenberger M. C16.5
Waldhoer T. P2.2.183
Waldhör T. P2.2.65
Waldron-Lynch F. C17.3
Walker J. C20.2
Walls T.A. C32.1
Walter S. C01.3
Wang A. P1.3.51
Wang Y. P4.5.184
Warner P. C48.1
Wason J. C17.4, C17.5, C24.2
Wassmer G. P1.2.88, P1.2.89
Wasunna M. P2.2.115
Wattanawong K. P4.1.131
Weber S. C45.2
Wedzicha W. P1.2.93
Wei W. P2.2.185
Weinhäusel A. P4.5.90
Weinmann A. C21.1, C27.3
Weir C.J. C48.1, P4.1.186
Weldeselassie Y. C12.5
Welton N. C35.3
Wentzel-Larsen T. P4.5.187
West J. P4.4.192
Weyer V. C08.5
Whelan T.J. P1.2.121
Whiley M. C04.2
Whitaker H. C03.5
White I. Course 1
White I.R. I2.1, C02.5, C14.1, C21.2
Whitehead J. C17.2, P1.2.146
Whittaker R. P4.1.80
Więcek A. P3.2.117
Wicker L. C17.3
Wieczorkowska-Tobis K. P3.2.117
Wilkinson I.B. C05.5
Witte S. C22.4, P2.3.188
Wittkop L. C43.3
Wolbers M. C08.2, C22.5, C45.5
Wolfe C. C46.5, P4.4.15
Wolfe C.D.A. C08.3
Wolfe R. C28.3
Wolfsegger M. P1.2.88
Wollenweber F.A. P2.3.196
Wright J. P4.4.192
Wright N. C23.3
Wu R. C06.5
Wynant W. P3.1.189
Wynants L. C19.4

X

Xu Y. C05.3
Xyrafas A. P1.2.64

Y

Yagel S. C34.6
Yamaoka K. P4.1.190
Yao W. C09.2
Yates T.A. C02.5
Yavuz Y. C18.5
Ybarra M. P4.1.80
Yeatts S. P1.2.195
Yehia S. P2.2.132
Yelland L.N. C23.2
Yeung W.Y. C17.2
Yildiz A. P4.5.90
Yilmaz Isikhan S. C16.1
Yim J.-J. P3.4.34
Yim J.J. P4.5.84
Yiu S. P3.1.191
Yokoyama T. P3.1.111
Yoo Y.H. P3.2.95
Yoon H.I. P3.4.34, P4.5.84
Young J. C04.5
Yu M. C39.1
Yuen H.M. P4.4.192

Z

Zach N. C20.2
Zamanzad Ghavidel F. P2.5.194
Zambon A. P1.2.142
Zamir E. C42.5
Zech A. P4.4.46
Zeilinger S. C16.5
Zelinka J. C41.1
Zetterberg H. P4.4.104
Zhang H. C06.5
Zhang Y. C34.1, C35.2
Zhao W. C29.3, C39.2, P1.2.195
Zhong T. P2.2.16
Ziegler A. C25.3, M2.4
Zietemann V.D. P2.3.196
Zohar S. P1.1.124
Zöller D. C21.1, C27.3
Zwahlen M. P1.1.1
Zwiener I. C42.1
Zwinderman A. C42.2, C46.4, P2.5.116, P3.1.69

www.iscb2014.info