41st Annual Conference
of the International Society
for Clinical Biostatistics

ISCB 41
KRAKÓW 2020

Programme and Abstract Book

Virtual Conference
co-organised by
Jagiellonian University
Kraków, Poland

23 – 27 August 2020

ISCB International Society for Clinical Biostatistics

UNIVERSITAS JAGELLONICA CRACOVIENSIS

Main Sponsor

AstraZeneca

Local Sponsor

Kraków

Partner

UNIVERSITAS JAGELLONICA COLLEGIUM MEDICUM

Exhibitor

CRC Press Taylor & Francis Group

CHAPMAN & HALL

ISBN: **978-83-942760-7-2**

Graphic design & computer typesetting:
Wojciech Szajnerman, **Agencja Reklamowa EURO GRAPHIC**

Table of Contents

Oral Contributed Sessions 91

Introduction

International Society for Clinical Biostatistics

The International Society for Clinical Biostatistics 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.

Membership is open to all interested individuals who share the Aims of the Society. ISCB's membership include clinicians, statisticians and members of other disciplines, such as epidemiologists, clinical chemists and clinical pharmacologists, working or interested in the field of clinical biostatistics.

ISCB has an Executive Committee and 6 sub-committees: Conference Organising, Early Career Biostatisticians, Education, National Groups, Statistics in Regulatory Affairs (SiRA), Student Conference Awards.

Executive Committee

President
Vana Sypsa, Greece

Vice-President
Zdenek Valenta, Czech Republic

Secretary:
Geir Egil Eide, Norway

Treasurer:
Thomas Jaki, United Kingdom

Members:
Dimitris Rizopoulos, Netherlands
Elaine Pascoe, Australia
Erik Cobo, Spain
Jan Terje Kvaløy, Norway
Jonathan Bartlett, United Kingdom
Lehana Thabane, Canada
Nadine Binder, Germany
Tomasz Burzykowski, Belgium
David W. Warne, Switzerland

Honorary Members

David W. Warne
Jørgen Seldrup
Stephen Senn
John Whitehead
Hans van Houwelingen
Emmanuel Lesaffre
Martin Schumacher
Lutz Edler
Harbajan Chadha-Boreham
Michael Schemper

Committees

Local Organising Committee

Chair:

Krystyna Szafraniec, Kraków

Vice-Chair:

Maciej Polak, Kraków

Treasurer:

Kinga Sałapa, Kraków

Małgorzata Ćwiklińska-Jurkowska, Toruń
Coraline Danieli, Montreal, Canada
Agnieszka Doryńska, Kraków
Aleksander Owczarek, Katowice
Andrzej Pająk, Kraków
Renata Wolfshaut-Wolak, Kraków

Scientific Programme Committee

Chair:

Michal Abrahamowicz, McGill, Montreal, Canada

John Carlin, Melbourne, Australia
Yingying Fan, USC Los Angeles, United States
Krista Fischer, Tartu, Estonia
Jessica Franklin, Harvard, United States
Laurence Freedman, Jerusalem, Israel
Mitchell H. Gail, NIH/NCI, United States
Andrzej Galecki, Michigan, United States
Dani Gamerman, Rio de Janeiro, Brazil
Lisa Hampson, Novartis, Basel, Switzerland
Ruth Keogh, London, United Kingdom
Zoltán Kutalik, Lausanne, Switzerland
Maia Lesosky, Cape Town, South Africa
Aleksander Owczarek, Katowice, Poland
(Mini-Symposia Coordinator)
Maja Pohar Perme, Ljubljana, Slovenia
Sven Ove Samuelsen, Oslo, Norway
Willi Sauerbrei, Freiburg, Germany
Krystyna Szafraniec, Kraków, Poland (LOC Chair)
Toshiro Tango, Tokyo, Japan

Dr. Michal Abrahamowicz is a James McGill Professor of Biostatistics at McGill University, in Montreal, Canada. His statistical research aims at development and validation of new, flexible statistical methodology, with main focus on time-to-event (survival) analyses of prognostic and pharmaco-epidemiological studies. He has developed also new methods to control for different sources of bias in observational studies. His collaborative research focuses on pharmaco-epidemiology, arthritis, cardiovascular and cancer epidemiology.

He is the co-Chair of the international STRATOS Initiative for enhancing analyses of observational studies www.stratos-initiative.org.

Welcome to ISCB41 Kraków 2020 VIRTUAL Conference!

The medieval city of Kraków and the 656-year old Jagiellonian University are proud to host the 41st annual conference of ISCB. The Local Organising Committee and Scientific Programme Committee have been working together to create an impressive scientific and social programme. However, due to the COVID-19 pandemic, we had to switch ISCB41 Kraków 2020 to a virtual format, which created a host of new challenges.

Given the Invited Speakers and participants come from all 6 continents and span 17 time zones, the programme was developed to maximize the opportunities for all attendees to participate in most activities. The conference retains the same scope as what we had originally planned but we will all participate remotely.

ISCB 2020 will span five days and will feature outstanding speakers, leading scientists and young people at the beginning of their career presenting their papers and joining discussions.

In **Plenary Sessions** eminent researchers will address topics hotly debated in modern biostatistics: in the President's Invited talk **Miguel Hernán** *(Harvard)* will discuss using Causal Inference to learn from real-world evidence, and the two Keynote Speakers **Frank Harrell Jr.** *(Vanderbilt)* and **Trevor Hastie** *(Stanford)* will exchange views on the relationships between Statistical Modelling and Machine Learning.

In the **13 Invited Sessions**, leaders in their fields will provide comprehensive views of recent methodological developments in both traditional areas of biostatistics (clinical trials, causal inference, survival analysis) and the emerging fields of big data and functional analysis, omics and personalized medicine, Bayesian modeling of infectious diseases, and many others. We are excited to have added a new Invited Session on "Statistical challenges in Covid-19 research" that will address the new reality and related methodological issues we all face in 2020.

The **Early Career Researchers' Day** will provide a unique opportunity for graduate students, post-doctoral fellows and biostatisticians at the beginning of their career to share experiences and advice, discuss opportunities and challenges, and practise their presentation skills in a less formal environment.

As in previous years, **pre-conference Short Courses** given by internationally renowned experts and **Mini-Symposia** on guidance for the analysis of observational data, statistical methods for pharmaco-epidemiology/-economics, and high-dimensional data will complete the programme.

In addition to 45 Invited presentations, more than 200 abstracts will be presented within 42 Oral Contributed Sessions and about 200 posters will be available for viewing during the entire conference. Almost 90 of these posters will be briefly presented within one of the 6 dedicated oral "Poster overview" sessions.

For more scientific details, see the program and abstracts provided in the following sections.

Meeting people from different geographical areas who share common methodological interests is an additional reason for, and the pleasure of, participating in the conference. In virtual circumstances this is not easy but we tried our best. During a Virtual Welcome Get Together on Monday and a Students' Get Together on Sunday, you will be able to meet, talk and relax with other conference participants.

We wish you strong scientific impressions and patience in overcoming potential connectivity problems,

Krystyna Szafraniec
Jagiellonian University, Krakow, Poland
Chair, Local Organising Committee

Michal Abrahamowicz
McGill University, Montreal, Canada
Chair, Scientific Programme Committee

Recipients of the Conference Fund for Developing Countries

Tanvir Ahammed, Shahjalal University of Science and Technology, Bangladesh
Moumita Chatterjee, Aliah University, India
Dip Das, University of Dhaka, Bangladesh
Vinay Gupta, JSS Medical Research India Pvt Ltd, India
Anower Hossain, University of Dhaka, Bangladesh
Samer Kharroubi, American University of Beirut, Lebanon
Elham Madreseh, Tehran University of Medical Sciences, Islamic Republic of Iran
Ihor Malyk, Yuriy Fedkovych Chernivtsi National University, Ukraine
Jerissa Samuel Pezhumkattil, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Navi Mumbai, India
Rubaiya Rubaiya, University of Dhaka, Bangladesh
Md Ariful Islam Sanim, University of Dhaka, Bangladesh
Leili Tapak, Hamadan University of Medical Science, Islamic Republic of Iran
Zijing Yang, Southern Medical University, Guangzhou, China

Award Winners

Student Conference Awards

Szymon Urbas	Lancaster University, United Kingdom	Inhomogeneous Poisson-gamma recruitment model for multi-centre clinical trials with model-averaged predictions	OC01: Clinical Trials Methodology 1 Mon 15.50 - 17.20
Wenyue Zhu	University of Liverpool, United Kingdom	Spatial predictive statistical model: detecting the need for treatment from retinal imaging data	OC04: Individualised Risk Prediction 1 Mon 15.50 - 17.20
Orlagh Carroll	London School of Hygiene and Tropical Medicine, United Kingdom	Combining bootstrap with multiple imputation to assess prognostic models with missing covariate data	OC22: Missing Data 1 Tue 14.00 - 15.30
Sabbir Ahmed Hemo	University of Dhaka, Bangladesh	Performance of inverse probability weighting compared to multiple imputation for missing binary outcome in CRT	OC35: Missing Data 2 Wed 12.10 - 13.40

Conference Awards for Scientists

Tasneem Fatima Alam	University of Dhaka, Bangladesh	Bias reduction and solution to separation in the AFT models for small or rare event survival data	OC14: Survival Analysis 2 Tue 12.10 - 13.40
Md Mynul Islam	University of Dhaka, Bangladesh	A spatial multilevel model using conditional autoregressive processes	OC23: Bayesian Methods Tue 14.00 - 15.30
Nayma Hossain	University of Dhaka, Bangladesh	Iterative Least-squares Regression with Censored Data: A Survival Ensemble of Learning Machine	OC41: Machine Learning Wed 14.00 - 15.30

Plenary Sessions

President's Invited Session

Organiser / Chair: **Vana Sypsa** (ISCB President)

Medical School, National and Kapodistrian University of Athens, Greece

Learning what works: Causal inference from observational data

Miguel Hernán

Harvard T.H. Chan School of Public Health, Boston, United States

Making decisions among several courses of action requires knowledge about their causal effects. Randomized experiments are the preferred method to quantify those causal effects. When randomized experiments are not feasible or available, causal effects are often estimated from non-experimental or observational databases. Therefore, causal inference from observational databases can be viewed as an attempt to emulate a hypothetical randomized experiment-the target experiment or target trial-that would quantify the causal effect of interest. This talk outlines a general algorithm for causal inference using observational databases that makes the target trial explicit. This causal framework channels counterfactual theory for comparing the effects of sustained treatment strategies, organizes analytic approaches, provides a structured process for the criticism of observational analyses, and helps avoid common methodologic pitfalls.

Miguel Hernán conducts research to learn what works to improve human health. Together with his collaborators, he designs analyses of healthcare databases, epidemiologic studies, and randomized trials. Miguel teaches clinical epidemiology at the Harvard-MIT Division of Health Sciences and Technology, and causal inference methodology at the Harvard T.H. Chan School of Public Health, where he is the Kolokotrones Professor of Biostatistics and Epidemiology. His edX course „Causal Diagrams" and his book „Causal Inference: What If", co-authored with James Robins, are freely available online and widely used for the training of researchers. Miguel is an elected Fellow of the American Association for the Advancement of Science and of the American Statistical Association, Editor Emeritus of Epidemiology, and past Associate Editor of Biometrics, American Journal of Epidemiology, and the Journal of the American Statistical Association.

Keynote Session: Machine Learning vs. Statistical Modeling

Organiser / Chair: **Michal Abrahamowicz** (SPC Chair, ISCB 2020)
McGill, Montreal, Canada

Predictive Models in Health Research

Trevor Hastie

Stanford University, USA

Random Forests, Boosting, SVMs and especially Deep Neural Networks are very popular in data science applications. Do they have a role in health research, and are they likely to replace more traditional statistical models? In this talk I will argue that it depends on the application, the amount of data available, and the purpose of the modeling, with some guidance from the „Occam's razor" principle. Musing on Statistical Models vs. Machine Learning in Health Research.

Musing on Statistical Models vs. Machine Learning in Health Research

Frank Harrell Jr.

Vanderbilt University School of Medicine, Vanderbilt, USA

Health researchers and practicing clinicians are with increasing frequency hearing about machine learning (ML) and artificial intelligence applications. They, along with many statisticians, are unsure of when to use traditional statistical models (SM) as opposed to ML to solve analytical problems related to diagnosis, prognosis, treatment selection, and health outcomes. And many advocates of ML do not know enough about SM to be able to appropriately compare performance of SM and ML. ML experts are particularly prone to not grasp the impact of the choice of measures of predictive performance. In this talk I attempt to define what makes ML distinct from SM, and to define the characteristics of applications for which ML is likely to offer advantages over SM, and vice-versa. The talk will also touch on the vast difference between prediction and classification and how this leads to many misunderstandings in the ML world. Other topics to be convered include the minimum sample size needed for ML, and problems ML algorithms have with absolute predictive accuracy (calibration).

Trevor Hastie is the John A Overdeck Professor of Statistics at Stanford University. Hastie is known for his research in applied statistics, particularly in the fields of statistical modeling, bioinformatics and machine learning. He has published six books and over 200 research articles in these areas. Prior to joining Stanford University in 1994, Hastie worked at AT&T Bell Laboratories for 9 years, where he contributed to the development of the statistical modeling environment popular in the R computing system. He received a B.Sc. (hons) in statistics from Rhodes University in 1976, a M.Sc. from the University of Cape Town in 1979, and a Ph.D from Stanford in 1984. In 2018 he was elected to the National Academy of Sciences.

Dr. Frank Harrell received his PhD in Biostatistics from UNC in 1979. Since 2003 he has been Professor of Biostatistics, Vanderbilt University School of Medicine, and was the department chairman from 2003-2017. He is Expert Statistical Advisor for the Office of Biostatistics for FDA CDER. He is Associate Editor of Statistics in Medicine, and a member of the Scientific Advisory Board for Science Translational Medicine. He is a Fellow of the American Statistical Association and winner of the Association's WJ Dixon Award for Excellence in Statistical Consulting for 2014. His specialties are development of accurate prognostic and diagnostic models, model validation, clinical trials, observational clinical research, cardiovascular research, technology evaluation, pharmaceutical safety, Bayesian methods, quantifying predictive accuracy, missing data imputation, and statistical graphics and reporting.

Programme overview

ISCB41 Conference - Program at a glance

Time Zone CET+1 Kraków	Sunday 23 August 2020	Monday 24 August 2020	Tuesday 25 August 2020	Wednesday 26 August 2020	Thursday 27 August 2020
10.20 - 11.50				Poster Overview Session PO4-PO6	
20-min break					Start 11.30
12.10 - 13.40	Start: 13.00		IS05 / IS13 New IS on Covid-19 Research / OC13-OC18	IS10 / OC31-OC36	Early Career Biostatisticians' Day / MS3
20-min break					MS1 MS2
14.00 - 15.30	Pre-conference short courses	Conference Opening President's Invited Speaker **Miguel Hernán**	IS06 / IS07 / OC19-OC24	IS11 / IS12 / OC37-OC42	
20-min break					
15.50 - 17.20	SC1 SC2 SC3 SC4 SC5	IS01 / IS02 / OC01-OC06	IS08 / IS09 / OC25-OC30	ISCB Annual General Meeting	
20-min break					Closing ISCB41 Invitation ISCB42
17.40 - 19.10		IS03 / IS04 / OC07-OC12	Poster Overview Session PO1-PO3	Keynote Speakers **Trevor Hastie & Frank Harrell**	
20-min break					
19.30 - 20.30	Students' Get Together	Welcome Get Together, Conference AWARDS Ceremony & Kraków virtual tour			

Sunday, 23 August 2020

Time Zone CET+1 Kraków	Pre-conference short courses				
13.00 - 15.00	SC1: Regression Modelling Strategies Frank Harrell Jr., Vanderbilt, US	SC2: Absolute Risk Methods and Applications in Clinical Management and Public Health Mitchell H. Gail and Ruth Pfeiffer, NIH/NCI, US	SC3: Understanding and tackling measurement error: A review of modern practical methods Pamela Shaw, U Penn, US Ruth Keogh, LSHTM, UK	SC4: Analysis of high-dimensional data Jörg Rahnenführer and Michel Lang, TU Dortmund, Germany, and Lara Lusa, University of Ljubljana, Slovenia	SC5: Early Phase Clinical Trials Alexia Iasonos MSKCC, New York, US and John O'Quigley, UCL, UK
30-min break					
15.30 - 17.00	SC1: continued	SC2: continued	SC3: continued	SC4: continued	SC5: continued
30-min break					
17.30 - 19.00	SC1: continued	SC2: continued	SC3: continued	SC4: continued	SC5: continued
30-min break					
19.30 - 20.30	Students' Get Together				

Monday, 24 August 2020

Time Zone CET+1 Kraków								
14.00 - 15.30	**Conference Opening** President's Invited Speaker: **Miguel Hernán**, Harvard University, US **Learning what works: Causal inference from observational data** Organiser: **Vana Sypsa** (President ISCB), *National and Kapodistrian University of Athens, Greece*							
break								
15.50 - 17.20	IS01: Bayesian modelling of chronic & infectious diseases: Reverend Bayes going anti-viral	IS02: Tackling the challenges of imperfect-ions in longitudinal and survival data: missing-ness, measurement error and irregular sampling	OC01: Clinical Trials Methodology 1	OC02: Causal Inference 1	OC03: Multi-State models	OC04: Individualised Risk Prediction 1	OC05: Omics & Survival Analysis	OC06: Novel Applications of Survival Analysis
break								
17.40 - 19.10	IS03: Can Ying dance with Yang?: Combining evidence from RCT's and Pharmaco-epidemiology population-based studies of drug effects	IS04: New Frontiers: Big Data & Functional Data Analysis	OC07: Clinical Trials Methodology 2	OC08: Causal Inference 2	OC09: Survival Analysis 1	OC10: Personalised Medicine 1	OC11: Methods Validation and Comparisons	OC12: Meta-analysis 1
break								
19.30 - 20.30	Welcome Get Together, Conference AWARDS Ceremony & Kraków virtual tour							

Tuesday, 25 August 2020

Time Zone: CET+1 Kraków								
12.10 - 13.40	IS05: Causal Inference vs. Unmeasured Confounding Let's face the Dragon"	IS13: Statistical challenges in Covid-19 research	OC13: Clinical Trials Methodology 3	OC14: Survival Analysis 2	OC15: Individualised Risk Prediction 2	OC16: Functional Data Analysis	OC17: Omics & Genetics	OC18: Longitudinal Data
20-min break								
14.00 - 15.30	IS06: When non-inferiority trials are superior?	IS07: Can Cox model survive in the 21st century survival analysis?	OC19: Epidemiology: Methods & Applications	OC20: Multivariable Model Building	OC21: Bayesian Methods & Omics	OC22: Missing Data 1	OC23: Bayesian Methods	OC24: Meta-analysis 2
20-min break								
15.50 - 17.20	IS08: New developments in Longitudinal data analysis	IS09: Progress in providing analytic guidance for observational studies: the STRATOS initiative	OC25: Clinical Trials Methodology 4	OC26: Causal Inference for Survival Analysis	OC27: Survival Analysis 3	OC28: Personalised Medicine 2	OC29: Measurement Errors	OC30: Bayesian Methods for Clinical Trials
20-min break								
17.40 - 19.10	PO1: Clinical Trials: Methods & Applications	PO2: Longitudinal Data Biomarkers	PO3: Survival analysis Machine Learning					

Wednesday, 26 August 2020

Time Zone CET+1 Kraków								
10.20 - 11.50	PO4: Big Data Machine Learning	PO5: Causal Inference	PO6: Epidemiology: Methods and Applications Electronic Health Records					
20-min break								
12.10 - 13.40	IS10: Individualised Risk Prediction & Omics		OC31: Clinical Trials Methodology 5	OC32: Causal Inference 3	OC33: Survival Analysis 4	OC34: Personalised Medicine 3	OC35: Missing Data 2	OC36: Dynamic Prediction
20-min break								
14.00 - 15.30	IS11: Biostatistical inference in practice: moving beyond false dichotomies	IS12: Efficient Designs & cutting-edge analyses	OC37: Clinical Trials Methodology 6	OC38: Causal Inference 4	OC39: AFT Modelling & Survival	OC40: Big Data	OC41: Machine Learning	OC42: Meta-analysis 3
20-min break								
15.50 - 17.20	ISCB Annual General Meeting							
20-min break								
17.40 - 19.10	Keynote Session: **Machine Learning vs. Statistical Modelling** **Trevor Hastie** *(Stanford, US)* **Predictive Models in Health Research** **Frank Harrell Jr.** *(Vanderbilt, US)* **Musing on Statistical Models vs. Machine Learning in Health Research** Organiser: **Michal Abrahamowicz** (SPC Chair, ISCB2020) *McGill, Montreal, Canada*							

Thursday, 27 August 2020

Time Zone CET+1 Kraków				
11.30				MS3: 11.30 - 17.20 **Advanced statistical methods for the analysis of high-dimensional medical data** Organisers: Małgorzata Bogdan, Poland and Jaroslaw Harezlak, United States
12.00	**Early Career Biostatisticians' Day** Speakers: Emily Karahalios Havi Murad Stephen Senn			
14.00		MS1: 14.00 - 17.00 **Statistical methods in pharmacoepidemiology and armacoeconomics** Organiser: Robert Platt, Canada	MS2: 14.00 - 17.20 **STRATOS initiative – more on guidance for analysis of observational studies** Organisers: Georg Heinze, Austria and Willi Sauerbrei, Germany	
17.20	Closing ISCB41 & Invitation ISCB42			

Detailed Programme

Sunday, 23 August 2020

Course: **SC1: Regression Modelling Strategies**

Instructor: **Frank Harrell Jr.**, *Vanderbilt University School of Medicine, United States*

13.00 - 19.00

13.00 - 14.30	Part 1: Chapters 1 - 2
	Introduction Hypothesis testing, estimation, prediction, classification Planning for modeling General Aspects of Fitting Regression Models Notation and parameters Relaxing linearity assumption Avoiding categorization Splines Machine learning vs. statistical models
14.50 - 16.20	Chapters 3 - 4
	Multivariable Modeling Strategies Prespecifying predictor complexity Perils of variable selection Overfitting and shrinkage Data reduction Comparing models Improving practice Overall strategies Describing the Model
16.40 - 17.40	Part 3: Chapters 4 - 6
	Resampling and Validating the Model Brief Overview of R Software Introduction to Longitudinal models
18.00 - 19.00	Part 4: Chapters 6 - 9
	Case Studies

Course: **SC2: Absolute Risk Methods and Applications in Clinical management and Public Health**

Instructors: **Ruth Pfeiffer**, *Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, HHS, United States*
Mitchell H. Gail, *Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, HHS, United States*

13.00 - 19.00	
13.00 - 13.45	Introduction
13.45 - 14.30	Survival analysis and competing risks without covariate a. Survival ideas including hazard, Kaplan-Meier; time scale b. Define cause-specific hazard and absolute risk without covariates c. Gaynor formula
14.30 - 15.00	Estimating absolute risk from cohorts a. Fine Gray b. Cause-specific models c. Nested cc and case-cohort
15.25 - 15.30	Software for model building
15.30 - 16.00	Estimation by combining cohort or case-control data with registry data
16.00 - 17.00	Criteria for evaluation a. Calibration b. Accuracy c. AUC d. PCF/PNF e. Loss
17.30 - 18.00	Comparing two models
18.00 - 18.45	Applications a. Counseling b. Public health c. How good do models need to be: implications for prevention
18.45 - 19.00	Questions / review or special topic

Course: **SC3: Understanding and Tackling Measurement Error a Review of Modern Practical Methods**

Instructors: **Pamela Shaw**, *University of Pennsylvania Perelman School of Medicine, United States*
Ruth Keogh, *London School of Hygiene and Tropical Medicine, United Kingdom*

13.00 - 19.00

13.00 - 13.05	Preliminaries
13.05 - 13.30	Introduction to measurement error and its effects
13.30 - 14.00	Regression Calibration
14.00 - 14.30	Simulation Extrapolation (SIMEX)
14.30 - 15.15	Practical worked examples + R practicum: Session 1
15.45 - 16.15	Special considerations for categorical exposures
16.15 - 16.45	Bayesian Methods for measurement error correction
16.45 - 17.30	Practical worked examples + R practicum: Session 2
18.00 - 18.25	Outcome and other types of measurement error
18.25 - 18.50	Considerations for study design
18.50 - 19.00	Closing remarks
19.00	Class End

Course: **SC4: Analysis of High-Dimensional Data**

Instructors: **Jörg Rahnenführer**, *TU Dortmund, Germany*
Michel Lang, *TU Dortmund, Germany*
Lara Lusa, *University of Ljubljana, Slovenia*

13.00 - 19.00

13.00 - 13.10	Introduction and overview of the short course
13.10 - 14.40	Part I: Statistical methods for the analysis of HDD
14.40 - 14.50	Discussion
15.20 - 16.30	Part II: Identification of informative variables and multiple testing for HDD
16.30 - 16.40	Discussion
17.10 - 18.40	Part III: Prediction for HDD with the mlr3 package
18.40 - 19.10	Discussion

Course: **SC5: Early Phase Clinical Trials**

Instructors: **Alexia Iasonos**, *Memorial Sloan Kettering Cancer Center, New York, United States*
John O'Quigley, *Department of Statistical Science, University College London, United Kingdom*

13.00 - 18.45	
13.00 - 13.15	Welcome and Introduction
13.15- 14.00	Introduction to general Phase I designs
14.00 - 15.00	General background to dose finding approaches. Model based, interval and semi-parametric methods
15.15 - 15.45	Statistical theory (convergence, coherence principles, model parameterization): Application to design
16.15 - 17.00	Dose expansion cohorts: theory and practice
17.00 - 17.45	More complex designs (patient heterogeneity and bridging, toxicity and efficacy, Phase I/II, combination therapies)
18.00 - 18.30	Protocol development, practical and logistical concerns such as accrual rates and endpoint observation, review of case studies, software
18.30 - 19.00	Open flor

19.30 - 20.30	**Students' Get Together**

14.00 **Conference Opening:**

Welcome by: Krystyna Szafraniec & Maciej Polak (LOC Chairs)
Professor Tomasz Grodzicki, *Vice Rector of the Jagiellonian University for Medical College*
Michał Abrahamowicz, SPC Chair
Vana Sypsa, President ISCB

14.30 - 15.30 President's Invited Speaker:

Miguel Hernán Learning what works: Causal inference from observational data

Harvard T.H. Chan School of Public Health, USA

Organiser: **Vana Sypsa**, *Medical School, National and Kapodistrian University of Athens, Greece*

Session: **IS01: Bayesian modelling of chronic & infectious diseases: Reverend Bayes going anti-viral**

Organisers / Chairs: **Dani Gamerman**, *UFMG, Brazil*
Maia Lesosky, *University of Cape Town, South Africa*

15.50 - 17.20

IS01.1	15.50 - 16.20	Paula Moraga	Bayesian Geospatial Models for Tropical Disease Mapping
IS01.2	16.20 - 16.50	Jony Arrais Pinto Jr.	Point pattern analysis of cerebrovascular deaths in Rio de Janeiro with spatially varying covariate effects
IS01.3	16.50 - 17.20	Alexandra Schmidt	A Poisson-multinomial spatial model for the number of cases of vector-borne diseases in Rio de Janeiro, Brazil

Session: **IS02: Tackling the challenges of imperfections in longitudinal and survival data: missingness, measurement error and irregular sampling**

Organisers / Chairs: **Laurence S. Freedman**, *Gertner Institute for Epidemiology, Israel*
Ruth Keogh, *LSHTM, United Kingdom*

15.50 - 17.20

IS02.1	15.50 - 16.20	Eleni-Rosalina Andrinopoulou	Challenges and opportunities of combined analysis of multiple outcomes in longitudinal studies
IS02.2	16.20 - 16.50	Havi Murad	Missing time-dependent covariate values for the Cox model - multiple imputation versus joint models approach
IS02.3	16.50 - 17.20	Shaun Seaman	Structural nested cumulative survival time models to adjust for time-dependent confounding

Session: **OC01: Clinical Trials Methodology 1**

Chair: **Colin Begg**, *Memorial Sloan Kettering Cancer Center, United States*

15.50 - 17.20

OC01.1	15.50 - 16.08	Heinz Schmidli	Use of external control information in clinical trials
OC01.2	16.08 - 16.26	Babak Choodari-Oskooei	Treatment selection in multi-arm multi-stage (MAMS) designs: an application to primary postpartum haemorrhage
OC01.3	16.26 - 16.44	David Robertson	Controlling type I error rates in multi-arm clinical trials: a case for the false discovery rate
OC01.4	16.44 - 17.02	Alexandra Graf	Optimized multiple testing procedures for nested sub-populations based on a continuous biomarker
OC01.5	17.02 - 17.20	Szymon Urbas **StCA Winner**	Inhomogeneous Poisson-gamma recruitment model for multi-centre clinical trials with model-averaged predictions

Session: **OC02: Causal Inference 1**

Chair: **Robert Platt**, *McGill University, Canada*

15.50 - 17.20

OC02.1	15.50 - 16.08	Sander Roberti	Bias correction for estimates from linear excess relative risk models in small studies
OC02.2	16.08 - 16.26	Hege Michiels	Estimation of the treatment effect in case of switching to rescue medication in a randomised clinical trial
OC02.3	16.26 - 16.44	Sharon Lutz	Caution Against Examining the Role of Reverse Causality in Mendelian
OC02.4	16.44 - 17.02	Daniela Schlueter	Understanding pathways to health inequalities in cystic fibrosis using UK registry data
OC02.5	17.02 - 17.20	Johan Steen	Using health records to estimate mortality attributed to acute kidney injury: challenges and reflections

Session: **OC03: Multi-State models**

Chair: **Jeremy Taylor**, *University of Michigan, United States*

15.50 - 17.20

OC03.1	15.50 - 16.08	Charlotte Castel	Multi-state Markov model for estimating HIV incidence from HIV surveillance data in France, 2008-2018
OC03.2	16.08 - 16.26	Jan Feifel	Assessing the effect of time-dependent exposures on time-to-event endpoints with economical sampling designs
OC03.3	16.26 - 16.44	Jonathan Broomfield	Assessing and Relaxing the Markov Assumption in the Illness-Death Model
OC03.4	16.44 - 17.02	Pierre Joly	Regression model for epidemiological indicators: an alternative to the pseudo-values
OC03.5	17.02 - 17.20	Jerome Lambert	Analysis of recurrent hospital acquired infections and hospital discharge using marginal regression modelling

Session: **OC04: Individualised Risk Prediction 1**

Chair: **Torben Martinussen**, *University of Copenhagen, Denmark*

15.50 - 17.20

OC04.1	15.50 - 16.08	Laure Wynants	Developing risk models for multicenter data: a simulation study comparing regression techniques
OC04.2	16.08 - 16.26	Menelaos Pavlou	Sample size calculation for external validation of risk models
OC04.3	16.26 - 16.44	Wenyue Zhu **StCA Winner**	Spatial predictive statistical model: detecting the need for treatment from retinal imaging data
OC04.4	16.44 - 17.02	Solon Karapanagiotis	Tailored Bayesian risk prediction modelling
OC04.5	17.02 - 17.20	Francesca Gasperoni	Personalised screening schedules for optimal prevention of cardiovascular disease

Session: **OC05: Omics & Survival Analysis**

Chair: **Jaroslaw Harezlak**, *Indiana University, United States*

15.50 - 17.20

OC05.1	15.50 - 16.08	Mirko Signorelli	How to predict a survival outcome using longitudinal and high-dimensional omic data
OC05.2	16.08 - 16.26	Agnieszka Król	Improving the evaluation of COPD exacerbation treatment effects by accounting for treatment discontinuations
OC05.3	16.26 - 16.44	Laura Villain	Generalized Berk Jones test for Gene Set Analysis of time to event transcriptomics applied to recurrence risk
OC05.4	16.44 - 17.02	Krista Fischer	The controversial concept of biological age
OC05.5	17.02 - 17.20	Lars LJ Van Der Burg	Impact of genetic markers on time-to-event outcomes taking into account ambiguous measurements

Session: **OC06: Novel Applications of Survival Analysis**
Chair: **Georg Heinze**, *Medical University of Vienna, Austria*

15.50 - 17.20

OC06.1	15.50 - 16.08	Ping Liu	Estimating conditional cumulative incidence in the presence of an internal time-varying exposure
OC06.2	16.08 - 16.26	Nina Ruzic Gorenjec	Random Cancers as Supported by Registry Data
OC06.3	16.26 - 16.44	Michelle Samuel	Comparison of methods to prevent immortal time bias for frequently recurring outcomes
OC06.4	16.44 - 17.02	Siyana Kurteva	Flexible modelling of recency-weighted cumulative opioid exposure provides new insights on the opioid safety
OC06.5	17.02 - 17.20	Sofia Kanavou	Defining impaired olfaction in Parkinson's disease: assessing agreement between four published methods

Session: **IS03: Can Ying dance with Yang?: Combining evidence from RCT's and Pharmaco-epidemiology population-based studies of drug effects**
Organiser / Chair: **Jessica Franklin**, *Brigham and Women's Hospital / Harvard Medical School, United States*

17.40 - 19.10

IS03.1	17.40 - 18.10	Sara Lodi	A three-step approach to compare randomized trials and observational studies
IS03.2	18.10 - 18.40	Jonathan Sterne	How risk of bias assessments can inform comparisons between RCTs and observational studies
IS03.3	18.40 - 19.10	Shirley Wang	Unambiguously reported and replicable real-world evidence to support decision-making

Session: **IS04: New Frontiers: Big Data & Functional Data Analysis**
Organiser / Chair: **Yingying Fan**, *USC, United States*
Chair: **Michal Abrahamowicz**, *McGill, Montreal, Canada*

17.40 - 19.10

IS04.1	17.40 - 18.10	James Ramsay	Statistics on curved manifolds: The next frontier
IS04.2	18.10 - 18.40	Timothy Cannings	Classification with imperfect training labels
IS04.3	18.40 - 19.10	Yingying Fan	An empirical Bayes shrinkage method for functional data

Session: **OC07: Clinical Trials Methodology 2**
Chair: **Mimi Kim**, *Albert Einstein College of Medicine, United States*

17.40 - 19.10

OC07.1	17.40 - 17.58	Ruitao Lin	Time-to-event model-assisted designs to accelerate and optimize early-phase immunotherapy trials
OC07.2	17.58 - 18.16	Megan Othus	Dynamic balancing in platform trials
OC07.3	18.16 - 18.34	Maximilian Pilz	Optimal unplanned recalculation in adaptive two-stage designs
OC07.4	18.34 - 18.52	Caroline Kristunas	An evaluation of the use of covariate constrained randomisation for stepped-wedge cluster randomised trial
OC07.5	18.52 - 19.10	Alessandra Serra	Multi-Arm Multi-Stage Design for Ordered Treatments

Session: **OC08: Causal Inference 2**

Chair: **Saskia Le Cessie**, *Leiden University, the Netherlands*

17.40 - 19.10

OC08.1	17.40 - 17.58	Lucy Teece	Comparing and expanding potential-outcome models for average treatment effect estimation in cardio-oncology
OC08.2	17.58 - 18.16	Paul Meyvisch	On the relationship between association and surrogacy when both the surrogate and true endpoint are binary.
OC08.3	18.16 - 18.34	Wenxi Yu	Effect Estimation By A Boosted Doubly Robust Method
OC08.4	18.34 - 18.52	Jonas Béal	Causal inference with multiple versions of treatment and application to personalized medicine
OC08.5	18.52 - 19.10	Christiana Drake	Propensity Weighting in the Estimation of Direct Effects

Session: **OC09: Survival Analysis 1**

Chair: **Richard. J. Cook**, *University of Waterloo, Canada*

17.40 - 19.10

OC09.1	17.40 - 17.58	Georgios Kantidakis	Prediction models in bone sarcoma - a simulation study to compare Cox models with Survival Neural Networks
OC09.2	17.58 - 18.16	Aurelien Belot	An excess hazard regression model based on a general structure and with individual frailty
OC09.3	18.16 - 18.34	Jennifer Hellier	Estimating survival functions using growth curve modelling with thresholds
OC09.4	18.34 - 18.52	Camille Maringe	Prediction of cancer survival for cohorts of patients most recently diagnosed using multi-model inference
OC09.5	18.52 - 19.10	Aimilia Exarchakou	Use of pseudo-observations to quantify the impact of cancer due to socioeconomic inequalities

Session: **OC10: Personalised Medicine 1**

Chair: **Helene Jacqmin-Gadda**, *Inserm, France*

17.40 - 19.10

OC10.1	17.40 - 17.58	Holly Jackson	Rare Disease Trials: Using a continuous biomarker to allocate patients in a response-adaptive clinical trial
OC10.2	17.58 - 18.16	Paweł Morzywołek	Assessing the optimal time to start renal replacement therapy using Causal Machine Learning
OC10.3	18.16 - 18.34	Janie Coulombe	Simple optimal adaptive treatment strategies in studies subject to informative monitoring times
OC10.4	18.34 - 18.52	Maren Hackenberg	Learning individual trajectories from biomedical time-series data with deep generative models
OC10.5	18.52 - 19.10	Michail Belias	Spline-based modelling to investigate treatment effect differences in IPD-MA: a gentle introduction

Session: **OC11: Methods Validation and Comparisons**

Chair: **Aris Perperoglou**, *AstraZeneca, United Kingdom*

17.40 - 19.10

OC11.1	17.40 - 17.58	Stephanie Wied	Evaluation of the Fill-it-up design to use historical controls in randomized controlled clinical trials
OC11.2	17.58 - 18.16	Matthieu Faron	Indirect treatment effect estimation in presence of modifier in Network Meta Analyses of time-to-event data
OC11.3	18.16 - 18.34	Katy Morgan	What happens to sample size predictions for series of N-of-1 trials when outcomes aren't normally distributed?
OC11.4	18.34 - 18.52	Md Ariful Islam Sanim	On estimating average treatment effect of rare exposure: sample size consideration and bias correction
OC11.5	18.52 - 19.10	Małgorzata Ćwiklinska-Jurkowska	Modeling improvement of performance for combined procedures in classification of microarray data sets

Session: **OC12: Meta-analysis 1**

Chair: **Marianne Huebner**, *Michigan State University, United States*

17.40 - 19.10

OC12.1	17.40 - 17.58	Ralf Bender	Exploring the empirical distribution of tau from IQWiG reports to inform Bayesian meta-analyses
OC12.2	17.58 - 18.16	Loic Darchy	Multi-Arm Multi-Stage (MAMS) Designs in Clinical Development: A comparative review
OC12.3	18.16 - 18.34	Theodoros Evrenoglou	Network meta-analysis of rare events using penalized likelihood regression
OC12.4	18.34 - 18.52	Silvia Metelli	Detecting outlying studies in Network Meta-analysis using Bayes factors
OC12.5	18.52 - 19.10	Konstantina Chalkou	A Prediction Model of Heterogeneous Treatment Effects Using Randomized and Observational Data

19.30 - 20.30 Welcome Get Together, Conference AWARDS Ceremony & Kraków virtual tour

Session: **IS05: Causal Inference vs. Unmeasured Confounding: Let's face the Dragon**

Organiser: **Eric Tchetgen Tchetgen**, *University of Pennsylvania, United States*
Chair: **John Carlin**, *Murdoch Children's Research Institute & University of Melbourne, Australia*

12.10 - 13.40

IS05.1	12.10 - 12.40	Vanessa Didelez	On the (In)Validity of Multiple Instruments for Mendelian Randomisation Studies
IS05.2	12.40 - 13.10	Torben Martinussen	Competing risk data and unmeasured confounding
IS05.3	13.10 - 13.40	Qingyuan Zhao	Using sparsity to overcome unmeasured confounding: two examples

Session: **IS13: Statistical challenges in Covid-19 research**

Organiser: **Krista Fisher**, *University of Tartu, Estonia*

12.10 - 13.40

IS13.1	12.10 - 12.40	Els Goetghebeur	Emulating trials with observational data and effect of hydroxychloroquine in hospitalized COVID-19 patients
IS13.2	12.40 - 13.10	Christopher Jackson	COVID-19: Estimating the severity of the epidemic from hospitalised cohorts in the UK
IS13.3	13.10 - 13.40	Tom Britton	Basic reproduction numbers, effective reproduction numbers and herd immunity

Session: **OC13: Clinical Trials Methodology 3**

Chair: **Stephen Senn**, *Statistical Consultant, United Kingdom*

12.10 - 13.40

OC13.1	12.10 - 12.28	Shinjo Yada	An adaptive design of phase I/II clinical trials for precision medicine using oncogene information
OC13.2	12.28 - 12.46	Jessica Kasza	Information content of cluster-period cells in stepped wedge designs with unequal cell sizes
OC13.3	12.46 - 13.04	Anneke Grobler	How to handle drop-out in palliative care trials
OC13.4	13.04 - 13.22	Julia Niewczas	GO/NO-GO interim decision making incorporating short- and long-term endpoints
OC13.5	13.22 - 13.40	Andrew Vincent	Empirically derived prior using ClinicalTrials.gov for a Bayesian RCT.

Session: **OC14: Survival Analysis 2**

Chair: **Zdenek Valenta**, *Institute of Computer Science, Czech Academy of Sciences, Czech Republic*

12.10 - 13.40

OC14.1	12.10 - 12.28	Regina Stegherr	Survival analysis for AdVerse events with Varying follow-up times - The empirical study of the SAVVY project
OC14.2	12.28 - 12.46	Audinga-Dea Hazewinkel	Prediction models with survival data: comparing machine learning to the Cox proportional hazards model
OC14.3	12.46 - 13.04	Il Do Ha	Semi-parametric Copula Modelling Approaches for Clustered Survival Data
OC14.4	13.04 - 13.22	Tasneem Fatima Alam **CASc Winner**	Bias reduction and solution to separation in the AFT models for small or rare event survival data
OC14.5	13.22 - 13.40	Martin Posch	Testing procedures for the comparison of multiple characteristics of different survival functions

Session: **OC15: Individualised Risk Prediction 2**

Chair: **Mark van de Wiel**, *Free University, Amsterdam, the Netherlands*

12.10 - 13.40

OC15.1	12.10 - 12.28	Ben Van Calster	Risk prediction for ordinal outcomes: calibration and the proportional odds assumption
OC15.2	12.28 - 12.46	J. Hoogland	Flexible parametric survival modeling by means of penalized maximum likelihood
OC15.3	12.46 - 13.04	Rumana Omar	Evaluation of sample size requirements for the development of risk prediction models
OC15.4	13.04 - 13.22	Isao Yokota	Regression model for personalized chance of longer survival using pseudo-observations
OC15.5	13.22 - 13.40	Pablo Gonzalez Ginestet	Stacked Inverse Probability of Censoring Weighted Bagging: A Case Study in The InfCare HIV Register

Session: **OC16: Functional Data Analysis**

Chair: **Timothy Cannings**, *University of Edinburgh, United Kingdom*

12.10 - 13.40

OC16.1	12.10 - 12.28	Francesca Ieva	Modeling the effect of dynamic covariates on time-to-event processes via Functional Data Analysis
OC16.2	12.28 - 12.46	Marta Spreafico	Including dynamic covariates in survival models via Functional Data Analysis: an application to osteosarcoma
OC16.3	12.46 - 13.04	Mohammad Fayaz	Functional Random Forest for Mixed Data
OC16.4	13.04 - 13.22	Stanislav Katina	Comparism of different statistical models used in shape index calculation on human face
OC16.5	13.22 - 13.40	Nicolo Margaritella	Parameter clustering in Bayesian functional data models of neuroscientific recordings

Session: **OC17: Omics & Genetics**

Chair: **Zoltán Kutalik**, *University of Lausanne, Switzerland*

12.10 - 13.40

OC17.1	12.10 - 12.28	Munshi Imran Hossain	Identification of Genetic Biomarkers using Models of Network Interactions
OC17.2	12.28 - 12.46	Pedro Afonso	Efficiently analyzing big data with Bayesian joint models for longitudinal and time-to-event data
OC17.3	12.46 - 13.04	Rodolphe Thiebaut	Generalized random forests for high-dimensional longitudinal data
OC17.4	13.04 - 13.22	Agus Salim	RUV-NB: Removing Unwanted Variation from Single-Cell RNA-seq Data
OC17.5	13.22 - 13.40	Marine Gauthier	Conditional cumulative distribution function estimation for differential expression analysis in scRNA-seq data

Session: **OC18: Longitudinal Data**

Chair: **Tomasz Burzykowski**, *Hasselt University, Belgium*

12.10 - 13.40

OC18.1	12.10 - 12.28	Abigail Burdon	Joint Modelling of Longitudinal and Survival Data Applied to Group Sequential Trials
OC18.2	12.28 - 12.46	Nicole Erler	Dealing with missing values in multivariate joint models for longitudinal and survival data
OC18.3	12.46 - 13.04	Zsolt Lang	A new type of generalized linear mixed models with linear predictor linked to marginal mean
OC18.4	13.04 - 13.22	Denis Rustand	Joint modelling of a semicontinuous longitudinal biomarker and a terminal event
OC18.5	13.22 - 13.40	Roger Marshall	Prognostic stratification: using Boolean classification to identify ordinal levels of risk

Session: **IS06: When non-inferiority trials are superior?**

Organiser: **Toshiro Tango**, *Center for Medical Statistics, Tokyo, Japan*
Organiser / Chair: **Lisa Hampson**, *Novartis Pharma AG, Switzerland*

14.00 - 15.30

IS06.1	14.00 - 14.30	Eisuke Hida	Can we evaluate assay sensitivity in a non-inferiority trial ?
IS06.2	14.30 - 15.00	Mimi Kim & Melissa Fazzari	Subgroup analyses in non-inferiority trials
IS06.3	15.00 - 15.30	Andrew Copas	Non-inferiority trials to directly inform treatment recommendations

Session: **IS07: Can Cox model survive in the 21st century survival analysis?**

Organiser: **Maja Pohar Perme**, *Universtiy of Ljubljana, Slovenia*

14.00 - 15.30

IS07.1	14.00 - 14.30	Per Kragh Andersen	50 years with the Cox model - 80 more to come?
IS07.2	14.30 - 15.00	Richard J. Cook	Challenges in causal inference with survival and life history processes
IS07.3	15.00 - 15.30	Bianca de Stavola	The Cox model in the era of target trials emulations

Session: **OC19: Epidemiology: Methods & Applications**

Chair: **Andrzej Pająk**, *Jagiellonian University Medical College, Poland*

14.00 - 15.30

OC19.1	14.00 - 14.18	Qi Ge	A Robust discriminant framework based on qEEG on the diagnosis of Alzheimer's disease
OC19.2	14.18 - 14.36	Jeremie Lespinasse	Modeling of multivariate longitudinal markers in a latent disease timescale:application in Alzheimer's disease
OC19.3	14.36 - 14.54	Dominik Grathwohl	Optimal pooling of blood samples
OC19.4	14.54 - 15.12	Matthew Smith	Survival inequalities in non-Hodgkin lymphoma: excess hazard models with joint modelling multiple imputation
OC19.5	15.12 - 15.30	Frederique Chammartin	Risk of non-AIDS defining and AIDS defining malignancies of early versus delayed initiation of antiretroviral

Session: **OC20: Multivariable Model Building**

Chair: **Anne-Laure Boulesteix**, *LMU Munich, Germany*

14.00 - 15.30

OC20.1	14.00 - 14.18	Ihor Malyk	Optimal separation of normal distributed samples. Application for anxiety and alexithymia scales in cardiology
OC20.2	14.18 - 14.36	Lucinda Archer	Dichotomisation of continuous outcomes in prediction model research: (not such) a bad idea?
OC20.3	14.36 - 14.54	Sergio Sabroso-Lasa	Evaluating the role of correlations among markers in prediction models
OC20.4	14.54 - 15.12	Andreas Bender	Piece-wise exponential (additive mixed) modelling tools for survival analysis
OC20.5	15.12 - 15.30	Aris Perperoglou	Common spline bases for regression models in practice: a comparative simulation study

Session: **OC21: Bayesian Methods & Omics**

Chair: **Rodolphe Thiebaut**, *Bordeaux University / Inserm, France*

14.00 - 15.30

OC21.1	14.00 - 14.18	Sarah-Laure Rincourt	Nonparametric bayesian sparse factor analysis to model individual heterogeneity in gene-expression data
OC21.2	14.18 - 14.36	Sonja Zehetmayer	Adaptive filtering increases the power but not the False Discovery Rate in RNA-seq experiments
OC21.3	14.36 - 14.54	Alma Revers	A Bayesian approach for identifying gut-microbiome - diet associations
OC21.4	14.54 - 15.12	Miguel Rodo	Identifying rare cell types using an outlier-detection approach

Session: **OC22: Missing Data 1**

Chair: **Sven Ove Samuelsen**, *University of Oslo, Norway*

14.00 - 15.30

OC22.1	14.00 - 14.18	Shahab Jolani	Hierarchical imputation of categorical variables with both systematically and sporadically missing values
OC22.2	14.18 - 14.36	Orlagh Carroll **StCA Winner**	Combining bootstrap with multiple imputation to assess prognostic models with missing covariate data
OC22.3	14.36 - 14.54	Grigorios Papageorgiou	An alternative characterization of Missing at Random in Shared Parameter Models
OC22.4	14.54 - 15.12	Halehsadat Nekoee Zahraei	Cluster Analysis in Incomplete Data
OC22.5	15.12 - 15.30	Jonathan Bartlett	Bootstrap Inference for Multiple Imputation under Uncongeniality and Misspecification

Session: **OC23: Bayesian Methods**

Chair: **Alexandra Schmidt**, *McGill University, Canada*

14.00 - 15.30

OC23.1	14.00 - 14.18	Jonathan Roux	Bayesian reconstruction of inter-hospital propagation chains of antimicrobial resistant bacteria
OC23.2	14.18 -14.36	Md Mynul Islam **CASc Winner**	A spatial multilevel model using conditional autoregressive processes
OC23.3	14.36 -14.54	Steffen Ballerstedt	A Probability of Success Approach in Late Stage Drug Development
OC23.4	14.54 -15.12	Paweł Wiczling	The Bayesian population pharmacokinetic analysis of dexmedetomidine and clonidine in Stan and Torsten
OC23.5	15.12 -15.30	Erik van Zwet	Corpus based priors for calibrated Bayesian inference

Session: **OC24: Meta-analysis 2**

Chair: **Laure Wynants**, *Maastricht University, Netherlands*

14.00 - 15.30

OC24.1	14.00 - 14.18	Gerta Rücker	Model selection for component network meta-analysis in disconnected networks
OC24.2	14.18 - 14.36	Virginia Chiocchia	Framework for evaluating reporting bias in network meta-analysis
OC24.3	14.36 - 14.54	Tasnim Hamza	Bayesian dose-response network meta-analysis
OC24.4	14.54 - 15.12	Hugo Pedder	Connecting the dots: Linking disconnected networks using dose-response Model-Based Network Meta-Analysis

Session: **IS08: New developments in Longitudinal data analysis**

Organiser: **Andrzej Gałecki**, *University of Michigan, United States*

15.50 - 17.20

IS08.1	15.50 - 16.20	Helene Jacqmin-Gadda	Challenges in longitudinal analysis of pre-dementia cognitive decline
IS08.2	16.20 - 16.50	Tomasz Burzykowski	Design of longitudinal trials
IS08.3	16.50 - 17.20	Jeremy Taylor	Evaluation of predictive model performance of an existing model in the presence of missing data

Session: **IS09: Progress in providing analytic guidance for observational studies: the STRATOS initiative**

Organisers / Chairs: **Laurence Freedman**, *Gertner Institute for Epidemiology, Israel*
Willi Sauerbrei, *Institute of Medical Biometry and Statistics, Medical Center - University of Freiburg, Germany*

15.50 - 17.20

IS09.1	15.50 - 16.20	Saskia Le Cessie	Bridging the gap between causal inference and survival analysis: a censored edition
IS09.2	16.20 -16.50	Jörg Rahnenführer	Statistical and machine learning techniques: Which help in patient care and medical research?
IS09.3	16.50 -17.20	Georg Heinze & Marianne Huebner	Regression without regrets: data screening is needed before modeling

Session: **OC25: Clinical Trials Methodology 4**

Chair: **Sara Lodi**, *Boston University School of Public Health, United States*

15.50 - 17.20

OC25.1	15.50 - 16.08	Jian-Lun Xu	A new method to estimate the predictive value in a repeated screening program
OC25.2	16.08 - 16.26	Tra My Pham	Analysis of binary composite outcomes with partially observed components: a comparison of strategies
OC25.3	16.26 - 16.44	Leonhard Held	The harmonic mean chi-squared test to substantiate scientific findings
OC25.4	16.44 - 17.02	Juan Guillermo Gonzalez Maffe	An Application of Bayesian Hierarchical Model in Platform Phase I Oncology Dose Escalation studies
OC25.5	17.02 - 17.20	Damjan Manevski	Confidence intervals for the Mann-Whitney test

Session: **OC26: Causal Inference for Survival Analysis**

Chair: **Els Goetghebeur**, *Ghent University, Belgium*

15.50 - 17.20

OC26.1	15.50 - 16.08	Arthur Chatton	G-computation and Inverse Probability Weighting for time-to-event analyses
OC26.2	16.08 - 16.26	Suzy Van Sanden	Adjusting the estimate of overall survival benefit for treatment switching in a randomized clinical trial
OC26.3	16.26 - 16.44	Ruth Keogh	Emulation of target trials to investigate causal effects of lung transplant on survival in cystic fibrosis
OC26.4	16.44 - 17.02	Ellie John	Comparing instrumental variables approaches for time to event data
OC26.5	17.02 - 17.20	Mark Clements	„Plug-In" Estimation for Parametric and Penalised Multi-State Markov Models

Session: **OC27: Survival Analysis 3**

Chair: **Terry Therneau**, *Mayo Clinic, United States*

15.50 - 17.20

OC27.1	15.50 - 16.08	Reuben Adatorwovor	Relaxing the Independence Assumption in Relative Survival Analysis: A Parametric Approach
OC27.2	16.08 - 16.26	Bernard Rachet	Building robust specific life tables using multidimensional penalized splines
OC27.3	16.26 - 16.44	Helmut Küchenhoff	Flexible estimation of complex, lagged, cumulative effects in the context of competing risks
OC27.4	16.44 - 17.02	Manuela Quaresma	Flexible Bayesian hierarchical excess hazard models using low-rank thin plate splines
OC27.5	17.02 - 17.20	Michal Abrahamowicz	Adapting SIMEX to correct for bias due to interval-censored data in survival analysis

Session: **OC28: Personalised Medicine 2**

Chair: **Ruth Pfeiffer**, *National Cancer Institute, Bethesda United States*

15.50 - 17.20

OC28.1	15.50 - 16.08	Svetlana Cherlin	Developing and testing high-efficacy patient subgroups within a clinical trial using risk scores
OC28.2	16.08 - 16.26	Shaima Belhechmi	Favoring the hierarchy constraint of interactions in penalized survival models of randomized trials
OC28.3	16.26 - 16.44	Sarah Booth	Accounting for calibration drift due to improvements in baseline survival during prognostic model development
OC28.4	16.44 - 17.02	Bernard Francq	I'm not an average but a patient! A significant Holy Grail
OC28.5	17.02 - 17.20	Gabriela Czanner	Statistical and machine learning methods for uncertainty-informed decision referral

Session: **OC29: Measurement Errors**

Chair: **Pamela Shaw**, *University of Pennsylvania, United States*

15.50 - 17.20

OC29.1	15.50 - 16.08	Valentijn M.T. De Jong	Adjusting for misclassification of a predictor in an individual participant data meta-analysis
OC29.2	16.08 - 16.26	Tanja Buelow	Various distributed data affected by lower limits of quantification: Point estimates vs. confidence intervals
OC29.3	16.26 - 16.44	Steve Ferreira Guerra	On Improved SIMEX for Covariate-Dependent Measurement Error with Continuous Covariates
OC29.4	16.44 - 17.02	Hana Sinkovec	Optimizing the use of dynamic quality indicators as covariates in regression models

Session: **OC30: Bayesian Methods for Clinical Trials**

Chair: **Paula Moraga**, *King Abdullah University of Science and Technology (KAUST), Saudi Arabia*

15.50 - 17.20

OC30.1	15.50 - 16.08	Samer Kharroubi	Modeling a preference-based index for EQ-5D and EQ-5D+Sleep using a Bayesian framework
OC30.2	16.08 - 16.26	Emma Gerard	Dose-finding Bayesian design for toxicity-schedule assessment using pharmacokinetics and pharmacodynamics
OC30.3	16.26 - 16.44	Hongchao Qi	Incorporating historical controls in clinical trials with longitudinal outcomes using the modified power prior
OC30.4	16.44 - 17.02	Haiyan Zheng	Bayesian sample size estimation for exploratory basket trials that enable borrowing of information
OC30.5	17.02 - 17.20	Yasin Desai	Prior elicitation of the efficacy of Methotrexate and Mycophenolate Mofetil in Juvenile Localised Scleroderma

17.40 - 19.10

Session	**PO1: Clinical Trials: Methods & Applications**
Session	**PO2: Longitudinal Data / Biomarkers**
Session	**PO3: Survival analysis / Machine Learning**

Wednesday, 26 August 2020

10.20 - 11.50

Session	**PO4: Big Data /Machine Learning**
Session	**PO5: Causal Inference**
Session	**PO6: Epidemiology: Methods and Applications / Electronic Health Records**

Session: **IS10: Individualised Risk Prediction & Omics**
Organize/Chair: **Krista Fischer**, *University of Tartu, Estonia*

12.10 - 13.40

IS10.1	12.10 - 12.40	Kristel Van Steen	Precision Medicine at the Interface of translational Science & Systems
IS10.2	12.40 - 13.10	Zoltán Kutalik	Causes and health consequences of altered transcriptome profiles
IS10.3	13.10 - 13.40	Mark van de Wiel	Learning from a lot: Empirical Bayes for high-dimensional model-based prediction

Session: **OC31: Clinical Trials Methodology 5**
Chair: **David W. Warne**, *Switzerland*

12.10 - 13.40

OC31.1	12.10 - 12.28	Lizzi Pitt	Using dynamic programming to find an optimal dose escalation scheme for a phase I trial
OC31.2	12.28 - 12.46	Kathrin Moellenhoff	Equivalence tests for binary efficacy-toxicity responses
OC31.3	12.46 - 13.04	Andrew Copas	Designing cluster randomised trials with unequal allocation of clusters or measurements
OC31.4	13.04 - 13.22	Anais Andrillon	Dose-finding designs with right censored endpoints. Evaluation through benchmark
OC31.5	13.22 - 13.40	Antonios Daletzakis	Estimation of the response rate in Simon's two stage design with early termination

Session: **OC32: Causal Inference 3**
Chair: **Vanessa Didelez**, *Leibniz Insitute for Prevention Research and Epidemiology – BIPS, Germany*

12.10 - 13.40

OC32.1	12.10 - 12.28	Zhiyue Huang	Inference on Time-to-Event Data after Propensity Score Matching
OC32.2	12.28 - 12.46	Margarita Moreno-Betancur	Data-adaptive methods for high-dimensional mediation analysis: Application to a tuberculosis vaccine trial
OC32.3	12.46 - 13.04	Camille Nevoret	Estimating causal effects from the large observational, with an application to multiple sclerosis
OC32.4	13.04 - 13.22	Mathilde Lefort	Long-term effect of first-line multiple sclerosis treatments using time-dependent propensity score matching
OC32.5	13.22 - 13.40	Emily Granger	Propensity score diagnostics: assessing the accuracy of a propensity-adjusted effect estimate

Session: **OC33: Survival Analysis 4**

Chair: **Per Kragh Andersen**, *Biostatistics, University of Copenhagen, Denmark*

12.10 - 13.40

OC33.1	12.10 - 12.28	Ali Shariati	Statistical Intervals for the MRL Function via Empirical Likelihood Using Right Censored Length-biased Data
OC33.2	12.28 - 12.46	Lars Bugiera	An Analytic Approach on the Cumulative Mean Function for Recurrent and Terminal Events
OC33.3	12.46 - 13.04	Reinhard Meister	Stratifying event-history data by a binary time-dependent covariate without conditioning on the future
OC33.4	13.04 - 13.22	Francesca Graziano	Discordant sampling designs from a longitudinal cohort: efficiency comparison
OC33.5	13.22 - 13.40	Lubomir Stepanek	A robust alternative for comparing time-event survival curves: initial simulations and cancer data analysis

Session: **OC34: Personalised Medicine 3**

Chair: **Shaun Seaman**, *University of Cambridge, United Kingdom*

12.10 - 13.40

OC34.1	12.10 - 12.28	Michael Seo	Predicting real world effectiveness of interventions using randomized and non-randomized data
OC34.2	12.28 - 12.46	Orestis Efthimiou	An evaluation of the c-statistic for benefit
OC34.3	12.46 - 13.04	Michael Lauseker	Estimating the common distribution of two right-censored potential treatment responses given a biomarker
OC34.4	13.04 - 13.22	Bernhard Haller	Confidence interval estimation for the changepoint of treatment stratification
OC34.5	13.22 - 13.40	Myra McGuinness	Statistical methods to assess dynamic treatment regimens in observational settings: a scoping review

Session: **OC35: Missing Data 2**

Chair: **Havi Murad**, *Gertner Institute for Epidemiology and Health Research Policy, Israel*

12.10 - 13.40

OC35.1	12.10 - 12.28	Marie-Cecile Fournier	Methods to deal with values under limit of detection or quantification
OC35.2	12.28 - 12.46	Sabbir Ahmed Hemo **StCA Winner**	Performance of Inverse Probability Weighting Compared to Multiple Imputation for Missing Binary Outcome in CRT
OC35.3	12.46 - 13.04	Edouard Bonneville	Missing covariates for competing risks models: an evaluation of available imputation methods
OC35.4	13.04 - 13.22	Vincent Audigier	Clustering with missing values: what about congeniality?

Session: **OC36: Dynamic Prediction**

Chair: **Eleni-Rosalina Andrinopoulou**, *Erasmus MC, Netherlands*

12.10 - 13.40

OC36.1	12.10 - 12.28	Dimitris Rizopoulos	Minimizing the Burden of Invasive Procedures via Personalized Scheduling
OC36.2	12.28 - 12.46	Zijing Yang	Analysis of dynamic restricted mean survival time for right-censored data
OC36.3	12.46 - 13.04	Harry Parr	Proposal of Clinical Dynamic Prediction Models to Predict Prognosis of Post-Treatment Prostate Cancer Patients
OC36.4	13.04 - 13.22	Rana Dandis	Joint modeling vs discriminant analysis for dynamic prediction based on longitudinal data: A simulation study
OC36.5	13.22 - 13.40	Itai Dattner	A statistical methodology for data-driven partitioning of infectious disease incidence into age-groups

Session: **IS11: Biostatistical inference in practice: moving beyond false dichotomies**

Organisers / Chairs: **John Carlin**, *Murdoch Children's Research Institute & University of Melbourne, Australia*
Jonathan Sterne, *University of Bristol, UK*

14.00 - 15.30

IS11.1	14.00 - 14.25	Laure Wynants	Dichotomania and other challenges: a perspective on principles and responsibilities
IS11.2	14.25 - 14.50	Colin Begg	Reflections on the strengths and weaknesses of significance testing
IS11.3	14.50 - 15.15	John Carlin	Beyond dichotomous thinking in the analysis and reporting of clinical trials: why, how and who's responsible?
IS11.4	15.15 - 15.30		Discussion, lead by J. Sterne

Session: **IS12: Efficient Designs & cutting-edge analyses**

Organisers / Chairs: **Mitchell H. Gail**, *National Cancer Institute, United States*
Sven Ove Samuelsen, *Department of Mathematics, University of Oslo, Norway*

14.00 - 15.30

IS12.1	14.00 - 14.30	Sven Ove Samuelsen	Use of auxiliary data for efficient analysis of two-phase survival studies
IS12.2	14.30 - 15.00	Pamela Shaw	Methods to address correlated covariate and time-to-event error in electronic health records
IS12.3	15.00 - 15.30	Ruth Keogh	Using negative controls to estimate causal effects of treatment in an entirely treated cohort

Session: **OC37: Clinical Trials Methodology 6**

Chair: **Andrew Copas**, *MRC Clinical Trials Unit at University College London, United Kingdom*

14.00 - 15.30

OC37.1	14.00 - 14.18	Isabelle Smith	The impact of misclassification in a clinical trial with discrete longitudinal outcomes
OC37.2	14.18 - 14.36	Heiko Goette	Adaptive design with interim decision for single arm trial with external controls or randomized trial
OC37.3	14.36 - 14.54	Cornelia Ursula Kunz	Extending the updating algorithm for the Graphical Approach
OC37.4	14.54 - 15.12	Philip Hougaard	EPAD: The European prevention of Alzheimer's dementia platform trial
OC37.5	15.12 - 15.30	Guillaume Mulier	Phase II trial with a trinomial efficacy-toxicity endpoint

Session: **OC38: Causal Inference 4**

Chair: **Bianca De Stavola**, *University College London, United Kingdom*

14.00 - 15.30

OC38.1	14.00 - 14.18	Maxime Léger	G-computation: a robust alternative for drawing causal inference when the positivity assumption does not hold
OC38.2	14.18 - 14.36	Rik van Eekelen	Correcting for selection bias when comparing separate treated and untreated cohorts
OC38.3	14.36 - 14.54	Linda Nab	Quantitative bias analysis for a misclassified confounder: marginal structural models vs conditional models
OC38.4	14.54 - 15.12	Klaus Groes Larsen	Treatment effects in the principal stratum of patients who would comply, if treated with a specific treatment
OC38.5	15.12 - 15.30	Stefanie Von Felten	Outcomes truncated by death in RCTs: a simulation study on the survivor average causal effect

Session: **OC39: AFT Modelling & Survival**

Chair: **Roch Giorgi**, *Aix-Marseille Univ. Inserm IRD / SESSTIM, France*

14.00 - 15.30

OC39.1	14.00 - 14.18	Moumita Chatterjee	Effects of covariates on alternately occurring recurrent events in Accelerated Failure Time models
OC39.2	14.18 - 14.36	Dip Das	Embedded Likelihood based Estimation under Accelerated Failure Time Models for Clustered Censored Data
OC39.3	14.36 - 14.54	Thomas Klausch	Bayesian estimation of three-state accelerated failure time models for interval censored surveillance outcomes
OC39.4	14.54 - 15.12	Menglan Pang	Flexible Modeling of Time-dependent and Non-linear covariate effects in Accelerated Failure Time model

Session: **OC40: Big Data**

Chair: **Jörg Rahnenführer**, *Technische Universität Dortmund, Germany*

14.00 - 15.30

OC40.1	14.00 - 14.18	Hugo Terrisse	Comparing predictive models for continous outcomes with proper scoring rules : a simulation study
OC40.2	14.18 - 14.36	Veronica Sciannameo	Deep Learning for predicting hospitalizations in elderly population using Electronic Health Records
OC40.3	14.36 - 14.54	Stephen Senn	Forward to the Past: The analysis of designed experiments and lessons for big data
OC40.4	14.54 - 15.12	Nicolas Ngo	Construction and comparison of feature selection methods using a new measure of separability: the gamma-metric
OC40.5	15.12 - 15.30	Jixian Wang	Propensity score matching and stratification using multiple sources without pooling individual level data

Session: **OC41: Machine Learning**

Chair: **Stanislav Katina**, *Masaryk University, Czech Republic*

14.00 - 15.30

OC41.1	14.00 - 14.18	Lilith Faucheux	Multi Objective Semi-Supervised Clustering with a right censored end point and in presence of missing data
OC41.2	14.18 - 14.36	Ido Azuri	Discriminating between healthy and malaria-diseased cytoskeleton red blood cells by fast-AFM and deep learning
OC41.3	14.36 - 14.54	Ilaria Gandin	Making time-series deep learning models more interpretable: a study in patients with cardiovascular disease
OC41.4	14.54 - 15.12	Nayma Hossain **CASc Winner**	Iterative Least-squares Regression with Censored Data: A Survival Ensemble of Learning Machine
OC41.5	15.12 - 15.30	Clemence Leyrat	Covariate adjustment in randomised trials: when and how?

Session: **OC42: Meta-analysis 3**

Chair: **Gerta Rücker**, *University of Freiburg, Germany*

14.00 - 15.30

OC42.1	14.00 - 14.18	Daniela Zöller	Federated regression modeling for selecting biomarkers under data protection constraints
OC42.2	14.18 - 14.36	Alessandra Meddis	Test of informative cluster size with survival data
OC42.3	14.36 - 14.54	Sylwia Bujkiewicz	Use of real world data in bridging disconnected networks of first and second lines of therapies
OC42.4	14.54 - 15.12	Pablo Verde	jarbes: An R Package for Combining Randomized and Observational Studies in Meta-Analysis
OC42.5	15.12 - 15.30	Iman Jaljuli	Quantifying replicability and consistency in systematic reviews

15.50 - 17.20 **ISCB Annual General Meeting**

17.40 - 19.10

Keynote Session: **Machine Learning vs. Statistical Modeling**

Organiser / Chair: **Michal Abrahamowicz**, SPC Chair, ISCB 2020
McGill, Montreal, Canada

17.46 - 18.22 **Trevor Hastie** **Predictive Models in Health Research**
Stanford University, USA

18.22 - 18. 58 **Frank Harrell Jr.** **Musing on Statistical Models vs. Machine Learning in Health Research**
Vanderbilt University School of Medicine, Vanderbilt, USA

18.58 - 19.10 Discussion

Session: **MS3: Advanced statistical methods for the analysis of high-dimensional medical data**

Organisers / Chairs: **Malgorzata Bogdan**, *University of Wroclaw, Poland*
Jaroslav Hazerlak, *Indiana University, United States*

11.30 - 17.20

MS3.1	11.30 - 11.40	Malgorzata Bogdan	Advanced technologies, large data and modern statistics in medicine
MS3.2	11.40 - 12.10	Armando Teixeira-Pinto	Recycling genome-wide association studies with Mendelian Randomisation
MS3.3	12.10 - 12.40	Florian Frommlet	Selecting predictive biomarkers from genomic data
MS3.4	12.40 - 13.10	Hernando Ombao	Modeling Dependence in Multivariate Time Series
MS3.5	13.10 - 13.40	Anna Dudek	(Dual-Frequency)-dependent dynamic functional connectivity analysis of visual working memory capacity
MS3.6	14.00 - 14.30	Julie Josse	Treatment effect estimation with missing attributes
MS3.7	14.30 - 15.00	Jaroslaw Harezlak	Regularization methods for multimodal brain imaging
MS3.8	15.00 - 15.30	Damian Brzyski	Revealing the common structure of brain networks by combining the nuclear and L1 norms
MS3.9	15.50 - 16.20	Marta Karas	Novel approach for precise walking cadence estimation from high - density tri - axial accelerometry data
MS3.10	16.20 - 16.50	John Kornak	Bayesian image analysis in transformed spaces (BITS) and the BIFS/BIWS Python packages
MS3.11	16.50 - 17.20	Grzegorz Rempala	Dynamical Survival Analysis for COVID-19 Predictions in Ohio

Session: **MS1: Statistical methods in pharmacoepidemiology and pharmacoeconomics**

Organiser / Chair: **Robert Platt**, *McGill University, Canada*

14.00 - 17.00

MS1.1	14.00 -14.20	Robert Platt	Overview lecture – state of the art and current problems
MS1.2	14.20 - 15.00	Jason Roy	Machine learning in pharmacoepidemiology
MS1.3	15.20 - 16.00	Thomas Debray	The importance of informative visit patterns
MS1.4	16.00 - 16.40	Menglan Pang	Targeted learning in pharmacoepidemiology
	16.40 - 17.00		Question period

Session: **MS2: STRATOS initiative - more on guidance for analysis of observational studies**

Organisers / Chairs: **Georg Heinze**, *Medical University of Vienna, Austria*
Willi Sauerbrei, *University of Freiburg, Germany*

14.00 - 17.15

	14.00 - 14.10	Georg Heinze	Introduction
MS2.1	14.10 - 14.35	Willi Sauerbrei	Outstanding issues in selection of variables and functional forms in multivariable analysis
MS2.2	14.35 - 15.00	Ben van Calster	Calibration of risk prediction models: making decisions with the lights on or off?
MS2.3	15.00 - 15.25	Helmut Kuechenhoff and Veronika Deffner	Measurement error and misclassification of variables in observational epidemiology - an overview
MS2.4	15.45 - 16.10	Maja Pohar Perme	Analysis of time-to-event for observational studies: Guidance to the use of intensity models
MS2.5	16.10 - 16.35	James Carpenter and Katherine Lee	Framework for the Treatment And Reporting of Missing data in Observational Studies: The TARMOS framework
MS2.6	16.35 - 17.00	Anne-Laure Boulesteix	A replication crisis in methodological research? On the design of comparison studies
	17.00 - 17.15		General discussion

Early Career Biostatisticians' Day

12.00 - 17.15

	12.00 - 12.15	Myra McGuinness	Welcome address

Session: **1**
Chair: **Michael Grayling**, *Newcastle University, United Kingdom*

	12.15 -13.00	Emily Karahalios	Working as a biostatistician – effectively planning, organizing and monitoring the progress of a project
	13.00 -13.15	Rushani Wijesuriya	Bachelors to PhD: taking the leap
	13.15 -13.30	Monsurul Hoq	A biostatistician's journey from aid work to academia
	13.30 -13.40		Discussion

Session: **2**
Chair: **Camila Olarte Parra**, *Ghent University, Belgium*

	14.00 -14.45	Havi Murad	Working in public health versus academia or pharmaceutical industry
	14.45 - 15.00	Mohammad Fayaz	From wrong analysis to the wrong decisions; the lift chart and the others
	15.00 - 15.15	Anna Schritz	Possible pitfalls when performing sample size calculation with online tools
	15.15 - 15.30		Discussion

Session: **3**
Chair: **Laure Wynants**, *Maastricht University, Belgium*

	16.00 - 16.15	Phillip Awodutire	Survival analyst in Nigeria: my experience
	16.15 - 17.00	Stephen Senn	The seven habits of highly effective statisticians
	17.00 - 17.15		Discussion

17.20 **Closing ISCB41 & Invitation ISCB42**

Tuesday, 25 August 2020

Session: **PO1: Clinical Trials: Methods & Applications**

Chairs: **Kinga Sałapa**, *Astra Zeneca Pharma, Poland*
Alexia Iasonos, *Memorial Sloan Kettering Cancer Center, New York, United States*

17.40 - 19.10

PO1.01	17.40 - 17.46	Maria Elisabeth Frey	Nonparametric Limits of Agreement for small to moderate sample sizes - a simulation study
PO1.02	17.46 - 17.52	Milan Geybels	Association rule mining to find correlations between adverse events and drugs in clinical trials
PO1.03	14.52 - 17.58	Monica Sofia Maria Persson	Using electronic health records to inform the development of clinical trials in an uncommon skin disease
PO1.04	17.58 - 18.04	Jacek Kopec	Development and preliminary validation of a novel risk and life expectancy calculator
PO1.05	18.04 - 18.10	Markus Pfirrmann	A practical approach to blinded sample size re-estimation in a randomized trial with a survival endpoint
PO1.06	18.10 - 18.16	Lamis Alamoudi	Testing treatment effects for Split-plot design for Clinical trials
PO1.07	18.16 - 18.22	Martin Reugels	Allocation bias in randomized clinical trials with binary response
PO1.08	18.22 - 18.28	Nan van Geloven	What happens when you switch the analysis scale in a non-inferiority trial?
PO1.09	18.28 - 18.34	Linda Sharples	Design and analysis of trials including subgroups defined by a biomarker
PO1.10	18.34 - 18.40	Jake Emmerson	Recommendations for analysing multi-arm non-inferiority trials and controlling family-wise error rate
PO1.11	18.40 - 18.46	Florian Halbeisen	Using routine data to conduct a pragmatic RCT to reduce antibiotic prescriptions in Swiss primary care
PO1.12	18.46 - 18.52	Julien St-Pierre	Comparison of penalized linear mixed models for high-dimensional data.
PO1.13	18.52 - 18.58	Elias Laurin Meyer	The evolution of Master Protocol Clinical Trial Designs - A Systematic Review
PO1.14	18.58 - 19.04	Derek Dinart	Sample size estimation for cancer randomized trials in the presence of heterogeneous data
PO1.15	19.04 - 19.10	Dawn Teare	The impact of primary outcomes measures with ceiling effects on non-inferiority designs

Session: **PO2: Longitudinal Data / Biomarkers**

Chairs: **Aleksander Owczarek**, *Medical University of Silesia, Sosnowiec, Poland*
Orlagh Carroll, LSHTM, *London, United Kingdom*

17.40 - 19.10

PO2.01	17.40 - 17.46	Celline Almeida-Brasil	Personalized Comparisons of Flare Rates after Hydroxychloroquine Tapering or Discontinuation in Systemic Lupus
PO2.02	17.46 - 17.52	Adam Korczyński	Imputation techniques for longitudinal data - applications in clinical trials
PO2.03	14.52 - 17.58	Brian Mosier	Estimation and Construction of CIs for the Cutoffs of Continuous Biomarkers in Trichotomous Settings
PO2.04	17.58 - 18.04	Natasa Kejzar	Three zone diagnostic decisions for numerical data
PO2.05	18.04 - 18.10	Wencan Zhu	A variable selection approach for highly correlated predictors in high-dimensional settings
PO2.06	18.10 - 18.16	Felix J. Clouth	Prediction of recovery in trauma patients using latent Markov models
PO2.07	18.16 - 18.22	Tore Wentzel-Larsen	Transition between living arrangements among elderly Norwegians 2011-2016. Use of the IPLOS register
PO2.08	18.22 - 18.28	Alice Bonomi	Coronary Heart Disease Risk Map according to Mediterranean Diet Score and Dietary variety
PO2.09	18.28 - 18.34	Lisanne Gitsels	Does having amblyopia affect school readiness and cognitive performance?
PO2.10	18.34 - 18.40	Rocio Aznar	Using multilevel models to estimate weight percentiles in twin pregnancies
PO2.11	18.40 - 18.46	Becki Green	Investigating the use of blood metabolites as biomarkers of early cognitive changes relevant to dementia
PO2.12	18.46 - 18.52	Zeynep Yavuz	Assessing the correlation and agreement of repeated measures: application to BIS data
PO2.13	18.52 - 18.58	Lamprini Syrogiannouli	Incorporating baseline outcome data in individual participant data meta-analysis of non-randomized studies
PO2.14	18.58 - 19.04	Willi Sauerbrei	Structured reporting to improve transparency of analyses in prognostic biomarker studies
PO2.15	19.04 - 19.10	Jeppe Christensen	Perturbation of proteomic biomarker data for sharing: gaining privacy while preserving utility?

Session: **PO3: Survival analysis / Machine Learning**

Chairs: **Janie Coulombe**, *McGill University, Montreal, Canada*
Steve Ferreira Guerra, *McGill University, Montreal, Canada*

17.40 - 19.10

PO3.01	17.40 - 17.46	Marina Machado	Pre-existing autoimmune disease and immune-related adverse events with checkpoint inhibitors in melanoma
PO3.02	17.46 - 17.52	Phillip Awodutire	Mixture and Non-Mixture Cure Rate Models Using the Beta Type I Generalized Half Logistic Distribution
PO3.03	14.52 - 17.58	Liping Tong	A New Method to Evaluate Post-Discharge Intervention on Risk of Adverse Events with Observational Data
PO3.04	17.58 - 18.04	Gustavo Soutinho	Estimation of the transition probabilities conditional on repeated measures in Multi-state models
PO3.05	18.04 -18.10	Marion Procter	Summarising a Global Interaction Test, Quantitative Interaction Terms and Qualitative Interaction Tests Summar
PO3.06	18.10 - 18.16	Anna Giczewska	Application of machine learning algorithms for recurrent musculoskeletal pain prediction
PO3.07	18.16 - 18.22	Anna Lecka	Missing Data: Imputation methods for survival analysis
PO3.08	18.22 - 18.28	Elvire Roblin	On the use of neural networks for survival models with censored data
PO3.09	18.28 - 18.34	Paola M.V. Rancoita	Enhancing the optimal cut-point identification of continuous covariates for predicting survival outcomes
PO3.10	18.34 - 18.40	Catherine Welch	Using multiple imputation to impute missing cancer stage in the estimation of relative survival
PO3.11	18.40 - 18.46	Marion Kerioui	Multilevel nonlinear joint model to characterize the tumor response variability under immunotherapy
PO3.12	18.46 - 18.52	Sonia Grandi	The Influence of Repeat Pregnancies on the Generalizability and Accuracy of Risk Prediction Models for Time-to
PO3.13	18.52 - 18.58	Nele Taba	Dietary patterns as markers for lifestyle - prediction of long-term mortality
PO3.14	18.58 - 19.04	Jadwiga Borucka	Comparison of Semiparametric and Parametric Survival Models with the Use of Time-Dependent ROC Curve
PO3.15	19.04 - 19.10	Mohammad Kaviul Anam Khan	Variable importance metric using modern machine learning techniques for clustered data

Wednesday, 26 August 2020

Session: **PO4: Big Data / Machine Learning**

Chairs: **Małgorzata Ćwiklińska-Jurkowska**, *Nicolaus Copernicus University, Torun, Poland*
Lara Lusa, *University of Primorska, Koper/Capodistria, Slovenia*

10.20 - 11.50

PO4.01	10.20 - 10.26	Jinheum Kim	Comparing survival functions with interval-censored data in the presence of an intermediate clinical event
PO4.02	10.26 - 10.32	Antoniya Georgieva	Fetal health risk assessment around the onset of labour: the role of an hour of cardiotocography monitoring
PO4.03	10.32 - 10.38	Adnan Karaibrahimoglu	An Innovative Study to evaluate the treatment efficiency of Percutaneous Nephrolithotomy by DEA
PO4.04	10.38 - 10.44	Francesco Innocenti	Sample size planning and optimal design for estimating regression-based reference values
PO4.05	10.44 - 10.50	Enzo Cerullo	Estimating Diagnostic Test Accuracy Whilst Adjusting for Imperfect Interrater Agreement in the Reference Stand
PO4.06	10.50 - 10.56	Alice Davis	Predicting patient engagement in IAPT services: A statistical analysis of electronic patient records
PO4.07	10.56 - 11.02	Emeline Courtois	Adaptive lasso approaches for pharmacovigilance signal detection: new penalty weights for variable selection
PO4.08	11.02 - 11.08	Balazs Dobi	Markov chart: an R package for cost-optimal patient monitoring and treatment
PO4.09	11.08 - 11.14	Rubaiya Rubaiya	Modeling and Mapping Low Birth-Weights: Role of Interactions among Determinants and Spatial Adjustment
PO4.10	11.14 - 11.20	Tomomi Takeshima	A dynamic treatment regime model to enhance adherence of patients with type 2 diabetes with Q-learning
PO4.11	11.20 - 11.26	Thomas Cowling	Selecting small sets of diagnosis codes with high prediction performance in large electronic medical datasets
PO4.12	11.26 - 11.32	Matteo Di Maso	A posteriori maternal dietary patterns and human milk: an application of the principal component analysis
PO4.13	11.32 - 11.38	Parinaz Mehdipour	Bayesian within-host modelling of red blood cell dynamics and primaquine induced haemolysis in G6PD deficiency
PO4.14	11.38 - 11.44	Maria Petropoulou	Shift the mean or the study variance to detect outliers in network meta-analysis

Session: **PO5: Causal Inference**

Chairs: **Agnieszka Pac**, *Jagiellonian University Medical College, Chair of Epidemiology and Preventive Medicine, Krakow, Poland*
Myra McGuinness, *University of Melbourne, Australia*

10.20 - 11.50

PO5.01	10.20 - 10.26	Nadia Dardenne	How to deal with multidimensional mediators? Overview and application with health literacy
PO5.02	10.26 - 10.32	Emilia Gvozdenovic	Causal inference concepts applied to three observational studies in the context of vaccine development
PO5.03	10.32 - 10.38	Maria Alvarez Hernandez	Agreement among raters: revisiting Hubert's kappa.
PO5.04	10.38 - 10.44	L. Naga Rajeev	Comparison of weight-for-height and BMI-for-age for estimating under-five thinness burden: Policy implications
PO5.05	10.44 - 10.50	David Phillippo	Assessing the performance of population adjustment methods for indirect comparisons: a simulation study
PO5.06	10.50 - 10.56	Sebastian Nielsen	Non-Specific Effects of Oral Polio Vaccine Campaigns on Child Mortality? Triangulation of Epidemiological Data
PO5.07	10.56 - 11.02	Rushani Wijesuriya	Evaluation of substantive-model-compatible multiple imputation approaches for incomplete three-level data
PO5.08	11.02 - 11.08	Monsurul Hoq	Comparison of statistical methods for estimating age-specific paediatric reference intervals
PO5.09	11.08 - 11.14	Anastassia Kolde	Martingale residual based approach for Cox modeling from high-dimensional data
PO5.10	11.14 - 11.20	Katri Parna	Prediction versus transferability of betas: a PCA informed strategy
PO5.11	11.20 - 11.26	Nicola Fitz-Simon	Multiple mediation modalities for social class disparities in cardiovascular patients
PO5.12	11.26 - 11.32	Stella Preussler	Using independent cross-sectional survey data to approximate post-migration health trajectories among refugees
PO5.13	11.32 - 11.38	Mariella Gregorich	A comparison of methods for causal inference with a rare binary outcome

Session: **PO6: Epidemiology: Methods and Applications / Electronic Health Records**

Chairs: **Agnieszka Doryńska**, *Institute of Public Health, Jagiellonian University Medical College, Krakow, Poland*
Kristel Van Steen

10.20 - 11.50

PO6.01	10.20 - 10.26	Aleksandra Turkiewicz	Probabilistic quantification of bias - combine strengths of population-based register data and clinical cohort
PO6.02	10.26 - 10.32	Gareth McCray	Creation of WHO indicators of Infant and Young Child Development (IYCD): metadata synthesis and item mapping
PO6.03	10.32 - 10.38	Patricia Soares	Time until the diagnosis of pulmonary tuberculosis in Portugal: spatiotemporal clustering
PO6.04	10.38 - 10.44	Kai Zhao	Risk of severe infections and mortality in patients with newly diagnosis of systemic lupus erythematosus
PO6.05	10.44 - 10.50	Simon Grřntved	Classification trees applied to identify predictors of autism spectrum disorder diagnosed in early childhood
PO6.06	10.50 - 10.56	Jerissa Samuel	A Review on the Application of Adaptive Designs in Oncology Clinical Trials
PO6.07	10.56 - 11.02	Borja Manuel Fernandez-Felix	Internal bootstrap validation of logistic predictive models: Stata command bsvalidation
PO6.08	11.02 - 11.08	Doris Tove Kristoffersen	Hospital readmissions and death within 30 days and 1 year for patients with hip fracture – a multi-state model
PO6.09	11.08 - 11.14	Frissiano Honwana	Predictive discrimination models to diagnose malabsorption from routine clinical diagnostics tests in Africa
PO6.10	11.14 - 11.20	Isobel Ridler	Applications of Regularised Structural Equation Modelling to Psychometrics
PO6.11	11.20 - 11.26	Rafał Pawłowski	Heart Rate Asymmetry assessment during head up tilt table test for healthy women
PO6.12	11.26 - 11.32	Susanne Schmitz	Efficacy of treatments for patients with recurrent glioblastoma: A systematic review and network meta-analysis
PO6.13	11.32 - 11.38	Sho Komukai	Doubly robust inference procedure for relative survival ratio in population-based cancer registry data
PO6.14	11.38 - 11.44	Michail Katsoulis	Weight change and the incidence of cardiovascular diseases in healthy adults; emulating trials using EHR
PO6.15	11.44 - 11.50	Malgorzata Hirsz	Age-period-cohort analysis of colorectal cancer incidence in New Zealand 1994-2018

Session: **RP1: Applications**

RP1.1	Kazue Yamaoka	A cluster randomised controlled trial of lifestyle intervention for adolescents using 'SPRAT' program
RP1.2	Denisa Krejci	Trends in incidence, mortality and survival of childhood cancer in the Czech Republic, 1994–2016
RP1.3	Minsu Park	Absence of causal association of BMI with eGFR: One- and two-sample Mendelian randomization analyses
RP1.4	Lucie Pehalova	Is there a difference in survival of patients with first primary and subsequent primary cancer?
RP1.5	Kaya Miah	Competing risks analyses for evaluating autologous stem cell transplantations in elderly myeloma patients
RP1.6	Elżbieta Chełmecka	Comparison of new diagnostic methods for determination of ethyl alcohol in forensic toxicology
RP1.7	Jodie Lord	Disentangling causal relationships with Alzheimer's Disease using multivariable Mendelian Randomization
RP1.8	Jody Ciolino	Considerations in Choosing Methods of Randomization for Cluster-Randomized Studies
RP1.9	Caterina Gregorio	An application of marginal structural joint frailty models on arrhythmogenic cardiomyopathies
RP1.10	Jose Maria Quintana Lopez	COPD readmissions prediction model
RP1.11	Anita Lindmark	Socioeconomic status and survival after stroke – using mediation and sensitivity analyses to assess the effect
RP1.12	Jordache Ramjith	Modelling time-varying effects on recurrent pneumonia in children under two years
RP1.13	Michał Michalak	The use of recurrent models to assess the risk of subsequent PCI or death in ischemic heart disease patients.
RP1.14	Alexander Hapfelmeier	A Systematic Review of Subgroup Analyses in Randomized Clinical Trials about the Cardiovascular Disease
RP1.15	Jufen Zhang	Association of red meat consumption and cardiovascular diseases in participants with and without obesity
RP1.16	Ronald McDowell	Associations Between Commonly Prescribed Medications And Cancer Risk: A Series of Nested Case-Control Studies
RP1.17	Lara Lusa	A case study of initial data analysis for longitudinal studies
RP1.18	Christoph Roethlein	Determining the influence of individual metabolic profiles on drug safety in Germany- the EMPAR study
RP1.19	Essi Syrjälä	Joint modeling of growth, puberty and type 1 diabetes in a multi-cohort data
RP1.20	Funda Seher Ozalp Ates	Longitudinal Analysis of Medical Faculty Students'empathy Scores With Latent Growth Curve Models
RP1.21	Alberto Alvarez-Iglesias	Estimating modifiable stroke risk in multiple regions of the world: the INTERSTROKE Modifiable Risk Score
RP1.22	Heather Riley	MEDIATION ANALYSIS IN RANDOMISED CONTROLLED TRIALS WITH ORDINAL OUTCOMES
RP1.23	Yuntao Chen	Development and validation of a heart failure phenotype stratified prognostic model: an IPD meta-analysis
RP1.24	Saber Dini	Using elastic net regularised logistic regression to find key protective antibodies against placental malaria
RP1.25	Chen Qu	Evaluation of Sample Size Requirements for Developing a Risk Prediction Model for Clustered Data
RP1.26	Megan Skelton	Growth Mixture Modelling of Anxiety and Depression Symptoms from Electronic Records of Psychological Therapy
RP1.27	Vinay Gupta	On the hazard rate of death due to HIV/AIDS with missing covariates and lost to follow up as competing risk
RP1.28	Michal M Skoczylas	Application of Cronbach's alpha in scientific studies on diagnosing and treating of neurological problems
RP1.29	Milada Cvancarova Smastuen	Merging and analysing data from national registries – use of antibiotics (defined daily doses)
RP1.30	Agnieszka Pac	Parent-child discrepancies in the assessment of adolescent's emotional and behavioral problems
RP1.31	Calin Avram	The use of logistic regression to identify risk factors in a clinical study

RP1.32	Jędrzej Chrzanowski	Continuous Glucose Monitoring record length and minimum daily observations for clinical interpretation.
RP1.33	Stanislav Katina	Population-Based Cancer Survival in the Czech Republic
RP1.34	Jiachen Chen	swdpwr: A SAS Macro and An R Package for Power Calculation in Stepped Wedge Cluster Randomized Trials
RP1.35	Janet L. MacNeil Vroomen	Longitudinal complex survey methods: weighting for missing data
RP1.36	Ayano Takeuchi	Apply Joint model to life course epidemiological data -birth cohort in Japan-
RP1.37	Lauren McVicker	Use of Aromatase Inhibitors and Risk of Cardiovascular Disease: A Population-Based Study
RP1.38	Raymond Nhapi	Multi-stage evaluation of mdediation and effect modification in a Gambian birth cohort
RP1.39	Una McMenamin	Prediagnostic Circulating Sex Hormones and Risk of Gastrointestinal Cancers in the UK Biobank
RP1.40	Lorenzo Guizzaro	Estimation of effect regardless of treatment discontinuation in trials with non-compliance and missing data
RP1.41	Maya Mueller	Application of an SEIRD model to track the spread of COVID-19 in Nigeria
RP1.42	Awa Diop	Variable selection in mediation analysis
RP1.43	Christoph Gerlinger	Clinical trials impacted by the COVID-19 pandemic: Adaptive designs to the rescue?
RP1.44	Ryan M Andrews	A practical guide to causal discovery for cohort data
RP1.45	Cattram Nguyen	Using multiple indicator cluster surveys to determine pneumonia burden in Lao People's Democratic Republic
RP1.46	Lan Kelly	Comparison of national COVID-19 death rates after age adjustment with a weighting method
RP1.47	Laura Hofstee	Implementing ECG Features into a Clinical Prediction Model to Assess the Individualized Risk in Arrhythmogenic cardiomyopathy
RP1.48	Lillian Boe	Prediction equations for blood concentration markers in the HCHS/SOL Nutrition and Physical Activity Study
RP1.49	Boris Hejblum	Multi-level modeling of early COVID-19 epidemic dynamics in France to estimate lockdown impact on infection
RP1.50	Renata Paprocka	Modeling of effectiveness of potential antibacterial and anti-inflammatory
RP1.51	Ana-Lucia Mayen	A life-course model to estimate associations between measures of alcohol exposure and colorectal cancer risk
RP1.52	Katarzyna Jagielnicka	Seeking possible cross-allergens to C4MGH0 from common mugwort
RP1.53	Paweł Zagożdżon	Antidepressants and the risk of death in northern Poland: registry-based cohort study
RP1.54	Ornella Wafo Noubissie	Evaluating the Establishment-based Risk Assessment Model for Salmonella spp. at Canadian Hatcheries (ERA – H)

Session: **RP2: Development of New Methods**

RP2.1	Alexia Sampri	Probabilistic data aggregation in healthcare data in order to solve content differences
RP2.2	Masahiro Kojima	Adjustments to network meta-analysis in the small number of studies by Bartlett-type corrections
RP2.3	Afiq Daniel	Rank-ordered logit model for confounder adjustment in the study of continuous outcomes
RP2.4	Konstantinos Pateras	Outliers that don't out-lie: Identifying and eliminating implausible birthweight for gestational age
RP2.5	Petra Ofner-Kopeinig	Allocation-Adaptive Randomization Methods – Biased Coin Probability and Unequal Allocation Rates
RP2.6	Dimitrios Mavridis	Using multidimensional scaling techniques to rank interventions taking multiple outcomes into account
RP2.7	Akimitsu Miyake	One small clinical trial design to provide additional information than single arm trials.
RP2.8	Philippe Wagner	Partitioning the family history relative risk into genetic and environmental effects
RP2.9	Leili Tapak	Multilevel Zero-inflated Censored Beta Regression Modeling for Proportions and Rate Data with Extra-zeros
RP2.10	Helene Colineaux	Decomposition of an interaction effect between two independent exposures through a mediator
RP2.11	Michael Seo	The Kilim plot: a tool for visualizing network meta-analysis results for multiple outcomes
RP2.12	Marie Alexandre	Comparison of AUC in clinical trials with follow-up censoring: Application to HIV therapeutic vaccines

RP2.13	Grethe Albrektsen	Adverse effect of diabetes on risk of myocardial infarction by gender and age: A methodological challenge
RP2.14	Lucie Biard	Bayesian models for early dose finding in phase I trials with multiple treatment courses
RP2.15	Akalu Banbeta Tereda	The Power Prior with Multiple Historical Studies for the Linear Regression Model
RP2.16	Agnieszka Kamedulska	Prediction of chromatographic retention using Bayesian hierarchical modeling
RP2.17	Shuo Sun	Causal Inference for Extreme Quantiles
RP2.18	Denis Talbot	Double robust estimation of partially adaptive treatment strategies
RP2.19	Daniel Rodriguez Duque	Semi-parametric Bayesian Inference for Optimal Dynamic Treatment Regimes via Marginal Structural Models
RP2.20	Zhujie Gu	Statistical integration of heterogeneous omics datasets using Group Sparse O2PLS (GO2PLS)
RP2.21	Said el Bouhaddani	Statistical integration of multi-omics data on Multiple System Atrophy
RP2.22	Gursimran Dhaliwal	Spatiotemporal transmission dynamics of Ebola in Congo
RP2.23	Mark W Donoghoe	A semiparametric joint binomial–Poisson model for the stimulation phase of an IVF cycle
RP2.24	Janine Witte	Multiple imputation for constraint-based causal discovery with cohort data
RP2.25	Claudia Coscia	Use of genetic variants as Instrumental Variables to be applied in mediation analysis: a new methodological pr
RP2.26	Werner Vach	Comparison regions to visualize uncertainty in two-parameter estimation problems in medical research
RP2.27	Anna Fowler	Machine learning approaches for the diagnosis of coeliac disease
RP2.28	Eddymurphy Akwiwu	A Progressive Three-State Model to Estimate Time to Cancer: a Likelihood-Based Approach
RP2.29	Fredrik Öhrn	A Case Study of a Successful Design Adaptation based on an improved Group Sequential Holm procedure

Session: **RP3: Method(s) Validation and/or Comparison**

RP3.1	Yuta Hamaguchi	Actual predictive performances of Bayesian prediction intervals for meta-analysis
RP3.2	Barbara Więckowska	Comparison of five spatial cluster detection methods – the most likely and secondary cluster
RP3.3	Cedric Laouenan	Evaluation by simulation of clinical trial designs for treatment efficacy during infectious diseases outbreak
RP3.4	Fabiola Del Greco M	Machine Learning Algorithm to Identify a Metabolic Profile Able to Predict Biomarker Levels
RP3.5	Ikuko Funatogawa	Autoregressive linear mixed effects models in structural equation modeling
RP3.6	Hiroya Hashimoto	Transformations for estimation of restricted mean survival times in small sample size
RP3.7	Jungyeon Choi	An overview on how measurements affected by medication use are handled and reported in observational research:
RP3.8	Miriam Hattle	A comparison of methods for meta-analysis of continuous outcomes from randomised trials
RP3.9	Patrick Hannan	A Comparison of Survival Methods for the Analysis of Recurrent Childhood Wheeze
RP3.10	Michel vaillant	Modelling egg counts of Helminths to evaluate the influence of multiple infections and compare treatments
RP3.11	Michael Grayling	Interim adaptation in individually randomised cross-over trials
RP3.12	Sayem Borhan	Effects of stratification on performance of methods for analyzing continuous data from stratified CRTs
RP3.13	Kim Luijken	Backward elimination to select covariates for confounding adjustment: harmful practice or sometimes useful?
RP3.14	Rose Sisk	Informative Presence and Observation in Clinical Risk Prediction: A review of methods
RP3.15	Katrina Sharples	Robustness of latent class models for diagnostic testing with no gold standard
RP3.16	Masataka Igeta	Incorporating follow-up lengths in sample size re-estimation to compare over-dispersed count data
RP3.17	Akiko Kada	Evaluation of treatment effect using propensity score in single-arm clinical trial and external control group
RP3.18	Elham Madreseh	A Bayesian Multivariate Joint Model for Multiple longitudinal and Multistate Processes
RP3.19	Aris Perperoglou	Prediction of Site Opening Times in Clinical Trials using Cox-type Survival Models and Random Forests

RP3.20	Mutamba Tonton Kayembe	Dealing with missing outcome and covariates in randomized trials: When and when not to use simple methods
RP3.21	Simon Thornley	Learning Bayesian networks and regression for studying the relationship between scabies and rheumatic fever
RP3.22	Diane Uschner	Patient-centric outcomes in clinical trials: The Desirability of Outcome Ranking
RP3.23	Tanvir Ahammed	Analysis of complex survey data: A comparison between SURVEY LOGISTIC and GLIMMIX with random effects
RP3.24	Marek Maly	A simple heuristics for detection of age-related increase of risk of rare congenital anomalies
RP3.25	Katsuhiro Iba	Re-evaluation of bootstrap-based optimism correction methods in multivariable clinical prediction models
RP3.26	Lucy Mosquera	Properties of Adjusted Per-Protocol Effect Estimators to Address Treatment Non-Adherence in Pragmatic Trials
RP3.27	Ryuji Uozumi	How to effectively recruit patients for adaptive enrichment clinical trials
RP3.28	Md Belal Hossain	Review of statistical methods to address treatment nonadherence in the pragmatic trial context
RP3.29	Tessa Lloyd	Challenges in integrating the different arms of the immune response against Mycobacterium tuberculosis
RP3.30	Lexy Sorrell	Copula-based modelling and analysis of semi-competing risk data, with application to renal transplant
RP3.31	Johan Verbeeck	Generalized Pairwise Comparison statistics: comparison of non-parametric and UMVUE estimators
RP3.32	Hae-Won Uh	Prediction of vascular ageing based on smartphone acquired PPG signals
RP3.33	Yi Cao	Comparison of Imputation Methods for Multivariate Longitudinal data with Mixed-type Incomplete Variables
RP3.34	Coraline Danieli	Preserving data privacy when using multi-site data to estimate individualized treatment rules
RP3.35	Federico Ambrogi	Machine learning for variable selection with high dimensional data and time-to-event outcomes
RP3.36	Alana Cuthbert	The impact of competing risks on individualised prediction of hip replacement outcomes using registry data
RP3.37	Micki Hill	Mind the gap: A simulation study on the performance of naive imputation approaches for interval-censored data
RP3.38	Moses Mwangi	Evaluation of a Piecewise Linear Mixed-effects Model in the Analysis of Randomized Cross-over Trial
RP3.39	Attila Csala	Functional principal component analysis for high dimensional longitudinal omics data
RP3.40	Laura Quinn	Review of interobserver variability studies in diagnostic imaging

Invited Sessions

IS01: Bayesian modelling of chronic & infectious diseases: Reverend Bayes going anti-viral

Organisers: **Dani Gamerman**
UFMG, Brazil

Maia Lesosky
University of Cape Town, South Africa

[IS01.1] Bayesian Geospatial Models for Tropical Disease Mapping

Paula Moraga

King Abdullah University of Science and Technology (KAUST), Saudi Arabia

Bayesian geospatial models provide a flexible and robust approach for spatial modeling of disease. Taking point observations of disease prevalence from a spatially sparse set of surveys, and using a suite of demographic, environmental and climate risk factors, the formulation of a geospatial model enables the prediction of disease prevalence at unsampled locations, and the generation of continuous disease risk maps. These maps are useful to understand the geographic and temporal patterns of disease, identify risk factors, and generate new hypotheses about disease transmission. In this talk, I will present two case studies where Bayesian predictive inference has been applied to model tropical diseases. The first study uses demographic and environmental data to construct a map of the prevalence of lymphatic filariasis in sub-Saharan Africa prior to the implementation of large-scale control interventions. The second study analyzes leptospirosis in a Brazilian urban slum. This community is generally characterized by low social status and poor sanitation infrastructure, but it is also highly inhomogeneous with wide spatial variability in social and environmental characteristics that are associated with leptospirosis risk. In both studies, the output of the geospatial risk models is key to inform better surveillance strategies and targeted interventions for disease elimination.

[IS01.2] Point pattern analysis of cerebrovascular deaths in Rio de Janeiro with spatially varying covariate effects

Jony Arrais Pinto Jr.

Statistic, Fluminense Federal University, Brazil

This work proposes a modeling approach for handling spatial heterogeneity present in the study of the geographical pattern of deaths due to cerebrovascular disease. The framework involves a point pattern analysis with components exhibiting spatial variation. Preliminary studies indicate that mortality of this disease and the effect of relevant covariates do not exhibit uniform geographic distribution. Our model extends a previously proposed model in the literature that uses spatial and nonspatial variables by allowing for spatial variation of the effect of nonspatial covariates. The methodology is applied to the study of the geographical death pattern of cerebrovascular deaths in the city of Rio de Janeiro. The results compare well against existing alternatives, including fixed covariate effects. Our model is able to capture and highlight important data information that would not be noticed otherwise, providing information that is required for appropriate health decision making.

[IS01.3] A Poisson-multinomial spatial model for the number of cases of vector-borne diseases in Rio de Janeiro, Brazil

Alexandra Schmidt

McGill University, Canada

Dengue-fever, zika, and chikungunya are arboviral infection diseases transmitted mainly by two vectors: Aedes aegypti and Aedes albopictus. During April 2016 the city of Rio de Janeiro experienced the peak of the first joint epidemic of the three diseases. As these diseases are transmitted by the same vectors, and the notified cases are either confirmed by laboratory exam or clinical-epidemiological criteria we propose a model that allows for uncertainty in the allocation of the number of cases per disease per borough. We propose a Poisson model for the total number of cases of arboviral infection diseases and, conditioned on the total number of cases, we assume a multinomial model for the number of cases of the three diseases. We discuss different parametrizations of the log-relative risk of the total number of cases and the parameters of the multinomial distribution. We have available the number of cases across the n = 160 boroughs of the city, the percentage of green area of the borough and a social-economic index. Inference is performed under the Bayesian framework. Our analysis suggests that as the percentage of green area increases the relative risk for the total number of cases decreases. The odds of a borough having chikungunya instead of dengue decreases as the social index increases, whereas the odds of a borough having zika instead of chikungunya increases with the social index. The odds ratio of zika or chikungunya with respect to dengue fever is not affected by the percentage of the green area of the borough. This is joint work with Laís P. Freitas, Marília S. Carvalho, and Oswaldo Cruz (Oswaldo Cruz Foundation, Brazil).

IS02: Tackling the challenges of imperfections in longitudinal and survival data: missingness, measurement error and irregular sampling

Organisers: **Laurence S. Freedman**
Gertner Institute for Epidemiology, Israel

Ruth Keogh
London School of Hygiene & Tropical Medicine, United Kingdom

[IS02.1] Challenges and opportunities of combined analysis of multiple outcomes in longitudinal studies

Eleni-Rosalina Andrinopoulou

Biostatistics, Erasmus MC, Netherlands

The increasing availability of clinical measures (e.g., electronic medical records) leads to collecting different types of information. This information includes multiple longitudinal measurements, and sometimes, also survival outcomes. The motivation comes from several clinical applications. In particular, patients after a heart valve replacement have a higher risk of dying or requiring a reoperation. These patients are followed echocardiographically, where several biomarkers are collected. Another example comes from patients after stroke, where measurements to assess recovery are taken over time.

Each outcome of interest is mainly analyzed separately, although it is biologically relevant to study them together. Previous work has focused on investigating the association between longitudinal and survival outcomes; however, less work has been done to examine the association between multiple longitudinal outcomes. In that case, it is common to assume a multivariate normal distribution for the corresponding random effects. This approach, nevertheless, will not measure the strength of association between the outcomes. Including longitudinal outcomes, as time-dependent covariates, in the model of interest would allow us to investigate the strength of the association between different outcomes.

Several challenges arise in both the analysis of multiple longitudinal data and longitudinal-survival data. In particular, different characteristics of the patients' longitudinal profiles could influence the outcome(s) of interest. Using extensions of multivariate mixed-effects models and joint models of longitudinal and survival outcomes, we investigate how different features (underlying value, slope, area under the curve) of the longitudinal predictors are associated with the primary outcome(s).

We obtain a strong association between death/reoperation with the underlying value and rate of change of the aortic valve gradient (difference in pressure on each side of the valve). Moreover, there is a strong association between the limb functioning with the rate of change in body function impairments for patients after stroke. Further results from simulations studies show important bias when not using the appropriate characteristic of the longitudinal profile. In this new era of rich medical data sets, it is often challenging to decide how to analyze all the available data appropriately.

[IS02.2] Missing Time-Dependent Covariate Values for the Cox Model - Joint Model Approach Versus Multiple Imputation

Havi Murad[1], Dankner Rachel[2], Laurence Freedman[1]

[1] Biostatistics and Biomathematics, Gertner Institute, Sheba Medical Center, Tel-Hashomer, Israel
[2] Cardiovascular Epidemiology, Gertner Institute, Sheba Medical Center, Tel-Hashomer, Israel

We have developed a procedure for imputing missing values for time-dependent covariates in a discrete time Cox model using the chained equations method [1]. The procedure multiply imputes the missing values for each time-period in a time-sequential manner, using completed covariates from previous time-periods as well as the survival outcome. It can be applied using the MI procedure in SAS with FCS statement or using similar packages in other software, e.g. the mice package in R. In this presentation we will compare this multiple imputation method to the approach of jointly modelling longitudinal and survival data [2]. This latter approach can be executed using the packages JointModel or JointModelBayes in R.

Although these two methods deal with missing values they do so in an entirely different manner. The joint modeling approach specifies a longitudinal model for each subject's series of time-dependent covariate values. Any deviation from the model is treated as measurement error. The time-sequential imputation approach specifies an imputation model for each missing value of the time-dependent covariate and the observed values are taken to be measured exactly. Thus the joint model yields an estimate that is adjusted for the assumed "measurement error", whereas the time–sequential imputation method yields an estimate based on there being no measurement error.

We use simulations based on data of glucose control variables among diabetics, from the Clalit Health Maintenance Organization database (n=546,000), using these methods to evaluate the association of glucose control with the risk of cancer. We examine different patterns of missing data in the glucose control variables (completely missing at random, missing at random and non-missing at random) and the impact of these patterns on the performance of the two methods.

References:

1. *Murad H., Dankner R., Berlin A., Olmer L and Freedman LS. Imputing missing time-dependent covariate values for the discrete time Cox model (2019). Statistical Methods in Medical Research. DOI: 10.1177/0962280219881168*
2. *Rizopoulos D. Joint models for longitudinal and time-to-event data: With applications in R. Chapman & Hall/CRC, 2012.*

[IS02.3] Structural nested cumulative survival time models to adjust for time-dependent confounding

Shaun Seaman

University of Cambridge, United Kingdom

Using data from observational studies to estimate the causal effect of a time-varying exposure, repeatedly measured over time, on an outcome of interest requires careful adjustment for confounding. Standard regression adjustment for observed time-varying confounders is unsuitable, as it can eliminate part of the causal effect and induce bias. Inverse probability weighting, g-computation and g-estimation have been proposed as being more suitable methods. G-estimation has some advantages over the other two methods, but until recently there has been a lack of flexible g-estimation methods for a survival time outcome. The recently proposed Structural Nested Cumulative Survival Time Model (SNCSTM) is such a method. I shall show how the SNCSTM can be fitted efficiently via g-estimation using standard software for fitting generalised linear models. The ability to implement g-estimation for a survival outcome using standard statistical software greatly increases the potential for uptake of this method. I shall illustrate the use of this method of fitting the SNCSTM by reanalysing data from the UK Cystic Fibrosis Registry.

IS03: Can Ying dance with Yang?: Combining evidence from RCT's and Pharmaco-epidemiology population-based studies of drug effect

Organiser: **Jessica Franklin**
Brigham and Women's Hospital / Harvard Medical School, United States

[IS03.1] A three-step approach to compare randomized trials and observational studies

Sara Lodi

Biostatistics, Boston University School of Public Health, United States

Randomized trials and observational studies are used to estimate the comparative effectiveness and safety of clinical strategies. When a randomized trial and an observational study address a similar question, discrepancies between their effect estimates tend to be attributed to uncontrolled confounding (due to imbalance of prognostic factors between the treatment groups) in the observational study. However, such discrepancies can also be also explained by differences in study design and data analysis. Therefore, effect estimates from randomized trials and observational studies may not be directly.

I will outline a three-step procedure to facilitate meaningful comparisons of effect estimates from randomized trials and observational studies: 1) harmonization of the study protocol (eligibility criteria, treatment strategies, outcome, start and end of follow-up, causal contrast) so that the studies target the same causal effect, 2) harmonization of the data analysis to estimate the causal effect, and 3) sensitivity analyses to investigate the impact of discrepancies that could not be accounted for in the harmonization process. In this method, the randomized trial is regarded as the target trial to be emulated with observational data. To illustrate the approach, I will compare estimates of the effect of immediate with deferred initiation of antiretroviral therapy in individuals positive to the human immunodeficiency virus from the START randomized trial and the observational HIV-CAUSAL Collaboration. Finally, I will discuss possible extensions of the approach and areas of application.

[IS03.2] How risk of bias assessments can inform comparisons between RCTs and observational studies

Jonathan Sterne

University of Bristol, United Kingdom

In recently published tools to assess the risk of bias in studies of interventions, bias is assessed in relation to the causal effect of the experimental intervention (versus the comparator intervention) on the outcome, defined as a comparison of counterfactuals. When using the Cochrane RoB 2 tool, results of many randomized trials will be assessed as at low risk of bias. However, the potential for unmeasured confounding means that few non-randomized studies of interventions assessed using the ROBINS-I tool will be assessed as at low risk of bias. I will describe both similarities and differences between these two tools, and conclude that combining results of randomized and non-randomized studies of interventions in a single meta-analysis will rarely be appropriate.

[IS03.3] **Unambiguously reported and replicable real-world evidence to support decision-making**

Shirley Wang

Brigham and Women's Hospital / Harvard Medical School, United States

Real world data analyses (non-randomized administrative and clinical healthcare databases) have become a vital tool, providing key insights for regulators, payers, and other healthcare decision makers. However, the credibility of database analyses suffers when 1) the results are not reproducible, 2) conflict with results from other database studies addressing the same question or 3) diverge from randomized clinical trial (RCT) results.

Conducting causal inference in database studies can be complex. Implementation of a database study can involve a great deal of data manipulation to create appropriately temporally anchored analytic cohorts from data that were not collected for research purposes. The many subtle design and analysis decisions made during study implementation, have left some stakeholders concerned about selective reporting of study specifications that arrive at a favorable result.

Healthcare decision-makers have emphasized the need greater transparency in research that make secondary use of databases. The REPEAT Initiative has embarked on several projects aimed at improving transparency, reproducibility and validity of database research. These projects include

large scale replication of 150 published database studies, evaluation of the robustness of results to alternative design and implementation parameters, and development of a structured reporting template with design visualization to increase transparency of reporting and minimize misinterpretation.

IS04: New Frontiers: Big Data & Functional Data Analysis

Organiser: **Yingying Fan**
USC, United States

[IS04.1] Statistics on curved manifolds: The next frontier

James Ramsay

Mc Gill University, Montreal, Canada

[IS04.2] Classification with imperfect training labels

Timothy Cannings[1], Yingying Fan[2], Richard Samworth[3]

[1] *School of Mathematics, University of Edinburgh, United Kingdom*
[2] *Marshall School of Business, University of Southern California, United States*
[3] *Statistical Laboratory, University of Cambridge, United Kingdom*

We study the effect of imperfect training data labels on the performance of classification methods. In a general setting, where the probability that an observation in the training dataset is mislabelled may depend on both the feature vector and the true label, we bound the excess risk of an arbitrary classifier trained with imperfect labels in terms of its excess risk for predicting a noisy label. This reveals conditions under which a classifier trained with imperfect labels remains consistent for classifying uncorrupted test data points. Furthermore, under stronger conditions, we derive detailed asymptotic properties for the popular *k-nearest* neighbour (*k*nn), Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) classifiers. One consequence of these results is that the *k*nn and SVM classifiers are robust to imperfect training labels, in the sense that the rate of convergence of the excess risks of these classifiers remains unchanged; in fact, it even turns out that in some cases, imperfect labels may improve the performance of these methods. On the other hand, the LDA classifier is shown to be typically inconsistent in the presence of label noise unless the prior probabilities of each class are equal. Our theoretical results are supported by a simulation study.

[IS04.3] **An Empirical Bayes Shrinkage Method for Functional Data**

Joshua Derenski, Yingying Fan, Gareth James

USC, United States

Shrinkage estimators are increasingly common in practice with examples dating back to the James-Stein shrinkage estimator. The issue of shrinkage arises, albeit with additional complications, for functional data. Consider, for example, the following scenario: given a set of observed functions, each with its own unknown mean, one wishes to select for further analysis those which are most extreme according to some metric such as the maximum value of the function. Let us also assume that the error distribution of these curves is known. Given that functions are often noisy realizations of some underlying mean process, these outliers are likely to generate biased estimates of the quantity of interest. In this paper we propose an empirical Bayes approach, using a variant of Tweedie's formula, to obtain shrinkage estimates of the true mean functions. Our approach has several advantages. It is non-parametric in nature, but automatically shrinks back towards a James-Stein type estimator in low signal situations. We demonstrate through extensive simulations and real data analyses that our approach can produce significant improvements in estimation accuracy relative to possible competitors.

IS05: Causal Inference vs. Unmeasured Confounding: Let's face the Dragon

Organiser: **Eric Tchetgen Tchetgen**
University of Pennsylvania, United States

[IS05.1] On the (In)Validity of Multiple Instruments for Mendelian Randomisation Studies

Vanessa Didelez

Leibniz Institute for Prevention Research and Epidemiology – BIPS, and Faculty of Mathematics and Computer Science, University of Bremen, Germany

Mendelian randomisation (MR) refers to situations where, in the known or suspected presence of unobserved confounding, a genetic predisposition can be exploited as an instrumental variable (IV) to estimate the causal effect of a modifiable risk factor or exposure on an outcome of interest. For example, the ALDH2 gene is associated with alcohol consumption, and has successfully been used as an IV to estimate the causal effect of alcohol on outcomes related to coronary heart disease.

MR analyses have become very popular recently with the increased availability of GWAS data. This gives rise to a number of challenges, especially around the topic of multiple IVs: it is common that several SNPs are found to be associated with an exposure of interest. However, the validity of such multiple IVs can often not be established in a convincing way and numerous methods that claim to allow for multiple - but partially invalid - IVs have been put forward in the last few years.

In this talk I will propose and investigate formal notions of „valid IV" within the context of multiple and potentially invalid IVs - this has been neglected by all of the previous literature but turns out to be crucial to assess the plausibility of various suggested methods. Using causal diagrams and the formal properties linking marginal and conditional independencies is especially useful to clarify some of the misunderstandings in the field. Amongst others, it can be shown that a notion of marginal validity of multiple IVs is not implied by nor implies a notion of joint validity.

We conclude that novel and increasingly popular method such as IV-Egger regression or 2-stage hard thresholding should be re-evaluated in view of these different notions of validity of IVs, and care has to be taken with typical analyses based on GWAS data when little subject matter background knowledge is available.

[IS05.2] Competing risk data and unmeasured confounding

Torben Martinussen

Section of Biostatistics, University of Copenhagen, Denmark

The analysis of competing risk data is notoriously difficult. It is well understood that one should be very careful if the focus is the cumulative incidence function corresponding to a primary cause. It is tempting to believe that shifting focus to cause specific hazards alleviates this problem, but this is false if interest is causal understanding of a given treatment. We study a new estimand that has a causal interpretation and develop a double robust estimation procedure, where we are also able to deal with unmeasured confounding in case a valid instrument is available. We illustrate the method with simulations and also a real data analysis.

[IS05.3] **Using sparsity to overcome unmeasured confounding: two examples**

Qingyuan Zhao

Statistical Laboratory, University of Cambridge, United Kingdom

Sparsity is often used to improve the interpretability of a statistical analysis and/or reduce the variance of a statistical estimator. This talk will explore another aspect–the utility of sparsity in model identifiability through two problems motivated by genetics applications.

The first problem is about removing „batch effects" or latent confounders in multiple hypothesis testing. I will present a general framework called Confounder Adjusted Testing and Estimation (CATE) we proposed to unify several widely used but ad hoc proposals. If the latent confounders are strong enough and the signals are sparse enough, CATE can be as powerful as the oracle estimator which observes the latent confounders. The second problem is about tackling invalid instrumental variables in Mendelian randomization. I will describe a new statistical framework called Robust Adjusted Profile Score (MR-RAPS) which can provide efficient and robust inference in such problems, by exploiting weak genetic instruments and limiting the importance of invalid instruments. Finally, connections to related work and potential future work will be discussed.

IS06: When non-inferiority trials are superior?

Organiser: **Lisa Hampson**
Novartis Pharma AG, Switzerland

Toshiro Tango
Center for Medical Statistics, Teikyo University, Tokyo, Japan

[IS06.1] Can we evaluate assay sensitivity in a non-inferiority trial ?

Eisuke Hida[1], Toshiro Tango[2]

[1] *Graduate School of Medicine, Osaka University, Japan*
[2] *Center for Medical Statistics, Japan*

Background: A new treatment that offers a better safety profile or a more preferable compliance with therapy compared to standard treatment may be beneficial even if somewhat less effective than standard treatment. A non-inferiority trial aims to demonstrate that the efficacy of the new treatment is not worse than that of the standard treatment by more than a prespecified small amount. However, non-inferiority trials have well-known issues, "the choice of the non-inferiority margin" and "the assessment of assay sensitivity". In particular, the following 3 points are described in some guidelines for the assessment of assay sensitivity; (a) historical evidence of the efficacy of treatment effects, (b) similarity of the current non-inferiority trial to historical trials (the "constancy assumption") and (c) quality of the non-inferiority trial (FDA guidance, 2016). Because there is no internal validity in a non-inferiority trial, assessing of assay sensitivity is more difficult than superiority trials. Consequently, three-arm non-inferiority trial (3NI), including a placebo is strongly recommended by some guidelines for assessing assay sensitivity.

Objective(s): In this presentation, we will introduce our proposed non-inferiority testing procedures with assay sensitivity to compare with the new treatment and the standard treatment in 3NI (Hida & Tango, Stat Med 2011; J Biopharm Stat 2013; Pharma Stat 2018).

Method(s): To test the non-inferiority with the assay sensitivity in 3NI, we have to show (i) the non-inferiority of the new treatment to the standard treatment with a prespecified non-inferiority margin and (ii) the substantial superiority of the standard treatment to a placebo, simultaneously. It is important that substantial superiority is not "simple" superiority for (ii). We derive a testing procedure for these hypotheses, and a procedure to calculate the required total sample size.

Results and Conclusions: We confirmed the validity of our proposed method from the type I error rate, the joint power and the accuracy of the calculated sample size by simulations in various scenarios. We conclude that our proposed procedure allows non-inferiority testing with assay sensitivity.

[IS06.2] **Subgroup analyses in non-inferiority trials**

Mimi Kim, Melissa Fazzari

Albert Einstein College of Medicine in New York, United States

Approaches and guidelines for performing subgroup analysis to assess heterogeneity of treatment effect have been the topic of numerous papers in the statistical and clinical trials literature, but have been discussed predominantly in the context of conventional superiority trials.

Concerns about treatment heterogeneity should be the same if not greater in non-inferiority (NI) trials, especially since overall similarity between two treatment arms in a successful NI trial could be due to the existence of qualitative interactions that are more likely when comparing two active therapies. Even in unsuccessful NI trials, subgroup analyses can yield important insights about the potential reasons for failure to demonstrate non-inferiority of the experimental therapy.

Recent NI trials have performed subgroup analyses using standard statistical tests for interaction, but there is increasing interest in more flexible machine learning approaches for post-hoc subgroup discovery. The performance and practical application of such methods in NI trials have not been systematically explored, however.

We adapted the Virtual Twin method (Foster et al., 2011), an algorithm combining random forest with CART for subgroup identification, to the NI setting and conducted extensive simulation studies to examine its performance under different NI trial conditions and to devise decision rules for selecting the final subgroups that merit further consideration.

We illustrate the utility of the method with data from a NI trial that compared two treatments for chronic musculoskeletal pain and offer some guidelines for performing subgroup discovery in NI trials.

[IS06.3] **Non-inferiority trials to directly inform treatment recommendations**

Andrew Copas

MRC Clinical Trials Unit at University College London, United Kingdom

Background: Non-inferiority (NI) trials compare new and current treatments, or treatment strategies, generally where the new treatment is thought to be similarly effective to the current but has additional benefits such as reduced side-effects. However they vary in their objectives. A trial may seek only to demonstrate that a new treatment is an acceptable alternative to the current i.e. (at least) almost as good. Alternatively a trial may seek to demonstrate that a new treatment can be recommended in place of the current because the additional benefits offset any possible reduction in effectiveness. In either case the trial may also be designed with licensing of the new treatment in mind. The methodological literature lacks clarity on how different objectives affect the design analysis and reporting of NI trials.

Objectives and Methods: We aim to explain how NI trials can best match their objective. We critically reviewed the methodology of NI trials, identifying variation according to the objective of the trial. Using example trials we illustrate possible approaches for trials to allow a new treatment to be recommended in preference to the current.

Results: When seeking to demonstrate a new treatment can be recommended in place of the current, the margin should be primarily informed by the magnitude of the new treatment's additional benefits. Consideration of preservation of effect may only be important if licensing is required. Consequently outcomes that measure the additional benefits are important to include and can inform a revision of the NI margin during the trial and also play a key role in interpretation of the final results.

Discussion: It is often recommended that the NI margin should be based on a minimum important clinical difference and/or preservation of effect but for new treatments with substantial additional benefits the trial can be designed from the outset to maximise impact by seeking recommendation in place of the current treatment. General treatment recommendations may sometimes be appropriate but risks and benefits will be valued differently by patients leading to personalised choices.

IS07: Can Cox model survive in the 21st century survival analysis?

Organiser: **Maja Pohar Perme**
University of Ljubljana, Slovenia

[IS07.1] 50 years with the Cox model - 80 more to come?

Per Kragh Andersen

Biostatistics, University of Copenhagen, Denmark

What do you want to do?

The Cox regression model has been a tremendous success and has for several years been one of the most widely used tools in biostatistics. One of the reasons behind this success is, undoubtedly, its ability to provide a treatment contrast that is a one-number summary of two survival curves. We will give a review of the way in which the model has influenced biostatistics over the past almost 50 years.

But will the model continue to be important - also for the next 50 (or 80 years)? The Danish writer and cartoonist Storm Petersen once, wisely, said that „prediction is difficult, especially when it deals with the future". However, we will anyway try to discuss the future of the Cox model at a time where it is being critizied for being too restrictive for practical applications, where the hazard ratio is being criticize d for not providing a causal contrast, and where targeted learning, including machine learning, is being put forward as methods with superior performances.

The preliminary and brief answer to the question raised in this session: „Can Cox model survive in the 21st century survival analysis?" is

„Yes, I think so!".

[IS07.2] Challenges in causal inference with survival and life history processes

Richard J. Cook

Statistics and Actuarial Science, University of Waterloo, Canada

Clinical trials should be designed based on a target estimand conveying the causal effect of an experimental treatment compared to a control intervention on an outcome of interest. When treatment comparisons are based on a Cox regression model, risk sets comprised of event-free individuals are defined at successive failure times. These risk sets form an improper subgroup of individuals since they are defined by an outcome observed post-randomization (i.e. survival to the corresponding failure time). In causal terminology, this survival status is a collider and conditioning on it induces a confounding in the assessment of treatment effect, even despite the baseline randomization. The phenomenon is discussed and illustrated, and similar settings where the phenomenon arises are presented. Alternative approaches to causal reasoning in these settings are then discussed.

This talk is based in part on joint work with Odd Aalen and Kjetil Roysland.

[IS07.3] **The Cox model in a world of target trial emulations**

Bianca De Stavola

Great Ormond Street Institute of Child Health, University College London, United Kingdom

The increasing availability and breadth of linked administrative and medical databases is mirrored by an expansion in causal investigations relying on these data. Investigations of the effectiveness of treatments and interventions as they occur "in the real world" are, in particular, becoming important additions (or sometimes alternatives) to those involving experimental studies.

Much has been written about the challenges of attempting causal inference from observational data. The framework of target trial emulation offers guidance in this respect, especially with regards to the designing principles that should direct the manipulation of the available information in ways that avoid introducing systematic biases (e.g. by avoiding immortal time bias). There are also additional challenges that are specific to time-to-event outcomes, as the follow-up one can construct from administrative sources is likely to be restricted because of data access issues and/or changes in coding and completeness. In general, this leads to shorter follow-up periods than anticipated, with the mechanisms leading to their censoring being informative.

There is also the issue of which format the causal question should take, i.e. which estimand to target. Others will discuss the relative merits of comparing hazards, or cumulative incidences, or survival probabilities. Here I will focus on the definition of exposure. Several options may be available: for example, we could define exposure in terms of received prescription, sustained prescription, or prescription change. Similar definitions could be envisaged for educational interventions, such as provision of special education. Further complications that may influence the definition of exposure arise from the temporal trends in medical/educational practice that affect the probability of receiving, sustaining or switching such exposures, as well as the consistency of their definitions over time.

In this talk I will review the choice of causal estimands within the context of linked administrative data, and consider the role that the Cox's proportional hazards model is likely to play in their estimation.

IS08: New developments in Longitudinal data analysis

Organiser: **Andrzej Galecki**
University of Michigan, United States

[IS08.1] Challenges in longitudinal analysis of pre-dementia cognitive decline

Helene Jacqmin-Gadda

Bordeaux Population Health, Inserm, France

Dementia, of which Alzheimer's disease is the main cause, is preceded by a very long cognitive decline. This decline, which extends over more than 10 years, is known to be heterogeneous between subjects and according to the cognitive functions. Investigating the characteristics of this pre-diagnosis phase gives insight on the underlying pathological process and can provide keys to the early management of these patients. This decline may be studied through longitudinal analysis of repeated measures of cognitive test in cohorts of elderly subjects with flexible mixed models to account for non linear time-trend. Among possible models, random changepoint models are particularly useful to identify key phases of the pre-diagnosis period. In this talk, I will discuss and illustrate various statistical challenges in the study of cognitive decline through random changepoint models. Due to competing risk of death and lost to follow-up, several designs and modelling strategies may be considered such as a nested case-control study or the analysis of the entire cohort with joint modelling of cognitive decline and dementia onset. As the time-scale differs according to the strategy, they do not allow to investigate the same clinical hypotheses. Inferential issues in random changepoint models will be also discussed.

[IS08.2] Design of longitudinal trials

Tomasz Burzykowski

Hasselt University, Belgium

In this presentation we evaluate the potential benefits of using biomarkers to enrich interim analyses of a clinical endpoint in randomized clinical trials. In particular, we consider the setting of a longitudinal study, in which a late-time measurement is the clinical endpoint of interest, and early-time measurement(s) are used in interim analysis to help in taking the decision to stop or continue the study. We distinguish between the cases when late-time measurements are available for some patients at the interim analysis and when only early-time measurements are available. It appears that in the latter case, the early-time measurements have to fulfil conditions for a good trial-level surrogate.

[IS08.3] Evaluation of predictive model performance of an existing model in the presence of missing data

Jeremy Taylor, Pin Li, Mattew J. Schipper

Department of Biostatistics, University of Michigan, United States

Background and Objectives: In medical research, the Brier score (BS) and the area under the receiver operating characteristic (ROC) curves (AUC) are two common metrics used to evaluate prediction models of a binary outcome, such as using biomarkers to predict the risk of developing a disease in the future. The assessment of prediction models using data with missing covariate values is challenging.

Methods: We propose inverse probability weighted (IPW) and augmented inverse probability weighted (AIPW) estimates of AUC and BS to handle the missing data. An alternative approach uses multiple imputation (MI). IPW requires a model for the missingness, MI requires a model for the distribution of the missing variable and AIPW requires both. We evaluated the performance of IPW and AIPW in comparison with MI in simulation studies under missing complete at random (MCAR), missing at random (MAR), and missing not at random (MNAR) scenarios.

Results and Conclusions: When there are missing observations in the data, MI and IPW can be used to obtain unbiased estimates of BS and AUC if the imputation model for the missing variable or the model for the missingness is correctly specified. MI is more efficient than IPW. AIPW can improve the efficiency of IPW, and also achieves double robustness from miss-specification of either the missingness model or the imputation model. The outcome variable should be included in the model for the missing variable under all scenarios, while it only needs to be included in missingness model if the missingness depends on the outcome.

IS09: Progress in providing analytic guidance for observational studies: the STRATOS initiative

Organisers: **Laurence Freedman**

Gertner Institute for Epidemiology, Israel

Willi Sauerbrei

Institute of Medical Biometry and Statistics, Medical Center - University of Freiburg, Germany

[IS09.1] Bridging the gap between causal inference and survival analysis: a censored edition

Saskia le Cessie[1,2], Bianca De Stavola[3], Vanessa Didelez[4], Els Goetghebeur[2], Ingeborg Waernbaum[5]

[1] *Leiden University Medical Centre*
[2] *Ghent University*
[3] *University College London*
[4] *University of Bremen*
[5] *Uppsala University*

STRATOS topic group 7 on causal inference recently wrote a tutorial on causal inference, focusing on point exposures and continuous outcomes (1). In this talk we go a step further and address causal questions for time-to-event outcomes. We consider the role of censoring and competing risks and discuss different estimands for survival outcomes. We stress the importance of a well defined time zero, and discuss the limitations of hazard ratios as causal parameters. We also discuss some issues in estimation, in particular the role of censoring. Topics are illustrated by some 'causal simulations', inspired by real data.

Reference:

1. Formulating causal questions and principled statistical answers. 2020, Goetghebeur et al. Accepted for publication in Statistics in Medicine

[IS09.2] Statistical and machine learning techniques: Which help in patient care and medical research?

Jörg Rahnenführer

Technische Universität Dortmund, Germany

Background: For diagnosis, prognosis and treatment selection for human diseases, various statistical and machine learning techniques are available. Methods that provide predictions are an important tool in all three areas. The diagnosis is typically associated with uncertainty that is relevant also for consequential treatment decisions. For prognosis, the uncertainty is part of the prediction, since the further development of the disease depends also on unknown developments in the future. Treatment selection aims to help a patient receive the treatment, which is most likely to lead to a positive outcome.

New methods often produce a hype cycle where an initial peak of high expectations is followed by disillusionment, before the real value becomes clear in the end. An exceptionally strong hype is associated with machine learning techniques, also in medical research. This hype is intensified by some spectacular successes of so-called deep learning techniques in applications with large numbers of observations, which is typically not the case in medical research.

Objective(s): Our goal is to give recommendations for the application of statistics and machine learning techniques for diagnosis, prognosis and treatment selection.

Method(s): We discuss the question, how much we can trust the claimed successes of a prediction method, especially from the practical viewpoint of a physician that is under the pressure of making a decision for a concrete patient that has to be treated.

Results and conclusions: Critical points for the assessment of the prediction method are the following: Clarity of the goal, suitability of the data, suitability of the analysis methods and algorithms, suitability of the evaluation measures, interpretability (also of the contribution of variables), availability of the explanatory variables (for the analysis of data from future patients), transportability across patient cohorts, and practical usefulness for the medical goal.

[IS09.3] **Regression without regrets: data screening is needed before modelling**

Marianne Huebner[1], Georg Heinze[2], Mark Baillie[3]

[1] Michigan State University, United States
[2] CeMSIIS - Section for Clinical Biometrics, Prognosis research group, Medical University of Vienna, Austria
[3] Biostatistical Sciences and Pharmacometrics, Novartis Pharma AG, Switzerland

In the analysis of observational studies multivariable regression models are popular for describing associations between variables, for risk stratification, or to facilitate informed treatment decisions.

Initial data analysis (IDA) focuses on informing a data analyst about features in the data that should be known to the analyst in order to a) properly interpret results of an analysis, b) make decisions on how to present the results of an analysis, and c) adapt the statistical analysis plan to avoid analysis errors.

Our aim is to define necessary steps and to stress the importance of IDA before performing regression analysis, assuming that a data set has already passed an initial data cleaning stage.

Following a previously published conceptual framework, we describe some aspects of IDA[1] in the context of regression modeling.

Appropriate graphical and analytical tools enable a researcher to perform IDA in order to avoid misinterpretation, poor presentation and analysis errors. These necessary preparations are too often forgotten, even by experienced data analysts.

Reference

1. *Huebner M, le Cessie S, Schmidt C, Vach W on behalf of the Topic Group "Initial Data Analysis" of the STRATOS Initiative: A Contemporary Conceptual Framework for Initial Data Analysis. Observational Studies 4 (2018):171-192.*

IS10: Individualised Risk Prediction & Omics

Organiser: **Krista Fischer**
University of Tartu, Estonia

[IS10.1] Precision Medicine at the Interface of translational Science & Systems

Kristel Van Steen
Liege, Belgium

[IS10.2] Causes and Health Consequences of Altered Transcriptome Profiles

Zoltán Kutalik
University of Lausanne, Switzerland

Genome-wide association studies (GWAS) identified thousands of variants associated with complex traits, but their biological interpretation often remains unclear. Co-localisation studies indicated that many of these mechanisms may be shared with gene expression regulation. At the same time, a plethora of high-throughput studies compared transcript levels between healthy and diseased individuals in order to pinpoint key genes underlying health conditions. However, neither approaches can reveal whether the identified genes are causes, consequences or mere correlates of the disease under scrutiny.

To elucidate these causal links, we adapted an instrumental variable approach, termed Mendelian Randomisation (MR), where the association strength of exposure-associated genetic markers with a given outcome can inform us about the causal effect of the exposure on the outcome. First, we demonstrate that the blood expression level of hundreds of genes has detectable causal effect on a wide range of human traits. Strikingly, third of the identified causal genes harbour no trait-associated genetic markers. Moreover, rare mutations in the implicated causal genes tend to cause monogenic forms of the same complex diseases. Finally, we observed that the disease-causing effects of gene expression are not mediated through protein levels.

Harnessing well-powered trans eQTL studies as input for MR, we could also examine the transcriptome-wide causal impact of common diseases. The analysis revealed that the (whole blood) gene expression-trait correlation is primarily driven by the causal effect of the phenotype to the expression level rather than the reverse. For example, BMI- and triglycerides-gene expression correlation coefficients robustly correlate with trait-to-expression causal effects ($r=0.1$, $P=1.2x10^{-43}$ and $r=0.09$, $P=4.2x10^{-37}$, respectively), but not detectably so with expression-to-trait effects.

In summary, our bidirectional MR approach revealed that complex traits have more pronounced impact on gene expression than the reverse and thus studies comparing gene expression levels of diseased and healthy subjects are more prone to reveal disease-induced changes in the transcriptome rather than disease causing genes. Finally, I will discuss the implications of the emerging multi-layer genetic architecture on biomarker discovery and drug repurposing.

[IS10.3] **Learning from a Lot: Empirical Bayes for High-Dimensional Model-Based Prediction**

Mark van de Wail

Amsterdam University Medical Centers, Netherlands

We consider generalised ridge regression in clinical prediction settings, in particular binary and survival, for high-dimensional data. We use complementary data („co-data", e.g. related studies, genomic annotation or cell line data) to define possibly overlapping or hierarchical covariate groups (e.g. gene sets, known signatures, Gene Ontology trees) that may differ considerably in terms of predictive strength. If so, penalising these groups by different ridge penalties likely improves prediction.

We present an Empirical Bayes approach to efficiently estimate hyper-parameters, which correspond to the the group penalties. Here, we provide an extra level of shrinkage to obtain stable group parameter estimates and to account for structure of the co-data. Any type of shrinkage can be used at this level, rendering a new, flexible framework to improve predictions. The flexibility is enhanced by also allowing non-penalized (e.g. known clinical) variables in the model. Moreover, the framework allows for integration and weighting of multiple co-data sets, plus posterior variable selection. In particular in our framework, which allows very different penalty parameters per variable, this approach (dense model + posterior selection) can be competitive or superior to sparse methods.

We demonstrate the method on an application to cancer genomics, in which we combine various sources of co-data and shrinkage types of the group parameters. Besides, we compare predictive performance with other commonly used methods, such as group lasso, which account for one single grouping structure, but which are not able to both shrink and estimate multiple group penalties from multiple sources. We show that the multi-group penalties stabilise variable selection, and improve the performance of parsimonious prognostic models. Finally, we emphasize that the method is not just another `learner', but also provides useful information on what external sources of information, the co-data, is relevant for the problem and data at hand. For this, we provide interpretation of the hyper-parameter estimates.

This is joint work with Mirrelijn van Nee and Lodewyk Wessels.

IS11: Biostatistical inference in practice: moving beyond false dichotomies

Organisers: **John Carlin**
Murdoch Children's Research Institute & University of Melbourne, Australia

Jonathan Sterne
University of Bristol, United Kingdom

[IS11.1] Dichotomania and other challenges: a perspective on principles and responsibilies

Laure Wynants

Epidemiology, Maastricht University, the Netherlands

A comment in Nature, signed by over 800 researchers, called for the scientific community to "retire statistical significance". The responses included a call to halt the use of the term „statistically significant", and changes in journal's author guidelines. The leading discourse among statisticians is that inadequate statistical training of clinical researchers and publishing practices are to blame for the misuse of statistical testing. In this presentation, we search our collective conscience by reviewing ethical guidelines for statisticians in light of the p-value crisis, examine what this implies for us when conducting analyses in collaborative work and teaching, and whether the ATOM (accept uncertainty; be thoughtful, open and modest) principles can guide us.

[IS11.2] Reflections on the strengths and weaknesses of significance testing

Colin Begg

Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, United States

Recently, a controversy has erupted regarding the use of statistical significance tests and the associated P values. Prominent academic statisticians have recommended that the use of statistical tests be discouraged or not used at all. This has naturally led to a lot of confusion among research investigators about the support in the academic statistical community for statistical methods in general. In fact, the controversy surrounding the use of P values has a long history. Critics of P values argue that their use encourages bad scientific practice, leading to the publication of far more false-positive and false-negative findings than the methodology would imply. The thesis of this commentary is that the problem is really human nature, the natural proclivity of scientists to believe their own theories and present data in the most favorable light. This is strongly encouraged by a celebrity culture that is fueled by academic institutions, the scientific journals, and the media. The importance of the truth-seeking tradition of the scientific method needs to be reinforced, and this is being helped by current initiatives to improve transparency in science and to encourage reproducible and replicable research. Statistical testing, used correctly, has an important and valuable place in the scientific tradition.

[IS11.3] Beyond Dichotomous Thinking in the Analysis and Reporting of Clinical Trials: Why, How and Who's Responsible?

John Carlin

Clinical Epidemiology & Biostatistics, Murdoch Children's Research Institute & University of Melbourne, Australia

Why is it so universally believed that the results of clinical trials must be interpreted dichotomously, forcing triallists, editors and readers into accepting a false dichotomy ("statistically significant") as real? I will provide a brief historical review and speculate on the many reasons, both psychological and institutional, for this entrenched belief, which leads to numerous widely recognised difficulties in scientific reasoning. A specific example is the ubiquitous phenomenon whereby the stated "alternative hypothesis" of interest is no less compatible with the data than the null hypothesis, but the conclusion enforced by many statisticians and editors alike is "no effect". The inherent lack of a sound logical basis for the conventional application of significance thresholds arises from the almost century-old confusion between the Fisher and Neyman-Pearson views of hypothesis testing. Clearly, dichotomous conclusions have a strong psychological appeal but I will argue that the role of the statistician should be to emphasise and explain the nature of uncertainty, which is inherently non-dichotomous. Truly null effects are rare and we as statisticians need to take responsibility for helping our collaborators embrace uncertainty rather than sweeping it under the rug with procedures that produce inevitable illogicalities and worse, feeding into the wider problems of current scientific practice. Taking up these challenges is not easy but we may find it increasingly difficult as a profession if we continue to ignore them.

IS12: Efficient Designs & cutting-edge analyses

Organisers: **Mitchell H. Gail**
National Cancer Institute, United States

Sven Ove Samuelsen
Department of Mathematics, University of Oslo, Norway

[IS12.1] Use of auxiliary data for efficient analysis of two-phase survival studies

Sven Ove Samuelsen
Department of Mathematics, University of Oslo, Norway

Case-cohort studies are commonly analyzed using inverse probability weighting (IPW) techniques as are sometimes nested case-control studies and other studies based on outcome dependent sampling. Such studies can be considered as two-phase studies where the first phase is the sampling of the cohort from a superpopulation and the second phase as the sampling of the actual study from the cohort.

Often IPW analysis of two-phase studies simply employs design-based inclusion probabilities. When auxiliary data are known for the entire cohort this practice will not use the data efficiently. One way to improve the analysis is through methods developed for survey sampling referred to as calibration or raking. This method is still IPW but with inclusion probabilities modified such that totals known from the cohort are exactly replicated by the weighting and at the same time staying close to the design inclusion probabilities.

In the talk I review survey sampling calibration and discuss implementations for two-phase survival studies. The potential for efficiency improvements is demonstrated. The calibration technique is furthermore compared to alternative methods for using auxiliary data.

[IS12.2] Methods to address correlated covariate and time-to-event error in electronic health records

Pamela Shaw[1], Eric Oh[1], Bryan Shepherd[2], Thomas Lumley[3]

[1] Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, United States
[2] Biostatistics, Vanderbilt University, United States
[3] Statistics, University of Auckland, New Zealand

Background: Biomedical studies that rely on electronic health records (EHR) data for inference are often subject to bias due to measurement error. The measurement error present in EHR data is typically complex, consisting of errors of unknown functional form in covariates and an outcome, which can be dependent. To address the bias resulting from such errors in analyses of survival outcomes, we seek a robust method that yields consistent estimates without the need to correctly model the error structure.

Methods: We consider generalized raking estimators, which will be consistent when an unbiased estimating equation is available on a validation subset. Raking is a standard method in survey sampling that makes use of auxiliary information on the population to improve upon the Horvitz-Thompson (HT) estimator applied to the validation subset. We propose multiple imputation to approximate the optimal auxiliary variable and compare this to previously proposed raking estimators that treat the error-prone data as surrogates. We also consider outcome-dependent sampling designs and investigate their impact on the efficiency of the raking estimators, either with or without multiple imputation. Detailed simulation studies are presented to examine the relative performance of the proposed estimators. We also apply the methods to analyze EHR data on HIV outcomes from the Vanderbilt Comprehensive Care Clinic.

Results: For problems with misclassification in the observed event indicator, the proposed raking estimator that made use of multiple imputation to approximate the optimal auxiliary variable showed appreciable gains in efficiency over the other estimators studied for a variety of settings. In cases without misclassification, simpler raking estimators achieved efficiency gains close to that of those using multiple imputation. As expected, all raking estimators improved over the HT; however, the improvement was smaller for more efficient study designs.

Conclusions: In two phase designs, raking offers a robust, efficient, and easy to implement method to address errors in studies relying on EHR. Further work is needed to improve the overall efficiency of the raking estimators through the selection of the phase two subset.

[IS12.3] Using negative controls to estimate causal effects of treatment in an entirely treated cohort

Ruth Keogh[1], Simon Newsome[2], Daniel Rodriguez Duque[3], Diana Bilton[4], Siobhan Carr[5]

[1] London School of Hygiene & Tropical Medicine, United Kingdom
[2] London School of Hygiene & Tropical Medicine and Novartis Pharma AG, Switzerland
[3] Cardiff University, United Kingdom
[4] Faculty of Medicine, National Heart and Lung Institute, Imperial College, United Kingdom
[5] Department of Paediatric Respiratory Medicine, Royal Brompton and Harefield NHS Foundation Trust, United Kingdom

Background: Observational data are increasingly used to gain understanding of treatment effects, by making use of using causal inference methods. A key assumption underpinning most such analyses is that of positivity. One situation where this assumption is not met is when an entire cohort of patients receives treatment. Then, it is difficult to estimate a treatment effect, because we do not observe contemporary individuals who did not receive the treatment.

Objectives: We describe how negative control outcomes can be used to estimate treatment effects in an entirely treated cohort and apply this to estimate the impact of the treatment ivacaftor on long term clinical outcomes for people with cystic fibrosis (CF). Ivacaftor was introduced in the UK in 2012 for a subset of the CF population with a specific genotype.

Methods: Previous observational studies have compared people receiving ivacaftor to either people in the time period prior to its availability or to people not eligible due to their genotype, resulting in a 'na%u010Fve' treatment effect. We discuss the conditions under which the na%u010Fve treatment effect estimates are unbiased and describe how negative control outcomes (the outcome measured in a time period when the treatment was not available and in a genotype group which is not eligible) can be used in combination with the difference-in-differences approach to obtain unbiased treatment effect estimates under weaker assumptions. Causal diagrams and the potential outcomes framework are used to explain the methods and assumptions. We apply the methods using longitudinal data from the UK CF Registry in the pre- and post-ivacaftor eras and on individuals in different genotype groups.

Results: Ivacaftor was found to improve lung function, to slow the lung function decline, and to reduce the rate of IV antibiotic use over a four-year period.

Conclusions: We have shown how negative control outcomes can be used to assess whether a control group is suitable for estimating the treatment effect in a group of people where everyone receives treatment, and to provide a more robust estimate of the treatment effect. Our results support evidence of a long-term clinical benefit of ivacaftor for people with CF.

IS13: Statistical challenges in COVID-19 research

Organiser: **Krista Fisher**
University of Tartu, Estonia

[IS13.1] Emulating trials with observational data and effect of hydroxychloroquine in hospitalized COVID-19 patients

Els Goetghebeur[1], Joris Hautekiet[2], Lucy Catteau[2], Nicolas Dauby[3], Marion Montourcy[2], Emmanuel Bottieau[4], Sabrina van Ierssel[5], Els Duysburgh[2], Herman Van Oyen[2], Chloe Wyndham-Thomas[2], Dominique Van Beckhoven[2]

[1] Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium
[2] Epidemiology and Public Health, Sciensano, Brussels, Belgium
[3] Department of Infectious Diseases, CHU Saint-Pierre, Brussels, Belgium
[4] Department of Clinical Sciences, Institute for Tropical Medicine, Antwerp, Belgium
[5] Department of General Internal Medicine, Infectious Diseases and Tropical Medicine, University Hospital Antwerp (UZA), Edegem, Belgium

The emerging COVID-19 pandemic came with large numbers of hospitalized patients experiencing in-hospital deaths and/or discharge alive in the relatively short term. In Belgium low dose hydroxychloroquine had been the recommended treatment. Data were available from across hospitals through an existing nationwide voluntary registration at hospital admission and discharge. We were asked to analyze the effect of the hydroxychloroquine treatment as implemented on hospital mortality. Trial emulation with a competing risks endpoint was adapted to handle several complications including the need to adjust for baseline confounders and any missing data. We considered informative censoring that likely entered since especially the earlier hospital admissions were limited to the sickest patients, as one feared for overcrowded hospitals. We further accounted for the timing of treatment initiation post hospital admission to avoid immortal time bias. Sensitivity analyses were performed to evaluate the impact of choices and assumptions made. Most analyses arrived at the same qualitative answers. We discuss the choices made and compare our approach and results with methods and results presented in some of the high profile early publications.

[IS13.2] COVID-19: Estimating the severity of the epidemic from hospitalised cohorts in the UK

Christopher Jackson, Tom Britton, Peter Kirwan, Paul Birrell, Anne Presanis, Daniela De Angelis

MRC BSU COVID-19 Working Group, PHE Joint Modelling Cell, United Kingdom

Estimates of the numbers of infections occurring at different levels of severity, and the risks of hospitalisation and death, are crucial to understand and predict the burden of the epidemic and its impact on healthcare services.

We use national cohorts of hospitalised COVID-19 cases in the UK to estimate various hospitalised case-severity risks and distributions of times to outcomes during the course of the epidemic. We use multi-state models, implemented in either cure-rate mixture or competing risks frameworks, or a combination of the two, to simultaneously estimate: the probabilities of intensive care unit (ICU) admission, recovery with and without an ICU stay, death with and without an ICU stay, all given hospital admission; and the times from hospital or ICU admission to each competing outcome.

Preliminary results suggest the probability of ICU admission given hospital admission decreases with age, from 25% (21-28%) in those under 45 to 1% (1-2%) in those aged 85+; and is larger in men than in women. The probability of dying in hospital before an ICU admission increases with age, from 2% (2-4%) in those aged <45 to 40% (37-42%) in those aged 85+; and is similar in men and women. The probability of dying given an ICU admission also increases with age, from 18% (12-25%) in those under 45 to 66% (34-86%) in those over 85. Such estimates have been crucial inputs to the scientific advisory groups informing policy-makers of the current evidence on COVID-19 severity.

[IS13.3] **Basic reproduction numbers, effective reproduction numbers and herd immunity**

Tom Britton

Department of Mathematics, Stockholm University, Sweden

We start by defining some basic concepts for the simplest epidemic model and then extend the results to a wider class of epidemic models allowing for some population heterogeneities. We end by showing a new result: the disease-induced herd immunity level h_D is smaller than the classical herd immunity level h_C=1-1/R.

Oral Contributed Sessions

OC01: Clinical Trials Methodology 1

Chair: **Colin Begg**

Memorial Sloan Kettering Cancer Center, United States

[OC01.1]

Use of external control information in clinical trials

Heinz Schmidli[1], Dieter A. Haering[2], Marius Thomas[2]

[1] *Statistical Methodology, Novartis, Switzerland*
[2] *Biostatistics, Novartis, Switzerland*

Background: Randomized controlled clinical trials are the gold standard to investigate new treatments. In rare diseases or with vulnerable patients such as children, randomizing patients to a control treatment may be difficult for ethical or feasibility reasons.

Objectives: Borrowing strength from external trials using relevant individual patient data on the control treatment may allow us to reduce the concurrent control group of a planned randomized controlled clinical trial.

Methods: Naive direct use of external control data is not valid due to differences in patient characteristics and other confounding factors. Instead, we suggest the rigorous application of robust Bayesian meta-analytic-predictive methods to use external controls in a principled way (Schmidli et al., 2014, Biometrics; Schmidli et al., 2019, Clinical Pharmacology and Therapeutics). This consists of an evidence synthesis of the relevant external control data, and the prediction of the control effect in the planned study. To robustify the inference, mixture priors are used which adaptively reduce the trial external information in case of conflict with the data from the new trial.

Results: A planned randomized adaptive trial in children with multiple sclerosis illustrates the approach. Evidence from past trials on the control treatment in adults and children is used to derive the prior information in the Bayesian primary analysis. The proposed trial design has been accepted for evaluation under US FDA's Complex Innovative Designs pilot program (https://www.fda.gov/media/129256/download).

Conclusions: Combined with tailored study designs, this approach allows us to reduce control groups and limit the burden to patients, which facilitates recruitment, reduces costs, shortens development time and potentially brings new medications to patients earlier.

[OC01.2]

Treatment selection in multi-arm multi-stage (MAMS) designs: an application to primary postpartum haemorrhage

Babak Choodari-Oskooei[1], Alexandra Blenkinsop[1], Soe Soe Thwin[2], Fernando Althabe[2], Mahesh K. B. Parmar[1]

[1] *MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology / UCL, United Kingdom*
[2] *Department of Reproductive Health and Research, WHO, Switzerland*

Background: Multi-arm multi-stage trials are an efficient, adaptive approach for testing many treatments simultaneously within one protocol. In settings where resources are limited, such as primary postpartum haemorrhage, it may be necessary to select a pre-specified subset of arms at interim stages even if they are all showing some promise against the control arm. This will guarantee the maximum number of patients required and reduce the associated costs.

Objectives: Motivated by WHO Tamponade Devices for Obstetric Haemorrhage (TADEOH) trial in postpartum haemorrhage, we explored the properties of such a selection design in a randomised phase III setting and compared it with other alternatives. The objectives are: i) to investigate how the selection of treatment arms affects the operating characteristics; ii) to explore the use of an information-rich (continuous) intermediate outcome, such as volume blood loss, to select the best performing arm, out of 4 options, compared with using the primary (binary) outcome for selection at the interim stage; and iii) to identify factors which can affect the efficiency of the design.

Methods: We conducted simulations based on the TADEOH (MAMS) trial to investigate the impact of the timing of treatment selection and applying an adaptive allocation ratio on the probability of correct selection, power and familywise type I error rate. Simulations were also conducted to explore how other design parameters will affect both the maximum sample size and trial timelines.

Results: The results indicates that the probability of correct selection plays a key role in the operating characteristics of the trial. The results showed that good operating characteristics are achieved if the treatment selection is conducted at around 17% of information time. Our results also showed that although randomising more patients to research arms before selection will increase the probability of selecting correctly, this will not increase the overall efficiency of the (selection) design compared with the fixed allocation ratio of 1:1 to all arms.

Conclusions: MAMS selection designs are efficient and flexible with desirable operating characteristics. Basing selection on an intermediate outcome measure can adversely affect the probability of correct selection if only one arm is selected to continue.

[OC01.3]

Controlling type I error rates in multi-arm clinical trials: a case for the false discovery rate

David Robertson, James Wason

MRC Biostatistics Unit, University of Cambridge, United Kingdom

Multi-arm trials are an efficient way of simultaneously testing several experimental treatments against a shared control group. As well as reducing the sample size required compared to running each trial separately, they have important administrative and logistical advantages. There has been controversy over whether multi-arm trials should correct for the fact that multiple null hypotheses are tested within the same experiment. Previous opinions have ranged from no correction being required, to a stringent correction (controlling the probability of making at least one type I error) being needed. We propose that controlling the false-discovery rate (FDR) provides a suitable compromise, with an appealing interpretation in multi-arm clinical trials.

We investigate the properties of the different correction methods in terms of the positive and negative predictive value (respectively how confident we are that a recommended treatment is effective and that a non-recommended treatment is ineffective). We show that controlling the FDR provides good properties. It retains the high positive predictive value of FWER correction in situations where a low proportion of treatments is effective. It also has a good negative predictive value in situations where a high proportion of treatments is effective. In a multi-arm trial testing distinct treatment arms, we recommend that sponsors and trialists consider using the FDR.

[OC01.4]

Optimized multiple testing procedures for nested sub-populations based on a continuous biomarker

Alexandra Graf[1], Dominic Magirr[2], Alex Dmitrienko[3], Martin Posch[1]

[1] *Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria*
[2] *Advanced Methodology and Data Science, Novartis Pharma AG, Switzerland*
[3] *Mediana, United States*

An important step in the development of targeted therapies is the identification and confirmation of sub-populations where a treatment has a positive treatment effect as compared to a control. These sub-populations are often based on continuous biomarkers, measured at baseline. Often patients are classified into biomarker low and biomarker high subgroups, which are defined via a threshold on the continuous biomarker. For example, there is a large discussion whether biomarkers have an influence on the outcome of treatments in patients with depression. Although a number of treatment options for such patients are available, no single treatment is universally effective. However, if insufficient information on the biomarker is available, the a priori choice of the threshold can be challenging and it has been proposed to consider several thresholds and to apply appropriate multiple testing procedures to test for a treatment effect in the corresponding subgroups controlling the family-wise type 1 error rate.

We consider the problem of how to design a trial with multiple nested subgroups, more precisely, how to select optimal thresholds and corresponding optimized multiple testing procedures that maximize the expected power to identify at least one subgroup with a positive treatment effect. Optimization is performed over a prior on a family of models, modelling the relation of the biomarker with the expected outcome under treatment and under control.

[OC01.5] **StCA Winner**

Inhomogeneous Poisson-gamma recruitment model for multi-centre clinical trials with model-averaged predictions

Szymon Urbas[1], Chris Sherlock[2], Paul Metcalfe[3]

[1] *STOR-i CDT, Lancaster University, United Kingdom*
[2] *Department of Mathematics and Statistics, Lancaster University, United Kingdom*
[3] *AstraZeneca, United Kingdom*

We introduce a general framework for monitoring, modelling, and predicting the recruitment to multi-centre clinical trials during interim analyses. The work is motivated by optimistic and narrow prediction intervals encountered when using standard recruitment models. The framework builds upon the commonly used Poisson-gamma recruitment by modelling the site-level enrolments using an inhomogeneous Poisson process,with the general structure of the framework permitting any parametric intensity models. The inference is carried out in the Bayesian paradigm, using weakly informative priors which can be modified based on historical data or expert opinion. The predictions are made using Bayesian model averaging which accounts for the uncertainty in both the parameters and the model. We test the validity of the method and its robustness to misspecification on simulated datasets, and then apply the method to prostate-cancer enrolment data, comparing forecast quality to those produced by existing accrual models.

OC02: Causal Inference 1

Chair: **Robert Platt**

McGill University, Canada

[OC02.1]

Bias correction for estimates from linear excess relative risk models in small studies

Sander Roberti[1], Flora van Leeuwen[1], Michael Hauptmann[2], Ruth Pfeiffer[3]

[1] *Department of Epidemiology and Biostatistics, the Netherlands Cancer Institute, Netherlands*
[2] *Institute of Biostatistics and Registry Research, Brandenburg Medical School Theodor Fontane, Germany*
[3] *Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, United States*

Background: Cancer patients treated with radiotherapy are known to be at increased risk of developing secondary tumors of irradiated healthy tissue later in life. A commonly used model in studies of second tumors after radiotherapy is the linear excess relative risk (ERR) model, where the dose-related hazard ratio (HR) is given by HR = 1+b Dose, with b the ERR per Gy of radiation dose. Commonly used dose metrics are mean organ dose and the dose to the site of the second tumor. Despite known numerical challenges in fitting this model, it is routinely applied. Its small-sample properties when data from cases and matched controls are analyzed have not been studied.

Objective: We evaluate small-sample biases under the linear ERR model in simulations and a data example and investigate several approaches to correct this bias.

Methods: We study first and second order jackknife bias correction, and also derive a set of modified score equations, following the approach by Firth (Biometrika, 1993). Uncorrected model parameters are estimated using conditional logistic regression fitted to matched case-control data, conditioning on different events under the linear ERR model. We perform a simulation study based on data from a published case-control study of second stomach cancer among testicular cancer survivors.

Results: Substantial upward bias was observed for realistic sample sizes of 75-150 matched sets in all investigated parameter settings. Neither first nor second order jackknife bias correction performs well. Using Firth's method of a modified version of the score equations improved convergence of the models but did not completely correct the biases.

Conclusion: The widely used linear ERR model may produce upwardly biased results, that were not corrected with jackknife in our simulations. Results based on Firth's correction were less biased.

[OC02.2]

Estimation of the treatment effect in case of switching to rescue medication in a randomised clinical trial

Hege Michiels[1], An Vandebosch[2], Stijn Vansteelandt[1, 3]

[1] *Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium*
[2] *Janssen R&D, Janssen Pharmaceutica NV, Belgium*
[3] *LSHTM, London, United Kingdom*

The interpretation of randomised clinical trial results is often complicated by intercurrent events. For instance, rescue medication is sometimes given to patients in response to worsening of their disease, either in addition to the randomised treatment or in its place. The use of such medication complicates the interpretation of the intention-to-treat analysis. Ignoring all data after switching to rescue medication is likely to deliver biased results since rescued patients form a highly selected group.

We aim to estimate the intention-to-treat effect that would have been observed, had patients on the active arm been switched to rescue medication only if they had been switched when randomised to control. As such, we are able to distinguish the treatment effect from the effect of rescue medication on a patient's outcome, while avoiding the strong extrapolations that are typically needed when inferring what the intention-to-treat effect would have been in the absence of rescue medication.

We express the targeted treatment effect in terms of potential outcomes and show that it has connections to so-called `interventional effects' from mediation analysis. We propose g-computation methods, inverse probability weighted methods and double robust methods for estimating this effect. The latter relies on working models for the outcome and time-varying confounders, as well as a model for the probability of switching, but only requires one of these to be correctly specified. Our proposal relies on untestable assumptions, in view of which we propose a sensitivity analysis.

We use the methods for the analysis of a clinical trial conducted by Janssen Pharmaceuticals, in which chronically ill patients can switch to rescue medication for ethical reasons. Monte Carlo simulations confirmed that the proposed estimator is easy to calculate, asymptotically unbiased when at least one set of models is correctly specified and efficient when all models are correctly specified.

The proposed new estimand evaluates what the effect of treatment would have been if the same rescue treatment decisions were made on both arms. It thereby minimises the need for extrapolation, on which competing methods rely.

[OC02.3]

Caution Against Examining the Role of Reverse Causality in Mendelian

Sharon Lutz[1], Ann Wu[1], Stijn Vansteelandt[2], Christoph Lange[3]

[1] *Population Medicine, Harvard Medical School, United States*
[2] *Applied Mathematics, Ghent University, Belgium*
[3] *Biostatistics, Harvard T. H. Chan School of Public Health, United States*

Recently, Mendelian Randomization (MR) has gained in popularity as a concept to assess the causal relationship between phenotypes in genetic association studies. One such method, the MR Steiger approach, has been proposed as a tool that can infer the causal direction between 2 phenotypes. Through simulation studies, we examine the ability of the MR Steiger approach to correctly determine the exposure and outcome (i.e. effect direction). Our results show that the Steiger approach may fail to correctly identify the direction of causality, especially in the presence of pleiotropy and unmeasured confounding. We also applied several popular MR approaches and the Steiger method to the COPDGene study, a case-control study of Chronic Obstructive Pulmonary Disease (COPD) in current and former smokers, to examine the role of smoking on lung function. We have created an R package on Github called reverseDirection which runs simulations for user-specified scenarios in order to examine when the MR Steiger approach can correctly determine the causal direction between the exposure and outcome in any user specified scenario.

[OC02.4]

Understanding pathways to health inequalities in cystic fibrosis using UK registry data

Daniela Schlueter[1], Ruth Keogh[2], Schadrac Agbla[3], Rhian Daniel[4], David Taylor-Robinson[1]

[1] *Public Health and Policy, University of Liverpool, United Kingdom*
[2] *Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom*
[3] *Biostatistics, University of Liverpool, United Kingdom*
[4] *School of Medicine, Cardiff University, United Kingdom*

Background: Cystic fibrosis (CF) is an inherited, progressive condition affecting over 10 000 individuals in the UK. Symptoms of CF include poor growth, lung infection, poor lung function and reduced survival. Outcomes are worse for people growing up in disadvantaged circumstances, but it is not clear to what extent inequalities in early growth can explain inequalities in later lung function and survival.

Objective: To assess how the association between socioeconomic circumstances (SECs) and subsequent lung function measured around age 6 is mediated by weight trajectory in early childhood.

Methods: We did a causal mediation analysis using data from the UK CF registry, which captures 99% of all people with CF in the UK and records clinical information including weight and infection status at annual review visits. All children born between 2000 and 2010 and diagnosed by newborn screening were included in the analysis if they had at least one lung function measure between ages 6 and 9, at least one weight and infection measure between birth and age 6, and complete data on SECs and baseline confounders. We imputed missing data using multiple imputation by chained equations. We used the parametric mediational g-formula to estimate the total effect of SECs on lung function, and the indirect effect mediated by weight trajectories in the first six years of life, accounting for potential time-varying confounding by infection status. Confidence intervals were estimated using non-parametric bootstrap.

Results: Using data from 853 children, we found a total effect of deprivation on lung function, measured by percent of predicted FEV1, of 4.53 percentage points (95% CI 3.44 to 5.77). Our results showed that if we could improve the weight of the most disadvantaged children to have the same distribution as that of the least disadvantaged children, their lung function would improve on average by 0.74 percentage points (95% CI 0.36 - 1.1).

Conclusion: Only 16% (95% CI 8%-25%) of the inequalities in early lung function for people with CF were explained by weight trajectories in the first 6 years of life, suggesting that other important pathways to inequalities need exploration.

[OC02.5]

Using health records to estimate mortality attributed to acute kidney injury: challenges and reflections

Johan Steen[1], Stijn Vansteelandt[2], Johan Decruyenaere[3], Wim Van Biesen[4]

[1] *Department of Internal Medicine and Pediatrics, Ghent University, Belgium*
[2] *Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium*
[3] *Department of Intensive Care, Ghent University Hospital, Belgium*
[4] *Renal Division, Ghent University Hospital, Belgium*

Acute kidney injury (AKI) is a common complication in the intensive care unit (ICU), estimated to affect up to 60% of critically ill patients. However, it remains unclear to what extent AKI is a true cause of excess ICU mortality or merely a marker of underlying disease severity.

We estimate the time-dependent population attributable fraction of ICU mortality due to ICU-acquired AKI in a cohort of patients admitted to the Ghent University Hospital ICUs (2013–2017). To highlight challenges related to estimation from electronic health records, we assess the sensitivity of our results to (i) availability and granularity of relevant criteria for retrospective AKI diagnosis and staging from electronic health records, (ii) different methods for accounting for timing of AKI onset to eliminate immortal time bias and (iii) adequate adjustment for measured time-varying confounding.

For each patient, we retrospectively calculate AKI stages, along with timing of their onset, based on repeated serum creatinine and urine output measurements extracted from the Intensive Care Information System database (applying predefined thresholds for either one of these criteria or for both, and according to different choices of baseline creatinine measures). The AKI-attributable fraction of ICU mortality is estimated as a function of time from admission using (i) the progressive disability model approach which treats AKI onset as a competing event for ICU death and discharge, and (ii-iii) two inverse probability of censoring weighted competing risk analyses which treat AKI onset as a censoring event for ICU death and discharge and where the censoring mechanism is assumed either uninformative or ignorable conditional on baseline and time-varying confounders.

Results will be discussed with a particular focus on the interplay between (i) different statistical estimation approaches and (ii) varying levels and temporal dynamics of AKI incidence according to criteria used for diagnosis and staging. The gained insights may foster reproducibility and help to disentangle causes of replication failures in a more systematic way. We conclude with some reflections on the required assumptions needed to interpret our estimates as those obtained from a hypothetical randomized controlled prevention trial.

OC03: Multi-State models

Chair: **Jeremy Taylor**

University of Michigan, United States

[OC03.1]

Multi-state Markov model for estimating HIV incidence from HIV surveillance data in France, 2008-2018

Charlotte Castel[1], Cécile Sommen[1], Yann Le Strat[1], Ahmadou Alioum[2]

[1] *Data Analysis Department, French Institute for Public Health Surveillance, France*
[2] *Biostatistics team, Inserm Center U1219-Bordeaux Population Health, University of Bordeaux, ISPED, France*

Thirty-five years after the discovery of the human immunodeficiency virus (HIV), the epidemic still going in France. Assessing the dynamics of the epidemic, through the estimation of incidence is a statistical challenge. This estimation is essential to guide HIV prevention strategies and monitor their impact.

The main objective was to develop a method for estimating the number of new HIV infections and improve the estimates of the last three years. Knowing that HIV is mainly transmitted by undiagnosed HIV-positive individuals and that the earlier they are treated the less they transmit HIV, we also estimated the number of infected people who are unaware of their HIV-positive status. Finally, the distribution of time between infection and diagnosis is estimated.

Since 2003, the French Institute for Public Health Surveillance has set up a surveillance system for HIV diagnoses which collects the clinical stage at diagnosis time. We adapted the approach proposed by Sommen et al. (2009) to the French data by taking into account the primary infection stage. We used a non-homogeneous multi-state Markov model describing the progression of the HIV disease. We proposed a penalized likelihood approach to estimate the HIV incidence curve as well as the diagnosis rates. The HIV incidence curve was approximated using cubic M-splines and an approximation of the cross-validation criterion was used to estimate the smoothing parameter.

The method is first illustrated from a thorough simulation study, then applied to the HIV mandatory notification data. From the simulation study, theoretical value is comprise between the pointwise Bayesian confidence limits of the estimated value for 9 years among 11 for the HIV incidence and all over for the HIV diagnosis for the period 2008-2018. From the HIV mandatory notification data, the obtained estimates are consistent with previous estimates using other approaches.

Our model is a new tool for estimating the HIV incidence giving more precise estimates in the last three years that previous methods. The non-homogeneity of the Markov model allows to introduce changes in testing behavior over time which is not the case for other methods used currently in France.

[OC03.2]

Assessing the effect of time-dependent exposures on time-to-event endpoints with economical sampling designs

Jan Feifel[1], Maja von Cube[2], Jan Beyersmann[1], Martin Schumacher[2]

[1] *Institute of Statistics, University of Ulm, Germany*
[2] *Institut für Medizinische Biometrie und Statistik, Albert-Ludwigs-Universität Freiburg, Germany*

Background: Hospital-acquired infections accompanied by antimicrobial resistance increase both morbidity and mortality of hospital patients. Researchers interested in the effect of these time-dependent infections on the length-of-hospital stay, as a measure of disease burden, face large cohorts with possibly rare exposures.

Objective(s): For large cohort studies with rare outcomes nested case-control designs are favorable due to the efficient use of limited resources. Here, nested case-control designs apply but do not lead to reduced sample sizes, because the outcome is not necessarily rare, but the exposure is. Recently, exposure density sampling (EDS) [1] and nested exposure case-control design (NECC) [2] have been proposed to sample for a rare time-dependent exposure in cohorts with a survival endpoint. The two designs differ in the time point of sampling.

Method(s): We outline the differences between the two designs, outline their merits and shortcomings and investigate their performance using observational data on the impact of hospital-acquired infection on length-of-stay.

We also introduce an enhanced version of the NECC, forcing discordant risk sets for each contribution to the partial likelihood. The performance of this design will be investigated using simulation studies. All presented methods will be compared to the gold-standard Cox model on the full cohort.

Results: The nested exposure case-control design appears to be inferior when only comparing statistical power. However, when considering study design and flexibility stratified case-control design are profitable. The enhanced NECC uses the available information more efficiently and, therefore, improves the estimation performance while keeping the number of required individuals minimal. We will discuss unbiasedness as well as asymptotic properties of the design.

Conclusions: Both EDS and enhanced NECC are capable of analyzing time-to-event data by simultaneously accounting for rare time-dependent exposure and result in affordable sample sizes. However, to asses their full potential a clear specification at the design stage of the study is required. For antimicrobial resistance data, both approaches are applicable and their performance under strong stratification appears is competitive.

References:

1. *Ohneberg et al. (2019), Statistics in Medicine*
2. *Feifel et al. (2020), Lifetime Data Analysis*

[OC03.3]

Assessing and Relaxing the Markov Assumption in the Illness-Death Model

Jonathan Broomfield, Michael Crowther

Department of Health Sciences, University of Leicester, United Kingdom

Background: Multi-state survival analysis considers several potential events of interest along a disease pathway. In the era of personalised medicine, such analyses are becoming crucial to model complex patient trajectories and are increasingly being used across a variety of applications, such as health technology assessments. Multi-state models often make the Markov assumption, whereby an individual's future trajectory is dependent only upon their present state, not their past. In reality, there may be transitional dependence upon either previous events, or more than one timescale (e.g., time since entry to a current or previous state).

Objective: The primary aim of this study was to evaluate the performance of a Markov illness-death model when the Markov assumption was violated, in a set of plausible simulation scenarios. The model was assessed using baseline hazard rates, covariate hazard ratios and transition probabilities as estimands, with bias and coverage as measures of performance.

Methods: Guided by a study in breast cancer, data were simulated from Weibull baseline distributions, with hazard functions dependent on single and multiple timescales. Markov and non-Markov models were fitted to account for/ignore the underlying data structure. The Stata commands survsim and merlin were used to simulate data and fit models. An application to the breast cancer study was made to evaluate the conclusions of the simulation.

Results and Conclusions: Ignoring true time dependencies led to bias in terms of underlying transition rates between states and covariate effects on transition rates, while the transition probabilities were fairly robust. The observed bias increased as the magnitude of the unaccounted-for time-dependencies increased. Further work is needed to evaluate different estimands, such as length of stay, and more complex multi-state models. We also describe software implementations in Stata for simulating and estimating general multiple timescale multi-state models.

Keywords: multi-state, Markov, multiple timescale, simulation.

[OC03.4]

Regression model for epidemiological indicators: an alternative to the pseudo-values

Pierre Joly[1], Camille Sabathé[2], Catherine Helmer[1], Hélène Jacqmin-Gadda[1], Daniel Commenges[1]

[1] *Inserm, Université de Bordeaux, France*
[2] *Université de Nantes, France*

Estimating the effect of explanatory variables on epidemiological indicators is useful from a public health perspective. However, generally these indicators are functions of several transition intensities when using an Illness-death model. Pseudo-values are a well-known and useful approach in this case. In this work, we propose an alternative to the pseudo-values to deal with both interval-censored and left-truncated data. Indeed, in data from cohort study, interval censorship of the disease onset can occur when the disease is not monitor continuously. In addition, subjects are included at different ages, resulting in left truncation of the age of entry into the cohort. Moreover, the competition with death requires to use an Illness-death model. We propose a two-step approach to estimate the effect of explanatory variables on epidemiological indicators. In a first step, we estimate parameters θ from the illness-death model and the quantities of interests are computed for each subject as functions of these estimates. In the second step, the quantities of interests are regressed on covariates using a linear model. Confidence intervals are computed empirically by repeating the estimation of the linear models with quantities of interest computed from random draws of θ from its posterior distribution. The behaviour of the linearization approach was evaluated in a simulation study. We illustrated this approach by estimating the effect of age at menopause on the lifetime risk of dementia, the life expectancy without dementia and the total life expectancy of non-demented women alive at a given age. The data came from the French PAQUID cohort.

[OC03.5]

Analysis of recurrent hospital acquired infections and hospital discharge using marginal regression modelling

Jerome Lambert[1], Lars Bugiera[1], Ksenia Ershova[2], Martin Wolkewitz[1]

[1] *Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Germany*
[2] *Department of Anesthesiology, Keck School of Medicine - University of Southern California, United States*

Background: Hospital acquired infections (HAI) are a common problem, especially among more severe patients, and one patient can encounter several HAI during hospitalization. Most of the study about predictive factors of HAI only focus on the first episode, resulting in a loss of information. Moreover, studies on HAI often present two specificities: the absence of censoring and the presence of a non-ignorable competing terminal event, discharge from hospital.

Objective: This work illustrates how the mean cumulative count (MCC) can be used to describe recurrent HAI, assess the effect of predictive factors, and contrast it with classical methods used to analyze HAI.

Method: The MCC is the mean expected number of recurrent events per patients across time. This marginal estimate takes into account the competing event of discharge from hospital, and a corresponding semi-parametric model has been developed by Gosh & Lin. It can be seen as an extension of the cumulative incidence function of first HAI and the corresponding Fine & Gray modelling of sub-distribution hazard. Other alternative strategies are simpler models that do not take into account the timing of events (log binomial or Poisson regression), and models for conditional estimates such as the cause-specific Cox model for the cause-specific risk of first HAI and its extension to multiple events, the Anderson-Gill (AG) model. We illustrate all these methods using data from a retrospective cohort of patients admitted to a neurological intensive care unit.

Results: The 2,281 patients experienced a total number of 2,145 HAI: at the end of follow-up, MCC was 0.94. Vascular disease and traumatic brain injury were associated with a higher average number of HAI than tumors. Conditional modelling also showed an effect of disease type on the competing event of discharge, which is captured as an indirect effect in the modelling of MCC.

Conclusions: The mean cumulative count is a clinically meaningful quantity that can usefully describe HAI. It takes into account both all the information on recurrent HAI and the importance of the competing event of hospital discharge.

OC04: Individualised Risk Prediction 1

Chair: **Torben Martinussen**

University of Copenhagen, Denmark

[OC04.1]

Developing risk models for multicenter data: a simulation study comparing regression techniques

Nora Falconieri[1], Ben Van Calster[1], Dirk Timmerman[1], Laure Wynants[2]

[1] *Development and regeneration, KU Leuven, Belgium*
[2] *Epidemiology, Maastricht University, Netherlands*

Background: Multicenter data is common, but most prediction model studies (estimated 78% of cardiovacular models) ignore the clustered nature of data during model development. Some studies have indicated random intercept models may predict better than standard logistic regression models, but it is not known whether prediction is also superior when model assumptions are violated. The objective of this study is to evaluate the predictive performance of regression methods for developing clinical risk prediction models using multicenter data.

Methods: We compared standard logistic regression, generalized estimating equations, random effects logistic regression (with and without random slope, with and without „poor man's method"), and fixed effects logistic regression (with and without center interactions) in a case study on the diagnosis of ovarian cancer (n=5909, 1929 malignancies, 24 centers) and in simulated data. Predictve performance was evaluated in terms of discrimination (c-index) and calibration (intercept and slope) at the center level and population level, including old and new centers. We varied the amount of clustering (ICC 5 vs. 20), development sample size and composition (5 vs. 50 centers, 50 vs.200 patients in each), distribution of center-specific intercepts (normal, uniform, extreme value), the presence of a center-predictor interaction, and the presence of a dependency between center effects and predictors.

Results: Small sample sizes led to overfitting and unreliable predictions for all models. When sample sizes were sufficiently large, conditional models yielded calibrated predictions at the center level, whereas marginal models yielded miscalibrated predictions at the center level. This miscalibration was worse with more heavily clustered data. Calibration of random intercept logistic regression was better than that of the routinely used standard logistic regression even when center-specific intercepts were not normally distributed, a center-predictor interaction was present, center effects and predictors were dependent, and when the model was applied in a new center.

Conclusion: We recommend random intercept logistic regression, ideally followed by local recalibration of the intercept in new centers.

[OC04.2]

Sample size calculation for external validation of risk models

Gareth Ambler[1], Menelaos Pavlou[1], Chen Qu[2], Shaun Seaman[3], Ian White[4], Rumana Omar[2]

[1] *University College London / Statistical Science, United Kingdom*
[2] *Statistical Science, University College London, United Kingdom*
[3] *MRC Biostatistics Unit, University of Cambridge, United Kingdom*
[4] *MRC Clinical Trials Unit at UCL, University College London, United Kingdom*

Background: Risk prediction models for health outcomes are used in clinical practice as part of medical decision-making. Given their importance in health care, it is essential that the performance of a risk model be assessed in an external validation dataset. An important aspect in the design of a validation study is the sample size, and broad recommendations suggest that 100 to 200 events are required.

Objective: In this work we investigate the sample size required to estimate measures of predictive performance with sufficient precision, or to detect differences from target performance values with sufficient power. We also explore how these requirements change with model strength and outcome prevalence.

Methods: We focus on the c-statistic and calibration slope, which are common measures of predictive performance. We first derive approximate variance estimators for these measures that only depend on the sample size and the anticipated values of the c-statistic and outcome prevalence. No patient-level data are required. Simulation is then used to investigate the performance of these estimators across a range of scenarios.

Results: Simulation results suggest that our variance estimators provide accurate estimates in most scenarios, even when key assumptions are violated. Smaller sample sizes are required with increasing model strength and decreasing outcome prevalence.

Conclusions: We have derived analytical expressions for precision and power-based sample size calculations based on the c-statistic and calibration slope. These may be used to determine a suitable sample size for a prospective external validation study. For example, to estimate the c-statistic with a confidence interval of +/- 0.05, we would require an external validation dataset with between 50 (high model strength) and 175 (low model strength) events.

[OC04.3] **StCA Winner**

Spatial predictive statistical model: detecting the need for treatment from retinal imaging data

Wenyue Zhu[1], Jae Yee Ku[1], Yalin Zheng[1], Paul Knox[1], Ruwanthi Kolamunnage-Dona[2], Gabriela Czanner[3]

[1] *Department of Eye and Vision Science, University of Liverpool, United Kingdom*
[2] *Department of Biostatstics, University of Liverpool, United Kingdom*
[3] *Department of Applied Mathematics, Liverpool John Moores University, United Kingdom*

Background: Retinal images are essential in the diagnosis and management of diabetic retinopathy. They contain spatial information and it is often important to identify the location of pathological change to guide treatment. However, statistical methods to reveal significant retinal locations are under-studied.

Objectives: To develop a spatial prediction framework which adjusts for clinically relevant covariates and identifies the most informative locations in retinal images.

Methods: We developed a spatial mixed-effects predictive model to analyse optical coherence tomography (OCT) retinal thickness data from the prospective observational Early Detection of Diabetic Macular Oedema study. The model considers the spatial correlations of different locations and the correlation between each participant's two eyes. Using Bayes theorem, we calculated the posterior probabilities using the marginal probabilities and prior information (the proportion of eyes receiving treatment). A discrimination rule was derived based on the posterior probabilities for each eye. The model was used to predict whether an eye needed treatment or not based on the retinal thickness from nine locations in the Early Treatment of Diabetic Retinopathy Study (ETDRS) grid centred on the fovea. Our model was validated using leave-one-out cross validation; Decision Curve Analyses (DCA) was used to select the most predictive locations in the retinal images.

Results: Data from 287 diabetic participants was included; 33 eyes received ocular treatment, 487 eyes did not. We compared performance of our spatial model to the currently used method, a Receiver Operating Characteristics (ROC) analysis based on central subfield thickness (CST) data only. We achieved a sensitivity of 0.81, a specificity of 0.78, and an area under the ROC curve (AUROC) of 0.86 (95% Delong's confidence interval: 0.78-0.95), while ROC analysis based on CST only has an AUROC of 0.58 (95% Delong's confidence interval: 0.48-0.68). We found that the temporal EDTRS subfields were the most informative area.

Conclusion: Our spatial predictive model showed promising levels of discriminative ability and identified the most clinically significant locations from OCT scans. The proposed method adjusts for spatial correlations and clinical covariates within the framework, and be used as a robust quantitative tool to support clinical decision making.

[OC04.4]

Tailored Bayesian risk prediction modelling

Solon Karapanagiotis, Paul Kirk, Paul J. Newcombe

MRC Biostatistics Unit, University of Cambridge, United Kingdom

Background: Risk prediction models are widely used in healthcare. The most common type of risk prediction is based on binary classifiers. They are often constructed to minimise the expected classification error; that is the proportion of incorrect classifications. The disadvantage of this approach is to implicitly assume that all errors cost the same. However, equality is but one choice, and an arbitrary one, which we suspect is in fact rarely appropriate. For example, in cancer diagnosis, a false negative (that is, misdiagnosing a cancer patient as healthy) could have more severe consequences than a false positive (that is, misdiagnosing a healthy patient with cancer); the latter may lead to extra medical costs and unnecessary patient anxiety but will not result in loss of life. For these applications, a prioritised control of asymmetric classification errors is desirable.

Objective: Develop a framework to incorporate misclassification costs into Bayesian inference.

Method: We use a decision theoretic approach to summarise the costs and benefits of (mis-)classifications into a single number, the target threshold. This is subsequently incorporated into a 2-stage model. In the first stage, we identify the most informative observations. An observation is treated as informative, if it is close to the target threshold of interest. We assign each observation a weight proportional to its distance from the target threshold. In the second stage, we use these weights to weight each individual's likelihood contribution when fitting the model.

Results and Conclusion: We conduct simulation studies to demonstrate the improvement in predictive performance of tailored Bayesian inference over standard classification paradigm. We then apply the methodology to two real data case studies and further demonstrate its potential for the clinical translation of prediction.

[OC04.5]

Personalised screening schedules for optimal prevention of cardiovascular disease

Francesca Gasperoni[1], Paul Newcombe[1], Christopher Jackson[1], Angela Wood[2], Jessica Barrett[1]

[1] *MRC-Biostatistics Unit, University of Cambridge, United Kingdom*
[2] *Cardiovascular Epidemiology Unit, University of Cambridge, United Kingdom*

Background: Cardiovascular disease (CVD) population screening strategies aim to identify and treat people at high risk of CVD. Current UK guidelines for CVD risk assessment recommend screening adults over 40 years old every 5 years and prescribing statins for those with a predicted 10-year CVD risk greater than 10%.

Objectives: The main goal of this work is to improve upon current screening practices, by providing a personalised screening schedule for each person, considering their specific risk profile.

Method: We develop a problem-specific utility function, accounting for: (i) event-free life years, (ii) cost to the health service of providing statins and cost of screenings. The idea of measuring benefit and costs of statin prescription in terms of event-free life years was proposed in [1], but they aimed at comparing prognostic models.

We optimize the previously defined utility function at different ages and according to 5-year CVD risk. To assess 5-year CVD risk for each person and to properly adjust for time varying endogenous covariates, we use a two-stage dynamic landmark model [2]. The first stage consists in fitting at each landmark age (i.e., 40,45,..,80 years) a multivariate linear mixed effect model with random intercepts and slopes. Using this model, we are able to predict risk factor values (i.e. cholesterol, blood pressure and smoke) at different ages. The second stage consists in predicting the CVD risk through a Cox model, adjusted for the risk factor values estimated at stage one.

Results and Conclusions: We apply the proposed model to data from the Clinical Practice Research Datalink (CPRD), comprising primary care Electronic Health Records from the UK. From preliminary analyses, baseline characteristics play a significant role on the optimal schedule. In particular, people labelled as high-risk seem to require more frequent visits, while low risk people seem to require visits less frequently than every 5 years.

References:

1. *Rapsomaniki, E., White, I. R., Wood, A. M., & Thompson, S. G. (2012). Statistics in medicine, 31(2), 114-130.*
2. *Paige, E., Barrett, J., Stevens, D., Keogh, R. H., Sweeting, M. J., Nazareth, I., Petersen, I. & Wood, A. M. (2018). American journal of epidemiology, 187(7), 1530-1538.*

OC05: Omics & Survival Analysis

Chair: **Jaroslaw Harezlak**

Indiana University, United States

[OC05.1]

How to predict a survival outcome using longitudinal and high-dimensional omic data

Mirko Signorelli[1], Roula Tsonaka[1], Pietro Spitali[2]

[1] *Medical Statistics, Leiden University Medical Center, Netherlands*
[2] *Human Genetics, Leiden University Medical Center, Netherlands*

Background: Longitudinal and high-dimensional measurements have become increasingly common in biomedical research. In survival analysis, longitudinal covariates are usually accounted for using joint models for longitudinal and time-to-event outcomes, while penalized survival models are employed to tackle high-dimensional sets of predictors. However, methods to predict survival outcomes using covariates that are both longitudinal AND high-dimensional are currently lacking.

Objectives: To overcome this limitation, we develop a method that allows to predict time-to-event outcomes using longitudinal and high-dimensional covariates. Our work is motivated by data from the MarkMD study, which was designed to understand if longitudinal biomarker data can be used to improve the prediction of time to loss of ambulation (LoA) in dystrophic patients. This study quantified 118 serum proteins in a longitudinal cohort of 157 dystrophic patients through an antibody-based bead array comprising 240 antibodies. Important features of these data are 1) the longitudinal nature of the predictors, 2) the presence of strong correlations between antibodies that target the same protein and 3) high-dimensionality.

Results: We propose a modelling approach whereby the longitudinal trajectories of antibodiesmeasuring the same protein are modelled using multivariate latent process mixed models (MLPMMs), which allow us to deal with problems (1) and (2). We derive from the MLPMM predicted protein- and antibody-specific random effects that we use to summarize the trajectories of the biomarkers. Then, we employ the (high-dimensional) summaries thus derived as predictors in an elastic-net penalized Cox model, from which we derive predictions of time to LoA. We first illustrate simulation studies to compare the predictive performance of our model to that of simpler prediction strategies, showing the advantage of fully exploiting the available information on the dynamic evolution of a large number of biomarkers. Then we apply our model to the MarkMD data, showing that predictions of time to LoA can be considerably improved using our approach.

Conclusions: We have developed a method for prediction of survival outcomes that can handle covariates which are both longitudinal and high-dimensional. The proposed method is computationally efficient and has been successfully applied to improve the prediction of time to LoA in dystrophic patients.

[OC05.2]

Improving the evaluation of COPD exacerbation treatment effects by accounting for treatment discontinuations

Agnieszka Król[1], Robert Palmér[2], Virginie Rondeau[3], Stephen Rennard[4], Ulf Eriksson[2], Alexandra Jauhiainen[2]

[1] *BioPharmaceuticals R&D, AstraZeneca, Poland*
[2] *BioPharmaceuticals R&D, AstraZeneca, Sweden*
[3] *University of Bordeaux, France*
[4] *BioPharmaceuticals R&D, AstraZeneca, United Kingdom*

Background: Chronic obstructive pulmonary disease (COPD) clinical trials aimed at evaluating long-term treatment effects on exacerbations frequently suffer from a high rate of early treatment discontinuations. This loss of information should ideally be considered in the statistical analysis of study results, particularly if the discontinuations are related to the disease or treatment itself.

Objectives: In this work, we quantify an association between the risks of exacerbation and treatment discontinuation in COPD clinical trials. By comparing different statistical methods for the analysis of repeated exacerbations, we also investigate whether disregarding this association can cause bias in exacerbation treatment effect estimates.

Methods: A joint frailty model describing the hazards of recurrent episodes of exacerbations and discontinuations was applied to analyse the association between the risks in data from five Phase III-IV COPD clinical trials. Specifically, the importance of the association when estimating exacerbation treatment effects was assessed by comparing the estimates of the joint frailty model to the negative binomial and shared frailty models, which both assume discontinuations to be unrelated to the trial outcome. This comparison was also made using simulated data.

Results: Simulations showed that, in the presence of early treatment discontinuations, the models disregarding the association risk produced biased results (>5 percentage point difference in exacerbation risk reduction). In the analysis of the clinical data, a statistically significant ($p<0.0001$) association between exacerbation and discontinuation risks was found in all trials. The differences in treatment effect estimates between the joint frailty and the other models was as high as 10-15 percentage points for some treatment comparisons. The important factors for the difference in estimates were the strength of the exacerbation-discontinuation association, the population heterogeneity in exacerbation risk, and the difference in treatment discontinuation rates between treatment arms.

Conclusions: We have found a significant association between the risks of exacerbation and treatment discontinuation in five COPD clinical trials and showed that this association can cause bias when estimating treatment effects. We recommend using the joint frailty model to account for this association in the estimation of exacerbation treatment effects, particularly when targeting the hypothetical estimand (ideal treatment effect).

[OC05.3]

Generalized Berk Jones test for Gene Set Analysis of time to event transcriptomics applied to recurrence risk

Laura Villain[1], Rodolphe Thiébaut[2], Andreas Bikfalvi[2], Boris Hejblum[3]

[1] *Inria, France*
[2] *Inserm, France*
[3] *Université de Bordeaux, France*

Background: Gene Set Analysis (GSA) is an analysis framework for genomics data that exhibits greater statistical power, and that is often biologically more relevant and easier to interpret than gene by gene analysis by leveraging prior biological knowledge in the form of pre-defined gene sets. The Generalized Berk-Jones (GBJ) testing method performs GSA while accounting for the intra-gene set correlation. Thus, it is particularly appealing when gene expressions are particularly correlated. However, this method is not yet applicable to the analysis of time to event data.

Objective: We extend the GBJ testing method to perform gene set analysis for time to event data, identifying gene sets whose expression is associated with survival (or any other event of interest).

Method: For each patient, the RNA-seq counts represent the gene expression of each gene from a biological sample as highly correlated high-dimensional data. The GSA approach tackles this high-dimensionality by combining gene level statistics together within a gene set. In particular, the GBJ test statistic relies on the individual score statistics computed for each gene across a gene set of interest, and uses a plugin estimate of the covariance matrix to account for correlation among genes of the same gene set. The GBJ testing method has been previously developed in the generalized linear model framework, and we extend it to GSA for time-to-event data by using the individual gene score statistics from the proportional hazard Cox model.

Results: We apply sGBJ to identify molecular biological signals associated with patient recurrence risk from tumor RNA-seq measurements in 71 patients having Lower Grade Glioma. Glioma are the most common primary brain tumors. Lower Grade Glioma mainly occur in the younger population with an invariable progression to a more malignant phenotype. Pilot data on glioma from TCGA have shown particularly high correlations among gene expressions from tumors, making our new survival GBJ (sGBJ) test particularly suited to this context.

Conclusion: Our extended sGBJ test unlocks state-of-the-art GSA methodology for time to event data and accounts for correlation within gene sets. We demonstrate its performance in a simulation study and on real data.

[OC05.4]

The controversial concept of biological age

Krista Fischer[1], Anastassia Kolde[1], Nele Taba[2], Erin Macdonald-Dunlop[3], Peter Joshi[3]

[1] *Institute of Mathematics and Statistics, University of Tartu, Estonia*
[2] *Institute of Genomics, University or Tartu, Estonia*
[3] *Centre for Global Health Research, Usher Institute, University of Edinburgh, United Kingdom*

Background: In the -omics data era, an increasing number of studies are focussed on the biological age (also called "omics clocks"). Given a set of age-related biomarkers, an intuitive explanation is that a biological age of an individual is conventionally defined as the average age of all individuals with a similar biomarker profile (the expected value of the age, conditional on the observed values of the biomarkers), estimated, for instance by classical multiple regression (sometimes combined with the principal component analysis for the biomarkers). Differences of the individual's actual age from the predicted biological age would mean that the person is "biologically older" or "younger" than his/her real age, that is naturally interpreted as having shorter or longer residual life expectancy, respectively, compared to an average person of the same age.

Objectives: In this talk it will be studied, under which assumptions the abovementioned interpretation could be plausible. We will also propose an alternative approach that enables to relax some of the assumptions.

Methods: First we will describe some realistic causal association structures where there are time-dependent observed or unobserved risk factors that affect the biomarkers. We will show how such confounders affect the estimation and interpretation of the "biological age". We propose an alternative definition of the biological age, using the concepts of survival analysis. Using simulations as well as the example of the Estonian Biobank, we compare both parametric (assuming the Gompertz distribution for survival time) and semiparametric approaches to the biological age estimation with the conventional method.

Results and Conclusions: We will show that when there are time-dependent risk factors that affect the biomarkers, younger or older "biological age" (when estimated as a regression prediction for age) does not necessarily correspond to lower or higher hazard levels and therefore is difficult to interpret. The approach based on survival analysis would lead to estimates that are more straightforward to interpret – however, this would require availability of survival data with sufficient number of cases.

[OC05.5]

Impact of genetic markers on time-to-event outcomes taking into account ambiguous measurements

Lars L. J. Van Der Burg[1], Henning Baldauf[2], Johannes Schetelig[3], Hein Putter[1], Stefan Böhringer[1], Liesbeth C. de Wreede[1]

[1] *Biomedical Data Sciences, LUMC, Netherlands*
[2] *CTU, DKMS, Germany*
[3] *Medizinische Klinik und Poliklinik I, University Hospital Dresden, Germany*

Background: A topic currently under investigation in the field of allogeneic hematopoietic stem cell transplantation is whether outcomes can be improved by matching unrelated donors to patients based on their KIR (killer-cell immunoglobulin-like receptors) genes. The KIR region is genetically complex, exhibiting copy number variations and a huge allelic diversity. Studies in this field are hampered by the large number of genetic predictors and ambiguous genotype calls, presenting a missing data problem.

Objectives: The objective of the current work was to explore several methods (1) to impute ambiguous KIR alleles to improve estimation of the total distribution, (2) to use these imputations in Cox models to estimate the impact of specific alleles. Several methods are investigated both in a simulation study and in a real data set, with data on donor KIR, patient characteristics and outcomes.

Methods: As reference, we took simple, maximum probability imputation. Next, the probability vector for ambiguous alleles was estimated by an Expectation-Maximization (EM)-algorithm. This was the basis for either a multiple imputation-based (MI) approach where draws based on posterior probabilities were generated, or the posterior probabilities were used as weights in the Cox models for both overall survival and relapse. Standard errors of the regression coefficients were calculated by Rubin's rules, ignoring weighting, the jackknife estimator and bootstrap. Simulations were based on the real data set.

Results: Maximum probability imputation led to bias in the estimated genotype distributions and regression coefficients. The EM-based methods performed better, but were not satisfactory in the case of many ambiguities and only single genes to base the estimation on; dilution due to imputation tended to bias the regression coefficients in our setting. The performance of the EM-based methods was similar in most cases, both in the real and simulated data.

Conclusions: The bias in regression coefficients associated with imputing alleles can only be removed if a rich set of auxiliary information is taken into account. Analyzing several genes simultaneously to borrow information is a possible approach. Extensive and realistic simulation studies are necessary to assess the impact of ambiguity in clinical data sets.

OC06: Novel Applications of Survival Analysis

Chair: **Georg Heinze**

Medical University of Vienna, Austria

[OC06.1]

Estimating conditional cumulative incidence in the presence of an internal time-varying exposure

Ping Liu[1], Giuliana Cortese[2], Robert Quinn[1], Pietro Ravani[1]

[1] *Departments of Medicine and Community Health Sciences, University of Calgary, Canada*
[2] *Department of Statistical Sciences, University of Padova, Italy*

Background: Cause-specific cumulative incidence functions (CIFs) are the preferred method for time-to-event data in the presence of competing risks. CIFs are often used to summarize the effect of an exposure measured at baseline on each cause of failure. When the exposure changes over time, simply ignoring the time-varying nature of the exposure during follow-up time is misleading. One strategy is to estimate conditional CIFs.

Objective: To estimate conditional CIFs for kidney failure and death by receipt of nephrology consultation in an illustrative example study of adults with severe chronic kidney disease (CKD).

Methods: We did a population-based cohort study of adults with stage 4 CKD, diagnosed between 2002 and 2014 in Alberta, Canada. The exposure was the receipt of outpatient nephrology consultation. The events of interest were kidney failure and death without kidney failure. We excluded patients who had a history of kidney failure or an outpatient nephrologist encounter in the 2 years prior to study entry. We obtained conditional CIFs indirectly from cause-specific Cox models to summarize absolute risks over time, under two conditions: the risk to develop an event before or at time t, given that a patient did not have nephrology consultation and was event-free until time s (s<t); and the risk to develop an event before or at t, given that a patient did have nephrology consultation at or before s and was event-free until time s.

Results: Of the 14,382 participants, 33% saw a nephrologist as an outpatient, 9% developed kidney failure, and 53% died over a median follow-up of 2.6 years. Compared with participants who did not see a nephrologist before or at 7 months (median time to consultation), those who did experienced a higher cumulative incidence of kidney failure [5-year risk (95% CI) 0.23(0.21-0.24) vs 0.07(0.065-0.075), and a lower cumulative incidence of death [0.35(0.33-0.37) vs 0.57(0.56-0.58)], regardless of age and other key prognostic factors.

Conclusions: Our study estimates conditional CIFs in the presence of an internal time-varying exposure and suggests that people who see a nephrologist are less likely to die and more likely to develop kidney failure.

[OC06.2]

Random Cancers as Supported by Registry Data

Janez Stare[1], Robin Henderson[2], Nina Ruzic Gorenjec[1]

[1] *Institute for Biostatistics and Medical Infromatics, University of Ljubljana, Faculty of Medicine, Slovenia*
[2] *Newcastle University, School of Mathematics, Statistics and Physics, United Kingdom*

Background: There has been considerable interest in recent years in quantifying the probability of unavoidable or so-called random cancers, as opposed to cancers linked to environmental, genetic or other factors. Our work was particularly motivated by Tomasetti et al. (2017) with estimates obtained from a random mutations argument which depend on biological assumptions and knowledge related to risk factors.

Objectives: The actual proportion of cancers that are unavoidable cannot be determined from observational data. However, our aim is to find an upper bound to the incidence of cancers that are random and unavoidable by analysis of multiple registry data, and to compare these bounds to values tabulated by Tomasetti et al.

Methods: We propose a data-based approach to estimate an upper bound to this probability, based on an analysis of multiple registry data. The argument is that the cumulative hazards for random cancers cannot exceed the minimum reliable cumulative hazard observed across the registries. We propose a Monte Carlo method to identify this upper bound and provide simulation results. We apply the method to registry data on nine different cancer sites recorded by 423 registries using IARC's (International Agency for Research on Cancer) Cancer Incidence in Five Continents (CI5), Volume X, database.

Results: Simulation results show that our method is valid though conservative, in the sense that it overestimates the true probability of random cancer. While for six of nine cancer sites our estimates are in line with the Tomasetti et al. values, in the sense that they are higher, our upper bound estimates for prostate, breast and colon cancers are much lower.

Conclusions: Differences in reporting are almost certainly responsible for some of the variability in the reported incidences, nevertheless, Tomasetti et al. conclusion that high numbers of prostate, breast and colon cancers are completely random is not supported by the data. More importantly, we provide a valid method for estimating an upper bound to the probability of random cancer that is based solely on the data without biological assumptions or risk factor estimates.

[OC06.3]

Comparison of methods to prevent immortal time bias for frequently recurring outcomes

Michelle Samuel[1], Vidal Essebag[2], Marie-Eve Beauchamp[3], Louise Pilote[3], Michal Abrahamowicz[3]

[1] *Cardiology, Montreal Heart Institute, Canada*
[2] *Cardiology, McGill University Health Center, Canada*
[3] *Center for Outcomes Research and Evaluation, Research Institute of McGill University Health Center, Canada*

Background: Left truncation (LT) and time-varying exposure (TVE) methods are established approaches to prevent resulting immortal time bias (ITB). Simulation and real-world design comparisons suggest both approaches yield similar effect estimates, however, these studies only investigated one-time or long-term events, such as mortality or cancer. Whether these approaches result in similar estimates in the setting of frequently recurring outcomes has yet to be investigated.

Objectives: To compare incidence rates and effect estimates between ITB prevention methods, LT and TVE, for fatal and frequently recurring non-fatal outcomes illustrated with a real-world example evaluating the association between treatment with catheter ablation (CA) for atrial fibrillation (AF) and 1) all-cause mortality and 2) first heart failure (HF) hospitalization among AF-HF patients.

Methods: Quebec administrative data (2000-2017) was used. In the LT design, CA patients were matched 1:2 to controls on time and frequency of hospitalizations. Follow-up began at treatment initiation or matched date (controls). In the TVE model, follow-up for all patients began at disease onset. Both analyses employed multivariable Cox regression models.

Results: Of 101,931 patients in the AF-HF cohort, all were included in the TVE cohort and 1,350 patients were included in the LT cohort. The number of CA patients (N=451), incidence rate (1.2 per 100 person-years), and aHRs [LT aHR 0.4 (95% CI 0.2-0.7) and TVE aHR 0.5 (95% CI 0.3-0.7)] were similar between approaches for all-cause mortality. However, for the non-fatal outcome of first HF hospitalization, there were fewer CA patients and the incidence rate was reduced with the TVE approach (367 CA patients; 3.6 per 100 person-years) compared to LT (451 CA patients; 4.6 per 100 person-years). Further, only the TVE modeling resulted in a protective effect of CA for HF hospitalizations [LT aHR 1.2 (95% CI 0.8-1.7) and TVE aHR 0.8 (95% CI 0.6-1.0)].

Conclusions: In the setting of frequently recurring outcomes, the use of a TVE variable overestimates the effectiveness of treatment due to selection bias from healthier patients remaining event-free prior to treatment. Therefore, a LT design is more appropriate for frequently recurring outcomes. Researchers should consider the outcome being investigated when deciding on a prevention method for ITB.

[OC06.4]

Flexible modelling of recency-weighted cumulative opioid exposure provides new insights on the opioid safety

Siyana Kurteva[1], Robyn Tamblyn[1], Michal Abrahamowicz[1], Marie-Eve Beauchamp[2]

[1] *Epidemiology, McGill, Canada*
[2] *McGill, Canada*

Background: Multiple efforts from policy makers, public health agencies and the medical community have been made to stem the opioid epidemic. However, research linking opioid prescribing to adverse events suffers major methodological limitations and there is a need to strengthen the current opioid safety evidence.

Objectives: Follow a prospective cohort of hospitalized patients within one year post-discharge to study the association between opioid use and the risk of acute healthcare events (ED visits/re-admission or death) by using novel modelling techniques.

Methods: Marginal structural Cox PH (MSM Cox) models were used to determine the association between time-varying opioid current use and the risk of the outcome while adjusting for time-varying confounders that may be affected by prior exposure to opioids.MSM Cox models were estimated with truncated stabilized time-varying inverse probability treatment (IPT) weights. We also estimated the association of interest with flexible weighted cumulative exposure (WCE) models, adapted to and validated for MSM Cox analyses, defined as the weighted sum of past exposure. The weight function was estimated using cubic splines to avoid a priori assumptions regarding its shape.

Results: A cohort of 1511 patients with at least one opioid dispensation during the follow-up was formed (mean age 69.6, SD = 10.3). The conventionalMSM model showed current daily opioid use to be associated with an almost 70% increased risk of opioid-related acute healthcare events, adjusted hazard ratio (aHR): 1.66, 95% CI (1.17–2.36). On the other hand, the WCE model showed a more than two-fold increase (aHR: 2.30). The estimated weight function suggests that opioid use in the last 30 to 40 days has the highest impact on the current risk of opioid-related healthcare events. The best-fitting model was the WCE accounting for opioid exposure over the past 120 days, with a Akaike information criterion(AIC) difference of 59.6 when compared to the conventional MSM model.

Conclusion: The method of recency-weighted cumulative opioid use showed how careful consideration of modelling current and past exposure, and their timing, may improve the model's fit and enhanced our understanding of the mechanism underlying potential adverse events of exposure to opioids.

[OC06.5]

Defining impaired olfaction in Parkinson's disease: assessing agreement between four published methods

Sofia Kanavou[1], Vanessa Pitz[2], Michael A. Lawton[1], Naveed Malek[3], Katherine A. Grosset[2], Yoav Ben-Shlomo[1], Donald G. Grosset[2]

[1] *Population Health Sciences, University of Bristol, United Kingdom*
[2] *Institue of Neuroscience and Psychology, University of Glasgow, United Kingdom*
[3] *Department of Neurology, Queen's Hospital, Romford, United Kingdom*

Background: Impaired olfaction (hyposmia) is a common non-motor feature in patients with Parkinson's disease (PD) that develops at an early stage. However, there is no standard method to define hyposmia in PD, making comparative research problematic.

Objective: To compare four published methods of defining hyposmia using 2,097 cases of early Parkinson's disease (PD) from 2 prospective studies, and examine why standardisation has proven difficult.

Methods: Olfaction was measured using findings from the 40-item University of Pennsylvania Smell Identification Test (UPSIT) or the 16-item Sniffin' Sticks test (Sniffin), harmonised using the Item Response Theory method. The proportion of hyposmic patients from UPSIT was calculated using 4 previously reported methods, corrected by age and/or gender. Normative (control) olfactory data were simulated, based on the age and gender structure of our PD cases, and published normal ranges. Association with age, gender and disease duration was explored by method and study cohort. Between method agreement was measured using Cohen's kappa and Gwet's AC1.

Results: The proportion classified as hyposmic ranged from 69.1%-97.9% of cases in Tracking Parkinson's, and 62.2%-90.8% of cases in the Parkinson's Progression Marker Initiative, depending on test method. Agreement among methods varied (kappa 0.09-0.80, AC1 0.55-0.86). Gwet's AC1 was often substantially higher than the kappa coefficient, e.g. 0.74 vs 0.11 (Methods 1 vs 2). The absolute difference in hyposmia between PD cases and simulated controls was similar for men and women across methods. Age and male gender were positively associated with hyposmia for each method ($p<0.001$). Increased odds of having hyposmia were detectable in relation to age (odds ratio 1.06, 95% CI 1.03, 1.10, $p<0.001$). Although longer disease duration had a negative impact on overall olfactory performance, no association with hyposmia was found.

Conclusion: The proportion of PD patients classified as hyposmic varies significantly according to the definition applied. Correction for age and gender is required when assessing hyposmia in early PD, to account for the background decline in olfactory performance with ageing, especially in men.

OC07: Clinical Trials Methodology 2

Chair: **Mimi Kim**

Albert Einstein College of Medicine, United States

[OC07.1]

Time-to-event model-assisted designs to accelerate and optimize early-phase immunotherapy trials

Ruitao Lin, Ying Yuan

Biostatistics, The University of Texas MD Anderson Cancer Center, United States

Immunotherapies and molecularly targeted agents have revolutionized cancer treatment. Unlike chemotherapies, these novel agents often take a longer time to show responses. This causes major logistic difficulty for implementing existing adaptive trial designs, which require the observance of the outcome early enough to apply data-adaptive decisions for new patients. In this talk, I will introduce a novel class of Bayesian adaptive designs, known as time-to-event model-assisted designs, to address this practical challenge in phase I dose-finding trials with late-onset toxicity. A unified methodology based on a novel formulation and approximation of the observed data likelihood will be introduced to facilitate seamless, real-time decision making. The dose escalation/de-escalation rules of the proposed designs can be tabulated before the trial begins, which greatly simplifies trial conduct in practice compared to that under existing methods. I will present some theoretical and numerical results to show the desirable properties of the proposed designs. Last, I will introduce user-friendly software for implementing the designs.

[OC07.2]

Dynamic balancing in platform trials

Megan Othus

Biostatistics, Fred Hutchinson Cancer Research Center, United States

In oncology, the use of basket, umbrella, and platform trials is becoming more common. The efficiency gains from platform trials, with a shared control arm, have been previously quantified in the setting of experimental arms being added and dropped in a dynamic manner. The SWOG Cancer Research Network Leukemia Committee designed a platform trial for patients with acute myeloid leukemia. The experimental treatments have varying mechanisms of action and varying toxicity profiles. To maximize potential accrual for this rare cancer, the study team decided to allow eligibility to vary across experimental arms. The control arm has no additional eligibility criteria, and patients are required to be eligible for at least one experimental arm. The study team wished to use dynamic balancing for randomization. Here we will present an algorithm to perform dynamic balancing in a setting of a platform trial with a shared control arm and eligibility criteria that vary across experimental arms. We will also describe how block randomization can be accomplished in this setting. In addition, we will discuss randomization choices when new arms are added. We will describe the efficiency trade-offs encountered in the trial to date.

[OC07.3]

Optimal unplanned recalculation in adaptive two-stage designs

Maximilian Pilz[1], Carolin Herrmann[2], Geraldine Rauch[2], Meinhard Kieser[1]

[1] *Institute of Medical Biometry and Informatics, University of Heidelberg, Germany*
[2] *Institute of Biometry and Clinical Epidemiology, Charité Berlin, Germany*

Adaptive designs for clinical trials are an attractive option to modify the design during the ongoing study. The conditional error principle allows modification of the trial design at any trial-internal time without type I error rate inflation. However, recent research addresses the question of how "adaptive" flexible designs actually should be or whether „potential decisions made at interim stages might not be better placed to the upfront planning stage" [1].

It has been shown that during the planning phase of a clinical trial, an adaptive two-stage design can be constructed by optimizing a trial-specific objective criterion (e.g., expected sample size) under specified constraints (e.g., type I error rate, power) [2]. Those designs are then fully specified before the trial starts, and any deviation from the optimally planned designs implies a subpar performance. Therefore, modifying an optimal design for trial-internal reasons should be avoided. It may, however, happen that new trial-external information arises during the ongoing trial and makes the initial planning assumptions obsolete. In those situations, recalculating the trial design may become inevitable.

In this talk, we apply the benefits of the conditional error principle to ensure type I error rate protection for the case of unplanned interim analyses. The situation that new trial-external information makes an update of the planning assumptions necessary is embedded in the framework of Bayesian theory and can, therefore, be interpreted as an update of the prior knowledge that was available when initially planning the trial design. We demonstrate how such an unplanned interim analysis can be performed in an optimal manner. This is done by formulating a new optimization problem that incorporates control of the conditional type I error rate. The problem and its solution are shown for both the cases that the unplanned interim analysis is performed before or after the initially planned interim analysis. By means of a clinical trial example, the concrete conduct of an optimal unplanned interim analysis is illustrated.

References:

1. *Bauer et al (2016). Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. StatMed; 35:325–347.*
2. *Pilz et al (2019). A variational approach to optimal two-stage designs. StatMed; 38:4159–4171.*

[OC07.4]

An evaluation of the use of covariate constrained randomisation for stepped-wedge cluster randomised trial

Caroline Kristunas[1], Laura Gray[2], Fan Li[3], Karla Hemming[4]

[1] *Institute of Clinical Sciences, University of Birmingham, United Kingdom*
[2] *Department of Health Sciences, University of Leicester, United Kingdom*
[3] *Department of Biostatistics, Yale University, United States*
[4] *Institute of Applied Health Research, University of Birmingham, United Kingdom*

Background: When only a small number of clusters are being randomised, as is often the case in stepped-wedge cluster randomised trials (SW-CRTs), simple randomisation (SR) may struggle to balance key cluster-level covariates across intervention and control conditions. Covariate constrained randomisation (CCR) scores the balance of the covariates for each randomisation scheme, from all possible schemes (randomisation space), and restricts the space to those with the best balance. CCR has been shown to be beneficial for parallel CRTs, but has not previously been evaluated for SW-CRTs.

Objective: To compare CCR to SR for SW-CRTs, in terms of power and maintenance of the type I error rate.

Methods: Data were simulated from a linear mixed model for two SW-CRT designs, each with four prognostic cluster-level covariates. The randomisation space for each SW-CRT was generated, and two balance metrics used to score the balance of each scheme in the space. Data were then analysed under each scheme in the randomisation space using a linear mixed model, either adjusted, or unadjusted, for the cluster-level covariates. To evaluate power, the proportion of true positives was calculated either over the entire randomisation space (SR), or over a candidate set of schemes with the best scores (CCR). To evaluate the type I error rate, multiple randomisation spaces were simulated, a scheme selected by SR or CCR, and the proportion of false positives calculated.

Results: The type I error rate was generally well maintained under SR and CCR. Although, when CCR balanced all four covariates, and the analysis fully adjusted, the error rate was inflated (not seen in larger SW-CRTs). Under CCR, there was a small gain in power over SR, provided the analysis adjusted for all covariates that were balanced in the randomisation. The greatest power gain from CCR was observed when all four covariates were balanced on, compared to if a scheme with the worst balance was selected.

Conclusions: CCR with an adjusted analysis can maintain the type I error rate for SW-CRTs, whilst protecting against the loss of power observed if a randomisation scheme with poor balance is selected under SR.

[OC07.5]

Multi-Arm Multi-Stage Design for Ordered Treatments

Alessandra Serra, Pavel Mozgunov, Thomas Jaki

Mathematics and Statistics, Lancaster University, United Kingdom

In infectious diseases such as Tuberculosis or Hepatitis B, the period of treatment with standard regimes is typically around 6 and 12 months, respectively. Novel treatments, however, offer the opportunity for higher efficacy and shorter treatment periods. Phase II designs that efficiently compare different treatment durations and reduce the overall sample size, would accelerate the progress of new drugs by ensuring that only the optimal doses and durations are selected.

One family of designs that can noticeably improve the decision-making in Phase II trials is Multi-Arm Multi-Stage (MAMS) designs. They allow for studying several arms in parallel, and for dropping treatment arms or stopping the trial earlier. MAMS designs, however, were conventionally developed for the setting of independent treatment arms. Therefore, they can be suboptimal when an order among the treatment effects is known. For example, in the trial with different treatment durations, in which it can be assumed that longer treatment durations correspond to higher efficacy.

In this work, we propose a MAMS design that employs information about the treatment durations and can identify the shortest promising treatment duration with high probability while maintaining control of the family-wise error rate. The design takes into account the ordering of treatments in the set of decisions that could be made at each interim analysis and does not use any parametric model for the duration-efficacy effect.

Via a comprehensive simulation study, the proposed design was compared to the fixed sample and standard MAMS designs in terms of the power to find the shortest treatment duration and the expected sample sizes.

It was found that the developed design can provide noticeable advantages in power and/or expected sample sizes required in the trial compared to the fixed sample and standard MAMS designs. For example, compared to the fixed sample design, the MAMS design that accounts for the order of treatments was found to provide a 5-10% increase in power to find the shortest treatment duration.

Therefore, we conclude that the inclusion of the information about the ordered treatments can inform better decision-making and requires fewer patients in the trial.

OC08: Causal Inference 2

Chair: **Saskia Le Cessie**

Leiden University, Netherlands

[OC08.1]

Comparing and expanding potential-outcome models for average treatment effect estimation in cardio-oncology

Lucy Teece[1], Michael Sweeting[1], Catherine Welch[1], David Adlam[2]

[1] *Biostatistics, University of Leicester, United Kingdom*
[2] *Cardiovascular Sciences, University of Leicester, United Kingdom*

Background: Evidence on whether the management of patients following acute myocardial infarction (AMI) differs between cancer survivors and those without cancer is inconclusive. The Virtual Cardio-Oncology Research Initiative (VICORI) links national cancer registration data with the UK Myocardial Ischaemia National Audit Project (MINAP) registry, allowing a unique opportunity to investigate outcomes and management of cardio-oncology patients in England.

Potential differences in patient management following AMI between cancer and non-cancer patients may be due to confounding, i.e. differences in patient demographics (age, sex, social economic factors etc.). Under certain assumptions, potential-outcome models eliminate the effects of these confounders on patient management, enabling estimation of the direct effect of a previous cancer diagnosis.

Objectives: investigate existing potential-outcome models to incorporate complex exposures and multiple correlated outcomes; determine whether measures of cancer severity are directly related to differences in patient management following AMI.

Methods: Potential-outcome models, such as multivariable regression adjustment, inverse-probability weighting (IPW), augmented IPW and IPW regression adjustment, enable estimation of average treatment effects in a causal framework. These models have been extended to incorporate more complex exposure information and to consider multiple outcomes simultaneously. We compare these approaches and investigate the impact of multi-valued exposures, including cancer type and stage, on multiple correlated quality indicator outcomes.

We identified 512,388 patients with AMI hospitalisations recorded in MINAP between January 2010 and March 2018. Of these 42,187 (8.2%) had a previous cancer diagnosis recorded in the national cancer registry. Median age of patients with no previous cancers and cancer survivors were 69.4 (58.6,79.9) and 77.3 (69.8,83.6) respectively.

Results: We present a comparison of average treatment effect estimates and E-values of cancer exposure and severity on the management of patients following an AMI, measured using multiple quality indicator measures. Strengths and limitations of the potential-outcome models are discussed.

Conclusions: Extensions to existing potential-outcome models allow for the assessment of complex exposures on multiple outcomes. These findings highlight differences in access to guideline-indicated care following an AMI between cancer and non-cancer patients. Improvements to the management of cancer patients following an AMI has the potential to reduce variation in death following AMI.

[OC08.2]

On the relationship between association and surrogacy when both the surrogate and true endpoint are binary.

Paul Meyvisch[1], Ariel Alonso[2], Wim Van der Elst[3], Geert Molenberghs[4]

[1] *Biostatistics, Galapagos NV, Belgium*
[2] *L-Biostat, Katholieke Universiteit Leuven, Belgium*
[3] *Statistical Decision Sciences, Janssen Pharmaceutica (J&J), Belgium*
[4] *I-Biostat, University of Hasselt, Belgium*

The relationship between association and surrogacy has been the focus of much debate in the surrogate marker literature. Recently, the individual causal association (ICA) has been introduced as a metric of surrogacy in the causal inference framework when both the surrogate and the true endpoint are normally distributed or, alternatively, when both endpoints are binary. Earlier work on the normal case has demonstrated that, although the ICA and the adjusted association are related metrics, their relationship strongly depends on unidentifiable parameters and, consequently, the association between both endpoints conveys little information on the validity of the surrogate. In addition, in the normal setting, the magnitude of the ICA does not depend on the observed mean outcomes. The latter implies that identifiable parameters such as mean responses and treatment effects provide no information on the validity of the surrogate. In the present work, it is shown that this is fundamentally different in the binary case. We demonstrated that the observed association between the outcomes, as well as the success rates in both treatment groups are quite predictive for the ICA. It is shown that finding a good surrogate will be more likely when the association between the endpoints is large, there are sizeable treatment effects and the success rates for both endpoints are similar in both treatment groups. These results are demonstrated using extensive simulations and are illustrated on a case study in multi-drug resistant tuberculosis.

[OC08.3]

Effect Estimation By A Boosted Doubly Robust Method

Wenxi Yu[1], Greg Ridgeway[2], William Marder[3], William Finkle[4]

[1] *Department of Biostatistics, University of California, Los Angeles, United States*
[2] *Department of Criminology and Department of Statistics, University of Pennsylvania, United States*
[3] *IBM Watson Health, United States*
[4] *Consolidated Research Inc., United States*

Background: Estimating treatment effects in observational studies with numerous correlated covariates is challenging. The choice of a covariate adjustment method will affect statistical results.

Objectives: We illustrate a method, the boosted doubly robust regression (BDR), that can effectively handle numerous correlated covariates and reduce bias and variance in the effect estimates.

Methods: BDR combines the merits of boosted propensity scoring (BPS) and doubly robust (DR) estimation. BDR first uses BPS to reweight the patients in the study to adjust for pre-treatment covariate imbalances. Then, BDR uses an outcome model to adjust for confounding left uncontrolled by BPS. The BDR estimates are consistent if the exposure or the outcome model are correctly specified. Statistical arguments and simulation experiments suggest that BPS reduces the bias in the treatment effects estimates compared with pure outcome-modeling methods, and reduces variance compared to pure propensity-scoring methods. To illustrate effects of adjustment methods, we examined the established associations with injury of diazepam, a long-acting benzodiazepine, and alprazolam, a short-acting benzodiazepine, comparing BDR and conventional adjustment methods including Poisson regression, propensity scoring using logistic regression (LPS), and high dimensional propensity scoring (HDPS).

Results: The study included 78,829 and 118,579 patients with a prescription for diazepam or alprazolam respectively in the IBM MarketScan Database from 2011 to 2018. We compared risk of injury in the 1-15 days following the initial prescription (post-treatment) with the risk of injury in the 1-365 days prior to treatment (pre-treatment) and computed the post- to pre-treatment rate ratio (RR) in each treatment cohort. We computed the ratio of the RRs (RRR) for the treatment and comparison cohorts using all methods. BPS successfully balanced the distributions of pre-treatment covariates between the diazepam and alprazolam groups. Also, the 95% compatibility ("confidence") intervals for BDR were narrowest, consistent with theoretical arguments that BDR reduces variance. The RRRs were 2.85 (0.82, 9.90) with no adjustment, 2.86 (0.89, 9.15) for Poisson regression, 3.27 (1.66, 6.44) for LPS, 3.28 (1.67, 6.46) for HDPS and 2.27 (1.17, 4.39) for BDR.

Conclusions: Compared with conventional methods, BDR improves balance in the pre-treatment covariates, reduces bias and variance with implications for evaluation of drug safety and regulatory policy.

[OC08.4]

Causal inference with multiple versions of treatment and application to personalized medicine

Jonas Béal, Aurélien Latouche

U900, Institut Curie, France

Background: The development of high-throughput sequencing and targeted therapies has led to the emergence of personalized medicine: a patient's molecular profile and the presence of specific biomarker(s) of drug response will correspond to a treatment recommendation made either by a physician or by a treatment assignment algorithm.

Objective: The growing number of such algorithms raises the question of how to quantify their clinical impact knowing that a personalized medicine strategy will intrinsically include different versions of treatment.

Methods: We thus specify a causal framework with multiple versions of treatment to define the causal effects of interest for precision medicine strategies and estimate them emulating clinical trials with observational data. Therefore, we determine whether the treatment assignment algorithm is more efficient than different control arms: gold standard treatment, observed treatments or random assignment of the same targeted treatments. These methods are implemented in available R scripts and interactive RShiny application.

Results: Standardized estimates of the precision medicine effects are first evaluated on simulated data and they demonstrate a lower bias compared with naive estimation of the difference in outcome between treatment arms. The various simulations scenarios also point out the different bias sources depending on the clinical situation (biomarker prevalence, assignment of observed treatments etc.).

The method is then applied to public data from patient-derived xenografts (PDX): each patient tumour is implanted in several immunodeficient cloned mice later treated with different version of treatment/drugs, thus providing access to all corresponding drug sensitivities for each patient. Such data provides access to treatment response values otherwise considered as counterfactual. Availability of these data ensures that the positivity condition holds and provides the opportunity to validate the reliability of our estimates. Examples are provided for the causal evaluations of different plausible treatment assignment strategies, based on mutations of tumours and involving more than 100 PDX models.

Conclusion: The need for estimation methods adapted to precision medicine is therefore supported by this work, which highlights the biases of simpler methods and opens the door to a better use of pre-clinical and clinical data already generated to evaluate future precision medicine strategies.

[OC08.5]

Propensity Weighting in the Estimation of Direct Effects

Christiana Drake[1], Julie Smith-Gagen[2]

[1] *Department of Statistics, University of California, Davis, United States*
[2] *School of Community Health Sciences, University of Nevada at Reno, United States*

Cancer of unknown primary (CUP) is diagnosed when metastases are detected but a primary tumor cannot be found. Some oncologists believe CUPs are not simply tumors where the primary tumor is not detectable but a different type of tumor altogether. We propose to investigate survival following as a direct effect following a diagnosis of CUP with treatment as a mediator. A direct effect of a CUP diagnosis would support such a theory. We will compare confirmed CUPs to CUP diagnosed without all the required diagnostics. We propose to use the tools of mediation analysis using principal stratification (Rubin and Frangakis 2002,2004) and propensity weighting. A comparison of the methods is carried out with the counterfactual approach by Vanderweele (2015) in a simulation study. The data analysis carried out on a SEER data set shows a moderate direct effect. The primary advantage of principal stratification with propensity weighting is the need for only one counterfactual dependent on treatment. It does not require a counterfactual for the mediator.

OC09: Survival Analysis 1

Chair: **Richard J. Cook**

University of Waterloo, Canada

[OC09.1]

Prediction models in bone sarcoma - a simulation study to compare Cox models with Survival Neural Networks

Georgios Kantidakis[1], Hein Putter[1], Saskia Litiere[2], Hans Gelderblom[3], Marta Fiocco[4]

[1] *Biomedical Data Sciences, Leiden University Medical Center (LUMC), Netherlands*
[2] *Statistics, European Organization for Research and Treatment of Cancer (EORTC), Belgium*
[3] *Medical Oncology, Leiden University Medical Center (LUMC), Netherlands*
[4] *Statistical Science, Mathematical Institute Leiden University (MI), Netherlands*

Background: Osteosarcoma is the most common primary bone cancer and the 3rd most frequent cancer in adolescents. Fortunately, the introduction of neoadjuvant chemotherapy in 1970's increased survival remarkably with a 5-year overall survival rate currently above 65%. Selecting the best prediction model is of great clinical interest because of the stagnation in standard of care for more than a decade. New prediction models such as neural network extensions to right-censored survival data should be explored to investigate their potential.

Objective(s): There is a strong discussion in the medical field about machine learning and whether it has greater potential than traditional statistical models for clinical trials. Criticism is related to unsuitable performance measures and lack of validation.

This study's primary aim is to compare Cox models with survival neural networks in terms of prediction and estimator properties through a simulation study.

Method(s): Data are generated based on a European Osteosarcoma Intergroup study (MRC BO06/EORTC 80931). Neural networks are applied to simulated datasets (250 or 1000 observations) with 5 prognostic factors selected to predict survival. Comparison is performed between Cox models and Partial Logistic Artificial Neural Networks (PLANNs) for different amount of censoring (20, 40, 61.5 and 80%) with survival times generated from a log-normal or an exponential distribution. For PLANNs, novel extensions are provided. Clinical endpoint is overall survival (defined as the time to death since surgery).

Results: A prognostic score is defined using PLANNs based on patient characteristics per time-interval. The simulation study shows how this score can improve C-index compared to Cox models. Furthermore, the methods are compared by using Brier score at 2 years and Integrated Brier Score. PLANNs exhibit different behaviour under erroneous assumptions on training data (removing patients or curtailing survival) and the performance is more sensitive to sample size and amount of censoring compared to Cox models.

Conclusions: Neural networks achieved better performance than Cox models for a number of scenarios. This work shows that survival neural networks can be a valuable tool for prediction as long as researchers are aware of the delicate aspect of parameter tuning and their behaviour in different settings.

[OC09.2]

An excess hazard regression model based on a general structure and with individual frailty

Aurelien Belot[1], Francisco Javier Rubio[2]

[1] *Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, United Kingdom*
[2] *Department of Mathematics, King's College London, United Kingdom*

Background: When analysing population-based cancer registry data, the cause of death is usually missing and/or unreliable. In this situation, excess mortality hazard models enable estimating cause-specific quantities. The main idea is to borrow information from life tables to estimate an excess mortality hazard, which can be attributed to the disease under study. Many regression models have been developed for excess mortality hazard but regression models with individual frailty to account for unobserved individual heterogeneity have not been studied.

Objective: We present an excess hazard regression model based on a general structure, using flexible parametric distributions, and which includes an individual frailty term in order to account for unobserved heterogeneity.

Method: We consider a general structure for the hazard and use the Power Generalized Distribution for modelling the baseline hazard, which enables capturing all basic shapes of the hazard (Iincreasing, decreasing, unimodal (up-then-down) and bathtub (down-then-up), and constant). We assume a Gamma distribution for the individual frailty term, thus leading to a closed-form Laplace transform and tractable expression for the likelihood function. Parameter estimates are obtained through the maximum likelihood method. We investigate the performance of the model in a simulation study, which includes one scenario where one covariate was used to simulate event times but was omitted in the analysis step.

Results: We demonstrate good performance of the proposed model under a variety of simulation scenario. We apply our approach to a real data set on women diagnosed in 2012 in England with a Non-Small Cell Lung Cancer cancer, which provides new insights into these data. Estimation and interpretation of these new insights are discussed.

Conclusions: This methodological development gives the analyst an additional tool in the statistical toolkit for analysing population-based cancer registry data, accounting for unobserved heterogeneity, and using a parametric but flexible distribution and relying on a general hazard structure that encompasses many different models as special cases (notably AFT and PH regression models).

[OC09.3]

Estimating survival functions using growth curve modelling with thresholds

Jennifer Hellier, Richard Emsley, Andrew Pickles

Biostatistics and Health Informatics, Kings College London, United Kingdom

Background: Time to event outcomes are typically modelled using survival based models. Growth curve models provide an alternative flexible framework to model repeated measures, including time to reach a threshold. We have developed an approach to estimate a survival function from a growth curve model. This approach captures the clinical complexity of repeated observations and also benefits from the flexible growth curve modelling framework.

Objective: We evaluate whether this novel application of growth curve modelling to model time to reach a threshold provides a valuable tool by (1) estimating an unbiased survival function compared to standard survival approaches and (2) provides a more efficient measure of treatment effects on the survival function.

Methods: We demonstrate an implementation of the method in Stata and validate its properties by Monte Carlo simulations that test various study designs, disease progression and measurement error rates. We illustrate the method using a clinical trial (GENDEP), analysing time to reach a threshold of depression remission.

Results: We show that the model and estimation procedure can generate comparable survival functions and estimate coefficients with better precision relative to classic survival methods. The marginal growth estimator makes use of data on the intermediate trajectory available from repeated measures of the outcomes and benefits from (1) not assuming an absorbing barrier, so those observed to have reached a threshold are not removed from observation, a realistic approach in the clinical setting and (2) increased efficiency of the growth curve estimator which has more power to demonstrate potential treatment effects and moderation if present in the dataset. These results are supported by the application in the GENDEP data, demonstrating the link between covariates and the time to depression remission and illustrating notably more precise estimates of treatment effect.

Conclusion:

- We have developed an innovative time to threshold based estimator.
- This approach to modelling survival provides a valuable tool for modelling time to reach threshold in repeated measures data.
- Growth models provide a flexible framework to model change over time, gaining advantage from the option of inclusion of polynomial time terms.

[OC09.4]

Prediction of cancer survival for cohorts of patients most recently diagnosed using multi-model inference

Camille Maringe, Aurélien Belot, Bernard Rachet

Department of Non-Communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, United Kingdom

Background: Despite a large choice of models, functional forms and types of effects, clear strategies to select excess hazard models for prediction of population cancer survival are not widespread in the literature. Objective: We propose multi-model inference based on excess hazard model(s) selected using AIC or BIC for prediction and projection of cancer survival.

Method: We adapt model selection strategies based on significance tests to select model(s) using information criteria. More than one model may show equivalent support from the data. We evaluate the properties of multi-model inference using empirical data of patients diagnosed with breast, colon or lung cancer in 1990-2011 in England. We artificially censor the data on 31/12/2010 and predict 5-year survival for the 2010 and 2011 cohorts. We compare these predictions to the observed 5-year non-parametric cohort estimates of cancer survival (Pohar Perme estimator). We also contrast our multi-model predictions to predictions from an a-priori selected simple model, and from the non-parametric period approach. We illustrate multi-model inference by replicating this on recently diagnosed cohorts of patients, for which stage at diagnosis and other important prognosis factors such as mode of diagnosis and performance status are available.

Results: We find that model-averaged predictions and projections of survival have close to minimal root mean integrated square differences with the Pohar Perme estimation of survival in many instances, particularly in subgroups of the population.

Conclusion: Advantages of Information-Criterion based model selection include (i) Transparent model-building strategy, (ii) Accounting for model selection uncertainty, (iii) No a-priori assumption for effects, (iv) Projections for patients outside of the sample.

[OC09.5]

Use of pseudo-observations to quantify the impact of cancer due to socioeconomic inequalities

Aimilia Exarchakou, Bernard Rachet, Aurelien Belot, Dimitra-Kleio Kipourou

Non-communicable Diseases Epidemiology, London School of Hygiene and Tropical Medicine, United Kingdom

Background: In population-based studies, alternative survival indicators such as the number of life years lost (NLYL) are useful for reflecting different dimensions in terms of prognosis and treatment choice, or for developing a control strategy. Reporting them along with other survival measures, provides a multi-perspective approach that allows the exploration of complicated topics such as the impact of socioeconomic inequalities on cancer.

Objective: Our aim is to estimate the socioeconomic deprivation differences in NLYL due to cancer using the pseudo-observation method and translate the years that a patient loses (due to cancer) to indirect societal economic loss through paid employment. The method was adapted to the relative survival setting, i.e. in the absence of information on cause of death.

Methods: We estimated the effect of deprivation on NLYL via direct modelling using the pseudo-observation approach. We predicted the individual NLYL and applied sex-, age-, and deprivation-specific average earnings to estimate the indirect economic loss in the society. We demonstrated this approach using data on different cancers from the England population-based cancer registry.

Results: Whilst NLYL due to cancer increased with increased deprivation, the economic loss in productivity was higher in the less deprived. We also found variation in NLYL by cancer type. Cancers with peak incidence in older age (>65years) and poor prognosis (eg, lung) contributed more in NLYL due to increased mortality. However, the largest impact on economic loss was due to cancers that affect younger individuals (<45 years) such as cervical cancer.

Conclusions: NLYL is a useful survival indicator that can be modelled directly with pseudo-observations within the relative survival setting. Quantifying the effect of deprivation on NLYL is essential for exploring the complicated mechanism between socioeconomic inequalities and cancer. Predicting the individual NLYL and translating those to economic loss is a useful approach to summarize the disease and economic burden in the society.

OC10: Personalised Medicine 1

Chair: **Helene Jacqmin-Gadda**

Inserm, France

[OC10.1]

Rare Disease Trials: Using a continuous biomarker to allocate patients in a response-adaptive clinical trial

Holly Jackson

Lancaster University, United Kingdom

Randomised controlled trials (RCTs) are the current gold standard of clinical trials. They assign each patient to either the experimental treatment or the control treatment (placebo or standard of care) with equal probability. However, rare disease trials include a large proportion of the patient population in the trial and hence, only a relatively small proportion of patients will be outside the trial to benefit from the results of the trial. Therefore, where rare disease treatments are being investigated, there should be a larger emphasis on the benefit to the trial population.

Response-adaptive randomisation (RAR) designs use information from previous patients within the study, to alter the probability of the next patient(s) receiving each treatment. Many RAR designs have optimal treatment allocation, some maximise patient success, some minimise patient failures. Targeting an optimal allocation can cause correlation between treatment assignments, which leads to a loss in power. Additionally, they assume every patient who receives the same treatment will also respond in the same way. This is not always the case.

We use a patient's biomarker to influence the allocation probability, to increase the number of patients who receive their individual best treatment. As personalized medicine becomes more favourable, we can use this method to select treatments, which may only be effective for a subgroup of the patient population.

The developed method starts with an initial burn-in period of ten patients, who with equal probability, are allocated to each treatment. We then use a regression method to predict the best outcome of the next patient, using their biomarker and the information from the previous patients. This estimated best treatment is then assigned to the next patient with high probability.

In this work, we investigate different regression methods such as Gaussian processes, splines and polynomial regression for use in the response adaptive scheme and evaluate them in simulations and an application for the treatment of malignant ascites. We find that the novel method yields notably higher patient benefit than a traditional RCT while maintaining adequate power.

[OC10.2]

Assessing the optimal time to start renal replacement therapy using Causal Machine Learning

Paweł Morzywołek[1], Johan Steen[2,3], Wim Van Biesen[2,3], Stijn Vansteelandt[1,4]

[1] *Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium*
[2] *Department of Internal Medicine and Pediatrics, Ghent University, Belgium*
[3] *Renal Division, Ghent University Hospital, Belgium*
[4] *LSHTM, London, United Kingdom*

There is general optimism and enthusiasm that the introduction of Big Data and related technologies will substantially change and improve delivery of medical care and healthcare decision-making. In this research, we explored how the use of routinely collected data (RCD) can be of real benefit for healthcare decision-making.

In particular, we applied statistical techniques on optimal treatment regimes developed in the literature on causal inference, to help medical doctors in making optimal patient-centered decisions on treatment initiation. For this, we made use of linked databases from the Intensive Care Unit (ICU) and dialysis center of the Ghent University Hospital, containing longitudinal, highly granular records from all adult patients admitted to the ICU from 2013 to 2017. In particular, we inferred whether or not to initiate renal replacement therapy (RRT) for individual patients suffering acute kidney injury (AKI) on each day of their ICU stay, based on their level of potassium, pH and urinary output until that day.

Our analysis makes use of state-of-the-art causal inference methodology to adjust for time-varying confounding. To alleviate concerns about model misspecification, confounding control was based on machine learning techniques. In particular, we made use of targeted minimum loss based estimation (TMLE) and double/debiased machine learning (DML) methods to ensure valid data-adaptive inference.

Our study involves a performance evaluation of TMLE and DML, while also evaluating the benefits of procedures that aim to deliver stable inverse probability weights, e.g. covariate balancing propensity score (CBPS) and kernel optimal weighting (KOW) procedures. Based on the analysis results, we propose a decision strategy for the optimal timing of initiation of RRT for patients with AKI.

[OC10.3]

Simple optimal adaptive treatment strategies in studies subject to informative monitoring times

Janie Coulombe, Erica E. M. Moodie, Robert W. Platt

Epidemiology, Biostatistics and Occupational Health, McGill University, Canada

Background: Citalopram and Fluoxetine are two antidepressant drugs commonly prescribed for the treatment of unipolar depression in adults. The literature is sparse on reasons for prescribing one of these two drugs rather than the other. In particular, it is not clear which of these two drugs may lead to greater weight fluctuations in patients and whether any patients' feature may modify their effect on weight. Observational longitudinal data can be used to study that question. However, analyses with these data are often subject to confounding and informative monitoring times, which, when not considered, may lead to biased estimates of treatment effects.

Objectives: In this work, we propose and demonstrate a new statistical methodology for building simple adaptive treatment strategies in settings with confounding and informative monitoring times. Using the Clinical Practice Research Datalink in the UK, we apply that new methodology to the study of the effects of two selective serotonin reuptake inhibitors on weight, and their modification by age.

Methods: We first extend an estimator for the marginal effect of treatment that we proposed in previous work. That estimator is further used to estimate an optimal treatment decision rule. We demonstrate the robustness of the new proposed methodology using results from extensive simulation studies. We then use the methodology to develop a simple one-stage treatment decision rule to choose between two antidepressants, which rule aims to improve weight outcomes and reduce fluctuations, based on a patient's age. We account for informative monitoring times, and assume they depend on patients' features measured at the previous visit. Potential confounders of the association between the antidepressant drugs and weight are considered.

Results: Several simulation studies show that the new method outperforms methods that do not account for informative monitoring times or confounders. Moreover, the application to the effect of antidepressants on weight fluctuations shows important differences in estimated optimal decision rules when accounting for informative monitoring times.

Conclusion: The statistical methodology we propose can be used to develop simple adaptive treatment strategies which consider informative monitoring times and confounding variables.

[OC10.4]

Learning individual trajectories from biomedical time-series data with deep generative models

Maren Hackenberg, Harald Binder

Faculty of Medicine - Medical Center Freiburg, Institute of Medical Biometry and Statistics, Germany

Background: When considering biomedical time-series data from individuals, personalised modelling means that each individual receives a customised dynamic model. For example, the dynamics of a latent health status, underlying the observed data, might be based on individual characteristics measured at baseline in a cohort setting. Yet, the corresponding modelling task is further complicated by a typically sparse, highly irregular time grid of measurements.

Objective: Inspired by recent advances on combining black-box deep learning with explicit mechanistic modelling by differential equations, we develop a deep generative model that captures individual dynamics in a low-dimensional latent representation as solutions of ordinary differential equations (ODEs) in such a challenging setting.

Methods: Motivated by a data scenario from a large epidemiological cohort, we consider an extensive characterisation of individuals at a baseline time point and measurements of a smaller subset of variables at an individually differing second time point, resulting in a very sparse (only two time points) and irregular time grid. We employ deep learning, specifically a variational autoencoder (VAE), to obtain a low-dimensional representation of the central factors of variation governing the development patterns in the data in a non-linear, unsupervised way. We constrain the representation to model smooth trajectories by imposing an ODE system on the latent space and infer an individual-specific set of ODE parameters from the variables measured only at baseline. In an extension of our model, we enrich each individual's information by assigning to it a batch of individuals with similar underlying development patterns, whose second time point measurements serve as proxy information on the common dynamics at multiple time points.

Results: Using simulated data, we show that our model recovers individual trajectories from two-dimensional ODE systems with two or four unknown parameters in linear and non-linear systems and accurately infers groups of individuals with similar trajectories. We additionally illustrate our method on data from a human cohort.

Conclusion: In conclusion, our model provides an individual-level understanding of the underlying dynamics governing individuals' developments rather than estimating average effects. This has the potential to assess the effect of interventions based on the knowledge of full individual-specific dynamical systems.

[OC10.5]

Spline-based modelling to investigate treatment effect differences in IPD-MA: a gentle introduction

Michail Belias[1], Jeroen Hoogland[2], Maroeska M. Rovers[3], Johannes B. Reitsma[2], Thomas P. A. Debray[2], Joanna IntHout[1]

[1] *Health Evidence, Radboud UMC, Netherlands*
[2] *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Netherlands*
[3] *Evidence based surgery, Radboud UMC, Netherlands*

Background: Modelling personalised treatment effects is an opportunity offered by individual participant data(IPD) meta-analysis(MA). However, analysing associations between outcomes and continuous patient characteristics may be challenging when non-linear associations are present. In this setting, spline approaches offer great flexibility, but they are rarely applied.

Objective: To provide an introduction on how to model the association between an outcome and a continuous variable using popular splines approaches; restricted splines, B-splines, smoothing splines, and P-splines, in settings of single and multiple studies.

Methods: We describe the splines and illustrate their performance in an artificial single study, consisting of 500 participants. Subsequently, we describe three IPD-MA methods to combine the results of multiple studies:two two-stage approaches based on either point-wise or multivariate meta-analysis and an one-stage approach based on generalised additive mixed effects models(GAMMs). Subsequently, we illustrate their performance using three scenarios: a setting with between-study differences in the regression lines, a setting where the domains of the independent variable are different across the studies, and a combination of those two. Each scenario consists of five studies with 500 participants each. Subsequently, we evaluated the aforementioned splines and pooling approaches, in an empirical IPD-MA example, modelling the risk of fever and/or ear pain 3-7 days after treatment with antibiotics or placebo in relation to age and bilaterality of ear pain.

Results: The four splines approaches showed similar results in the artificial single study. However, across the multiple studies scenarios results varied. Although, in the first scenario all IPD-MA pooling methods showed similar results, in the second and in the third scenario multivariate meta-analysis showed unreliable results and unrealistically wide confidence intervals; pointwise meta-analysis showed discontinuous regression lines and confidence intervals, while one-stage GAMMs performed best showing unbiased regression lines and reasonable confidence intervals. In the empirical example, results varied across the pooling methods, while within each pooling method the four splines approaches provided similar results.

Conclusion: Adjusting for non-linearities whilst combining multiple studies needs careful modelling. Different pooling methods may show different results some of which may be unreliable. We show that in 3 common IPD-MA scenarios one-stage GAMMs seems more robust and reliable than multivariate and point-wise meta-analysis.

OC11: Methods Validation and Comparisons

Chair: **Aris Perperoglou**

AstraZeneca, United Kingdom

[OC11.1]

Evaluation of the Fill-it-up design to use historical controls in randomized controlled clinical trials

Stephanie Wied, Ralf-Dieter Hilgers, Nicole Heussen

Department of Medical Statistics, RWTH Aachen University, Germany

Context: The most appropriate method to investigate the effects of interventions in clinical research is to carry out a randomized clinical trial (RCT). For small population groups, as in the field of rare diseases, conducting an RCT is challenging. Historical control data may be incorporated into the new RCT to reduce the sample size. Several approaches exist on how to combine historical controls with randomized controls. However, the approaches usually do not take into account their equivalence to justify the inclusion.

Objectives: The aim is to reduce the required sample size of the new RCT. To prevent biased estimates of the treatment effect, the historical controls should be similar to the randomized controls to a reasonable extent and with a certain confidence.

Method: The combination of historical and randomized controls should only be considered if their equivalence is confirmed in an equivalence pre-test. We therefore propose to stop the originally planned study when a certain sample size is reached in order to perform the equivalence pre-test. If equivalence is confirmed, the historical control data is included in the new RCT. If equivalence cannot be confirmed, the historical controls are not considered at all and the randomization of the original trial continues. We investigate the performance of this study design in terms of the adherence of the overall type-I-error probability and the overall power.

Results: We show how the significance levels of the separate tests need to be adjusted to maintain the overall type-I-error probability and overall power of our design within acceptable limits while reducing the total randomized sample size.

Conclusions: With our design, we are able to verify the inclusion or ignorance of historical control data. The combination of historical and randomized data leads to a reduction of the sample size of the randomized groups, permitting more RCTs to be performed in small populations.

[OC11.2]

Indirect treatment effect estimation in presence of modifier in Network Meta Analyses of time-to-event data

Matthieu Faron[1], Pierre Blanchard[2], Jean-Pierre Pignon[1], Stefan Michiels[2], Gwenael Le Teuff[1]

[1] *Department of Biostatistics and Epidemiology, Gustave Roussy, University Paris-Saclay, Gustave Roussy, France*
[2] *Oncostat U1018, Inserm, University Paris-Saclay, labeled Ligue Contre le Cancer, Gustave Roussy, France*

Background: Aggregated Data (AD) and Individual Patient Data (IPD) Network Meta-Analysis (NMA) give similar Treatment Effect (TE) estimation in absence of effect modifier / treatment by covariate interaction. Yet, the bias in its presence is seldom studied, notably when the distribution of effect modifier differs across comparison.

Objectives: We aimed to propose a frequentist one-step approach for an IPD NMA of time-to-event data and to evaluate the bias and mean squared error in TE estimation when comparing to AD NMA approach.

Methods: A simple, 3-nodes (A-B-C) network was simulated: ten trials of 200 patients each, were simulated with the same TE for A-C and B-C, and no trial for A-B. Age was simulated as continuous variable. No interaction between age and the TE was simulated in B-C, but was in A-C according to 3 scenarios (1) no interaction, same age; (2) interaction (25% variation in TE for each SD of age); (3) interaction and different ages (-1SD in A-C; +1 SD in B-C). Analyses aimed to indirectly estimate the A-B TE via comparator C either by AD (R package Netmeta) or by one-step-IPD using a Poisson generalized linear model with random effects. 1000 replications of different TE were done. We illustrate the comparison AD vs. IPD from a head and neck cancer IPD-NMA (MACH-NC: 64 trials including 12,129 patients).

Results: In scenario (1) A-B's TE was estimated accurately. In (2) the mean TE (at mean age) was estimated correctly by both methods but a bias was observed at other ages in the AD approach (no interaction). In (3) a bias was found in the AD approach at all ages including the mean TE incorrectly estimated by the model at -0.12 for a simulated value at 0 (mean squared-error (MSE)=0.022). Conversely IPD approach correctly estimated TE at all ages (mean TE= 0.03, MSE=0.014). An interaction with age was seen in one comparison of the MACH-NC NMA.

Conclusions: When the distributions of effect modifier differ across comparison, AD-NMA can produce biased indirect TE estimation. When feasible, IPD-NMA should therefore be the preferred approach.

[OC11.3]

What happens to sample size predictions for series of N-of-1 trials when outcomes aren't normally distributed?

Katy Morgan[1], Elizabeth Williamson[1], Liam Smeeth[2], Andrew Forbes[3]

[1] *Medical Statistics, LSHTM, United Kingdom*
[2] *Non-communicable Disease Epidemiology, LSHTM, United Kingdom*
[3] *Biostatistics Unit, Monash University, Australia*

Background: In series of N-of-1 trials, participants receive two or more treatments in a random order over several treatment periods. Such trials allow estimation of treatment effects for a particular individual, and results can be combined to estimate the population-averaged treatment effect. A sample size calculation exists for series of N-of-1 trials comparing two treatments that measure a continuous normally distributed outcome once per treatment period (Senn, SMMR (2019) 28(2)372-383). Series of N-of-1 trials with a patient-reported outcome, particularly a subjective outcome such as pain, may not have normally distributed outcomes. The StatinWISE trial, for example, investigates the effect of statins on patient reported muscle side effects (Herrett et al. BMJ Open (2017) 7:e016604).

Objective: To assess how the sample size formula performs when outcomes are not normally distributed, using a simulation study.

Methods: We performed a simulation study to assess the performance of the sample size formula under various data generation mechanisms: outcomes are in fact normally distributed; person-level treatment effects are normally distributed but residuals are skewed; person-level effects and residuals are skewed; non-normal outcome based on data from StatinWISE (pain scores generated from uniform distributions with latent classes of people who always/never/sometimes experience pain); ordinal outcome.

The sample size for each data type was obtained from the formula with Type I error set to 5%, and a power of 80% or 90%. Other parameter values for the sample size calculation were set according to the particular data generation mechanism. 5000 datasets were generated for each outcome type, for a trial with 4, 6, 8 or 10 treatment periods. Each dataset was analysed using the model assumed by the sample size calculation (a mixed effects model on the differences between treatments in each pair of periods, with a random person intercept), and the empirical power calculated.

Results and Conclusions: Preliminary results suggest that the sample size formula is generally fairly robust to deviations from normality. However, the power does differ from the nominal level when the treatment effects and residuals are both skewed, and the variance of the person-level treatment effects is large relative to the residual variance.

[OC11.4]

On estimating average treatment effect of rare exposure: sample size consideration and bias correction

Md Ariful Islam Sanim, M. Shafiqur Rahman

Institute of Statistical Research and Training (ISRT), University of Dhaka, Bangladesh

Background: Propensity score methods are increasingly being used for estimating average treatment effect (ATE) of exposures or treatment on outcome in observational studies. The propensity score (PS), an estimated probability of being exposed/treated obtained from fitting logit or probit model, is used for estimating the ATE by using inverse probability weighting (IPW). This approach performs well when the prevalence of exposure is relatively high. However, there is much doubt about the performance of the IPW method when the prevalence of exposure is low. This is because the PS model estimated using maximum likelihood approach may provide biased estimate of the PS, which in turn may provide biased estimate of the ATE in the outcome model.

Objectives and Methods: In this research, we firstly investigated, using an extensive simulation study, how many events (or sample) required per variable in the PS model for estimating the ATE in outcome model with negligible amount of bias. Secondly, while the required sample size (or EPV) is unavailable, we investigated if the amount of bias can be reduced by providing correct estimates of the PS using penalized likelihood approach (Firth's intercept corrected logit model). We also explored both bootstrap and Jacknife bias-corrections for ATE in the outcome model and compare the results using another extensive simulation study.

Results: The simulation results show that the standard maximum likelihood based PS provide substantial amount of upward bias and high MSE in the estimate of ATE when exposure is rare (low prevalence, <15%, and/EVP <10). The amount of bias can be slightly reduced by estimating PS using penalized method, which is however not sufficient. Both the bootstrap and Jackknife bias correction showed comparable results by reducing substantial amount of bias in the estimate of ATE. Application of the method is provided in estimating ATE of severe-malnutrition on acute respiratory infection in children of under five-years after controlling for several baseline covariates.

Conclusion: For estimating ATE of the binary treatment/exposure with low prevalence, one needs EPV>10 in PS model, otherwise bootstrap or Jackknife-based bias correction is required in outcome model.

[OC11.5]

Modeling improvement of performance for combined procedures in classification of microarray data sets

Malgorzata Cwiklinska-Jurkowska

Department of Biostatistics and Theory of Biomedical Systems, Nicolaus Copernicus University, Poland

Background: Combining procedures have a significant role in classification tasks. Combined classifiers should make different decisions on the same data sets to increase performance.Additionally, the compromise between diversity and accuracy of base classifiers is thusly important.

In high-dimensional classification problems, the initial stage of dimension reduction is vital.

Objective: The aim of the work was modeling of performance improvement of heterogeneous combined classifiers by generalized estimating equations (GEE) models.

Methods: Heterogenous combining procedures were applied for both genes selection and ensemble learning.The basis for genes selections was the ranking of importance obtained on the basis of 26 individual constituent methods. Combined methods of dimension reduction are based on examining the frequency of variable choices by the basic selection methods. Examined gene subsets are monotonic increasing sets. The tenfold CV technique was used- repeatedly 9 folds used in gene selection and building classifiers, 10-th fold for assessment of classification error.

Due to the increasing number of genes in the analysis, the number of genes is a longitudinal variable. Autoregressive type of working correlation matrix was applied. Such GEE models were used for 28 different measures of diversity and for monotonic increasing sets of 5, 7,9 and 11 heterogeneously combined classifiers. Constituent classifiers were of different types: 1NNeigh, kNNeigb, RDA, SVM with different kernels, nearest mean classifier, SDA and diagonal SDA.

Results: Five microarray high-dimensional data sets were applied for examinations, where the real classification problem is the recognition of medical case into one of two groups.

The diversity derived from the heterogeneous combination of different types of classifiers and the number of genes are important explanatory variables in the GEE model for improving the efficiency of the classification of microarray data relative to the average error of the base classifiers.

The square of diversity measures and the square of the number of genes are significant in some cases. In addition, it appears that interactions between variable numbers and diversity are sometimes important.

Conclusions: The GEE models and significance of the diversity, number of genes and the interactions depend on the measure of diversity and the number of heterogeneously combined classifiers.

OC12: Meta-analysis 1

Chair: **Marianne Huebner**

Michigan State University, United States

[OC12.1]

Exploring the empirical distribution of tau from IQWiG reports to inform Bayesian meta-analyses

Ralf Bender, Sibylle Sturtz, Corinna Kiefer

Medical Biometry, IQWiG, Germany

Background: Meta-analysis is the method of choice in systematic reviews to summarize the effect estimates of the included studies. According to the IQWiGs methods paper (Version 5.0), the Knapp-Hartung method is the standard approach to conduct random-effects meta-analyses [2]. However, in the case of very few studies, the heterogeneity parameter tau cannot be reliably estimated leading to broad confidence intervals [1]. In such situations, meta-analysis applying Bayesian methods with an informative prior distribution is an option [1,3]. Different choices for prior distributions for tau are possible according to several proposals given in the literature [3-5].

Objective: The goal of the talk is to explore the empirical distribution of tau from IQWiG reports in order to inform future Bayesian meta-analysis in the case of few studies.

Methods: We collected all published meta-analyses from IQWiG reports for the years 2005 to 2019 and recalculated the estimates of tau by applying random-effects meta-analyses and the Paule-Mandel method. We summarized the empirical distributions of tau in various settings (comparison, endpoint category, effect measure) and compared these distributions with the proposals for prior distributions in the literature.

Results: Different empirical distributions of tau can be derived from IQWiG reports in various settings.

Conclusions: Based on empirical distributions of tau from IQWiG reports in various settings it should be discussed which prior distributions for Bayesian meta-analyses are appropriate in the framework of health technology assessment.

References:

1. *Bender, R. et al. (2018): Methods for evidence synthesis in the case of very few studies. Res. Syn. Methods 9, 382–392.*
2. *IQWiG (2017): General Methods, Version 5.0 of 10 July 2017. Institute for Quality and Efficiency in Health Care, Cologne.*
3. *Friede, T. et al. (2017): Meta-analysis of few small studies in orphan diseases. Res. Syn. Methods 8, 79-91.*
4. *Rhodes, K.M. et al. (2018): Between-trial heterogeneity in meta-analyses may be partially explained by reported design characteristics. J. Clin. Epidemiol. 95, 45-54.*
5. *Turner, R.M. et al. (2015): Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. Stat. Med. 34, 984-998.*

[OC12.2]

Multi-Arm Multi-Stage (MAMS) Designs in Clinical Development: A comparative review

Loic Darchy

Essonne, SANOFI R&D, France

Multi-Arm Multi-Stage (MAMS) designs have been receiving increasing interest over the last decade particularly but not exclusively in oncology. Platform trials are a good illustration. In platform trials, patients are first classified at entry according to their baseline characteristics (typically biomarker data) and a treatment selection then operates within each biomarker stratum across interim looks at the data. New treatments may also sometimes be added during the course of the study. Different statistical frameworks have been proposed to achieve this goal. Some of them allow information borrowing across biomarker strata with each treatment and also across treatments. The presentation gives an overview of the different MAMS designs and methods recently published and is an attempt to assess them with the key objective to provide statistical guidance depending on sponsor's objectives and constraints.

[OC12.3]

Network meta-analysis of rare events using penalized likelihood regression

Theodoros Evrenoglou [1], Dimitris Mavridis [2], Anna Chaimani [1]

[1] *Inserm Research Centre in Epidemiology and Statistics (CRESS-U1153), Université de Paris, France*
[2] *Department of Primary Education, University of Ioannina, Greece*

Background: Network meta-analysis (NMA) allows the synthesis of data on multiple interventions and, under certain conditions, provides the highest possible level of evidence for the development of clinical guidelines. However, the normal approximation made by the conventional NMA model can be poor and effect estimates are biased when studies are small or there are few events. In such individual studies, the penalized likelihood function proposed by Firth [1] for bias reduction of the maximum likelihood estimator (MLE) is a popular approach but has never been considered in the context of NMA.

Objective: To improve the accuracy and the precision of NMA relative effect estimates in the presence of rare events by extending Firth's bias reduction method into NMA.

Methods: We developed a new model for low event rate binary data forming a network of interventions. Following Firth's method, to reduce the bias of the MLE, we penalize the likelihood function using Jeffreys prior and we remove the first-order term in the asymptotic bias expansion of the MLE. We evaluate the performance of our approach using various simulation scenarios and two real datasets: a network comparing the safety of different drugs for chronic plaque psoriasis and a network comparing interventions for decreasing blood loss and blood transfusion requirements during liver transplantation.

Results and Conclusions: In comparison to three alternative NMA models that have been suggested for handling rare events, our method tended to give more precise relative effects estimates particularly for comparisons with only one or two studies, while all models did not perform very well in the presence of substantial heterogeneity. Our approach allows the inclusion of all available studies in the network no matter the number of events per arm and it reassures the connectedness of the network. As a result, our penalized regression NMA model offers a reliable and potentially more informative alternative to existing approaches for the analysis of networks with rare events.

References:

1. *Firth D. Bias reduction of maximum likelihood estimates.Biometrika. 1993;80(1):27-38.*

[OC12.4]

Detecting outlying studies in Network Meta-analysis using Bayes factors

Silvia Metelli [1], Dimitris Mavridis [2], Anna Chaimani [1]

[1] *Inserm Research Centre in Epidemiology and Statistics (CRESS-U1153), Université de Paris, France*
[2] *Department of Primary Education, University of Ioannina, Greece*

Background: Network meta-analysis (NMA), a method that generalizes standard meta-analysis by allowing for multiple treatment comparisons, has been increasingly used to inform public health decisions. Within this context, it becomes crucial to identify studies with markedly different or extreme effect sizes that may bias the summary results and strongly influence conclusions from NMA. Bayes factors are commonly used to select among competing models and offer a convenient way to capture deviating studies within Bayesian NMAs.

Objectives: To develop a model-based method for the identification of outlying studies in NMA on the basis of Bayes factors.

Methods: We define an outlier as a study with a 'shifted' effect size. Based on this concept, we introduce a Bayesian NMA mean-shifted model, which assumes a location-shifted vector for the effect size of each study. Then, we use Bayes factors to test whether each study is more compatible to the conventional or to the mean-shifted NMA model. In the latter situation the study is considered as an outlier. As outliers often tend to be grouped together, standard leave-one-out (LOO) procedures sometimes suffer from masking (i.e. a true outlier not detected as such). To mitigate this, we consider a modified LOO cross-validation scheme where we restrict our search to groups of studies comparing either the same treatments or the same class of treatments. We apply our outlier detection method to both simulated and real data sets.

Results and Conclusions: The method successfully identified existing outliers in the simulated data. When we used our approach in an NMA of 112 randomised controlled trials comparing second-line treatments for advanced non-small cell lung cancer, we identified one clear and two potential outliers corresponding to a very large and two moderate Bayes factors respectively. Such studies may threaten the validity of the NMA findings and need further investigation. To quantify the associated uncertainty of our results, a probabilistic score of being outlier for each study in the NMA may be used.

[OC12.5]

A Prediction Model of Heterogeneous Treatment Effects Using Randomized and Observational Data

Konstantina Chalkou [1], Pascal Benkert [2], Suvitha Subramaniam [2], Jens Kuhle [3], Fabio Pellegrini [4], Georgia Salanti [1]

[1] *Institute of Social and Preventive Medicine, University of Bern, Switzerland*
[2] *Clinical trial unit, University of Basel, Switzerland*
[3] *Departemente für Medizin, Biomedizin und Klinische Forschung, Universitätsspital Basel, Switzerland*
[4] *Biogen International GmbH, Switzerland*

Background: Treatment effects vary across different patients and estimation of this variability, known as heterogeneous treatment effects, is important for clinical decisions. A determinant of this variability is the baseline risk of patients often reflecting the severity of the condition. As treatment options are often numerous, comparative effectiveness research is vital for clinical decision-making, and evidence from randomised and observational studies is increasingly available.

Objectives: Our first objective is to develop a two-stage prediction model for heterogeneous treatment effects, by combining prognosis research and network meta-analysis methods and by synthesizing several sources of data; individual participant data, aggregated data, randomized clinical trials and observational data. Our second objective is to validate the performance of the developed model.

Methods: In a first stage, we develop a prognostic model and we predict the baseline risk of the outcome prior to treatment, using observational data. In the second stage, we use this baseline risk as candidate prognostic factor and effect modifier in a network meta-regression model with individual participant data and aggregated data from randomized clinical trials to make personalized predictions under different treatment options. Finally, we expand existing validation methods to evaluate the performance of the model, using measures related to clinical usefulness. We apply the methodology in the case of choosing between four treatments in relapsing-remitting multiple sclerosis with respect to relapse prevention.

Results: Our model indicates that baseline risk score modifies the relative and absolute treatment effects. Several patient characteristics such as age and disability status influence the baseline risk of relapse, and this in turn moderates the benefit within each one of the treatments. We will present an R-Shiny app that estimates the risk of relapse under all available treatments conditional on patients' characteristics and consequently, indicates which treatment is preferable for which patient

Conclusions: This two-stage model and its validation methods, bridge methods in prognosis research and evidence synthesis, and is applicable to a wide range of medical decision-making problems.

The HTx project is funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 825162.

OC13: Clinical Trials Methodology 3

Chair: **Stephen Senn**

Statistical Consultant, United Kingdom

[OC13.1]

An adaptive design of phase I/II clinical trials for precision medicine using oncogene information

Shinjo Yada[1], Ryuji Uozumi[2]

[1] *Biostatistics Department, A2 Healthcare Corporation, Japan*
[2] *Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine, Japan*

Background: Cancer has been treated according to its type and location in the body, such as lung, breast, gastric, or colorectal cancers. Cancer tissue samples obtained via biopsy or surgery are normally examined for the presence of specific gene mutations using genetic testing. Next-generation sequencing allows for the simultaneous analysis of many genes at faster speeds compared to conventional DNA sequencing, thus reducing the cost of genetic testing for cancer. The detected gene mutations vary based on the type of cancer, and the drugs administered to treat these cancers target specific genes. Taking into account not only the cancer type but also the genetic information, environment, and lifestyle of each patient, precision medicine is an approach that can be used for the prevention and treatment of disease in individual patients.

Objective: It is assumed that patient-specific characteristics, including biomarkers, increase over time and that these characteristics are highly correlated with outcomes for the patient. The sample size at the beginning of early-phase clinical trials is often limited. In conventional models that include baseline characteristics such as patient biomarkers as the covariates, estimation of model parameters is difficult. To overcome this issue, we proposed a dose-finding method that considered the background characteristics, including biomarkers, by building a deep learning model. Our goal was to estimate the efficacy and toxicity probabilities and to select the optimal dose for each patient with biomarker profiles for phase I/II oncology trials.

Methods: We built a feed-forward neural network in which the input variables were dose, biomarkers, and interactions between the dose and biomarkers for each patient, and the output variables were the efficacy and toxicity outcomes for each patient. We trained the feed-forward neural network to select the optimal dose based on the background characteristics, including biomarkers, for each patient.

Results and Conclusions: Simulation studies showed that the proposed method reduced the trial period without compromising performance when compared with conventional methods.

Reference:

Yada S. and Hamada C. Pharmaceutical Statistics, 16(2), 2017, 114-121.

[OC13.2]

Information content of cluster-period cells in stepped wedge designs with unequal cell sizes

Jessica Kasza, Andrew Forbes

Monash University, Australia

Background: Stepped wedge trials are being conducted with increasing frequency. Previous work has investigated the contribution of each cluster-period cell in a stepped wedge to the estimation of the treatment effect. This work has shown that cluster-periods near the time of the treatment switch tend to provide a great deal of information about the treatment effect, but assumes that all cluster-period cells contain data from the same number of participants. However, reports of stepped wedge trials show that clusters in these trials are often of widely differing sizes. The impact of differing cluster-period sizes on the amount of information contributed by cells is unknown.

Objective: To investigate the information contributed by each cluster-period cell of stepped wedge designs when cell sizes differ. Such an investigation is necessary to help researchers understand how to design efficient "incomplete" stepped wedge designs where not all clusters contribute measurements in all periods.

Methods: We measure the contribution of cluster-period cells to the estimation of treatment effects through the definition of a cell's "information content". We apply these results to a trial assessing an intervention aimed at preventing falls in rehabilitation units and investigate the contribution of each cell to the estimation of the intervention effect. We also explore several incomplete designs for this trial.

Results: We show that the pattern of information content of cells within stepped-wedge designs depends markedly on cluster-period sizes. In the rehabilitation unit trial example we show that the efficiency of incomplete designs depends on which cells are excluded: smaller incomplete designs may be more powerful than larger incomplete designs that include a greater total number of participants.

Conclusions: The burdensome nature of stepped wedge trials – all clusters must provide measurements for the entire trial duration – implies that efficient incomplete designs are appealing alternatives. We have shown that cluster-period sizes can have a large impact on the amount of information contributed by each cluster-period of a stepped wedge design, and that incomplete designs with a larger number of participants are not necessarily more powerful: there may be alternative incomplete designs with fewer participants that have greater power.

[OC13.3]

How to handle drop-out in palliative care trials

Anneke Grobler

MCRI, Australia

Background: Clinical trials are hard to do. Humans participate in trials and the ideal scenario of complete compliance with the protocol, no missed visits or assessments and complete follow-up is seldom realised.

These is issues are exacerbated in certain clinical areas, such as palliative care where endpoints are related to quality of life, symptoms or pain and not disease outcomes. In palliative care trials mortality and drop-out due to deterioration of the patient's clinical condition is high and it is not clear how these issues should be handled in the analysis.

Objectives: We review and evaluate the palliative modified intention to treat (ITT) analysis which has been proposed in the literature for this setting, and sits between the ITT and per protocol analysis, where patients who have missing data due to disease progression are excluded from the analysis, while other patients with missing data are included and their endpoints imputed.

Methods: We conducted a simulation study to investigate the bias and efficiency of the palliative modified ITT analysis. We propose alternative solutions to minimise missing data in this setting and highlight current well known missing data techniques which are appropriate in this setting. Finally, we describe alternative endpoints that could be more useful than the currently chosen endpoints for such patient populations.

Results: We develop a checklist to follow at the design stage in palliative care to provide protection against missing data. This checklist is derived from the recommendations by the Panel on Handling Missing Data in Clinical Trials recommendations.

We recommend choosing endpoints that can be assessed by the caregiver, rather than the patient, since these could still be assessed when the patient's condition worsens, thus reducing missing data. An alternative endpoint is reporting of observed effect of pain on the patient's functioning by a caregiver, rather than a self-reported visual analogue scale (VAS) score for pain.

Conclusions: Drop-out and missing data in trials in the palliative care setting create challenges that need to be addressed in both the design and analysis stages. We make some recommendations to improve the design and analysis of such trials.

[OC13.4]

GO/NO-GO interim decision making incorporating short- and long-term endpoints

Julia Niewczas, Fredrik Öhrn

Early Biometrics & Statistical Innovation, Data Science & AI, AstraZeneca, Sweden

In drug development interim analyses (IAs) are performed in order to enable earlier decision making. This can be done in late phases e.g. to stop the trial for futility or efficacy, or to perform other design adaptations. In Phase II, interim analyses may be performed to make a decision about whether to accelerate to Phase III. The amount of data on the primary endpoint might be limited at the time of the IA, so a common approach is to base the interim analysis on a short-term endpoint, which could be e.g. an objective response rate in oncology or a measurement of the primary outcome at an earlier time point, e.g. 3 months compared to the long term response for the primary endpoint.

In this presentation we consider a Phase II trial with an interim analysis incorporating both short-and long term information. We compare three ways to estimate the effect at the IA: primary endpoint only, short-term endpoint only and a combination of both and discuss in which situations it might be beneficial to use the different methods. We also compare different approaches to interim decision making, including the Lalonde framework which is commonly used in early clinical development. Two ways of IA assessment are compared: direct application of the Lalonde framework using pre-specified GO/NO-GO criteria and calculation of predictive power for reaching a GO/NO-GO at the end of the study. We further discuss pros and cons with these two approaches and provide overall recommendations for their use.

[OC13.5]

Empirically derived prior using ClinicalTrials.gov for a Bayesian RCT.

Andrew Vincent, Kylie Lange, Adel Aref, Leonie Heilbronn, Emma Knight

School of Medicine, University of Adelaide, Australia

Background: In 1994 Spiegelhalter et al proposed a Bayesian RCT methodology. However in contrast to the simplicity of the design, the choice of the prior distribution is not so straight forward. We wish to apply this RCT methodology to a parallel arm randomised controlled trial targeting HbA1c in adult Australians with type 2 diabetes (T2DM).

Objective: Estimate the between-study variance of treatment effects in trials targeting HbA1c in T2DM registered on ClinicalTrials.gov, and thereby propose a sceptical (zero-centred) prior for a Bayesian RCT.

Methods: ClinicalTrials.gov selection criteria included diabetes interventional phase III studies in adults started prior to 2016. Study inclusion criteria: multi-arm, randomised, registered prior to completion, targeting HbA1c, and an assessment within 3-9 months post randomisation. The treatment effect is modelled in STAN with a meta-regression (location-scale model) assuming Gaussian distributed study-specific random means. The location model predictors were year of start, baseline HbA1c, therapy duration, sample size, comparison-type (placebo vs additional vs replacement), registration post initiation (Y v N) and participant masking (Y v N); while the scale model included the first four predictors. Studies for which no publication could be retrieved were not included.

Results: We identified 496 studies that satisfied the inclusion criteria, of which 283 (57%) were extracted electronically and 101 (20%) manually. A further 112 (23%) appear currently unpublished or without results. The 384 published studies report a total of 682 experimental vs control HbA1c (%) comparisons. A preliminary analysis pooling groups within multi-arm studies provides evidence for associations with between-study mean (location) for baseline HbA1c, comparison type, registration time and masking, while baseline HbA1c and sample size influence the between-study variance (scale). The estimated standard deviation of treatment effects for medium sized studies in an average population (baseline HbA1c = 8.2%) is 0.23% (95%CrI=[0.19, 0.26]). A sceptical zero-effect centred prior with this variance is equivalent to a prior sample size of 32.

Conclusion: The use of ClinicalTrials.gov is a promising approach to the quantification of a prior distribution. However hurdles remain including data-linking issues due to database design, and the not insubstantial proportion of pre-registered studies that remain unpublished.

OC14: Survival Analysis 2

Chair: **Zdenek Valenta**

Institute of Computer Science, Czech Academy of Sciences, Czech Republic

[OC14.1]

Survival analysis for AdVerse events with Varying follow-up times - The empirical study of the SAVVY project

Regina Stegherr [1], Jan Beyersmann [1], Valentine Jehl [2], Friedhelm Leverkus [3], Kaspar Rufibach [4], Claudia Schmoor [5], Tim Friede [6]

[1] *Institute of Statistics, Ulm University, Germany*
[2] *Novartis Pharma AG, Switzerland*
[3] *Pfizer, Germany*
[4] *Methods, Collaboration, and Outreach Group (MCO), Department of Biostatistics, Hoffmann-La Roche Ltd., Switzerland*
[5] *Clinical Trials Unit, Faculty of Medicine and Medical Center, University of Freiburg, Germany*
[6] *Department of Medical Statistics, University Medical Center Göttingen, Germany*

Background: The SAVVY project aims to improve the analyses of adverse event (AE) data in clinical trials through the use of survival techniques appropriately dealing with varying follow-up times and competing events. Although statistical methodologies have advanced, in AE analyses often the incidence proportion, incidence densities or a non-parametric Kaplan-Meier are used, which either ignore censoring or competing events.

Objectives: In an empirical study including randomized clinical trials from several sponsor companies, these potential sources of bias are investigated. The main purpose of the empirical study is to compare the estimators that are typically used in AE analysis to the non-parametric Aalen-Johansen estimator as the gold standard.

Methods: The comparisons of the standard estimators to the Aalen-Johansen estimator are done with descriptive quantities, plots and with a more formal assessment using a random effects meta-analysis on AE level. Factors that influence the bias are investigated in a meta-regression. The comparisons are not only conducted at the end of follow-up but also at three earlier time points. Similar methods are applied for group comparisons.

Results: In the empirical study 10 sponsor companies provided 17 clinical trials including 186 investigated AEs. The 1-Kaplan-Meier estimator is on average about 1.2-fold larger than the Aalen-Johansen estimator and the probability transform of the incidence density overestimates the AE probability even more. Although here the average bias using the incidence proportion is less than 5\%, the bias should not generally be neglected as its size strongly depends on the amount of censoring. Furthermore, the adequate consideration of non-constant hazards is less an issue. But the decision on how a competing event is defined is important. As not only death but all treatment-related terminations of follow-up may be considered as a competing event, this influences the amount of censoring and of competing events which are the leading forces influencing the bias.

Conclusions: The choice of the estimator in the analysis of AEs is crucial. There is an urgent need to improve the guidelines of reporting AEs by finally replacing the Kaplan-Meier estimator and the incidence proportion by the Aalen-Johansen estimator with appropriate definition of competing events.

[OC14.2]

Prediction models with survival data: comparing machine learning to the Cox proportional hazards model

Audinga-Dea Hazewinkel [1], Marta Fiocco [2], Hans Gelderblom [3]

[1] *Population Health Sciences/MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, United Kingdom*
[2] *Mathematical Institute (MI)/Leiden University Medical Centre (LUMC), Medical Statistics, Department of Biomedical Data Science, Leiden University, Netherlands*
[3] *Leiden University Medical Centre (LUMC), Department of Medical Oncology, Leiden University, Netherlands*

Background: Recent years have seen increased interest in using machine learning (ML) methods for survival prediction, chiefly focused on big datasets with mixed datatypes and/or many predictors. Model comparisons have frequently been limited to performance measure evaluation, with the choice of measure often suboptimal for assessing survival predictive performance.

Objectives: We investigate ML model performance in an application to osteosarcoma data from the EURAMOS-1 clinical trial (NCT00143030, ISRCTN 67613327), restricting ourselves to 7 clinically relevant predictors. This comparatively simple data structure -able to accommodate the traditional Cox proportional hazard model- and the clinical knowledge available a priori allow us to establish an expectation of results, providing a unique opportunity for investigating ML model behaviour.

Methods: We compare the performance of survival neural networks (SNN), random survival forests (RSF) and the Cox proportional hazards model. Three performance measures suitable for assessing survival model predictive performance are considered: the concordance index, and the time-dependent Brier and Kullback-Leibler scores. The models are also compared on predictor importance and patient-specific survival predictions. Additionally, the effect of ML model hyper-parameters on performance is investigated. Results are confirmed in a data-based simulation study.

Results: The RSF and Cox models have similar performance as measured by the C-index, Brier scores and Kullback-Leibler score, while the SNN has better performance in the second half of follow-up. Cox and SNN are comparable in terms of predictor importance and individual survival predictions. RSF shows a tendency for according less importance to predictors with uneven class distributions and predicts clustered survival curves, the latter a result of tuning hyperparameters that influence forest shape through restrictions on terminal node size and tree depth.

Conclusions: SNN and Cox are comparable in terms of predictor importance, patient-specific survival prediction and performance measures, with SNN superior for later time points. RSF performs similarly to Cox in terms of performance measures. While possessing good discriminative ability, RSF is not suited to unbalanced predictors. We caution in particular against using RSF for predicting patient-specific survival, as standard model tuning practices may result in aggregated predictions, a quality that is not reflected in performance measure values.

[OC14.3]

Semi-parametric Copula Modelling Approaches for Clustered Survival Data

Il Do Ha, Sookhee Kwon

Statistics, Pukyong National University, Republic of Korea

Copula models have been widely used for analyzing clustered multivariate survival data. However, their modelling approaches are mainly based on two-stage estimation procedures. In this talk, we propose a semi-parametric Archimedian-copula modelling approach using a one-stage likelihood procedure. Here, marginal baseline hazards are non-parametrically modeled on a basis of a cubic M-spline. The performanc of the proposed method is evaluated using simulation study, with comparison of two-stage estimation methods. The new method is illustrated using a well-known clinical data set.

[OC14.4] **CASc Winner**

Bias reduction and solution to separation in the AFT models for small or rare event survival data

Tasneem Fatima Alam, M. Shafiqur Rahman

Applied Statistics, ISRT, University of Dhaka, Bangladesh

Background: The Accelerated failure time (AFT) model is widely used in medical science and reliability engineering for its intuitive interpretation. The model parameters are generally estimated by maximum likelihood estimation (MLE) which reports unbiased and consistent estimates when sample size is large and/or rate of censoring is low; however, its small sample performance is unknown.

Objectives: This study investigated the properties of MLEs of the regression parameters of the AFT models for small sample or rare-event (high rate of censoring) situation and introduced a penalized likelihood approach to address the problems.

Methods: The penalized likelihood function and the corresponding score equation were derived by adding a penalty term, equivalent to the Jeffreys invariant prior, to the existing likelihood function which was originally proposed by Firth (Biometrika, 1993) for reducing the first order bias in MLEs of the regression parameters of the exponential family models. The penalized method was illustrated for the most commonly applied log-location-scale family models such as Weibull, Log-normal and Log-logistic distributions. The illustration showed that the Jeffreys-prior based penalized likelihood succeeds to achieve convergence, providing finite estimates of the regression coeffcients and solves the problem of separation or monotone likelihood created by a covariate, which are not often possible by the MLE.

Results: Extensive simulation studies conducted separately for each of the log-location-scale models demonstrated the penalized approach to have a substantial improvement over MLE by providing smaller amount of bias, mean squared error (MSE) and narrower condence interval. An application of the methods using data with small sample and rare event supports the findings from the simulation.

Conclusion: When sample size is small ($N<50$) and even large but rate of censoring is high, the penalized method is recommended to be used in practice.

[OC14.5]

Testing procedures for the comparison of multiple characteristics of different survival functions

Robin Ristl [1], Martin Posch [2]

[1] *Medical University of Vienna, Austria*
[2] *Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria*

If the proportional hazards assumption does not hold the standard hazard ratio estimate is not a reliable measure of the treatment effect: it depends not only on the actual survival distributions but also on the censoring pattern, the study duration, and variations in the recruitment rates. Furthermore, in case of crossing hazard or survival functions estimated hazard ratios are hardly interpretable.

However, under non-proportional hazards, differences in the survival functions can be described by several parameters such as the difference between survival probabilities at predefined time points (as 1-year and 2-year survival) the difference between quantiles of the survival functions (as the difference in medians), or average hazard ratios computed up to a predefined time-point. Which of these parameters is best suited to quantify the difference in the survival functions can depend on their specific shape. Therefore, it can be desirable to specify more than one of these parameters as primary endpoint. However, if more than one primary endpoint is considered, a multiplicity problem arises and an inference approach controlling the family wise type I error rate as well as simultaneous coverage of confidence intervals are required for confirmatory conclusions.

Based on the counting process representation of the survival function estimators and the resulting asymptotical multivariate normality of the estimators, we derive an estimator of their asymptotic covariance matrix and construct corresponding simultaneous confidence intervals. Furthermore, we derive simultaneous confidence intervals based on the perturbation approach for survival function estimates which corresponds to a parametric bootstrap.

The finite sample properties of the proposed methods are investigated in a simulation study. We find that coverage probabilities are close to the nominal value, even for moderate sample sizes.

OC15: Individualised Risk Prediction 2

Chair: **Mark van de Wiel**

Free University, Amsterdam, Netherlands

[OC15.1]

Risk prediction for ordinal outcomes: calibration and the proportional odds assumption

Michael Edlinger[1], Maarten van Smeden[2], Hannes F. Alber[3], Ewout Steyerberg[4], Ben Van Calster[5]

[1] *Department of Medical Statistics, Informatics, & Health Economy, Medizinische Universität Innsbruck, Austria*
[2] *Julius Center for Health Science and Primary Care, University Medical Center Utrecht, Netherlands*
[3] *Department of Internal Medicine and Cardiology, Klinikum Klagenfurt am Wörthersee, Austria*
[4] *Leiden University Medical Center, Netherlands*
[5] *Development and regeneration, KU Leuven, Belgium*

Background: When evaluating the performance of risk prediction models, calibration is often underappreciated. There is little research on calibration for discrete ordinal outcomes.

Objectives: We aimed to compare calibration measures for risk models that predict a discrete ordinal outcome (typically 3 to 6 categories), investigate the impact of assuming proportional odds on risk estimates and calibration, and study the impact of assuming proportional odds.

Methods: We studied multinomial logistic, cumulative logit, adjacent category logit, continuation ratio logit, and stereotype logistic models. To assess calibration, we investigated calibration intercepts and slopes for every outcome level, for every dichotomized version of the outcome, and for every linear predictor (i.e. algorithm-specific calibration). Finally, we constructed calibration plots based on smoothed vector splines.

We used large sample simulations to study the behaviour of the logistic models in terms of risk estimates, and small sample simulations to study overfitting. As a case study, we used data from 4,888 symptomatic patients to predict the degree of coronary artery disease (five levels, from no disease to three-vessel disease).

Results: Models assuming proportional odds resulted in incorrect risk estimates when the assumption did not hold. Then, calibration slopes for specific outcome levels or for dichotomized outcomes deviated from unity even on the development data. Non-proportional odds models, however, suffered more from overfitting, because these models contain more parameters. The stereotype model fixed issues with the calibration slope, but still gave incorrect individual risk estimates. Algorithm-specific calibration for proportional odds models assumes that this assumption holds, and therefore did not fully evaluate calibration.

Conclusions: Assuming proportional odds can have a large impact on risk estimates and calibration. Therefore, proportional odds models should be used with care for risk prediction. More general models are recommended, but typically require larger sample sizes.

[OC15.2]

Flexible parametric survival modeling by means of penalized maximum likelihood

J. Hoogland[1], M. Belias[2], J. intHout[2], M. M. Rovers[2], K. G. M. Moons[1], T. P. A. Debray[1], J. B. Reitsma[1]

[1] *Julius Center, University Medical Center Utrecht, Netherlands*
[2] *Radboud Institute for Health Sciences, Radboud University Medical Center, Netherlands*

Context: Flexible parametric proportional hazards modeling for censored survival data has long been available in the form of the Royston-Parmar model. The same holds for penalized maximum likelihood methods that improve out of sample prediction.

Objectives: To capitalize on the advantages of fully parametric survival modeling in the context of prediction, we propose a penalized maximum likelihood alternative to the Royston-Parmar model. The main goal is to apply this method to the prediction of absolute individualized treatment effect over time.

Methods: The log hazard is modeled as a flexible function of time and the effects of the covariates of interest, where the latter may be non-linear or time-varying. To this purpose, time is modeled using restricted cubic splines and the covariates may interact with a function of time. Since the number of parameters that needs to be estimated grows quickly when loosening linearity or proportionality assumptions, model complexity is penalized during optimization of the likelihood. Cross-validation is used to estimate the optimal degree of penalization. A simulation study compares the accuracy of i) estimated survival probabilities as a function of time, and ii) estimated treatment effect on the probability of survival over time. We illustrate the implementation of the survival model in a clinical example where we develop a prediction model to estimate absolute treatment effect for individual patients. Hereto, we consider a large randomized trial on the secondary prevention of vascular events after ischemic stroke.

Results: Simulation study results show the advantage of penalization when model complexity increases relative to sample size. Simulation settings will cover varying baseline hazards (monotone versus non-monotone) and non-linear or time-varying covariate effects. Also, results from the case study will be presented.

[OC15.3]

Evaluation of sample size requirements for the development of risk prediction models

Rumana Omar[1], Chen Qu[1], Menelaos Pavlou[1], Shaun Seaman[2], Ian White[1], Gareth Ambler[1]

[1] *Statistical Science, University College London, United Kingdom*
[2] *MRC Biostatistics Unit, University of Cambridge, United Kingdom*

Background: Risk prediction models are routinely used in clinical practice. It is important that the dataset used to develop a risk model is of an appropriate size to avoid model overfitting problems and poor predictions in new data. Until recently, the main guideline was that at least 10 events per variable (EPV) are required. Recent work suggest that other factors rather than EPV are important for model performance.

Objective: We aim to find the sample size where there is a 90% probability that the risk model will have acceptable performance as assessed by the C-statistic, calibration slope and the accuracy of individual predictions (AIP). We consider performance to be acceptable if the estimated C-statistic is within 10% of true value, relative to the null value (0.5), for example for a true value of 0.75, performance is acceptable if the estimated C-statistic was within [0.725 to 0.775]. For the calibration slope, the performance is acceptable if the estimated slope was within [0.85 to 1.15] and if the true value of AIP is 0.05, performance is acceptable if the estimated AIP was within [0.03 to 0.08]. We also compared maximum likelihood (ML), ML with heuristic shrinkage and bootstrapping and ridge and lasso estimation methods. Additionally, we examine sample size requirements when bootstrap is used for internal validation as part of the model development process by examining how close the estimated performance (internal validation) is to actual performance (validation). We used simulation studies for our investigation and considered the effects of EPV, model strength and outcome prevalence on sample size.

Results and Conclusions: EPV and model strength are important for model performance. If the prognostic strength is high, an EPV of 15 provides a high probability of developing a risk model with good calibration, discrimination and predictive accuracy. For low prognostic strength an EPV of >30 is required. These EPV values are also acceptable for performing internal validation. There were small differences between the estimation methods with respect to the C-statistic, although lasso performed slightly worse. For calibration, ML was worse, and ridge was slightly better. ML, and sometimes lasso were slightly worse for AIP.

[OC15.4]

Regression model for personalized chance of longer survival using pseudo-observations

Isao Yokota[1], Kentaro Sakamaki[2]

[1] *Department of Biostatistics, Hokkaido University Graduate School of Medicine, Japan*
[2] *Center for Data Science, Yokohama City University, Japan*

Background: When patients choose a treatment, the personalized treatment effect is more important. Recently, Zhao et al. proposed personalized chance of longer survival which is the probability that the new treatment will lengthen survival time more than m-time in comparison to the standard treatment. In the existing method, the predicted personalized chance for each individual is estimated by using smoothing technique and bootstrap confidence interval is constructed for it.

Objective: We will predict the personalized chance based on regression models with pseudo-observations.

Method: Pseudo-observation for each individual is regarded as the contribution to marginal of personalized chance. An offset term in originally proposed personalized chance estimator is modified to apply pseudo-observations, while the definition of personal change is same. The regression of pseudo-observations on some covariates by using general linear model. To increase model flexibility, fractional polynomial terms could be included. Confidence interval could be constructed using model variance estimator.

Results: Our estimator showed little bias under finite samples as well as the original estimator through Monte Carlo simulation. Average length of confidence intervals of our method is 2-3% wider than the original method.

Discussions and Conclusions: The advantages of our method are (i) statistical test for the heterogeneity among subgroups, (ii) fewer amount of computation without bootstrap and (iii) easy application using standard analytical software. However, the statistical efficiency of our method is slightly poorer than the existing method.

Reference:

Zhao YQ, Redman MW, LeBlanc ML. Quantifying treatment effects using the personalized chance of longer survival. Stat Med. 2019; 38: 5317–5331.

[OC15.5]

Stacked Inverse Probability of Censoring Weighted Bagging: A Case Study in The InfCare HIV Register

Pablo Gonzalez Ginestet [1], Erin Gabriel [1], Julian Wolfson [2], David Vock [2], Ales Kotalik [2]

[1] *MEB, Karolinska Institutet, Sweden*
[2] *Division of Biostatistics, School of Public Health, University of Minnesota, United States*

Background: Rich data sources like population registers and electronic health care records are ideal for building prognostic predictions. Due to the many potentially highly correlated measures, non-linear associations and complex interactions that can be present in real-world, modern machine learning (ML) techniques can be better suited than more standard regression methods for developing predictive algorithms. However, right-censoring and competing risks are often present in these same data sources and can pose an obstacle to the proper application of many machine learning methods.

Objective: We propose a general approach called stacked Inverse Probability of Censoring Weighted (IPCW) Bagging as a pre-processing step that allows for all existing and any newly developed ML methods for classification to be applied to right-censored data and in presence of competing risks.

Method: Stacked IPCW Bagging incorporates the IPC weights into the resampling step of the bagging method and not in the algorithm itself. The procedure optimally stacks predictions from any set of IPCW bagged ML algorithms. We consider the IPCW area under the time-dependent ROC curve (IPCW-AUC) as a performance evaluation and calibration metric. The IPCW-AUC accounts for censoring and competing risk. In two simulation settings and over various censoring mechanisms, including dependent censoring, we compare our procedure to existing ML methods commonly available in R packages that allow for IPC weights as well as survival methods. Lastly, we illustrate our method using the Swedish InfCare HIV register to predict individuals for whom treatment will not maintain an undetectable viral load for at least two years following initial suppression.

Results and Conclusions: Finding suggests that predictions based on the IPCW bagging procedure preform similarly to the analogous directly or 'natively' weighted procedures. However, when weights cannot be directly applied to an ML procedure, IPCW bagging allows for the desired algorithm to account for censored observations. Moreover, the stacked IPCW bagging methods have the best performance on average.

OC16: Functional Data Analysis

Chair: **Timothy Cannings**

University of Edinburgh, United Kingdom

[OC16.1]

Modeling the effect of dynamic covariates on time-to-event processes via Functional Data Analysis

Francesca Ieva, Marta Spreafico, Davide Burba

Mathematics, Politecnico di Milano, Italy

Background: Many situations in clinical practice can be modeled in the framework of recurrent events. It is often the case where the association between the occurrence of events and final outcomes is of interest. A clear example is the study of how survival in chronic patients is affected by the drug assumption (or adherence to treatment) over time, as well as by other events processes, like re-hospitalizations or chemotherapy treatments.

Objectives: Aim of our study is to enrich the information available for modeling survival with relevant dynamic features, properly taking into account their time-varying nature, as well as to provide a new setting for quantifying the association between time-varying processes like re-hospitalizations or drugs purchases and time-to-event outcomes like survival.

Methods: Data were retrieved from the administrative databases of Lombardy Region (Italy). Patients hospitalized for Heart Failure between 2000 and 2012, assuming multiple drugs over time, were included considered. Compensators of marked point processes are estimated by the recurrent events (drugs purchases and re-hospitalizations). By means of Functional Principal Component Analysis (FPCA), a suitable dimensional reduction of these objects is carried out in order to plug them into a survival Functional Cox regression model.

Results & Conclusions: After adjusting for case mix, the models showed that proper Beta Blocker agents intake is correlated to longer life expectancy (HR=0.997, 95% CI = [0:9948; 0:9992]). Interestingly, the HR relative to the functional scores of hospitalization process was 0:756 (95% CI = [0:7039; 0:8119]), meaning that patients with many hospitalizations at the beginning of the observation period and few hospitalizations in the end had higher survival, standing as a proxy of patients' conditions.

In the end, we proposed an effective methodology to extract and summarize information from trajectories of compensators of suitable marked point processes for recurrent events intended as functional data. The introduction of this novel way to account for time-varying variables allowed for modeling self-exciting behaviors, for which the occurrence of events in the past increases the probability of a new event, and to make personalized predictions, quantifying the effect of personal behaviors and therapeutic patterns on survival.

[OC16.2]

Including dynamic covariates in survival models via Functional Data Analysis: an application to osteosarcoma

Marta Spreafico [1], Francesca Ieva [1], Marta Fiocco [2]

[1] *MOX - Department of Mathematics, Politecnico di Milano, Italy*
[2] *Mathematical Institute, Leiden University, Netherlands*

Background: Time-varying covariates are of great interest in clinical research since they represent dynamic patterns that could reflect patient's disease progression. In osteosarcoma, depending on patient's health status or development of toxicity, biomarker values evolve and chemotherapy treatment is modified by delaying a course or reducing the dose intensity. Models for time-to-event able to deal with the dynamic nature of time-varying responses during follow-up are necessary, still not well developed. The effect of these type of biomarkers on survival are still unclear.

Objectives: Studying the association between time-varying processes and a time-to-event outcome is a challenging task. In this work, innovative methods to represent dynamic time-varying covariates by means of Functional Data Analysis (FDA) and to include them into survival models are discussed.

Methods: Data from MRC BO06/EORTC 80931 randomised controlled trial for treatment of osteosarcoma were used. Time-varying covariates related to alkaline phosphatase (ALP) levels and chemotherapy dose during treatment were considered. Firstly, the joint modelling approach in order to investigate how patients' survival were influenced by ALP longitudinal values was estimated. Then, FDA techniques to represent time-varying ALP levels and chemotherapy dose as functions were exploited. Through Functional Principal Component Analysis (FPCA), dynamic covariates were finally included in functional regression models for survival data.

Results: Methodological approaches that deal with time-varying covariates into the time-to-event framework allowed to take into account important information about the dynamic behaviours of the generating processes underlying the data. This provide interpretative and forecasting tools in osteosarcoma research. High ALP levels during follow up reflected poorer overall survival. On the other hand, dose-intense profiles seemed not to be associated with a better prognosis. Functional representations were able to model individual realisations of the intended treatment, suggesting that other aspects of chemotherapy treatment, such as latent accumulation of toxicity, must be taken into account.

Conclusions: The complexity of the effect of chemotherapy data on survival outcome asks for the developments of new methodologies. This study shows that working in this direction it is a difficult though promising approach, which could lead to new improvements for subject-specific predictions and personalised treatment.

[OC16.3]

Functional Random Forest for Mixed Data

Mohammad Fayaz [1], Allireza Abadi [2]

[1] *Department of Biostatistics, Shahid Beheshti University of Medical Sciences, Iran, Islamic Republic of*
[2] *Faculty of Medicine, Department of Community Medicine, Shahid Beheshti University of Medical Sciences, Iran, Islamic Republic of*

Background: Functional Data Analysis (FDA) is a field in statistics which the building blocks are derived functions rather than observation points. In the high dimensional settings, the FDA reduces the dimensionally of the data. There are many options for estimating these functions such functional principal component analysis. But in many situations, we don't interest in only functional data and there are vectors or images. The regression models in the FDA include: Function-on-scalar regression, Scalar-on-Function Regression and Function-on-Function Regression. There exists different frameworks for estimating these models, for example classical statistical methods such as Generalized Additive Models (GAM) and statistical learning models such as Random Forest.

Objective(s): In this study, we first propose a functional random forest for mixed data with both continues and categorical response variables and mixed covariates. In this way, we first split functional data by a random mechanism and transform it to multiple observations. We also consider the function sequences in the splitting mechanism. In the second part, we first estimate the functional part with basis functions and FPCA and then run the random splitting procedures. And finally, we consider the vectors in FPCA and put it as hybrid FPCA, then we do functional random forests on it. This method is conduct in scalar-on-function (single and multiple functions), Function-on-scalar and function-on-function regressions.

Method(s): First, we compare our models in simulation scenarios with dense and sparse functions. Then, we adopt and compare our models on different datasets (DTI, EEG-EMG and …) which is published in the FDA literature.

Results: The proposed functional random forest models fit well on the mixed data and have higher prediction accuracy and lesser errors than non-functional random forest models. We also calculate and compare the variable importance measure.

Conclusions: In this study, the mixed data is a combination of functions and vectors. One of the most important statistical learning model, random forest, is modified to handle these type of dataset. The Shiny App with the R codes are available in GitHub.

[OC16.4]

Comparism of different statistical models used in shape index calculation on human face

Stanislav Katina, Vojtěch Šindlář

Institute of Mathematics and Statistics, Masaryk University, Czech Republic

Background: Spatial interpolation and smoothing is usually done for one surface. In our case, we have random samples of such surfaces represented by human faces captured by stereo-photogrammetry and characterised by about 150,000 points. These points are triangulated by about 300,000 triangles. The number of points is extremely high for the purpose of statistical analyses, therefore the 3D coordinates of (semi)landmarks on curves or surface patches sufficiently characterising the shape have to be automatically identified and this simplified model comprising about 1000 points is then used in further statistical modelling in functional data analysis setting. The identification of (semi)landmarks is a complex process during which B-splines, P-splines and thin-plate splines are used together with the measures of local surface topology, including principal curvatures and shape index [1].

Methods: Shape index is calculated in R using different linear statistical models of z coordinates on x and y coordinates, i.e. quadratic with interaction without/with intercept, cubic with interaction of x and y without/with intercept (without/with other interactions), and similar models of higher order. The estimates of regression coefficients related to the quadratic terms and their interaction are elements of Weingarten matrix from which the principal curvatures are calculated. These models are applied on sufficiently large neighbourhood of all points in local 3D coordinate system.

Objectives: Since the measures of local surface topology represent principal guide in estimating locations of ridge and valley curves across the face [2], we aim to compare different linear models used in shape index calculation on faces of patients with facial palsy and healthy controls.

Results: Problems related to the surface imperfections could be solved by trimming unimportant surface regions around face and by winsorisation of outliers of principal curvatures. It could also lead to spatial smoothing of these characteristics or shape index.

Conclusions: We suggest to use quadratic or cubic model with interaction of the first order without intercept.

Acknowledgment: Project No. MUNI/A/1418/2019.

References:

1. *Vittert L, AW Bowman, S Katina. A hierarchical curve-based approach to the analysis of manifold data. The Annals of Applied Statistics 13,4 (2019): 2539–2563.*
2. *Katina S, et al. The definitions of three-dimensional landmarks on the human face: an interdisciplinary view. Journal of anatomy 228,3 (2016): 355–365.*

[OC16.5]

Parameter clustering in Bayesian functional data models of neuroscientific recordings

Nicolo Margaritella

University of Edinburgh, United Kingdom

Background: The modelling of brain activity has been central to the field of neuroscience throughout the 20th century but only in recent years it has started benefiting from the extraordinary technological advances in the fields of neurophysiology and neuroimaging. Most of the recording tools in neuroscience today produce remarkable amounts of spatio-temporal data that are obtained simultaneously from several parts of the brain. These large and intricate datasets require new advanced statistical methods to efficiently extract meaningful patterns behind such complexity.

Objectives: We propose new models that merge ideas from Functional Data Analysis and Bayesian nonparametrics to obtain a flexible exploration of spatio-temporal data in neuroscience. We developed a Dirichlet process Gaussian mixture model to cluster functional Principal Component scores within the standard Bayesian functional Principal Component Analysis framework. This approach allows us to capture the structure of spatial dependence among smoothed curves and its interaction with the time domain. Moreover, by moving the mixture from data to functional Principal Component scores, we obtain a more general clustering procedure, allowing a substantially finer curve classification and higher level of intricate insight and understanding of the data.

Moreover, we extend our model to the challenging case of multilevel functional data where multiple curves are nested within subjects, and subjects are divided in groups. We develop a method based on a parsimonious trade-off between group behaviours and individual deviations, returning a comprehensive exploration of intricate multilevel functional data.

Results and Conclusions: We present results from a Monte Carlo simulation study showing improvements in curve and correlation reconstruction compared with different Bayesian and frequentist functional Principal Component Analysis models. Furthermore, we apply our method to a resting-state single-subject fMRI data analysis and to an EEG multilevel group analysis, both providing a rich exploration of the spatio-temporal patterns in brain time series. The proposed methods have the potential to answer important questions not only in the study of brain processes but also in the characterisation, early diagnosis and prognosis of brain diseases.

OC17: Omics & Genetics

Chair: **Zoltán Kutalik**

University of Lausanne, Switzerland

[OC17.1]

Identification of Genetic Biomarkers using Models of Network Interactions

Munshi Imran Hossain

CYTEL Inc., India

Background: Genes play an important role in regulation pathways. They regulate the synthesis of proteins which in turn regulate cell processes. Since a number of diseases can be traced back to errors in regulation pathways, knowledge of these pathways can help us understand critical points at which such pathways can be intercepted with drugs, for the treatment of diseases. Such studies are also becoming popular because of the easy availability of high-throughput technologies such as microarrays for gene expression.

Objective(s): The objective of this talk is to present some methods of analysing gene regulation pathways with the help of Boolean networks; the higher aim being an attempt to establish possible causal pathways in the networks. Boolean networks are models used for studying gene regulatory networks in which genes are modelled as binary variables; 0 indicates that a gene is not expressed and 1 indicates that it is expressed. I will also talk about the idea of attractors in such networks. These are stable states of the network and are important because they are manifested as phenotypes.

Method(s): Time-series data of gene expression levels is used for building these models. Using this data, we can fit Boolean networks of different types, such as - synchronous, asynchronous and probabilistic networks. In order to build Boolean networks, the expression levels are converted to binary states. This is done using methods like k-means clustering, edge detection etc.

Once the network has been constructed, we can analyse the transitions between its states and the stability of the networks.

Results: These methods create directed networks that show possible pathways of causality between the genes, explaining which genes regulate others in the network. Analysis of the stable states of these networks helps identify which genes are usually expressed and consequently the dominant phenotypes of the organism.

Conclusions: Boolean networks are an efficient way to capture the complex behaviour of gene regulation pathways and unlike other methods such as those that are based on ordinary differential equations, they do not require extensive knowledge of the underling biological process. With improvement in computing, these are becoming an integral part of the drug development process.

[OC17.2]

Efficiently analyzing big data with Bayesian joint models for longitudinal and time-to-event data

Pedro Afonso [1], Dimitris Rizopoulos [1], Rhonda Szczesniak [2], Weiji Su [3], Eleni-Rosalina Andrinopoulou [1]

[1] *Department of Biostatistics, Erasmus University Medical Center, Netherlands*
[2] *Department of Biostatistics & Epidemiology, Cincinnati Children's Hospital Medical Center, United States*
[3] *Department of Mathematical Sciences, University of Cincinnati, United States*

Background: Cystic Fibrosis (CF) is a lethal genetic lung disease that requires frequent monitoring. A commonly measured marker of lung function in CF is forced expiratory volume in 1 s of % predicted (FEV1). This sub-study stems from an international workgroup project to establish guidelines for longitudinal analysis of CF FEV1 using the U.S. CF Foundation Patient Registry, which contains health-related information for approximately 30,800 CF patients who collectively contributed 1,215,171 observations and 255,804 years of follow-up. It has been of clinical interest to model the association between FEV1 and time-to-death, but previous work neglected recurrent acute pulmonary exacerbation (PE) events and time-to-lung-transplantation, and mainly made use of smaller subsets of the registry data. Analyzing the entire dataset is expensive in terms of computation times. There is a need for algorithms that perform distributed analyses that simultaneously and unbiasedly investigate multiple correlated outcomes.

Objectives: Our primary goal was to investigate the association between FEV1 and the risk of PE, lung-transplant/death using all available registry data (2003-2016) for patients aged ≥ 6 years.

Methods: We relied on the joint modelling framework, under the Bayesian paradigm, to simultaneously fit a joint model for a longitudinal marker, a recurrent event, and a terminal event. Due to the high computational time required to evaluate the entire dataset, we split the dataset into smaller non-overlapping subsets, parallelized their analyses, and then averaged individual posterior samples together. We explored different data splitting and combining strategies to get a consensus posterior.

Results: Preliminary results suggest that FEV1 is negatively associated with the risk of experiencing death or transplantation. The model obtained from the split data estimates an association of -0.11 (95%CI -0.13,-0.10), while the model leveraging the entire dataset shows an estimate of -0.11 (-0.12,-0.10). The 10-fold reduction in sample size produced a 90% decrease in computing time. The results were also investigated through simulation studies.

Conclusions: The idea of distributing the Bayesian calculation is likely to be a useful solution to handle big datasets without compromising the amount of information taken into account or sacrificing model adequacy, thereby enhancing our understanding of CF FEV1 decline.

[OC17.3]

Generalized random forests for high-dimensional longitudinal data

Louis Capitaine, Robin Genuer, Rodolphe Thiebaut

Public Health, Bordeaux Univ / Inserm, France

Background: Random forests are one of the state-of-the-art supervised machine learning method which behaves well in high-dimensional settings where p, the number of predictors, is much larger than the number of observations n. Repeated measurements can help by offering additional information, so it is of particular interest to take them into account when analyzing high-dimensional data. Tree-based methods have already been adapted to clustered and longitudinal data by using a semi-parametric mixed-effects model, in which the non-parametric part is estimated using regression trees or random forests.

Objective: To propose a generalized random forest model for high-dimensional longitudinal data

Methods: We develop a flexible stochastic semi-parametric mixed effects model where the non-parametric mean function is a RF aggregating randomized Classification And Regression Trees (CART) and the stochastic part is including a general Gaussian process (Ornstein-Uhlenbeck process and fractional Brownian motion) in addition to the random effects. Parameters were estimated using an adaptation of the Maximum Likelihood (ML)-based EM algorithm. All available tree-based methods for longitudinal data were compared in an extensive simulation study inspired by the application with 100 datasets of n=20 and p=6 or p=8000. We apply the methods to a therapeutic vaccine trial including 17 HIV-infected patients with 10 repeated measurements per individual of plasma HIV viral load to be predicted by gene expression measured at the same times in whole blood. All existing and proposed methods have been implemented together in the R package longituRF.

Results: Simulation experiments show that RF-based methods, including the generalization proposed in this work, gave much better results than tree-based methods in term of bias (25% lower in average) and prediction error, especially in the context of high-dimensional framework. In the application, the best approaches leading to the lowest prediction error (as evaluated with 25 training/test sets random splits) were RF-based methods including a Brownian motion. The approach selected 21 gene transcripts for which the association with HIV viral load was fully relevant and consistent with results observed during primary infection.

Conclusions: We propose a flexible RF model for high-dimensional longitudinal data which presents the best performances as compared to previously existing methods.

[OC17.4]

RUV-NB: Removing Unwanted Variation from Single-Cell RNA-seq Data

Agus Salim [1], Terence P. Speed [2]

[1] *Mathematics and Statistics, La Trobe University, Australia*
[2] *Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, Australia*

Despite advances in other aspects of single-cell RNA-seq (scRNA-seq) processing and analyses, relatively little progress has been made in developing normalization methods specific to scRNA-seq data. Library size-based normalization still used quite frequently in scRNA-seq data analysis. Normalization can be seen as attempt at removing unwanted variation due to factors other than biology. Unwanted variation due to batch effects, library size, cell-cycle and other factors exist in scRNA-seq data. However, existing methods for batch effect correction return corrected data in lower dimensions, making them unsuitable for genewise analyses such as differential expression. Corrected data from existing methods often also fail to maintain the mean-variance relationship of count data, making the corrected data unsuitable for analyses using standard methods that use Negative Binomial (NB) or Zero Inflated Negative Binomial (ZINB) distributions.

We developed RUV-NB, a method for removing unwanted variation from scRNA-seq data. RUV-NB models the scRNA-seq count data directly using Negative Binomial regression. As such, it avoids the need to log transform the data and it maintains the natural mean-variance relationship inherent in scRNA-seq data. Using subset of control genes from which the unwanted variation can be estimated, we employ iterative reweighted least squares (IRLS) algorithm to estimate gene- and cell-specific normalization factors required to remove the unwanted variation. Using published scRNA-seq data, we demonstrate that RUV-NB can be used to successfully remove batch and cell-cycle effects. When there is no biological differences in the data, we show that unlike other methods, RUV-NB maintains type I error rates well when used in the context of differential expression (DE) analysis. The type I error rates also appear to be robust when the number of unwanted factors are overestimated. The method is implemented as an R package and implemented parallel computation to ensure scalability

[OC17.5]

Conditional cumulative distribution function estimation for differential expression analysis in scRNA-seq data

Marine Gauthier[1], Rodolphe Thiébaut[2], Véronique Godot[3], Boris Hejblum[1]

[1] *Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, team SISTM, UMR 1219, Bordeaux; INRIA Bordeaux Sud Ouest, Talence; Vaccine Research Institute, Créteil, France*
[2] *Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, team SISTM, UMR 1219, Bordeaux; INRIA Bordeaux Sud Ouest, Talence; CHU Bordeaux; Vaccine Research Institute, Créteil, France*
[3] *Vaccine Research Institute, Créteil; Inserm, U955, Team 16, Univ. Paris-Est, Créteil, France*

Background: Unlike bulk RNA-seq data analysis (using the average gene expression in a cell population), the analysis of single-cell RNA-seq (scRNA-seq) data makes it possible to study biological mechanisms at the cellular level. Differential Expression Analysis (DEA) consists in testing the association of a gene expression with one or more factors. State-of-the-art methods for scRNA-seq DEA face methodological issues, as they rely on strong distributional assumptions that are difficult to test in practice, questioning the validity of their results given the high rate of zeros in scRNA-seq data. While the increasing complexity of clinical and biological studies calls for greater tool versatility, the majority of existing methods only tackles the comparison between two conditions.

Objective: We propose a new distribution-free approach to test the association of gene expression to one or several variables of interest (continuous or discrete) potentially adjusted to additional covariates.

Method: If a group of factors is associated with the gene expression, the immediate consequence is that the conditional cumulative distribution (cCDF) function of gene expression would be significantly different than the marginal cumulative distribution (mCDF), which overlooks these conditioning. Because the cCDF can be rephrased as an expectation, it can be estimated through regression methods. This estimation of the cCDF can then be compared to the mCDF by computing a distance between the two (L1, L2 or Lsup norms can be considered). A permutation test then detects the differentially expressed genes among a population of cells.

Results: We compare the properties and performance of the various norms to differentiate the cCDF from the mCDF. While asymptotics remains elusive, we have developed a permutation procedure – accounting for additional (continuous) covariates –to estimate the null distribution of the distance between the two CDFs. We demonstrate the good performance in extensive numerical simulations and we apply this method on a real scRNA-seq data set studying subpopulations of blood dendritic cells.

Conclusion: We propose a new and versatile method to perform differential analysis of scRNA-seq data, that can accommodate complex designs, e.g. with more than two experimental conditions or with multivariate responses, while adjusting for additional covariates.

OC18: Longitudinal Data

Chair: **Tomasz Burzykowski**

Hasselt University, Belgium

[OC18.1]

Joint Modelling of Longitudinal and Survival Data Applied to Group Sequential Trials

Abigail Burdon

Mathematical Sciences, University of Bath, United Kingdom

Consider a Phase 3 clinical trial where the primary endpoint is overall survival. The investigators wish to define a group sequential stopping rule which allows for early stopping for efficacy or futility. Suppose that longitudinal data are observed on a biomarker which is assumed to be predictive of survival, and we wish to use this biomarker information to inform early stopping decisions. We shall present a joint model for survival and longitudinal data and a method which establishes the distribution of successive estimates of parameters in the joint model across interim analyses. Then, we are equipped to use the estimates to define a stopping rule. With the methodology in place, by simulation we can assess the potential benefits of including biomarker information, how this affects interim decisions and ultimately alters the trial.

[OC18.2]

Dealing with missing values in multivariate joint models for longitudinal and survival data

Nicole Erler

Biostatistics, Erasmus MC, Netherlands

Background: Chronic hepatitis C is a severe and increasing public health issue. Although nowadays most patients can be cured, the infection is often undetected until symptoms of permanent liver damage become apparent, putting patients at a considerably higher risk for liver diseases and, as a consequence, liver-related mortality.

Patients are therefore often monitored beyond the end of treatment, which provides us with both baseline and repeatedly measured data on patient and disease characteristics. Typically, joint models for longitudinal and time-to-event data are used to adequately model (and subsequently predict) the hazard of experiencing an adverse event utilizing time-varying covariate information.

A potentially serious additional issue in the analysis of our retrospective hepatitis C cohort is the large amount of missing data for several important covariates (for up to 58% of patients), and the restriction of currently available methodologies and software to complete case analysis.

Objective: To prevent severe loss of power and to reduce the possibly large bias that a complete case analysis would produce, we present a fully Bayesian approach that jointly models longitudinal and survival outcomes in the presence of missing data.

Methods: We factorize the joint distribution of outcome and covariates into a sequence of univariate distributions, which allows us to appropriately handle multiple incomplete baseline and time-varying covariates of different types. Moreover, non-linear associations between variables can be incorporated while assuring compatibility of the models involved. The approach is valid under Missing At Random, with the potential for further extension to non-random missingness, and is also applicable to other types of joint models (not involving time-to-event).

Results: In our hepatitis C cohort of 490 patients, the complete case analysis would have been based on 24% of the patients, of which only 28 experienced an event, and was therefore not feasible. Using our approach, we successfully implemented a joint model with four longitudinal and eleven baseline covariates. This approach is currently being implemented in an R package.

Conclusion: Our proposed approach provides a flexible way to handle complex joint models in the presence of incomplete data. Simulation studies are needed to empirically confirm that results are unbiased.

[OC18.3]

A new type of generalized linear mixed models with linear predictor linked to marginal mean

Zsolt Lang [1], István Fodor [2], László Ózsvári [2]

[1] *Department of Biomathematics and Informatics, University of Veterinary Medicine Budapest, Hungary*
[2] *Department of Veterinary Forensics, Law and Economics, University of Veterinary Medicine Budapest, Hungary*

Background: The marginal mean of a classical generalized linear mixed model is usually not immediately connected to the linear predictor expressing fixed effects. For instance, in the case of a logistic normal model fixed effects can be related to the logit-scaled means, which are equal to (assuming perfect model fit) the marginal medians typically differing from the marginal means. Although medians may better represent central tendency in skewed distributions, in some applications (e.g. when pooled prevalence of a disease in a population is estimated) marginal means are preferred to marginal medians.

Objective: We propose a new class of generalized linear mixed models where the marginal mean is expressed with a linear predictor, i.e. a linear expression of explanatory variables (fixed effects), and the normally distributed random effects part of the model reveals the hierarchical, clustered and correlated nature of data being evaluated.

Method: The basic idea of our approach is that a distribution in a selected family (e.g. beta, gamma, Weibull) having the same expected value as the marginal mean is fitted to the conditional mean of the outcome variable given the random effects of the model. The variance of the conditional mean is obtained by a suitably chosen function (e.g. a power function with estimated exponent) of the variance of the actual linear expression of the random effects involved. This framework allows for selecting separate modelling structures for the marginal means and the random effects.

Results and Conclusions: Our method is illustrated with estimation of prevalence of paratuberculosis infection using diagnostic test results of cattle bred in 53 Hungarian herds. Binomial distribution was fitted to the number of infected animals in each herd. Beta distribution was selected for the conditional mean proportion of infection given the random effect of the actual herd. Both frequentist and Bayesian analyses are planned to be outlined and compared.

[OC18.4]

Joint modelling of a semicontinuous longitudinal biomarker and a terminal event

Denis Rustand [1], Laurent Briollais [2, 3], Virginie Rondeau [1]

[1] *Biostatistic Team, Bordeaux Population Health Center, ISPED, Centre INSERM U1219, Bordeaux, France*
[2] *Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada*
[3] *Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, ON, Canada*

In cancer clinical trials, the sum of the longest diameter (SLD) of target lesions is a biomarker of interest which reflects both the tumor burden and its evolution over time. An excess of zero values and right skewness often characterize the SLD distribution. While a nonlinear transformation can easily handle the latter, the zero-inflation problem requires a more sophisticated approach. Left-censoring has been proposed as a way to handle the excess of zero values in a mixed-effects model (i.e. values below a detection limit are censored). Patients responding well to a treatment can reach the complete response state as defined by RECIST criteria, in which case the tumor size shrinks until reaching a ‚true zero' value (i.e. not censored). We propose a two-part joint model that decomposes the distribution of the biomarker into a binary outcome (zero values vs. positive values) and a continuous outcome, both outcomes being modelled by a mixed effects regression model. Therefore, the binary part captures the effect of covariates on the probability of zero value of the biomarker. We propose two forms for the continuous part: the conditional form captures the effect of covariates on the expected value of the biomarker among positive values and the marginal form captures the effect of covariates on the marginal mean of the biomarker. The survival times are modelled using a Cox proportional hazards model with splines approximation for the baseline hazard. We propose several association structures to link the biomarker to the risk of event. We illustrate with simulation studies the performances of the models in terms of bias, accuracy and coverage probabilities when the model assumptions are misspecified, with different rates of zero excess. A real data application to a colorectal metastatic cancer trial comparing two treatment strategies is also proposed, showing how the biomarker evolution over time can bring additional information to evaluate the treatment effect on the risk of death. The zero inflation is a common problem in biomedical research, e.g. when quantifying exposure or measuring symptoms of a disease, and our proposed model is also relevant in this wide spectrum of applications.

[OC18.5]

Prognostic stratification: using Boolean classification to identify ordinal levels of risk

Roger Marshall

Population Health, The University of Auckland, New Zealand

Background: Recent years has seen the growth in regression modelling to create models of individuals' risk. Typically individual probabilities are assigned into bands, or strata, (e.g. "low", "moderate", "high" risk) to determine patient management and interventions.

Objective: An alternative way to form bands is to directly identify prognostic strata, without modelling individual risk which are typically based on linear scores. Rather, Boolean logical constructs are considered to create strata that are homogeneous with respect to risk.

Method: A procedure is described, first suggested in 1999 (Marshall, Statist. Med 18), to form ordinal strata by identifying Boolean combinations of attributes. It is based on the idea of forming nested shells of risk, with shells identified using search partition analysis to optimize between group discrimination but within group homogeneity. Previously awkward to implement, the procedure is now a feature of an R package, spanr, developed by the author and which will be described.

Results: The approach is illustrated with analysis of breast cancer and cardiovascular disease for which simple strata, defined by Boolean combinations, are identified that give performance comparable to regression based rules.

Conclusions: The approach suggested here is more in line with clinical logic, in which demarcated subgroups are formed, than regression modelling. It gives simple and easy-to-understand prognostic strata.

OC19: Epidemiology: Methods & Applications

Chair: **Andrzej Pająk**

Jagiellonian University Medical College, Poland

[OC19.1]

A Robust Discriminant Framework Based on Qeeg on the Diagnosis of Alzheimer's Disease

Qi Ge, Zhuochen Lin, Yongxiang Gao, Jinxin Zhang

Medical Statistics, Sun Yat-Sen University, China

Background: Over the last two decades, quantitative electroencephalography (qEEG) has emerged as a diagnostic tool of Alzheimer's disease (AD), which is inexpensive and non-invasive. However, the original EEG signals are susceptible to various artifacts and robust markers are still lacking, qEEG-based automatic diagnosis of AD is therefore not widespread in the clinical.

Objective: The study aimed to propose a discrimination framework based on robust EEG features to distinguishing patients with AD from healthy controls (HC).

Methods: In the developed discrimination framework, the multi-channel EEG signals were preprocessed by filtering, segmenting and independent principal component analysis (ICA) firstly. After that, Maximal overlap discrete wavelet transform (MODWT) was used for extracting time-frequency information of the preprocessed signals. The complexity and synchronization of brain function were measured using Shannon entropy (SE), Variance (VA), Inter-quartile range (IQR), Pearson correlation coefficient (PCC) and Hoeffding's D correlation coefficient (HCC). Afterward, feature selection yielded a robust reduction to the number of features and channels needed and also improved classification performance. Lastly, the selected features were fed into a Linear discriminant Analysis (LDA) based classifier to distinguish the signal between two groups: AD vs HC. The performances of the proposed discrimination framework were evaluated by leave one subject out cross-validation (LOSOCV) on a dataset of 48 subjects (24 AD and 24HC), recorded during an eye-closed resting condition at the first affiliated hospital of Sun Yat-sen University (China).

Results: The results showed MODWT outperformed the traditional Fourier transform. The proposed parametric features (the combination of VA and PCC) and the nonparametric features (the combination of IQR, HCC, and SE) gave similar performance when discriminating HC and AD patients. And the combination of all five features achieved the best accuracy of 98.49%, the sensitivity of 98.21% and the specificity of 98.85%, separately.

Conclusions: The developed discrimination framework can make an automatic diagnosis of AD and MCI with high accuracy in a systematic way.

The work is supported by the National Natural Science Foundation of China (NSFC 81773545) and the Natural Science Foundation of Guangdong Province, China (2016A030313365)

[OC19.2]

Modeling of multivariate longitudinal markers in a latent disease timescale:application in Alzheimer's disease

Jeremie Lespinasse [1], Cécile Proust-Lima [2], Carole Dufouil [1]

[1] *Bordeaux University Hospital, France*
[2] *INSERM U1219, France*

Background: Alzheimer's disease and related disorders (ADRD) are characterized by a progressive deterioration of many biomarkers including imaging and cognitive functions. Yet the sequence and shape of successive deteriorations is very difficult to identify from longitudinal cohort data. One critical challenge is the timescale of interest, the disease time, which is unknown.

Objective: Our objective was to define a nonlinear multivariate mixed model incorporating a latent individual time shift to realign trajectories of multiple biomarkers into the latent disease timescale. We apply this methodology to characterize dynamics progression and temporal sequence of the ADRD biomarkers and cognitive domains in a French clinic-based cohort study (the Memento study).

Method: We defined a joint model to simultaneously describe the biomarker trajectories using a nonlinear multivariate mixed model based on a latent disease time which is defined as an individual linear function of observation time. The estimation procedure by maximum likelihood builds on latent process mixed modeling techniques implemented in the R package lcmm.

To better apprehend the latent disease time, we also considered a mixture of distributions which allows discriminating early from late disease onsets.

The model was applied on the Memento study. The sample included a total of 2190 participants followed up for a maximum of 5 years with repeated measurements of markers from neuropsychological assessments, cerebral MRI and FDG-PET and cerebrospinal fluid puncture. We considered the repeated data of 10 biomarkers (around 45000 observations) and adjusted the analyses for the major confounders in ADRD (age, sex, education, ApoE4 status).

Results: We identified that (1) the latent disease time spread over decades around the observed time, (2) all outcomes significantly progressed to more severe pathological conditions, (3) some covariates were differently associated with biomarkers (e.g., higher education).

Conclusions: By exploiting the mixed modeling framework, we were able to simultaneously model multivariate longitudinal data which had the particularity to be very heterogeneous in time. Beyond ADRD, this approach responds to a recurrent problem in the epidemiology of chronic diseases where the true timescale of interest is usually not observed.

[OC19.3]

Optimal pooling of blood samples

Dominik Grathwohl

Clinical Development Unit, Nestlé Research Center, Switzerland

Background: Laboratory costs of big trials (n > 1000) may approach costs of $ 1 million. Pooling of blood samples reduces costs and produces unbiased mean estimates. However, the problem arises on how to estimate the variability of the pooled samples? Pooling has a long tradition and goes back to Dorfman, 1943. It is used in human Genetics, see Gasthwirth 2000. However, these are applications to binary outcomes. For continuous outcomes interest arise at the dawn of omics. The high costs of Affymetrix chips lead people to develop pooling strategies: Peng 2003, Joly 2005, Zhang 2005, Kusonmano 2012. However, none of the authors have used the decomposition of the between and within subject's variability in order to reconstruct the test statistics resulting from individually measured data.

Objective: Optimal pooling of blood samples of dedicated sizes in order to obtain the maximum power.

Methods: We assume a normally distributed outcome measurement. The variance of the individual measurements can be decomposed in a between subjects' variance and a within subjects' variance:

VAR = VAR_b + VAR_w. The variance of m pools of k subjects is VAR_p = 1/k * VAR_b + VAR_w. Making use of different sized pools, 1/k can be regressed on VAR_p. The regression coefficients are VAR_b and VAR_w. Hence the variance of the individual measurements can be reconstructed.

Results: In a simulation study we assumed to have 500 subjects in some control group and 500 subjects in an intervention group. The intervention lead to an increase from 1.204 to 1.34 in alpha tocophenol. The between subjects sd is assumed with 0.71 and the within subjects sd is assumed with 0.29, all in mg/dl. A two-sample t-test has a power of 80% in order to demonstrate this difference as statistically significant on an alpha level of 0.05. A pooling strategy of 21 individually measured subjects and two pools of each 240 subjects in each group will lead to a power of 74%.

Conclusions: Instead of 1000 measurements, only 46 measurements are necessary. This corresponds to cost savings from $ 1 million to $ 46 000.

[OC19.4]

Survival inequalities in non-Hodgkin lymphoma: excess hazard models with joint modelling multiple imputation

Matthew Smith[1], Edmund Njeru Njagi[1], Bernard Rachet[1], Aurelien Belot[1], Matteo Quartagno[2], Sara Benitez-Majano[1], Audrey Bonaventure[1], Susan Gachau[3]

[1] *Cancer Survival Group, London School of Hygiene and Tropical Medicine, United Kingdom*
[2] *Institute of Clinical Trials and Methodology, University College London, United Kingdom*
[3] *Health Services Unit, KEMRI-Wellcome Trust Research Programme, University of Nairobi, Kenya*

Background: Universal healthcare in England aims to offer equitable access to facilities for all patients. However, inequalities in cancer survival persist, particularly between socioeconomic levels. Furthermore, the presence of comorbidities may influence the timing of a cancer diagnosis and the treatment received, leading to disparities in prognosis. While substantial missing explanatory variable data is a major problem in population-based cancer survival data, standard implementations of multiple imputation are incompatible with analysis models used. We aim to investigate the association between survival in non-Hodgkin lymphoma, the 6th most commonly diagnosed cancer in England, and patient comorbidity status, other prognostic factors and sociodemographic characteristics.

Methods: We analyse data on 84,029 adults diagnosed with non-Hodgkin lymphoma between 2005 and 2013 in England. To account for unreliability and unavailability of cause of death, and clustering of patients, we apply a multilevel excess hazard model. B-spline functions are used to model the time-dependent effects of age and stage at diagnosis. The latent normal joint modelling multiple imputation approach, which not only allows for imputation from different distributions based on variable type, but accounts for clustering, is used to account for missing data in stage and other explanatory variables. Multiple imputation is conducted using the jomo package in R, while the flexible excess hazard model is estimated via the mexhaz package.

Results: Patients with either severe or mild comorbidity status had significantly higher excess hazard of death in comparison to those without comorbidities: excess hazard ratio 1.48; confidence interval 1.44-1.53 and 1.26; CI 1.23-1.29 respectively, adjusting for other factors. Those living in most deprived areas had 1.25 times excess hazard of death compared to those in least deprived areas (CI 1.19-1.31).

Conclusions: Significant survival differences exist between those living in least and most deprived areas, even after accounting for comorbidity status. Greater public health efforts are needed to reduce the barriers in access to care for patients in most deprived areas. Further research will assess sensitivity of results to departures from the missing at random mechanism, by conducting multiple imputation under a missing not-at-random mechanism using prior distributions and delta adjustment methods.

[OC19.5]

Risk of non-AIDS defining and AIDS defining malignancies of early versus delayed initiation of antiretroviral

Frederique Chammartin[1], Sara Lodi[2], Roger Logan[3], Lene Ryom[4], Amanda Mocroft[5], Ole Kirk[4], Leah Shepherd[4], Jens Lundgren[4], Heiner C. Bucher[1]

[1] Basel Institute for Clinical Epidemiology & Biostatistics, Department of Clinical Research, University Hospital Basel and University of Basel, Switzerland
[2] Boston University School of Public Health, United States
[3] Depatment of Epidemiology, Harvard T H Chan School of Public Health,, United States
[4] Department of Infectious Diseases, CHIP, Section 2100, Rigshospitalet, University of Copenhagen, Denmark
[5] Department of Infection and Population Health, University College London Medical School, United Kingdom

Background: Data from large observational studies with long follow-up can be used to complement evidence from randomized controlled trials (RCT). In the START RCT immediate antiretroviral therapy (ART) decreased the risk of serious AIDS and non-AIDS related events compared with deferred initiation in persons living with HIV (PLWHIV). However, the effect of immediate initiation on the risk of non-AIDS and AIDS defining malignancies could not be reliably estimated due to the relatively small number of events and short follow-up.

Objectives: To estimate the long-term effect of immediate versus deferred ART initiation on the risk of non-AIDS and AIDS defining malignancies using data from the prospective observational D:A:D study, a multi-cohort study of 9 cohorts from Europe, Australia and the USA, and statistical methods to control for confounding by indication.

Methods: We used parametric g-formula to estimate the relative risk for non-AIDS and AIDS defining malignancies of immediate versus deferred ART at CD4 cell counts of 350 and 500 cells/µl. We adjusted our models for baseline (age, calendar year, body mass index, CD4 cell count and HIV-RNA, gender, mode of HIV acquisition, ethnic group, cohort, hepatitis C virus and ever smoking status) and time-dependent confounders (CD4 cell count, HIV-RNA, and visit frequency) and assumed no unmeasured confounders.

Results and Conclusions: The study population was predominately male (77.26%), with median [IQR] CD4 cell counts 410 [260-583] cells/µL, HIV-RNA 10.5 [9.1-11.5] log10 copies/ml, and age 36 [29-43] years. During a median follow-up of 8.4 years 489 PLWHIV developed any malignancy with 228 being non-AIDS defining malignancies. Compared to immediate ART initiation, the 10 years follow-up relative risk of non-AIDS malignancy and AIDS defining malignancies was 1.04 (95% CI 1.00 – 1.10) and 1.13 (95% CI 1.07 – 1.19) for starting ART at CD4 cell count of less than 500 cells/µl and 1.09 (95% CI 0.99 – 1.27) and 1.41 (95% CI 1.25 – 1.59) when starting ART at CD4 cell count of less than 350 cells/µl. Immediate ART initiation was associated with a decreased risk of non-AIDS and AIDS defining malignancies but estimates for non-AIDS defining malignancies fell short to be statistically significant.

OC20: Multivariable Model Building

Chair: **Anne-Laure Boulesteix**

LMU Munich, Germany

[OC20.1]

Optimal separation of normal distributed samples. Application for anxiety and alexithymia scales in cardiology

Ihor Malyk[1], Mariia Ivanchuk[2], Oleksandr Polishchuk[3], Pavlo Ivanchuk[3]

[1] *Mathematical Control Problems and Cybernetics, Yuriy Fedkovych Chernivtsi National University, Ukraine*
[2] *Biological Physics and Medical Informatics, Bukovinian State Medical University, Ukraine*
[3] *Internal Medicine, Physical Rehabilitation and Sports Medicine, Bukovinian State Medical University, Ukraine*

Background: Problem of finding the optimal separation hyperplane for two normal distributed samples is considered in the propose study. This task is important from a medical point of view, since there are a large number of researches in which studies are conducted on the characteristic that have a normal distribution or can be converted to a normal distribution

Objective: The main aim of the work is developing of the algorithm for building the optimal separating hyperplane for two normal distributed samples, with minimizing the probability of error. The paper presents theoretical results that will determine the „initial" separating set. The main characteristic for minimization in this work is the probability of classification error. The practical part of the paper is the analysing of the Spilberger-Hanin test results in patients with heart diseases, and finding the relationship between Spilberger-Hanin Scale and Toronto Alexithymia Scale in these patients.

Results: The main result of the work is an algorithm for finding the optimal separating hyperplane for binary classification. The optimal solution of the minimizing classification error for normal distributed samples was found in the paper. Theoretical results were checked by analysing Spilberger-Hanin Scale and Toronto Alexithymia Scale in patients with heart diseases. On the basis of this example it is possible to show that the probability of error is not higher according to the proposed method in comparison with the classical methods (Support Vector Machines, Linear Discriminant Analysis, Naive Bayes Classification, etc.).

Conclusions: The main advantage of the paper is development of the theoretical basis for building the optimal separation hyperplane for normal distributed samples. No doubt that proposed method may be extended for different distributions and for any separation objects in Euclidian space.

[OC20.2]

Dichotomisation of continuous outcomes in prediction model research: (not such) a bad idea?

Lucinda Archer, Kym Snell, Siobhán Stynes, Kate Dunn, Nadine Foster, Joie Ensor, Richard Riley, Gwenllian Wynne-Jones, Daniëlle van der Windt, Jonathan Hill

School of Primary, Community and Social Care, Keele University, United Kingdom

Background: It is well established that dichotomising continuous variables is inappropriate in prediction model research, as it reduces power and requires arbitrary cut-points. However, less attention has been given to dichotomisation of continuous outcomes, although some work suggests this too reduces power to detect predictor effects and may result in misleading conclusions.

Objectives: To describe and compare prediction models for pain intensity in patients with neck and/or low back pain (NLBP), using continuous and dichotomised versions of the same outcome.

Methods: Multivariable linear and logistic regression models were fitted with backward selection to data from two large, primary care based studies with patients consulting for musculoskeletal pain: i) Keele Aches and Pains Study (n=1,890), and ii) STarT MSK pilot randomised trial (n=524). Analysis was limited to patients with NLBP. The outcome was pain intensity at 6 months following consultation (0-10 numerical rating), dichotomised at the midpoint (≥5, pre-specified) to form a binary outcome.

Predictive performance was evaluated, with optimism in model performance assessed using internal validation by bootstrapping. Differences in the included predictors of the models were compared, as were their predictive performance and clinical usefulness of the predictions.

Results: Preliminary results suggest similar predictive performance in the linear model, which gave an R2 = 37%, and the logistic model, with a Nagelkerke's R2 = 33%, both after optimism adjustment.

The selected variables were reasonably consistent, with patients' long-term expectations regarding recovery being the strongest predictor in both models. Only the binary model included fear of harm from movement, with a protective effect, and omitted pain-site (back vs neck) – a clinically essential predictor.

Predictions of the dichotomised outcome were easily gained after modelling on a continuous scale (dichotomising predicted pain intensity at 5), with such a scale distinguishing patients with a prediction of 5.1 (borderline) and those with a prediction of 10 (maximum).

Conclusions: Prediction models that fail to keep continuous outcomes on their continuous scale may suffer from loss of information and reduced power to detect predictor effects. To discuss this more broadly, a simulation study and further comparison of performance statistic results will be presented.

[OC20.3]

Evaluating the role of correlations among markers in prediction models

Sergio Sabroso-Lasa[1], Luis Mariano Esteban[2], Rocío Aznar-Gimeno[3], Gerardo Sanz[4], Trinidad Serrano[5]

[1] *Instituto de Investigación Sanitaria Aragón, Spain*
[2] *Escuela Politécnica La Almunia, Universidad de Zaragoza, Spain*
[3] *Big Data y Sistemas Cognitivos, Intituto Tecnológico de Aragón, Spain*
[4] *Métodos Estadísticos, Universidad de Zaragoza, Spain*
[5] *Hepatología, Hospital Clínico Universitario Lozano Blesa, Spain*

Background: How to combine biomarkers in order to improve diagnostic accuracy is a widely analyzed issue that does not have a global answer. The improvement in discrimination power provided by the addition of a new marker to a model is not always clear, even if the new marker by itself has a good discrimination ability.

Objective(s): To analyze the effect of the sign and magnitude of the correlations between markers in the improvement of the discrimination ability of predictive models in different scenarios.

Method(s): We have considered situations with different correlations between a standard and a novel marker. Under multivariate normality assumption, using the formula provided by Su and Liu (1993), we have derived an expression for the maximum AUC as a function of the correlation between markers. Using surfaces, we show graphically the relationship between the increase of the AUC and the sign and the magnitude of the correlation.

Also, we have extended that idea for skewed data. We used cubic transformations of Normal distribution simulated data for this purpose. In this case, we estimate coefficients of the combination of two markers using a non parametric approach that maximizes AUC implemented by Esteban et al. (2011).

Results and Conclusions: Results show that in any case, the magnitude of the correlations should be taken into account in order to select new markers. Depending on the discrimination ability of each marker on its own, large positive or negative correlations are important to improve predictive models.

For markers with a similar discrimination ability only negative correlation aids to increase the AUC of the combination of markers, whereas for markers with different AUC large or positive correlation among markers provide new markers with a high discrimination ability.

The limitations of this study may come from the difficulty of finding real cases with highly negatively correlated variables. However, we remind that moderate correlation also favour the increase in AUC.

[OC20.4]

Piece-wise exponential (additive mixed) modelling tools for survival analysis

Andreas Bender, Fabian Scheipl, Philipp Kopper

Department of Statistics, LMU Munich, Germany

Background: In the context of survival analysis estimation of realistic processes requires flexible methods that allow for the inclusion of non-linear effects, strata, frailties, time-varying effects as well as covariates and additionally support settings that go beyond right-censoring, like competing risks. These topics have received a lot of attention in research papers and respective implementations are often available. However, many implementations are specialised for specific tasks, while lacking features available in other general purpose software packages, e.g., penalized splines, cumulative effects, big data methods, etc. From a research software engineering perspective it is therefore favorable to view the estimation of survival processes from a more generalised gradient based optimization perspective which allows to perform time-to-event data analysis within already available software implementations, utilizing their respective strengths.

Objective(s): We aim to provide an abstraction of the estimation of survival processes from specialised packages and instead provide a general interface that can perform time-to-event data analysis using any software or package that supports gradient based optimization.

Methods: The piecewise exponential model provides one possible framework for the abstraction of the estimation task from the specific computing engine such that any of the survival problems mentioned above can be generalized into three steps: (1) data transformation, (2) model estimation using any software or package that supports the optimization of the Poisson log-likelihood (3) post-processing (e.g., prediction and visualisation). In this framework, left-truncation, time-varying effects and competing risks can be realised by appropriate data transformation and/or inclusion of interaction effects. No additional adjustments are required for the computation engine in step (2).

Results and conclusions: The abstraction discussed above is implemented in the R package pammtools. Computation engines currently include generalized linear models (via glm), generalized additive mixed models (via mgcv), gradient boosted trees (via xgboost) and model based boosting (via mboost). Support for other computing engines only requires a thin additional layer for pre- and post-processing.

[OC20.5]

Common spline bases for regression models in practice: a comparative simulation study

Aris Perperoglou[1], Daniela Dunkler[2], Matthias Schmid[3]

[1] *Cambridge, AstraZeneca, United Kingdom*
[2] *Institute of Clinical Biometrics, Medical University of Vienna, Austria*
[3] *Institut für Medizinische Biometrie, Informatik und Epidemiologie, Universität Bonn, Germany*

Background: The splendour of splines relies on their flexibility and ability to model a variety of cases, from simple curves to rather complex patterns. That flexibility comes at a price: fitting a spline can be a daunting exercise for analysts with limited previous experience. In a recent STRATOS paper, Sauerbrei et al [1] highlighted seven topics for which evidence-supported guidance is needed. The second topic concerns comparison of spline procedures in both univariable and multivariable context. In a review paper Perperoglou et al [2] showcase that spline methods can differ substantially depending on the choice of parameters and software used. Although fine-tuning a spline function can result in a reasonable fit, in practice, optimising splines can be a complicated process, especially for non-experts.

Objectives: In this work we investigate complexity of functions and usability of procedures both in a univariate and multivariable cases. We aim to provide empirical evidence and practical guidance on the use of splines in a series of realistic scenarios by conducting informative simulation studies.

Methods: We focus on b-splines, natural splines, p-splines and thin-plate regression splines. In the univariate case we investigate whether results derived with suggested spline procedures differ substantially from the true function. For the multivariable case we compare performance in real and simulated data. We adapt the structured approach for simulation, as suggested by Morris et al [3] defining aims, data-generating mechanisms, estimands, methods and performance measures.

Results and Conclusions: We discuss findings and provide recommendations for splines fitting. We scrutinize fitting strategies as suggested by "experts" and highlight the importance of tuning relevant parameters, sample size, and the dangers of working with default software values.

References:

1. *Sauerbrei W, Perperoglou A, Schmid M, Abrahamowicz M et al: State-of-the-art in selection of variables and functional forms in multivariable analysis--outstanding issues. 2019; ArXiv preprint.*
2. *Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M on behalf of TG2 of the STRATOS initiative. A review of spline function procedures in R. BMC Medical Research Methodology. 2019; 19:46.*
3. *Morris T, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med 2019; 38: 2074–2102.*

OC21: BayesianMethods & Omics

Chair: **Rodolphe Thiebaut**

Bordeaux Univ / Inserm, France

[OC21.1]

Nonparametric bayesian sparse factor analysis to model individual heterogeneity in gene-expression data

Sarah-Laure Rincourt[1], Stefan Michiels[2], Damien Drubay[2]

[1] *Oncostat U1018, Inserm, University Paris-Saclay, France*
[2] *Department of Biostatistics and Epidemiology, Gustave Roussy, France*

Introduction: Cancer is a complex disease with a high degree of inter-individual heterogeneity in involved biological mechanisms. The characterization of this heterogeneity is particularly challenging in high dimensional omics data.

Objective: We propose a modified infinite sparse factor analysis to identify individual molecular specificities.

Methods: We imposed sparsity on the weights matrix of factor analysis (FA) model using a beta-Bernoulli process (BBP). The BBP generates an infinite row sparse binary matrix, avoiding the restriction of the model space to a finite number of factors. We introduced a baseline profile associated to all individuals, adding an all-ones row to the sparse binary matrix, the other factors being deviations from this profile which are common to different subgroups of individuals. Posteriors are obtained using mean-field variational inference maximazing the evidence lower bound (ELBO). We performed a simulation study to assess the ability of our method to recover latent structure. We used the accuracy of the estimated binary matrix to quantify the quality.

Results: We simulated specific sparse binary structures for N={50,200,200,500,500} subjects vs P={200,50,200,1000,1500} variables, and the corresponding weights and factors of infinite FA. The accuracy showed that we can obtain good recovery (accuracy = 0.98 for N=500 and P=1500). The accuracy increased with the number of individuals, and decreased with the dimension and complexity of the simulated binary structure (overlap of the sparsity structure of the factors) due to a lack of posterior sparsity (accuracy = 0.56 for the most complex scenario, N=50 and P=200). Our empirical experiment suggested that by tuning BBP hyperparameter values the sparsity can be increased using the ELBO as criterion.

Conclusion: We proposed a new approach to infer inter-individual heterogeneity modeling individual deviations from a common baseline profile. We will present an application of the breast cancer gene-expression data set.

[OC21.2]

Adaptive filtering increases the power but not the False Discovery Rate in RNA-seq experiments

Sonja Zehetmayer

Medical University of Vienna, Austria

In RNA sequencing studies a large number of hypothesis tests are performed to compare differential expression of genes between several conditions. To reduce the multiple testing burden and to increase power, filtering steps have been proposed to remove candidate genes with low counts and no or little chance of showing a difference between conditions.

We show in a simulation study that for RNA sequencing data filtering leads to some increase in power, too aggressive filtering, however, can lead to a decline. No uniformly optimal filtering method in terms of multiple power exists. We propose an adaptive filtering strategy, which selects one out of several filtering methods to maximize the number of rejections. We show that the adaptive filter leads to the highest power while controlling the FDR at a pre-defined level. Furthermore, it does not require to pre-specify a single filtering method.

[OC21.3]

A Bayesian approach for identifying gut-microbiome - diet associations

Alma Revers, A. H. Zwinderman

Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam UMC, Netherlands

Background: The gut-microbiome plays an important role in human health and is influence by food consumption. With 16S RNA gene sequencing the gut-microbiome is measured in fecal samples. Identified microorganisms are clustered into organizational taxonomic units (OTUs). OTUs are counts or relative abundances (RA) and are often overdispersed. OTUs have a hierarchical ordering of phylogenetic relationships, which is based on the RNA similarity. We assume that associations between diet-variables and an OTU abundances are similar for related OTUs.

Objective: We propose a hierarchical Bayesian model to identify associations between diet and the microbiome, incorporating the microbiome phylogenetic relationships.

Method: We specified hierarchical generalized linear mixture models, assuming the phylogenetic relationships between the OTU and the same diet-variable were comparable for all OTUs in the same phylogenetic family. We modelled OTU counts with negative binomial (NB) or Poisson distributions, and RAs with binomial or multinomial distributions. All models were developed in STAN using non-informative priors.

The data was from the Healthy Life in an Urban Setting (HELIUS) study, included 1036 subjects, 53 diet-variables and 1372 OTUs with an abundance of at least 20% and being member of a family with at least 4 other members.

Results: With Poisson, binomial and multinomial models many small associations were identified with small credibility intervals. In contrast with the NB model the number of identified associations was much smaller with realistic credibility intervals. Increased red meat consumption was related with decreased abundance of OTUs in the *Bifidobacterium* and *Bacteroidaceae* family.

Jointly estimating mean and dispersion parameters of the NB models had however multiple convergence issues. We modeled the dispersion parameter as inverse function of the mean, similar to the approaches implemented in the EdgeR and DESeq2 packages in R. This solved most convergence issues and yielded a faster result.

Conclusion: With our constrained Bayesian NB model we are capable of jointly analyzing a large number of microbiome OTU counts in a large sample of subjects with a large number of covariates taking account of phylogenetic clustering. Main finding is the decreased abundance of a large number of OTUs by increasing red meat consumption.

[OC21.4]

Identifying rare cell types using an outlier-detection approach

Miguel Rodo[1], Virginie Rozot[1], Carly Young[1], Phu Van[2], Evan Greene[2], Greg Finak[2], Raphael Gottardo[2], Francesca Little[3], Thomas Scriba[1]

[1] *South African Tuberculosis Vaccine Initiative, Department of Pathology, University of Cape Town, South Africa*
[2] *Vaccine and Infectious Disease Division, Fred Hutch Cancer Research Center, United States*
[3] *Department of Statistical Sciences, University of Cape Town, South Africa*

Background: Tuberculosis (TB) is the world's leading cause of death due to infectious disease. To better understand host-pathogen interactions, we compared cellular immune responses between individuals who developed TB and those who did not during a two-year observational study of Mycobacterium tuberculosis-exposed adolescents.

To count immune cells of various types for each participant, protein expression for individual cells was measured for forty-four proteins. Current clustering algorithms adequately identify major cell types, but struggle to isolate rare cell types. In this study, the rare cell types are especially important as they mediate the response to the pathogen. Rare cell types are typically identified manually by immunologists using prior biological knowledge and a visual assessment. However, this scales poorly and cannot identify new cell types.

Objective: Our objective is to provide a clustering algorithm that can automatically identify rare cell types for differential abundance analysis.

Method: We used a non-parametric clustering algorithm (Greene, 2019) to identify major cell types. Our task is to then identify rare cell types that are sub-types of these.

First, we identify co-expressed proteins using changepoint analysis. This detects combinations of proteins for which high values in one is associated with high values in at least one other. Combining information from these variables increases power to detect outliers.

Second, we use negative controls (samples of blood that do not typically contain the rare cells of interest) to define the null distribution of protein markers, i.e. their distribution in the absence of the rare cells that express these proteins at elevated levels.

We then use false-discovery rate control methods to isolate cells that express a sufficiently high level of each protein to be outliers from the null distribution and hence constitute a different sub-type.

We compare our counts to those attained manually on a subset of the data. The identified cell types are then analysed using longitudinal mixed-effects modelling.

Results: Our counts of rare cell types correspond to those identified manually. Rare cell types associated with disease progression are detected.

Conclusions: Our solution can identify rare cell types consistent with manual identification and that are associated with disease progression.

OC22: Missing Data 1

Chair: **Sven Ove Samuelsen**

University of Oslo, Norway

[OC22.1]

Hierarchical imputation of categorical variables with both systematically and sporadically missing values

Shahab Jolani

Methodology and Statistics, Maastricht University, Netherlands

Background: In development of prediction models, the data are often combined from different sources, the so-called individual participants data (IPD) sets. A common problem with IPD sets is the presence of missing values in predictors and/or outcome. Although multiple imputation using chained equations (MICE) is a popular approach to deal with missing data, its standard practice is inappropriate when the data have nested structure (e.g., patients within studies in IPD sets). As a result, the imputation model should take the nested structure of the data into account when missing values are imputed. We distinguish two types of missingness in IPD sets: systematically missing data and sporadically missing data. Recently, we have developed a hierarchical imputation methodology for continuous variables with systematically and sporadically missing values. However, the hierarchical imputation of categorical variables with systematically and sporadically missing data is challenging because of the categorized nature of data.

Objective(s): The main objective is to develop a hierarchical imputation methodology for categorical variables that accommodates the presence of both systematically and sporadically missing data in nested designs with an arbitrary pattern of missing data.

Method(s): A novel hierarchical imputation method is developed within the MICE framework. Here, we use a random effect imputation model, within which a Bayesian approximation method is engaged to simplify drawing from the random effects in order to speed up the imputation task. After some theoretical arguments, we evaluate the performance of the new methodology in a set of simulation studies and illustrate its application in an IPD set to predict kidney failure among patients.

Results: The Monte carlo simulation results showed that the new imputation methodology has good statistical properties in term of bias and coverage. The results from the application was also in line with the expectation.

Conclusions: By introducing the new hierarchical imputation methodology for categorical variables, we complete the puzzle of imputation for systematically and sporadically missing data in IPD sets within the MICE framework.

[OC22.2] **StCA Winner**

Combining bootstrap with multiple imputation to assess prognostic models with missing covariate data

Orlagh Carroll[1], Patrick Rockenschaub[2], Tim Morris[3], Ruth Keogh[1]

[1] *Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom*
[2] *Institute of Health Informatics at UCL, United Kingdom*
[3] *MRC Clinical Trials Unit at UCL, United Kingdom*

Background: The bootstrap (BS) is commonly used to assess the performance of clinical prediction models. The standard algorithm does so by developing a prediction model in a BS sample and testing it in both the full dataset and within the BS sample. This is repeated for many samples. Missing data is common in many studies and presents challenges for developing and evaluating models. Multiple imputation (MI) is a popular method for dealing with missing data. There is no clear guidance on how to combine MI with bootstrapping to assess model performance when both current and future patients may have missing data on covariates.

Objective: To investigate methods for combining BS with MI for prediction modelling.

Methods: We consider several ways of combining BS and MI to determine whether BS-then-MI or MI-then-BS is more appropriate. Considerations include whether to reuse the imputations of the original data for training and evaluating models in BS samples, or to perform MI within each BS sample separately. We considered both the standard approach and the 0.632 method. A complete-case approach in which missing data is ignored was also considered. A simulation study was conducted to assess the performance of the methods for a linear prediction model, with predictive performance measured using mean-squared error (MSE). Several parameters were varied in the simulation including sample size, R-squared, missing data mechanism and the number of imputations.

Results: When missingness does not depend on the outcome, a complete-case analysis performs poorly for small samples but well for large samples. When missingness is dependent on the outcome, the complete-case analysis severely underestimates the MSE. Underestimation is also seen in all variations of MI-then-BS and for BS-then-MI when reusing the original dataset imputations. With increasing sample size BS-then-MI approaches the expected MSE and the magnitude of over-estimation decreases.

Conclusions: We recommend BS-then-MI for the assessment of prediction models when there is missing covariate data. While it tends to overestimate the MSE this bias is small and BS-then-MI avoids data leakage. Performing the MI first results in over-optimistic measures of model performance and is prone to data leakage.

[OC22.3]

An alternative characterization of Missing at Random in Shared Parameter Models

Grigorios Papageorgiou, Dimitris Rizopoulos

Biostatistics, Erasmus MC, Netherlands

Background: Dropout is a common complication in longitudinal studies, especially since the distinction between random (MAR) and non-random (MNAR) dropout is intractable. Consequently, one starts with an analysis that is valid under MAR and then performs sensitivity analysis by considering MNAR departures from it. To this end, specific classes of joint models, such as pattern-mixture models (PMMs) and selection models (SeMs), have been proposed. Contrariwise, shared-parameter models (SPMs) have received less attention as they do not embody a characterization of MAR. A few approaches to achieve MAR in SPMs exist, but are challenging to implement in existing software.

Objectives: Motivated by randomized clinical trial data on the efficacy of drugs on HIV patients, we propose an alternative characterization of MAR in SPMs by exploiting the conditional independence between outcome and missingness given a set of common random-effects. By doing so, we can utilize the censoring distribution to cover various assumptions for the missing data generating mechanism on the subject-specific level. Our objective is to assess the plausibility of different assumptions concerning missingness under the SPM framework.

Methods: We use joint models for incomplete longitudinal and time-to-event data to investigate how considering all dropout cases to be MNAR, MAR, or partly MNAR and MAR impacts model coefficients and subject-specific predictions. The behavior of our approach under different settings for the amount of MAR and MNAR dropout is stressed in a simulation study.

Results: Our results suggest that MAR and MNAR models can lead to differences in model coefficients and subject-specific predictions. The magnitude of these differences and the performance of each model may be affected by the amount of MAR and MNAR dropout cases, as suggested by our simulation findings.

Conclusion: This intuitive approach offers substantial advantages and can be easily implemented in existing software. It offers flexibility over the missing data generating mechanism by allowing subject-specific perturbations of the censoring distribution, whereas, in PMMs and SeMs, dropout is strictly considered MNAR. Finally, the subject-specific nature of such a sensitivity analysis framework encompasses a toolkit for both individual-specific and risk-specific assumptions, with the latter allowing the consideration of competing dropout mechanisms.

[OC22.4]

Cluster Analysis in Incomplete Data

Halehsadat Nekoee Zahraei[1], Renaud Louis[2], Anne-Françoise Donneau[1]

[1] *Biostatistics Unit, Department of Public Health, University of Liege, Belgium*
[2] *Department of Pneumology, GIGA, University of Liege, Belgium*

Background: One method which helps researchers to find a structure in their dataset is cluster analysis. In cluster analysis, multi-dimensional data are divided into homogeneous groups such that subjects in each group have similar properties. However, missing value is an unavoidable part of multi-dimensional data. Even if missing data can be easily handled by several methods, clustering approaches have to account for this management of missing values.

Objective: Multiple imputation (MI) is a simple but powerful method in this field. However, there are several challenges for clustering when MI is applied. The objective of this present research was to introduce an efficient framework to apply cluster analysis on incomplete dataset by using MI. By simulating different scenarios inspired by real data, our proposed method addressed some limitations in statistical literature to find high discriminating clusters.

Method: In the first step of MI, m imputed datasets were generated. Variable reduction methods and cluster analysis strategies were then applied to imputed datasets. Finally, for each imputed dataset, cluster assignment was calculated. For that purpose, application of finite mixture of multivariate multinomial distribution was proposed to estimate number of clusters; final cluster result was assigned to observations by solving maximum likelihood via EM algorithm.

Results: Motivated by real datasets, 178 subjects with mixed continuous and categorical variables but with two known clusters were generated by normal and multivariate mixture distribution, respectively. Several scenarios were defined for different percentages of missingness (e.g. 25%, 50%, 75%) and overlap between two known clusters (e.g. 30%, 45%, 65%). In addition, different imputation, variable reduction and clustering methods were compared. The results showed that our proposed method had high discrimination and matching compared to other methods. The best method, based on MI, variable reduction and our proposed combination method, was then applied on real data from the Pneumology Department of University hospital of Liege, which aimed to identify clinical phenotypes among adults suffering from pulmonary disease.

Conclusions: Based on large simulation study, our proposed method yielded to the best discrimination with the highest matching between the final result of clustering and the known clustering from the simulated dataset.

[OC22.5]

Bootstrap Inference for Multiple Imputation under Uncongeniality and Misspecification

Jonathan Bartlett[1], Rachael Hughes[2]

[1] *Department of Mathematical Sciences, University of Bath, United Kingdom*
[2] *Bristol Medical School, University of Bristol, United Kingdom*

Background: Multiple imputation (MI) has become one of the most popular approaches for handling missing data in statistical analyses. Part of this success is due to Rubin's simple combination rules. These give frequentist valid variance estimates when the imputation and analysis procedures are so called 'congenial' and the corresponding Bayesian model is correctly specified, but otherwise may not. Uncongeniality can occur in many different situations, including fitting the analysis model in a subgroup of the data, unnecessarily including an interaction in either the imputation or analysis model, or reference-based imputation methods in clinical trials.

Recently, a large number of different ways of combining bootstrapping with MI have been proposed, but it is unclear which of these (if any) could give unbiased variance estimates under uncongeniality or misspecification.

Objective: We determine approaches to combining bootstrapping with MI can give valid variance estimates under uncongeniality or misspecification.

Methods: We investigated the validity of variance estimates obtained from several recent proposals which combine bootstrapping with MI in different ways, using both theoretical arguments and simulation studies.

Results: Methods which impute first then bootstrap generally result in biased variance estimates under uncongeniality or misspecification. Conversely, most methods which bootstrap then impute give valid variance estimates under either uncongeniality or misspecification.

Conclusions: Imputation followed by bootstrapping generally does not result in valid variance estimates under uncongeniality or misspecification, whereas bootstrapping followed by imputation does. We recommend a particular computationally efficient variant of bootstrapping followed by imputation.

OC23: Bayesian Methods

Chair: **Alexandra Schmidt**

McGill University, Canada

[OC23.1]

Bayesian reconstruction of inter-hospital propagation chains of antimicrobial resistant bacteria

Jonathan Roux[1], Finlay Campbell[2], Clément Massonnaud[1], Laura Temime[3], Pascal Crépey[1]

[1] *Univ Rennes, EHESP, REPERES - EA 7449, France*
[2] *Department of Infectious Disease Epidemiology, MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, United Kingdom*
[3] *Laboratoire MESuRS, Conservatoire national des Arts et Métiers, France*

Background: Carbapenemase-producing Enterobacteriaceae (CPE) threaten healthcare systems worldwide. While CPE are known to circulate within the community and with patient transfers between hospitals, their exact spreading routes often remain poorly understood.

Objective: This study aimed at adapting a Bayesian framework reconstructing inter-individual transmission chains during an outbreak to the reconstruction of inter-hospital propagation chains for antimicrobial resistant bacteria, and at applying it on CPE propagation.

Methods: The present work is an extension of the outbreaker2 R package. It introduces a new likelihood to account for transfers of patients between hospitals. For each episode of hospital contamination, the framework identifies the most likely "infector" hospital from which colonized patients may have been transferred. We applied it using Markov-Chain Monte-Carlo with 50,000 iterations on both simulated (1 year of episodes) and real data on CPE episodes in France (all 2010-2016 episodes). The framework was informed by 2016 data on all French patient transfers among 2,427 public and private hospitals. Ten simulated datasets allowed the identification of a minimal support, i.e. the threshold posterior probability of each link, maximizing the values of sensitivity and specificity used afterwards on the real data. We also performed a sensitivity analysis on the average generation interval (GI: 1 to 90 days) on the real data.

Results: Based on the simulated data (203 to 418 links to be identified), the identified minimal support was 0.135, allowing for a 79.3% sensitivity and a 43.0% positive predictive value (PPV). With the real dataset composed of 3,054 episodes (1,548 internationally imported episodes), we observed a rise and plateau of the number of linked episodes starting at an average GI of 38 days. This corresponded to 700 linked episodes (46.5%) with at least 84% of common links between two consecutive GIs. Considering sensitivity and PPV values, this result suggests that at most 25.2% of episodes may be explained by patient transfers.

Conclusions: This study is the first to use a Bayesian framework to reconstruct inter-hospital chains of bacterial spread. Our results suggest an important CPE circulation in the community leading to local carriage and introduction within hospital settings.

[OC23.2] **CASc Winner**

A spatial multilevel model using conditional autoregressive processes

Md Mynul Islam, Syed Shahadat Hossain, Mohammad Samsul Alam

Department of Applied Statistics, Institute of Statistical Research and Training, University of Dhaka, Bangladesh

Background: Multilevel Modeling (MLM), where the lower level units are nested into higher level ones based on a geographical hierarchy, helps us to understand individual and contextual effects of a complex survey data. As the availability of geographically hierarchical datasets are increasing day by day, multilevel models are widely employed to examine the outcomes of interest measured at the lower and higher levels simultaneously. However, the classical MLM does not involve spatial effect. It assumes that lower level units are correlated with the same correlation belonging to a higher level unit and takes into account the higher level unit as independent to each other. It does not consider directly the distance between them. As a consequence, spatial interaction effects remain unmodeled.

Objective: In this study, our main purpose is to separate the higher and lower level effects, to quantify higher and lower level effect, and to provide better estimates of the model parameters as well as their standard errors precisely in the presence of spatial interaction. To deal with this, the concept of multilevel modeling is extended to propose a spatial multilevel model to incorporate spatial interactions at both lower and higher levels in this study.

Method: In the proposed model, the outcome variable is modeled through spatially dependent random effect at both level and conditional autoregressive process (CAR) is used at both higher and lower level random effects. Estimation of the proposed model is done using Bayesian framework. To asses, the performance of the proposed model, a series of Monte Carlo simulations are conducted.

Results and Conclusions: The results exhibit a significant contrast between spatial MLM and classical MLM. The results show that proposed models perform better than classical MLM to retrieve true model parameters and the true scenario of the process. As a demonstration of the proposed model, we applied it to HYVboro rice production data in Bangladesh.

[OC23.3]

A Probability of Success Approach in Late Stage Drug Development

Steffen Ballerstedt[1], Björn Bornkamp[1], Björn Holzhauer[1], Markus Lange[1], Joseph Kahn[2], Wen-Lin Luo[2], Lisa Hampson[1]

[1] *Novartis Pharma AG, Switzerland*
[2] *Novartis Pharmaceuticals Corporation, United States*

There are several steps to confirming the safety and efficacy of a new medicine. A sequence of trials, each with its own objective, is usually required. Quantitative risk metrics can be useful for informing decisions about whether a medicine should transition from one stage of development to the next. Traditionally, pharmaceutical companies have used cross-industry success rates to estimate the probability of obtaining regulatory approval. Project teams then typically apply subjective adjustments to reflect project-specific information. However, this approach lacks transparency and fails to make full use of data from previous clinical trials. We describe a quantitative Bayesian approach for calculating the probability of success (PoS) at the end of Phase II, which incorporates internal clinical data, cross-industry success rates, and expert opinion or external data if needed. Using an example, we illustrate how PoS can be calculated accounting for differences between our Phase II data and future Phase III trials, and how the sensitivity of PoS to assumptions can be evaluated and communicated. We will also elaborate what the project teams can learn from this exercise.

[OC23.4]

The Bayesian population pharmacokinetic analysis of dexmedetomidine and clonidine in Stan and Torsten

Paweł Wiczling[1], Alicja Bartkowska-Śniatkowska[2], Justyna Mocarska[3], Paweł Sobczyński[4], Agnieszka Bienert[3]

[1] *Department of Biopharmacy and Pharmacodynamics, Medical Uniwersity of Gdansk, Poland*
[2] *Department of Pediatric Anesthesiology and Intensive Therapy, Poznan University of Medical Sciences, Poland*
[3] *Department of Clinical Pharmacy and Biopharmacy, Poznan University of Medical Sciences, Poland*
[4] *Department of Anesthesiology, Intensive Therapy and Pain Treatment, Poznan University of Medical Sciences, Poland*

Background: Dexmedetomidine and clonidine are alpha-2 agonists often used for long-term sedation in critically ill patients. Plasma drug levels can vary significantly among individuals who are following the same dosing regimen, especially in ICU settings. One of the challenges of modern pharmacology is to accurately describe the source of variability in drugs' PK and to treat outliers, thus allowing for better targeting of drug dose.

Objective(s): In this work a Bayesian population models with literature-based weekly informative priors were developed to describe two PK datasets observed after administration of clonidine or dexmedetomidine to a group of children and adults treated in four intensive care units and requiring continuous sedation.

Method(s): Data analysis was conducted using Stan software implementing the Bayesian inference with Markov-chain Monte Carlo sampling and Torsten functions for pharmacometric applications [1-3]. For both drugs a two compartment PK model, with allometrically and izometrically scaled clearances and volumes of distribution, maturation of clearance and t-student residual distribution on a log-scale (to ensure robustness to outliers) was used to describe the data. The prior information on dexmedetomidine and clonidine pharmacokinetics was elicited from published literature studies and visually assessed using prior predictive checks.

Results and Conclusions: This work demonstrates the usefulness of Bayesian approach in analyzing pharmacokinetic data. Stan and Torsten offers an interesting alternative to nonlinear mixed-effect and Bayesian analysis implemented in NONMEM®.

The project was supported by the Grant 2015/17/B/NZ7/03032 founded by the Polish National Science Centre.

References:

1. *Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. Stan: A probabilistic programming language. Journal of Statistical Software 76(1). DOI 10.18637/jss.v076.i01*
2. *Bob Carpenter, Matthew D. Hoffman, Marcus Brubaker, Daniel Lee, Peter Li, and Michael Betancourt. The Stan Math Library: Reverse-Mode Automatic Differentiation in C++. arXiv:1509.07164 [cs], September 2015. arXiv: 1509.07164.*
3. *Torsten: library of C++ functions that support applications of Stan in Pharmacometrics. https://github.com/metrumresearchgroup/Torsten.*

[OC23.5]

Corpus based priors for calibrated Bayesian inference

Erik van Zwet[1], Simon Schwab[2], Stephen Senn[3]

[1] *Leiden, Leiden University Medical Center, Netherlands*
[2] *University of Zurich, Switzerland*
[3] *University of Edinburgh, United Kingdom*

Background: Frequentist properties like unbiasedness and correct coverage are only meaningful before the data have been observed. Once the data are in, the estimate is just a number and the confidence interval either does or doesn't cover. However, if we condition only on statistical significance, some randomness remains and we can still talk about bias and coverage. Conditionally on significance, the estimator is biased away from zero which is known as the "winner's curse". The bias is especially severe when the power is low. In that case, also the confidence interval will undercover.

Objective: Bayesian inference is valid conditionally on the data, but requires a suitable prior. Researchers in the life sciences often believe that they have little or no prior information because their study is unique in many ways. We believe that it is a mistake to think like that. At the highest level of aggregation, theirs is just another study in the domain of the life sciences and that fact alone represents a lot of information.

We propose obtain prior distribution from large collections or "corpora" of studies that are similar in the sense that they meet certain criteria, e.g. phase 3, placebo controlled, RCTs. If we know only that a study meets certain criteria, then it is exchangeable with all other studies that meet them. So, the criteria of the corpus represent exactly the information that we are including in the prior. This implies that making the scope wider by removing certain criteria means putting less information into the prior. Importantly, there is no reason to expect that that would yield a more widely dispersed prior distribution.

Method: To obtain correct prior information, we need large collections of honest effect estimates and their standard errors. Examples are large replication projects (such as the 2015 replication study in psychology) and large curated databases (such as Cochrane).

Results and Conclusions: We have developed an app that allows researchers to use the Cochrane database to perform a Bayesian analysis of frequentist RCT results. Depending on the characteristics of the RCT, we find that considerable shrinkage is indicated.

OC24: Meta-analysis 2

Chair: **Laure Wynants**

Maastricht University, Netherlands

[OC24.1]

Model selection for component network meta-analysis in disconnected networks

Gerta Rücker, Guido Schwarzer

Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Germany

Component network meta-analysis (CNMA) is an extension of standard network meta-analysis (NMA). It can be used when treatments are composed of common components, such as combinations of drugs or combinations of a drug with a psychotherapy. In addition to knowledge of the structure of the network, a model is needed that describes how the effects of treatment components add in combination. The simplest and most parsimonious model is an additive model, which can be enriched by adding interaction terms.

A special property of CNMA models is that they can be applied also in cases where the network is disconnected, where standard NMA is impossible, provided that all subnetworks have at least one common component. One has to choose whether to apply the sparsest (that is, the additive) model or to add a number of interactions, and how many. The more interactions are added, the more the model resembles a standard NMA.

We suggest two model selection strategies (forward and backward selection) based on the model fit, measured by the Q statistics that is popular in meta-analysis. We investigate the strategies using simulations and demonstrate the methods on an example of multiple myeloma.

CNMA models, if feasible, are a promising alternative to matching approaches in disconnected networks.

[OC24.2]

Framework for evaluating reporting bias in network meta-analysis

Virginia Chiocchia[1], Adriani Nikolakopoulou[1], Dimitris Mavridis[2], Matthias Egger[1], Georgia Salanti[1]

[1] *Insitute of Social and Preventive Medicine, University of Bern, Switzerland*
[2] *Department of Primary Education, University of Ioannina, Greece*

Background: Reporting bias, or "non-reporting bias" as defined in the latest Cochrane Handbook for Systematic Reviews of Interventions (Version 6, 2019), can seriously compromise the results of systematic reviews and meta-analysis and, as a consequence, potentially affect clinical decision-making. Various graphical and statistical methods are available to assess the risk of reporting bias. However, these approaches have mostly been developed for pairwise meta-analysis, making it difficult to assess the impact of reporting bias on the results from network meta-analysis (NMA).

Objective: To develop a conceptual and methodological framework for evaluating the impact of reporting bias on NMA results.

Methods: The framework combines comparison-adjusted funnel plots, regression techniques, selection models and threshold analysis. We produce comparison-adjusted funnel plots where the direction of potential bias in each comparison is informed by pairwise contour-enhanced funnel plots and regression slopes for small-study effects. The limit meta-analysis model to adjust for small-study effects (Rücker et al, 2011) is extended to multiple treatment comparisons. To explore the impact of publication bias, we use the extension for NMA of the Copas selection model (Mavridis et al, 2014). For comparisons with less than 10 studies a qualitative assessment of the risk of bias is performed following the framework described in Chapter 13 of the Cochrane Handbook. The threshold analysis to assess the sensitivity of treatment recommendations to bias (Philippo et al, 2019) is also applied where, for each relative effect, a threshold is calculated indicating how much the pairwise evidence could change due to bias before a different treatment is favoured. Then, the plausibility of this change is judged qualitatively.

Results and Conclusions: We present the feasibility and applicability of the methods using illustrative examples of previously published NMAs accessed through the nmadb R package (Papakonstantinou, 2019). We plan to implement these strategies in the Confidence in Network Meta-Analysis (CINeMA) framework (Nikolakopoulou et al., 2019) and web-application (https://cinema.ispm.unibe.ch/). This will allow a more systematic evaluation of the reporting bias domain and produce better informed confidence ratings of the NMA findings.

This project is funded by the Swiss National Science Foundation under grant agreement No. 179158.

[OC24.3]

Bayesian dose-response network meta-analysis

Tasnim Hamza, Georgia Salanti

Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland

Background: Dose-response models express the effects of different doses or exposures levels on a specific outcome. Study-specific data on dose-response can be synthesized in a pairwise dose-response meta-analysis. When more than two interventions are compared, network meta-analysis models can be used. When the interventions are given at different dosages two approaches have been used in the literature so far: ignoring the differences in dose or consider each dose as a separate treatment. The first approach carries a higher risk of inconsistency for the network. while the second approach might produce disconnected networks. However, deciding which interventions is preferable and at which dose is important for taking reimbursement decision and issue treatment guidelines.

Methods: We propose a flexible Bayesian dose-response network meta-analysis model where the association between dose and outcome is modelled using restricted cubic splines. The approach needs only aggregated-level data as input. We employ various assumptions about the application of the consistency equations (the constraint that direct and indirect evidence about parameters are in agreement). Consistency can be assumed at the drug-level or at each drug-dose level and we explore various assumptions about the similarity of the dose-response shape across the involved interventions. The resulting series of models is programmed in R using JAGS.

Application: We apply the methods to synthesize data from 60 randomized trials (145 arms, 15,174 participants) examining the efficacy and tolerability of various doses of SSRI antidepressant drugs when these are given at various doses.

Results: Important differences are detected in the efficacy and tolerability between the different SSRIs and the dosages. The optimal dose, balancing efficacy and tolerability for nearly all SSRIs was between 30 and 40 mg per day Fluoxetine equivalent.

Conclusions: The proposed dose-response network meta-analysis model is very flexible and can naturally make dose-specific and comparison-specific predictions for an outcome. Extensions to more complex situation, e.g. involving different groups of drugs with different dose-response shapes are easy to implement.

The work is part of the HTx project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 825162.

[OC24.4]

Connecting the dots: Linking disconnected networks using dose-response Model-Based Network Meta-Analysis

Hugo Pedder[1], Sofia Dias[2], Martin Boucher[3], Meg Bennetts[3], Nicky Welton[1]

[1] *Population Health Sciences, University of Bristol, United Kingdom*
[2] *University of York, United Kingdom*
[3] *Pfizer, United Kingdom*

Background: Network Meta-Analysis (NMA) is often used as a source of clinical evidence with which to make cost-effectiveness decisions in Health Technology Appraisals (HTA). However, NMA relies on a connected network of treatments, which may not be available. Comparisons between disconnected treatments are not possible using NMA without making strong, untestable assumptions. In some circumstances, modellinga parametric dose-response function using Model-Based NMA (MBNMA) has the potential to re-connect the network via the dose-response relationship enabling comparisons between otherwise disconnected treatments.

Objectives(s): To identify scenarios in which MBNMA allows evidence synthesis in disconnected networks that could otherwise not be obtained using standard NMA, and to evaluate the performance of the method.

Method(s): Using an example dataset of triptans for migraine relief which included 70 studies investigating 6 agents and placebo, we removed agents and doses to generate different scenarios that illustrated disconnectedness in all possible combinations of pairs of agents. To assess performance, we examined agreement between MBNMA estimates from the disconnected network and NMA estimates from an "augmented" network connected by adding agents/doses back into the dataset.

Results: Within augmented networks, estimates of relative efficacy were more precise from MBNMA than from NMA models (ratio of posterior SDs for NMA vs MBNMA: median=1.13; range=1.03-1.20). In disconnected networks MBNMA was able to provide estimates for all treatment comparisons where NMA could not. MBNMA relative effects in disconnected networks were entirely in agreement with NMA estimates in augmented networks for 15 out of 18 treatment comparisons.

Conclusions: By modelling the dose-response in a statistically robust manner, MBNMA can provide some increase in precision over NMA in networks that are already connected. In disconnected networks, it can allow for estimation of treatment effects that cannot be estimated using NMA. MBNMA relies on the dose-response relationship being characterised correctly, but this assumption can be tested providing sufficient evidence is available at different doses of each agent to allow reliable estimation. MBNMA enables earlier phase evidence to strengthen recommendations based on phase-II studies in HTA, and we argue that earlier phase evidence should be included in systematic reviews to support HTA.

OC25: Clinical Trials Methodology 4

Chair: **Sara Lodi**

Boston University School of Public Health, United States

[OC25.1]

A new method to estimate the predictive value in a repeated screening program

Jian-Lun Xu

Bethesda, National Cancer Institute, United States

The goal of screening tests for a disease such as cancer is early detection and treatment with a consequent reduction in mortality from the disease. An important quantity one would like to know after the screening test is the predictive value (positive or negative predictive value) because the screening test might produce false-positive and false-negative results. The method of estimating this quantity in the literature typically assumes that the disease status and the verification indicator are conditionally independent when the testing result is given. In this talk we provide a method to determine when the conditional independence assumption doesn't hold. Furthermore, we introduce a new method which does not require the disease status and the verification indicator to be independent. The new method can easily be incorporated with covariate information. The estimator of the predictive value under the new method is investigated. The asymptotic normality of the estimator is also proved. In addition, simulation studies are conducted in order to see the performance of the new method when sample size is small or moderate. Finally, we apply the new method to the data from the ovarian component of Prostate, Lung, Coloretcal, and Ovarian cancer screening trial.

[OC25.2]

Analysis of binary composite outcomes with partially observed components: a comparison of strategies

Tra My Pham[1], Ian R. White[1], Brennan Kahan[1], Tim P. Morris[1], Gordon Forbes[2]

[1] *MRC Clinical Trials Unit at UCL, United Kingdom*
[2] *Biostatistics and Health Informatics, King's College London, United Kingdom*

Background: Composite outcomes are commonly used as primary endpoints in randomised trials. A participant is classified as meeting the endpoint if they experience an event in a combination of components (e.g. a composite of being alive and having negative culture results in trials assessing tuberculosis treatments). Missing data in the components can complicate the analysis of composite outcomes. A strategy often used in practice is derivable outcomes, where the composite is derived from observed components if possible (e.g. when there are events in the observed components), and participants whose composites cannot be derived (e.g. when there are no events in the observed components or all components are missing) are excluded from analysis. Alternatively, complete record analysis (excluding participants with partially observed components) or multiple imputation can be used.

Objective: We compare a set of strategies for analysing composite outcomes with partially observed components.

Methods: We consider the case of binary composite outcomes with binary components. In an analytic study, we show that derivable outcomes can be missing not at random even when the components are missing completely at random. In a simulation study, we compare analysis strategies for when the components are either missing completely at random or missing at random, including complete record analysis, derivable outcomes, multiple imputation of missing composites, and multiple imputation of missing components. The methods are evaluated in terms of bias in estimating the randomised treatment effect, efficiency, and coverage of 95% confidence intervals.

Results: When the components are missing completely at random, the derived composites can be missing not at random. Consequently, the treatment effect estimated from derivable outcomes is biased while complete record analysis results are valid. Under this missingness mechanism, multiple imputation of missing composites also produces an unbiased treatment effect. More complex missing at random mechanisms require multiple imputation at the component level.

Conclusions: Although appearing intuitive, derivable outcomes should be used with extreme caution. Multiple imputation of missing components is the preferred method in this study setting. In practice, choosing the imputation model for the incomplete components might not be straightforward; the choice depends on the interactions between components and with the randomised treatment.

[OC25.3]

The harmonic mean chi-squared test to substantiate scientific findings

Leonhard Held

Department of Biostatistics, University of Zurich, Switzerland

Statistical methodology plays a crucial role in drug regulation. Decisions by the FDA or EMA are typically made based on multiple primary studies testing the same medical product, where the two-trials rule is the standard requirement, despite a number of shortcomings. A new approach is proposed for this task based on the (weighted) harmonic mean of the squared study-specific test statistics. Appropriate scaling ensures that, for any number of independent studies, the null distribution is a chi-squared distribution with one degree of freedom. This gives rise to a new method for combining one-sided p-values and calculating confidence intervals for the overall treatment effect. Further properties are discussed and a comparison with the two-trials rule is made, as well as with alternative research synthesis methods. An attractive feature of the new approach is that a claim of success requires each study to be convincing on its own to a certain degree depending on the overall significance level and the number of studies. A real example with 5 clinical trials investigating the effect of Carvedilol for the treatment of patients with moderate to severe heart failure patients is used to illustrate the methodology.

Preprint available at https://arxiv.org/abs/1911.10633

[OC25.4]

An Application of Bayesian Hierarchical Model in Platform Phase I Oncology Dose Escalation studies

Juan Guillermo Gonzalez Maffe[1], Tiina Kirsilae[1], Andrew Bean[2], Alessandro Matano[1], Matt Whiley[1]

[1] *EDB-Oncology, Novartis, Switzerland*
[2] *EDB-Oncology, Novartis, United States*

Background: In oncology, combination therapies are often used to improve the anti-tumour activity compare to single agent therapy. Nowadays, platform dose escalation studies are designed where a backbone treatment is combined with different drugs, and treatments can be added or dropped in an adaptive fashion. For these studies, the common practice is to perform model-based dose escalation using separate models for each treatment. However, as all drugs in the study are combined with the backbone treatment, a more efficient approach would be to implement a single model which includes all the treatments and allows borrowing of safety information on the backbone agent(s).

Objective: The objective of this presentation is to illustrate an extension of the meta-analytic-combined (MAC) approach proposed by Neuenschwander et al [1] for a platform dose escalation study with one model.

Methods: Bayesian hierarchical models are often used to guide dose escalation studies. These models estimate the risk of toxicity for any dose level, and are used to identify the Maximum tolerated dose (MTD), and at each dose escalation provide an estimate of the risk to future patients. Specifically the MAC approach proposed by Neuenschwander et al (2016) allows the model to borrow information between strata (i.e. treatments) to estimate the model parameters. In addition, historical and/or external co-data associated with any of the treatments can be included in the model.

Results: The results will be shown during the presentation. The performance of the model is investigated by exploring its behaviour in hypothetical scenarios. Operating characteristics are computed using different dose-toxicity curves to assess the long term performance of the model in terms of sample size optimization and identification of the MTD.

Conclusion: The MAC approach is a powerful tool that allows borrowing information between strata from different data sources. It can help when dealing with small sample size as it makes efficient use of all the data. The MAC approach can be robustified to deal with situations when the assumption of exchangeability may be questionable.

Reference:

1. *Neuenschwander B, Roychoudhury S & Schmidli H. On the Use of Co-Data in Clinical Trials, Statistics in Biopharmaceutical Research (2016)*

[OC25.5]

Confidence intervals for the Mann-Whitney test

Damjan Manevski, Maja Pohar Perme

Institute for Biostatistics and Medical Informatics, Faculty of Medicine, University of Ljubljana, Slovenia

Background: The Mann–Whitney test is a commonly used non-parametric alternative of the two-sample t-test. Despite its frequent use, it is only rarely accompanied with confidence intervals of an effect size. If reported, the effect size is usually measured with the difference of medians or the shift of the two distribution locations. Neither of these two measures directly coincides with the test statistic of the Mann–Whitney test. Thus, the interpretation of the test results and the confidence intervals may be importantly different.

Objective: In this talk, we will focus on the probability that the value of the random variable X is lower than the random variable Y. This measure is often referred to as the degree of overlap or the probabilistic index; its estimator is in a one-to-one relationship with the Mann–Whitney test statistic. The measure also equals the area under the ROC curve. In our work, we focus on calculating confidence intervals for this measure.

Methods: Several methods have been proposed for the construction of the confidence interval, the Newcombe 5th method being the commonly suggested one. Here, we will review the most promising methods and explain their ideas.

Results: We will show some properties of different variance estimators and the small sample problems of the confidence intervals construction. We will identify scenarios in which the existing approaches yield inadequate coverage probabilities. We will conclude that the DeLong variance estimator is a reliable option regardless of the scenario, but confidence intervals should be constructed using the logit scale to avoid values above 1 or below 0 and the poor coverage probability that follows. A correction is needed for the case when all values from one group are smaller than the values of the other. We will propose a method that improves the coverage probability also in these cases.

Conclusions: Our simulations have revealed that the suggested approach gives expected results even in scenarios where the other methods fail. We believe that the Mann–Whitney test could be much more commonly reported together with the given measure and a reliable calculation of confidence intervals is the key step for making such reporting sensible.

OC26: Causal Inference for Survival Analysis

Chair: **Els Goetghebeur**

Universiteit Gent, Belgium

[OC26.1]

G-computation and Inverse Probability Weighting for time-to-event analyses

Arthur Chatton[1], Florent Le Borgne[1], Clémence Leyrat[2], Yohann Foucher[1]

[1] *UMR INSERM 1246 SPHERE, Nantes University, France*
[2] *Department of Medical Statistics & Cancer Survival Group, London School of Hygiene and Tropical Medicine, United Kingdom*

Background: Real-world evidence from observational data is growing valuable to influence health care policies. Unfortunately, observational studies are subject to bias, such as confounding. G-computation (GC) and Inverse Probability Weighting (IPW) allow us to draw causal inferences from these data. However, they rely on different modeling assumptions.

Objectives: We sought to compare the performances of the GC and the IPW to estimate the causal difference in restricted mean survival time (RMST) in the context of time-to-event outcome and time-fixed confounders. Furthermore, we investigated the choice of the covariates set to consider: all the covariates causing the outcome or only the subset of the common causes of the outcome and the exposure (true confounders).

Methods: For the GC, we used the joint likelihood proposed by Breslow from a Cox model to estimate the cumulative baseline hazard and the regression coefficients necessary as the Q-model, which we then used to predict the survival rates in counterfactual populations. For the IPW, we fitted the propensity score from a logistic model and we used the weighted Kaplan-Meier estimator to obtain the survival rates. We estimated the variance by bootstrapping. We compared the performances of the two methods through simulations with different censoring rates, sample sizes, and effect intensities. We have also illustrated these methods on real data-sets. We focused here on the difference in RMST because it is risky to draw clear causal conclusions from the hazard ratio.

Results: We reported that both GC and IPW led to a bias close to 0. However, the variance was lower for GC than IPW in all the scenarios. Besides this main result, the simulations also highlighted that the GC should preferentially consider all the covariates causing the outcome. When we only considered the true confounders, the variance and the bias increased. Both sets led to similar results for IPW.

Conclusions: The present study shows that the GC is associated with a higher power than the IPW when estimating the difference in RMST for a right-censored time-to-event outcome in the presence of time-fixed confounders. All the covariates causing the outcome should be considered in GC.

[OC26.2]

Adjusting the estimate of overall survival benefit for treatment switching in a randomized clinical trial

Suzy Van Sanden, Joris Diels

Market Access Analytics, Janssen, Belgium

Background: When patients in the control arm of a randomized clinical trial (RCT) are allowed to switch to the experimental treatment once treatment benefit based on short-term endpoints is demonstrated, ITT based estimates of long-term endpoints as OS can lead to substantial underestimation of the real treatment benefit.

Objective: We want to demonstrate the impact of treatment switching and the application of several methods to adjust for bias induced by treatment switching on the hazard ratio (HR) for overall survival (OS) across different interim analyses.

Method: We analyzed an RCT in a hematological disease which allowed patients in the control arm to switch to experimental treatment at time of progression. The rank preserving structure failure time model (RPSFTM), inverse probability of censoring weights (IPCW) and the two-stage method were considered [1]. The underlying assumptions behind all three methods were assessed.

Results: After a median follow up of approximate 40 months, over 65% of the patients in our trial had switched to the active arm. The estimated OS-benefit, based on the ITT HR [95%CI] of experimental treatment versus control, decreased from 0.39 in the first interim analysis (less than a year of follow-up) to 0.77. The assumptions behind RPSFTM and IPCW were judged to be reasonable. Both methods applied to the later datacut had a significant impact on the HR for OS (RPSFTM: 0.37; IPCW: 0.41). The adjusted HRs were similar to the ITT HR of the first interim analysis, where impact of treatment switching was still minor (smaller number and limited exposure time on the active treatment).

Conclusions: The current case illustrates ITT estimates of OS benefit within an RCT to become increasingly biased over time along with increasing switching and exposure of patients in the control arm to the experimental treatment and demonstrates the potential of advanced methods to adjust for the induced bias. This is especially relevant for health technology assessment of highly effective novel treatments in oncology, where initial OS-results based on early datacuts are often requested to be confirmed by more mature data with longer follow-up.

Reference:

1. *Latimer et al. 2014 NICE DSU Support Document 16.*

[OC26.3]

Emulation of target trials to investigate causal effects of lung transplant on survival in cystic fibrosis

Ruth Keogh[1], Daniela Schlueter[2], Nan van Geloven[3], Elisa Allen[4], Mo Al-Aloul[5], Freddy Frost[6], Dilip Nazareth[6], Oliver Rayner[7], Nicholas Simmonds[8], Thom Daniels[9]

[1] *London School of Hygiene & Tropical Medicine, United Kingdom*
[2] *University of Liverpool, United Kingdom*
[3] *Leiden University Medical Center, Netherlands*
[4] *NHS Blood and Transplant, United Kingdom*
[5] *Wythenshawe Hospital Manchester, United Kingdom*
[6] *Liverpool Heart and Chest Hospital, United Kingdom*
[7] *Cystic Fibrosis Trust, United Kingdom*
[8] *Royal Brompton Hospital London, United Kingdom*
[9] *University Hospital Southampton NHS Foundation Trust, United Kingdom*

Background: Cystic Fibrosis (CF) is a life-shortening condition affecting around 10,500 people in the UK. Lung transplantation is an option for people with CF who have end-stage lung disease. However, making decisions about transplant is difficult and there is a lack of information about its benefits. Understanding the impact of transplant on survival relies on observational data such as national transplant registers, as randomized trials are infeasible. Previous investigations have typically quantified the effect of transplant on survival using a hazard ratio, and the causal questions to be addressed have not been clearly articulated.

Objectives: To specify 'target trials' to articulate different causal questions and illustrate how to answer them using observational longitudinal and survival data from the UK CF Registry.

Methods: We focus on three causal questions: (i) the effect of joining the transplant wait-list in the 'real world', which may or may not result in a transplant; (ii) the effect of joining the wait-list in an 'ideal world' where patients receive a transplant within one year of joining the waiting list, and (iii) the impact of transplant in those who actually received one. Analyses are based on forming a sequence of emulated trials within the longitudinal data, and time-dependent confounding is handled using covariate adjustment and inverse probability weighting. Results are based on people with CF in the UK eligible to join the waiting list for transplant between 2007 and 2016.

Results: The results support a survival benefit of lung transplant in CF. In preliminary findings, the effect of joining the waiting list was found to be beneficial for survival up to 5 years post-transplant with an increased survival probability of 15%. In the 'ideal world' in which all individuals could receive a transplant within one year, the transplant was estimated to give a 25% improvement in 5-year survival. The impact of transplant in those who actually received one was estimated to be greater still.

Conclusions: The target trials framework is helpful in asking and answering causal questions about the benefit of organ transplantation. Different questions may be of interest to patients, clinical teams and policy makers.

[OC26.4]

Comparing instrumental variables approaches for time to event data

Ellie John, Michael Crowther, Nuala Sheehan

Health Sciences, University of Leicester, United Kingdom

Background: Instrumental variable (IV) approaches can yield causal effect estimates in the presence of unmeasured confounders. These methods have been well developed for linear models but their application to time-to-event data is more problematic due to non-collapsibility. Additive hazards models have been suggested for IV estimation in time-to-event data as they yield a non-collapsible effect measure. Methods for IV estimation in Cox models have been proposed but only provide approximations of the causal effect.

Objectives: The aim of this study was to assess the performance of different IV methods for Cox-proportional hazards and additive hazards models. The objective was to ascertain whether an approximate estimate of the hazard ratio obtained from a Cox IV model could be more informative than an unbiased estimate of the hazard difference from an additive IV model. Models were compared using survival probabilities since the methods estimate different, and thus incomparable, parameters.

Methods: Survival times were simulated using both exponential and Weibull baseline hazards alongside additive and multiplicative covariate effects. The predicted survival probabilities from each model were compared to the true survival probabilities for each scenario at several follow-up times.

Results and Conclusions: Cox IV methods performed well when the covariate effects were multiplicative, yielding survival probabilities similar to the true probabilities, when there was moderate unmeasured confounding. However, when covariate effects were additive, the Cox IV model performed very poorly with more bias than the naive regression models. Similarly, additive IV models performed fairly well when the underlying covariate effect was additive although they tended to slightly over predict survival, even with minimal unmeasured confounding. When the additivity assumption was violated, and the true model was multiplicative, the additive IV models were biased compared to the true predicted probabilities with more bias than the naive regression models. In general, additive IV models tended be more robust to violations of the additive model assumption.

Instrumental variable methods are known to be sensitive to violations of the IV assumptions. Here, even with a valid instrument, the methods are also sensitive to the underlying modelling assumptions. This poses a practical challenge for analyses of time-to-event data.

[OC26.5]

„Plug-In" Estimation for Parametric and Penalised Multi-State Markov Models

Mark Clements

Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden

Background: Multi-state models form a broad model class that includes standard survival models, competing risks and illness-death models. However, existing multi-state models are restricted either by the ‚plug-in' parameters that they can estimate, the dependence on simulations for parameter estimation, or the dependence on the bootstrap for variance estimation of those parameters.

Objective: To develop an efficient algorithm and implementation for ‚plug-in' maximum likelihood estimation for parametric and smooth penalised Markov multi-state models.

Methods: We restrict our attention to smooth parametric and penalised transition intensities for multi-state Markov models. We propose a new algorithm that uses a system of ordinary differential equations to calculate the parameters and their gradients, with standard errors calculated using the delta method. The algorithm supports ‚plug-in' parameter estimation for state occupancy probabilities, transition probabilities, prevalence, length of stay, relative survival, screening sensitivity, utilities and costs. We provide an implementation in R that allows for a wide range of parametric and penalised survival models, including Poisson regression, smooth generalised survival models and smooth accelerated failure time models.

Results: Using simulations, we demonstrate good coverage for a range of transition intensity models. As a demonstration of the approach, we apply these methods to an earlier multi-state analysis of the Rotterdam Breast Cancer Data by Crowther and Lambert (2017). We then extend their analysis to include regression standardisation.

Conclusions: We have provided a broad framework for ‚plug-in' parameter estimation for Markov multi-state models with smooth transition intensities. These methods have applications to a range of disciplines, including descriptive epidemiology, causal inference and health economics.

OC27: Survival Analysis 3

Chair: **Terry Therneau**

Mayo Clinic, United States

[OC27.1]

Relaxing the Independence Assumption in Relative Survival Analysis: A Parametric Approach

Reuben Adatorwovor[1], Aurelien Latouche[2], Jason Fine[1]

[1] *Biostatistics, University of North Carolina at Chapel Hill, United States*
[2] *Conservatoire National des Arts et Metiers, Paris, France, France*

With known cause of death, competing risk survival methods are applicable in estimating disease-specific survival. Relative survival analysis may be used to estimate disease-specific survival when cause of death is either unknown or subject to misspecification and not reliable for practical usage. This method is popular for population-based cancer survival studies using registry data and does not require cause of death information. The standard estimator is the ratio of all-cause survival in the cancer cohort group to the known expected survival from a healthy reference population. Disease-specific death competes with other causes of mortality, potentially creating dependence among the causes of death. The standard ratio estimate is only valid when death from disease and death from other causes are independent. To relax the independence assumption, we formulate dependence using a copula-based model. Likelihood-based method is used to fit a parametric model to the distribution of disease-specific death without cause of death information, where the copula is assumed known and the distribution of other cause of mortality is derived from the reference population. Since the dependence structure for disease related and other-cause mortality is nonidentifiable and unverifiable from the observed data, we propose a sensitivity analysis, where the analysis is conducted across a range of assumed dependence structures. We demonstrate the practical utility of our method through simulation studies and an application to French breast cancer data.

[OC27.2]

Building robust specific life tables using multidimensional penalized splines

Patricia Soares[1], Laurent Roche[2], Mathieu Fauvernier[3], Laurent Remontet[2], Bernard Rachet[3]

[1] *Public Health Research Centre, Universidade NOVA de Lisboa, Portugal*
[2] *Service de Biostatistique-Bioinformatique, Hospices Civils de Lyon, France*
[3] *Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, United Kingdom*

Background: General population life tables are typically defined by age, sex, calendar year and geography. Including other key characteristics (e.g. socioeconomic level and ethnicity) is nevertheless of great interest. Specific life tables, i.e. for sub-populations, are however rarely available and building them is challenging because of instability due to data sparseness or poor quality. We developed a flexible Poisson generalized linear model (GLM) with restricted cubic splines for estimating robust mortality hazards (Rachet et al, BMC Public Health 2015). Simulation studies showed good performances, but the simulation-based process for choosing the splines' knots remains complex and partly depends on the user.

Objective: To remedy these issues, we apply generalized additive models (GAMs) with tensor product smoothers.

Method: Both GAM and GLM approaches are applied to observed mortality rates from the general population of Portugal (1996-2017) and England (2011). Mortality is stratified by sex and respectively modelled by age, year and region (Portugal), or by age, deprivation index and region (England). For the GAM, we used a multidimensional penalized spline built from the tensor product of the marginal restricted cubic splines bases of the continuous variables. For the regions, initially modelled separately, the penalty is applied to the mortality hazard ratio (the largest region being the reference), allowing information to be drawn from this region. R-package mgcv (option te for tensor products) is employed.

Results: Preliminary results on real data show that the GAM approach provides mortality hazard rates by single year of age and other combination of covariables (year or deprivation, and region) as robust as with the GLM approach, even in situation of sparse data (based on AIC, residuals, MSE). Furthermore, compared to GLMs, GAMs naturally ease the model building process by accounting for complex interactions and provide estimates which are much less sensitive to the user's decisions.

Conclusions: When building specific life tables, preliminary results suggest that the penalized framework and in particular multidimensional penalized splines, a natural way to model simultaneously two or three variables, are advantageous compared to more conventional approaches. Projections, for example beyond the census year, are also easier to draw.

[OC27.3]

Flexible estimation of complex, lagged, cumulative effects in the context of competing risks.

Helmut Küchenhoff[1], Andreas Bender[1], Philpp Kopper[1], Fabian Scheipl[1], Wolfgang Hartl[2]

[1] *Statistical Consulting Unit StaBLab, LMU Munich, Germany*
[2] *Department of General, Visceral, and Transplantation, Surgery, University School of Medicine, LMU Munich, Germany*

Background: Critically ill patients require artificial nutrition. To study the association between different qualities of nutrition and survival, we analysed a large international database containing more than 10,000 patients treated on 400 intensive care units (ICUs) worldwide. Individual actual calorie and protein intake had been recorded for a maximum of twelve days after ICU admission. Actual intake had been compared to the prescribed nutrition targets. Survival time had been registered until the time of death or of hospital discharge, or up to a maximum of 60 days if a patient had remained in the hospital during this time. In an earlier study, we had developed piece-wise exponential additive mixed models (PAMMs) to estimate the complex, lagged, cumulative (functional) effects of time-dependent covariates, and we had used these models to calculate nutrition-associated hazard rates [1]. Patients discharged alive before day 60 had been treated as censored.

Objective(s): Since censoring a patient being discharged alive before day 60 is problematic, we want to develop competing risk models to get insides of the associations of nutrition with hospital discharge and with survival.

Methods: We extended the methodological foundations of PAMMs to the competing risk setting by estimating the cumulative incidence function using (a) cause-specific hazards and (b) subdistribution hazards. Both methods were used to evaluate the association between nutrition and hospital mortality.

Results and conclusions: We found that it is possible to extend complex lag response models to a competing risk analysis. We checked our new strategy by a simulation study and estimated the cumulative association of nutrition with survival and hospital discharge in our study. The outlined methods are implemented in the R package pammtools.

Reference:

1. *A. Bender, F. Scheipl, W. Hartl, A. Day and H. Küchenhoff (2018): Penalized estimation of complex,non-linear exposure-lag-response associations. Biostatistics, 20(2):315-331.*

[OC27.4]

Flexible Bayesian hierarchical excess hazard models using low-rank thin plate splines

Manuela Quaresma, James Carpenter, Bernard Rachet

Faculty of Epidemiology & Population Health, London School of Hygiene & Tropical Medicine, United Kingdom

Context: Regression models for the excess hazard are the preferred modelling tool for cancer survival research using population-based data. The models are commonly formulated as the additive decomposition of the overall hazard into two components: the excess hazard due to the cancer of interest and the population hazard due to all other causes of death.

Objectives(s): To introduce a flexible excess hazard model within the Bayesian framework that accommodates hierarchical data structures, and to demonstrate how to derive posterior distributions of net survival (the survival due to the cancer of interest) for different subgroups of the population.

Method(s): The proposed model was formulated on the log-excess hazard scale, using low-rank thin plate splines to model the baseline log-excess hazard and the smooth effect of any continuous covariates. As an application, we analysed data for 16,326 patients diagnosed with colon cancer between 2006-2013, who were living within a London Clinical Commissioning Group (CCG) at the time of their diagnosis and received cancer care in a hospital located within a London CCG. We estimated the variability in net survival at both CCG and hospital level after adjusting for the individual-level factors age at diagnosis, socioeconomic status and stage at diagnosis.

Results: Variability in survival between CCGs vanished after adjusting for hospital of care. This result contrasted with a much more pronounced variability between hospitals.

Conclusions: We have shown how a flexible Hierarchical Bayesian model for the log-excess hazard can be used for population-based research, making it possible to model more complex cancer data structures. In our experience, we found that using low-rank thin plate splines provides a good compromise between the achieved model flexibility and the retained tractability that reduces computational intensity.

Future work in the context of the presented application should aim (i) to investigate hospitals with poorer performance to understand its causes, including resources and organisation among other factors, and (ii) to examine more in depth (including qualitative studies) what determines the choice (or absence of choice) of patients for a given hospital, in order to suggest actions to correct such wide disparities.

[OC27.5]

Adapting SIMEX to correct for bias due to interval-censored data in survival analysis

Michal Abrahamowicz[1], Marie-Eve Beauchamp[2]

[1] *Department of Epidemiology and Biostatistics, McGill University, Montreal, Canada*
[2] *Centre for Outcomes Research and Evaluation, Research Institute of the McGill University Health Centre, Canada*

Background: In many real-life time-to-event analyses, the clinical endpoint of interest may be ascertained only at clinic visits, several months or years apart. Examples include cancer recurrence, cognitive deterioration or organ damage, where it is only known that the event occurred within an interval between two visits. For such 'interval-censored data', assigning an arbitrary event time at the mid-point or end of the interval may induce seriously biased estimates and incorrect inference [1]. Existing methods for interval-censored data rely on strong assumptions [1] and do not handle time-varying exposures.

Objectives: To propose a pragmatic, easy-to-implement 'generic' method that avoids restrictive assumptions and can be used in time-to-event analyses with time-varying data, and to evaluate its performance in simulations.

Methods: Our method adapts the simulation-extrapolation (SIMEX) approach [2] to interval-censored time-to-event data. The general idea is to use the DELEX approach proposed by Andersen and Liestol to deal with sparse measurements of time-varying covariates [3]. We first artificially decrease the visit frequency and estimate how this affects the parameter of interest. Then, the results are extrapolated to the 'ideal case' scenario with the outcome observed continuously during follow-up. In simulations, we evaluate the method in two practically relevant scenarios. The first, simpler scenario focuses on the effects of time-invariant prognostic factors on recurrence-free mortality, with cancer recurrence as an interval-censored censoring event. The second, more complex, simulated scenario deals with inaccurate event times in studies of the impact of time-varying exposures, such as e.g. current drug use or cumulative duration of use in the past 3 months, on the hazard of interval-censored adverse events.

Results: In both scenarios, our SIMEX-based estimates reduce the bias by 30%-70% relative to Cox model-based HRs and improve the mean squared errors. We will also present a pharmaco-epidemiology application with time-varying drug exposure.

Conclusions: Results suggest that our SIMEX-based method can substantially improve the accuracy of the analyses of interval-censored data.

References:

1. *Lindsey JC, Ryan LM. Stat Med 1998; 17(2): 219-38.*
2. *Cook JR, Stefanski LA. JASA 1994; 89(428): 1314-1328.*
3. *Andersen PK, Liestol K. Biostatistics 2003: 4(4); 633–649.*

OC28: Personalised Medicine 2

Chair: **Ruth Pfeiffer**

National Cancer Institute, Bethesda, United States

[OC28.1]

Developing and testing high-efficacy patient subgroups within a clinical trial using risk scores

Svetlana Cherlin, James M. S. Wason

Newcastle Upon Tyne, Newcastle University, United Kingdom

Background: There is the potential for high-dimensional information about patients collected in clinical trials (such as genomic, imaging and data from wearable technologies) to be informative for the efficacy of a new treatment in situations where only a subset of patients benefits from the treatment. The adaptive signature design (ASD) method allows a trial to develop and test efficacy of treatment in a high-efficacy patient group (the sensitive group) using genetic data. The method requires selection of three tuning parameters which may be highly computationally expensive. Additionally, the approach applies the same threshold for inclusion of the genes, therefore the same genes are included into the signature for all the patients. This has the result that the signature does not necessarily efficiently use all the information from the genetic profile.

Objective: We aim to develop a computationally efficient alternative to the existing ASD method by incorporating flexibility in the selection of the genes which has the potential to improve the predictive ability of the signature. Furthermore, we aim to extent the new method to developing signatures to classify patients on two outcomes.

Method: We propose a variation to the ASD method, the CVRS (cross-validated risk scores) design method, that does not require selection of any tuning parameters. The method is based on computing a risk score for each patient and dividing them into clusters using a non-parametric clustering procedure. Additionally, we propose an extension of the CVRS design (CVRS2) that considers two outcomes. We have implemented the new methods in an R package.

Results: Both CVRS and CVRS2 methods, as assessed for various sample sizes and response rates, have a substantial reduction in the computational time required. In many scenarios there is a substantial improvement in the ability to correctly identify the sensitive group and the overall power of the design. We illustrate the application of both CVRS and CVRS2 methods on a randomised psychiatry trial with 86 baseline covariates.

Conclusions: The new methods show a superior performance and drastically improve the computational time, in comparison to the existing ASD method. Further research will consider different types of outcomes.

[OC28.2]

Favoring the hierarchy constraint of interactions in penalized survival models of randomized trials

Shaima Belhechmi[1], Riccardo De Bin[2], Federico Rotolo[3], Stefan Michiels[1]

[1] *Villejuif, Université Paris-Saclay, Univ. Paris-Sud, UVSQ, CESP, INSERM, F-94805; Gustave Roussy, Service de biostatistique et d'épidémiologie, F-94805., France*
[2] *Oslo, 3Department of Mathematics, University of Oslo, Norway*
[3] *Marseille, Biostatistics and Data Management Unit, Innate Pharma, France*

Background: Randomized clinical trials (RCT) often assume that the treatment effect is homogeneous across patient characteristics whereas precision medicine assumes that the treatment effect is heterogeneous and seeks to identify biomarkers distinguishing patients that will benefit most from a targeted treatment. Previously, Ternes et al (Biom J, 2017) have shown that the adaptive lasso performs reasonably well to select biomarker by treatment interactions in a high-dimensional survival regression model, but its main drawback is that the hierarchy constraint – between the main effect of a biomarker and the interaction effect– is not respected.

Objective: Our objective was to evaluate different approaches to favor the hierarchy constraint in penalized survival regression models from RCTs and to evaluate whether they improve calibration.

Methods: We present several approaches: we propose to group the biomarker effect and its interaction with treatment and to evaluate two-level selection methods such as sparse group lasso (SGL), the composite minimax concave penalty (cMCP), the group exponential lasso (gel) and adaptive lasso with weights proportional to the strength of the interaction between the treatment effect and the biomarker. In order to assess the selection and prediction ability of the proposed methods, we use our previously proposed difference between trial-arm c-indices, the C-for-benefit and calibration for benefit with censored survival data.

Results and Conclusion: We conducted a simulation study with six scenarios, including three null scenarios and three alternative scenarios. In the null scenarios, no biomarker interacts with the treatment and in the alternative scenarios; at least one biomarker interacts with the treatment. For each scenario, the approaches proposed will be applied to the simulated data to evaluate the performance of these models in terms of true positive selection corresponds here to a predictive biomarker and in terms of C-for-benefit and calibration for benefit. Simulations will show the different calibration metrics of the proposed methods. To illustrate the proposed approaches, we will also use gene expression data (p=462 genes) from a randomized controlled trial, evaluating the effect of trastuzumab in 1574 breast cancer patients.

[OC28.3]

Accounting for calibration drift due to improvements in baseline survival during prognostic model development

Sarah Booth, Paul Lambert, Mark Rutherford

Department of Health Sciences, University of Leicester, United Kingdom

Background: Prognostic models which are used to produce long-term survival predictions are often developed using datasets which cover a long diagnosis period. If survival is continually improving during this time, it can lead to predictions which under-estimate the survival of recently diagnosed patients. The developments here focus on models which produce survival predictions following a cancer diagnosis, however these methods can be generalised to other outcomes.

Objectives: The main aim was to compare a range of methods which can be applied when developing prognostic models to account for calibration drift due to improvements in baseline survival. A further aim was to identify which approach may perform best in different settings.

Methods: An extensive comparison of approaches for accounting for improvements in survival was performed. This included modelling calendar time directly by including the year of diagnosis as a predictor, and if required, allowing for non-proportional hazards and interaction terms. Alternative period analysis based methods involved using delayed-entry techniques to analyse the most recent subset of data and temporal recalibration which uses this subset to re-estimate the baseline hazard of standard prognostic models. We compare approaches in a simulation setting under a range of assumptions and further show examples of models for colon cancer developed using US population-based registry data from the Surveillance, Epidemiology, and End Results (SEER) Program database.

Results: Accounting for changes in baseline survival often improved the calibration of the survival predictions for newly diagnosed patients compared to the standard model. However, when time-dependent effects for the year of diagnosis were included, this sometimes resulted in inaccurate survival predictions. This is likely due to the most recently diagnosed patients having a limited amount of follow-up which leads to an inappropriate extrapolation based on the long-term information from patients who were diagnosed earlier.

Conclusions: Care should be taken when modelling complex trends in calendar time; particularly when allowing the effect of calendar time to be time-dependent. A more conservative approach to produce up-to-date survival predictions can be achieved using period analysis based methods as less extrapolation is required.

[OC28.4]

I'm not an average but a patient! A significant Holy Grail.

Bernard Francq, Dan Lin, Walter Hoyer

GSK, Belgium

Background and Objective: Which of the following sentences would you prefer your surgeon to tell you?

1. Surgery A is significantly better than B, on average ($p<0.05$).
2. It might be possible that surgery A does not remove your tumor (>5%) while this is nearly impossible with surgery B (<0.1%).
3. Surgery A is successful for 80% of the patients and surgery B for 99% of the patients.

When investigating the effect of new drugs, medical papers usually report the estimated (difference of) treatment effects with their corresponding Confidence Intervals (CIs). Clinical trials are designed to achieve a given statistical power and afterwards researchers hope that the p-value will be tiny to reach the Holy Grail, namely the statistical significance.

On the other hand, Prediction Intervals (PIs) are very useful as they give a range where the measurement for a new patient may lie (under the same treatment condition). Furthermore, a confidence level can be added to obtain a Tolerance Interval (TI) which is a good extension especially for small sample sizes.

Methods, Results: Several examples will be used to illustrate the different statistical intervals: from bridging study, assay calibration study and orthopedic surgery study. In the last one, the risk of intralesional resection bone tumor is evaluated with linear mixed models. The concepts of p-values, CI, PI, TI and equivalence testing will be discussed under the significance crisis framework. The three statistical intervals together provide a good understanding of the medical study results but answer different questions. Several drawbacks of the classical CIs and the significance testing via p-values are mitigated by using PIs and TIs instead. If equivalence, non-inferiority or robustness is to be shown, then in addition turning to an equivalence approach is highly recommended.

Conclusions: In medicine, the focus must be on the patient level (and not "on average"), and this is where the PI or TI are useful. CIs have been widely accepted and used; while PIs should be equally addressed or even preferred.

Reference:

Francq, Lin, Hoyer. Confidence, Prediction and Tolerance Intervals in Linear Mixed Models, Stat in Med, 2019.

[OC28.5]

Statistical and machine learning methods for uncertainty-informed decision referral

Gabriela Czanner[1], Silvester Czanner[2], Ala Al-Kafri[2]

[1] *Department of Applied Mathematics, Liverpool John Moores University, United States*
[2] *Department of Computer Science, Liverpool John Moores University, United States*

Background: Glaucoma is the leading cause of irreversible blindness worldwide, with a strong need for effective earlier diagnosis to prevent sight loss. One important clinical feature is asymmetrical narrowing of the optic nerve rim, which is difficult for humans to quantify reliably, making diagnosis a challenge. Accurate and automated optic disc assessment is therefore needed to prevent sight loss. Such automated assessment has traditionally been proposed without methods to quantify their uncertainty in a decision. In contrast, a clinician knows when he is uncertain such as when an unusual case is presented, in such case a more experienced clinician is needed.

Methods: We have developed an algorithm that has two parts: machine learning (ML) algorithm to automatically detect boundaries of optic cup and disc, and a statistical predictive algorithm, which combines generative spatial-statistical shape modelling with Bayes to determine the likelihood that an optic disc is symptomatic of glaucoma. We proposed an uncertainty measure of the automated classification based on posterior confidence intervals and boostrapping. We then proposed a decision support system that holds the uncertain cases for further clinical investigation (such as more precise test or retaking the image). We trained on online datasets ORIGA (n=650) and tested externally on dataset RIMONE (n=159).

Results: The performance of the glaucoma detection improved from 91 to 94% (AUROC), 90 to 95% (sensitivity) and from 74 to 87% (specificity) when referring 13% (20 out of 159) of the most uncertain decision for further inspection.

Conclusion: The performance of the automated glaucoma detection improved due to uncertainty-informed decision referral. This has potential to make AI-automated glaucoma classification less uncertain and to make screening more accurate.

OC29: Measurement Errors

Chair: **Pamela Shaw**

University of Pennsylvania, United States

[OC29.1]

Adjusting for misclassification of a predictor in an individual participant data meta-analysis

Valentijn M. T. de Jong[1], Harlan Campbell[2], Thomas Jaenisch[3], Paul Gustafson[2], Thomas P. A. Debray[1]

[1] *Julius Center for Health Sciences and Primary Care, UMC Utrecht, Netherlands*
[2] *Department of Statistics, The University of British Columbia, Canada*
[3] *Department of Infectious Diseases, Universität Heidelberg, Germany*

Background: A common problem in the analysis of multiple data sources, including individual participant data meta-analysis (IPD-MA), is the presence of misclassification of binary variables. Misclassification may bias estimates of parameters (including covariate effects), even when the misclassification is entirely random. Available methods for addressing misclassification in the analysis of exposure-outcome associations do not account for between-study heterogeneity in IPD-MA.

Objective: We aimed to develop statistical methods that facilitate unbiased estimation of adjusted and unadjusted exposure-outcome associations and between-study heterogeneity in IPD-MA where the extent and nature of exposure misclassification may vary across or within studies.

Methods: We present Bayesian methods that allow misclassification of binary exposure variables to depend on study- and participant-level characteristics. We illustrate this in an example of the differential diagnosis of dengue using two variables, where the gold standard measurement for the exposure variable is unavailable for some studies which only measured a surrogate prone to misclassification. We present a simulation study to assess bias, root mean square error (RMSE), coverage and power in estimating an exposure-outcome association.

Results: In the example, our methods yielded estimates with less error than analyses naive with regard to misclassification or based on gold standard measurements alone. In our simulations, the evaluated misclassification model yielded valid estimates of the true exposure-outcome association, with less RMSE, greater power and similar coverage compared to an analysis restricted to available gold standard measurements.

Conclusions: Our proposed framework can appropriately account for the presence of binary exposure misclassification in IPD-MA. It requires that 1) some studies supply IPD for the surrogate and gold standard exposure and 2) misclassification is exchangeable across studies conditional on observed covariates (and outcome). Further work is needed to address other types of misclassification.

[OC29.2]

Various distributed data affected by lower limits of quantification: Point estimates vs. confidence intervals

Tanja Buelow, Ralf-Dieter Hilgers, Nicole Heussen

Department of Medical Statistics, RWTH Aachen University, Germany

Background: Single or multiple lower limits of quantification (LLOQ) appear in concentration measurement data due to one or more laboratories involved in the quantification of observations, in which some observations are too low to be quantified with required precision. As the missing data mechanism is not random, most statistical methodology to handle missing data is not applicable. In clinical practice, simple imputation methods are often used to receive substitute values for the missing observations. Nevertheless, they lead to severe bias in estimating parameters such as the mean and variance. Even procedures relying on the assumption of normally distributed concentration data are little robust against distributional model misspecification. Instead of building conclusions upon point estimates, interpreting confidence intervals (CI) for the intriguing parameters would broaden the applicationof parametric methods.

Objective(s): The objective is to propose parametrical methods for different distributional assumptions to estimate the parameters mean and variance as well as showing the advantage of interpreting CI's rather than point estimates regarding robustness.

Method(s): We transfer the existing maximum-likelihood based approach relying on the normal distribution assumption for LLOQ's to other distributional assumptions, where we distinguish between truncated and censored samples. With suitable approaches at hand for specific distributions, we not only investigate the robustness of the point estimates for mean and variance, but also compare bootstrap CI's with parametrical CI's to evaluate other ways of interpretation. The application of the proposed procedure will be demonstrated using data from a cohort study, in which the presence of multiple LLOQs are an issue.

Results: The variety of distributional assumptions for which the methods are applicable gives the applicant a broadly usable tool to handle LLOQ affected data with appropriate approaches. Under uncertainty regarding the underlying distribution, CI's prove to deliver robust interpretation possibilities for the intriguing parameters.

Conclusion: If the data is affected by LLOQs, the distribution of the present observations should be examined to find the most suitable approach in handling the estimation issue. Whenever the distribution cannot be specified, the interpretation of the results should be based on the presented confidence intervals rather than solely on point estimates.

[OC29.3]

On Improved SIMEX for Covariate-Dependent Measurement Error with Continuous Covariates

Steve Ferreira Guerra, Michal Abrahamowicz, Robert Platt

Epidemiology, Biostatistics and Occupational Health, McGill University, Canada

Background: Due to non-adherence, drug prescriptions or dispensations in electronic health records (EHR) may not accurately represent the actual drug intake. Such discrepancies induce measurement error (ME) in the assessment of the true drug exposure, resulting in biased inference in naive analyses. For classical additive ME, SIMEX is a bias reducing approach only requiring distributional assumptions about the variance of the ME, as it assumes errors have zero mean. However, there are many instances in which errors may not have zero mean or may depend on other observed covariates.

Objective: We propose to extend SIMEX to allow to parametrically model the mean of the unknown error distribution. This improves upon existing methods by accounting for dependencies of the ME on many, possibly continuous, variables, without any additional knowledge of the ME distribution.

Methods: We use a simulation study in which we mimic various real-life non-adherence patterns observed in EHR to evaluate the performance of the method to correct for this type of ME. We also assess the sensitivity of the method to various tuning parameters in the SIMEX procedure and compare it to a standard regression calibration approach.

Results: Preliminary results show that naive methods produce biased estimates. The proposed method performs well in simulation, by reducing the bias due to ME. Finally, we demonstrate the performance of the suggested SIMEX when the sample size is relatively small.

Conclusion: Accounting for ME is crucial in obtaining unbiased results in pharmaco-epidemiologic studies. Accounting for non-classical covariate-dependent ME, as is the case with ME due to non-adherence, is especially challenging. Methods that allow modelling the functional relationship between errors and covariates are a promising avenue to more appropriately account for this type of ME.

[OC29.4]

Optimizing the use of dynamic quality indicators as covariates in regression models

Hana Sinkovec, Georg Heinze

Section for Clinical Biometrics, CeMSIIS, Medical University of Vienna, Austria

Background: We consider the problem of evaluating suitability of quality indicators (QI) of health care providers (HCP) by their association with clinical outcomes. As examples, consider the adenoma detection rate commonly used to describe the quality of endoscopists in screening colonoscopy, and its association with post-screening cancer incidence, or prevalence of guideline-compliant treatment of patients with myocardial infarction in different health care regions and its association with patient outcomes such as re-hospitalization or death. In principle, multilevel regression models may be applied for such evaluations. However, a number of challenges arise in practice: (1) QI need to be well-interpretable, (2) QI need to be adjusted for different case-mixes of patients treated by HCP and (3) QI may change with increasing experience of HCP. Typically, QI are expressed as the probability of successful or guideline-compliant treatment and can be adjusted to a reference population using direct or indirect standardization.

Objective: To investigate how QI should enter a regression model as predictor in order to fulfill criteria (1)-(3). In addition, bias and mean squared error of estimated QI-outcome associations should be minimized.

Methods: In various plausible scenarios including those where QI may change over time we investigated four QI definitions that allow to assess their association with outcomes in survival models including one static and three dynamic versions based on a 'cumulative', 'rolling' and 'predicted' QI.

Results: Generally, small numbers of patients per HCP make estimation of QI difficult and obscures estimation of effects on outcomes. The most promising method was to use the predicted QI in a survival model, in which an HCP's current QI value is dynamically predicted from a model trained on all available data, straightforwardly accommodating direct standardization.

Conclusions: In practice underlying values of QI are never known and any estimate is error prone. Unfortunately, in order to fulfill criterion (1), criteria (2) and (3) have often been ignored. In our study we demonstrate that advanced but established statistical techniques allow to define QI that are able to fulfill all three criteria and facilitate unbiased and efficient estimation of associations of QIs with patient outcomes.

OC30: Bayesian Methods for Clinical Trials

Chair: **Paula Moraga**

King Abdullah University of Science and Technology (KAUST), Saudi Arabia

[OC30.1]

Modeling a preference-based index for EQ-5D and EQ-5D+Sleep using a Bayesian framework

Samer Kharroubi[1], Yara Bayh[1], Yaling Yang[2], John Brazier[3]

1 Beirut, American University of Beirut, Lebanon
[2] Nuttfield Department of Primary Care Health Sciences, Oxford University, United Kingdom
[3] Sheffield, University of Sheffield, United Kingdom

Background: The EuroQol five-dimensional (EQ-5D) questionnaire is a generic measure widely used for the assessment of health status, in order to generate quality adjusted life-years and hence to conduct cost utility analysis of health care interventions. The EQ-5D was criticized for being inappropriate in some specific medical conditions and so it was extended by adding a sleep dimension i.e. EQ-5D+Sleep. Conventionally, models used for health state valuation data have been frequentists. Recently a number of researchers have investigated the use of Bayesian methods in this area.

Objectives: The aim of this research is to present an alternative approach to modeling health state valuation data of the EQ-5D and EQ-5D+Sleep descriptive systems, using a Bayesian framework and to demonstrate its superiority to conventional frequentist methods.

Methods: The valuation study is composed of 18 EQ-5D health states and 18 EQ-5D+Sleep health states valued by 160 members of the general public in South Yorkshire, UK using the time tradeoff technique. Three different models were developed for EQ-5D and EQ-5D+Sleep accordingly using Bayesian Markov chain Monte Carlo simulation methods. Bayesian methods were applied to models fitted included a linear regression, random effect and random effect with covariates. The models are compared based on their predictive performance using mean predictions, root mean squared error (RMSE) and deviance information criterion (DIC). All analyses were performed using Bayesian Markov chain Monte Carlo simulation methods.

Results: The random effects with covariates model performs best under all criterions for the two preference-based measures, with RMSE (0.037) and DIC (637.5) for EQ-5D-3L and RMSE (0.019), DIC (416.4) for EQ-5D+Sleep. Compared with models previously estimated using frequentist approach, the Bayesian approach provided better predictions of observed values. It also provided the full distribution of the utility values as a direct output from the modeling process rather than simply providing the mean value and/or standard deviation as is the case with the conventional frequentist approach.

Conclusion: Bayesian methods provide a better way to model EQ-5D-3L valuation data with and without a sleep 'bolt-on' and provide a more flexible in characterizing the full range of uncertainty inherent in these estimates.

[OC30.2]

Dose-finding Bayesian design for toxicity-schedule assessment using pharmacokinetics and pharmacodynamics

Emma Gerard[1], Sarah Zohar[2], Marie-Karelle Riviere[3], Moreno Ursino[4]

[1] Inserm U1138/Sanofi, France
[2] Inserm U1138, France
[3] Sanofi, France
[4] Inserm U1138/APHP, France

Background: Phase I dose-finding trials in oncology seek to find the maximal tolerated dose of a drug under a specific schedule. Evaluating drug-schedules aims at improving treatment safety while maintaining efficacy. However, while we can reasonably assume that toxicity increases with dose for cytotoxic drugs, the relationship between toxicity and multiple schedules remains elusive.

Objective: The aim of this work was to develop a Bayesian dose-finding design for multiple schedules using pharmacokinetic/pharmacodynamic (PK/PD) information to estimate the maximal tolerated dose-sequence (MTDS) at the end of the trial. We propose to model the binary toxicity via a PD endpoint extracted from the continuous PD profile of a biomarker related to toxicity.

Methods: First, the relationship between the dose-sequence and the PD biomarker profile is modeled using non-linear mixed models. Secondly, we propose 2 Bayesian approaches to model the relationship between the PD endpoint and toxicity. For the first approach, we consider a Bayesian 2-parameter logistic regression model using a value of interest of the PD endpoint. For the second approach, we propose a Bayesian hierarchical model with a latent variable considering longitudinal values of the PD endpoint. Lastly, we integrate both models to estimate the posterior toxicity probability of each sequence in order to recommend the MTDS.

Results: We evaluated the operating characteristics of our methods through simulation studies under various scenarios. The results showed that our methods perform better than traditional model-based designs in terms of percentage of MTDS correct selection. Moreover, due to the additional PK/PD information, our methods estimate more precisely the entire dose-sequence-toxicity curve and can propose untested sequences for expansion studies. Our methods will be applied to an ongoing dose escalation trial for patients with relapsed or refractory acute myeloid leukemia (NCT03594955).

Conclusion: Our proposed dose-finding design for multiple schedules provides a reliable way to identify the MTDS when toxicity can be related to a PD biomarker. However, as our methods are applied at the end of the trial, they can be sensitive to the dose escalation design implemented. We plan to extend our approaches to sequential dose-allocation designs in the future.

[OC30.3]

Incorporating historical controls in clinical trials with longitudinal outcomes using the modified power prior

Hongchao Qi[1], Dimitris Rizopoulos[1], Emmanuel Lesaffre[2], Joost Van Rosmalen[1]

[1] *Department of Biostatistics, Erasmus University Medical Center, Netherlands*
[2] *I-Biostat, KU Leuven, Belgium*

Background: Several dynamic borrowing methods, including the power prior and the commensurate prior, have been proposed to increase statistical power and reduce the required sample size in clinical trials where comparable historical controls are available. Most methods have focused on one clinical endpoint, and there is a lack of appropriate methodology for longitudinal outcomes in clinical trials. The modified power prior (MPP) is a borrowing method that downweights the likelihood of the historical data with a power parameter estimated based on the observed data. Here we extend the MPP to longitudinal data analysis based on the linear mixed model (LMM). An important question is whether the MPP approach should be based on the conditional version of the LMM (given the random effects) or the marginal version (averaged over the distribution of the random effects). We refer to the two different implementations as the conditional MPP and the marginal MPP, respectively.

Objective: To extend the MPP to the analysis of clinical trials with longitudinal outcomes.

Method: We have derived some partial analytical results, but the computations involve necessarily MCMC computations. To evaluate the relative performance of the MPP against the commensurate prior, no borrowing, and pooling, a simulation study was performed. In addition, we evaluated our proposal when incorporating one historical control into the analysis of data from the UC San Diego Alzheimer's Disease Cooperative Study (ADCS).

Results: The conditional MPP borrowed most of the historical information among all the dynamic borrowing methods, and it led to more power than the marginal MPP did with low between-study heterogeneity (7.6-12% vs. 6-9%) at the expense of an inflated type I error rates with moderate or high between-study heterogeneity (10.6-20.8%). In contrast, the marginal MPP appropriately controlled the type I error rate. The commensurate prior yielded a limited power gain (0.4-2.8%) but controlled the type I error rate. For the analysis of the ADCS data, the borrowing methods improved the precision of estimates but provided the same conclusions.

Conclusions: The marginal MPP is useful to borrow historical controls in longitudinal data analysis, but the conditional MPP tends to over-borrow the historical data.

[OC30.4]

Bayesian sample size estimation for exploratory basket trials that enable borrowing of information

Haiyan Zheng, James Wason

Population Health Sciences Institute, Newcastle University, United Kingdom

Background: Basket trials have emerged as a new class of efficient approaches in oncology to evaluate a new treatment in several patient subgroups simultaneously. A number of analysis methods for borrowing of information across subtrials with similar treatment effects have been proposed. However, the sample size is often estimated considering the subtrials in isolation. This approach potentially exposes more patients to an unproven treatment than might be necessary, and causes challenges when powering subtrials of rare disease patient subgroups.

Objectives: To develop a sample size estimation (SSE) approach that formalises borrowing of information from subtrials with commensurate treatment effects.

Methods: Representing an external subtrial information into a prior for θ_k, the treatment effect of a contemporary subtrial (labelled k, k = 1, …, K), we introduce a new parameter to characterise the pairwise commensurability between the subtrials. A discrete mixture prior is placed on this commensurability parameter for robust borrowing. This leads to an approximately Normal predictive prior for θ_k, given information from any complementary subtrial $q \neq k$. When the basket trial has at least three subtrials ($K \geq 3$), we combine the commensurate predictive priors into a marginal prior, which is updated to a posterior for θ_k by the contemporary subtrial data. The subtrial-wise sample size nk is found based on the posterior, as a function of the commensurability parameter, information from complementary subtrials, and other standard SSE parameters. We solve nk by a search procedure over all possible values of $n_k > 1$ to find the smallest suitable integer solution.

Results: Simulations demonstrate that our approach converges to the approach of no borrowing, when the commensurability parameter is set to suggest dissimilarity between treatment effects. A smaller expected number of patients are required to establish treatment efficacy in other cases. However, if the postulated prior deviates severely from the true commensurability, our approach may lead to untrustworthy estimation.

Conclusions: We have proposed a Bayesian SSE for basket trials that preserves satisfactory operating characteristics in realistic scenarios. Our research encourages the development of adaptive basket trials with sample size re-estimation at an interim.

[OC30.5]

Prior elicitation of the efficacy of Methotrexate and Mycophenolate Mofetil in Juvenile Localised Scleroderma

Yasin Desai[1], Thomas Burnett[1], Pavel Mozgunov[1], Thomas Jaki[1], Clare Pain[2]

[1] *Mathematics and Statistics, Lancaster University, United Kingdom*
[2] *Alder Hey Children's Foundational Trust, United Kingdom*

Background: Juvenile localised scleroderma (JLS) is characterised by chronic inflammation within the skin and tissues leading to fibrosis. The incidence rate of JLS in the UK is 3.4 per one million children per year. Even though the disease is rare it is associated with significant complications including joint contractures, limb length discrepancy and facial atrophy that impact quality of life.

To date, there has only been one clinical trial in JLS: a randomised placebo controlled trial of methotrexate (MTX). At one year, a disease relapse rate of 32.6% was shown in the MTX group compared to 70.8% of the placebo group ($p<0.005$). This highlighted that MTX therapy may not be effective in one third of patients.

An emerging alternative to MTX is mycophenolate mofetil (MMF). A novel Bayesian trial design is proposed due to concerns of recruiting sufficient patients for a conventional trial. A requirement in this framework is to produce prior distributions.

Objective: To produce informative prior distributions through elicitation of expert opinion on the safety and efficacy of MTX and MMF. This would enable the comparison of the two treatments as well as provide a basis to be used for a subsequent Bayesian clinical trial in JLS.

Method: Prior distributions were produced through undertaking an international meeting to establish expert consensus prior opinion. Questions were asked to 12 clinical experts to elicit their opinion on the quantities of interest. A questionnaire alongside an R Shiny application was used to elicit expert opinion alongside providing live visual updates of results.

Results & Conclusions: The consensus distributions showed that the most likely value of the probability of successful treatment under MTX was 0.680 [95% credibility interval (0.416, 0.854)] whilst under MMF it was 0.705 [95% credibility interval (0.339, 0.899)]. Further the most likely value of the probability of successful tolerance under MTX was 0.620 [95% credibility interval (0.316, 0.842)] whilst under MMF it was 0.766 [95% credibility interval (0.300, 0.936)].

A common theme was that experts were more certain about their opinions on MTX than they were on MMF.

OC31: Clinical Trials Methodology 5

Chair: **David W. Warne**
Switzerland

[OC31.1]

Using dynamic programming to find an optimal dose escalation scheme for a phase I trial

Lizzi Pitt[1], Christopher Jennison[2]

[1] *SAMBa, University of Bath, United Kingdom*
[2] *Mathematical Sciences, University of Bath, United Kingdom*

Background: Many phase I trial designs involve dosing decisions that put subjects at greater risk than necessary, or dosing an excess of subjects at doses that do not contribute information that improves the estimation of the maximum tolerated dose.

Objectives:

A. Derive an optimal dose escalation scheme for a phase I trial with a binary safety endpoint.

B. Extend the framework to optimise the dose escalation scheme for a phase I trial with two binary endpoints.

Methods: We consider a one-parameter model to describe the relationship between dose and the probability of a dose limiting event occurring and work within a Bayesian framework. We apply the dynamic programming algorithm to obtain a dose escalation scheme that is globally optimal with respect to a loss function that quantifies the aims of the trial.

Dynamic programming requires a set of calculations to be performed for every possible data set at each stage of the trial. Even with a small sample size this state space is large and in particular, it is prohibitively large when it comes to considering phase I trials with a toxicity and an efficacy endpoint. We consider reformulating the state space as the space of posterior density functions for the dose-response model parameter and adapting the dynamic programming algorithm to a sample of this space.

Results and Conclusions: For a phase I trial with a binary toxicity endpoint we can apply dynamic programming to find an optimal dose escalation scheme with respect to a given loss function. Different loss functions lead to different schemes, thus this work provides a flexible framework that can be used to compare different trial designs.

We have created a modified dynamic programming algorithm that operates on a sample of the full state space. This produces a dose escalation scheme that is an approximation to the optimal rule produced by performing dynamic programming on the full space of all possible data sets. With this approximate version of the algorithm, we have extended the framework to tackle the problem of finding an optimal dose escalation scheme for a phase I trial with two endpoints.

[OC31.2]

Equivalence tests for binary efficacy-toxicity responses

Kathrin Moellenhoff[1], Holger Dette[2], Frank Bretz[3]

[1] *Institute of Medical Statistics and Computational Biology, Faculty of Medicine, University of Cologne, Germany*
[2] *Department of Mathematics, Ruhr Universitaet Bochum, Germany*
[3] *Novartis Pharma AG, Switzerland*

Clinical trials often aim to compare a new drug with a reference treatment in terms of efficacy and/or toxicity depending on covariates such as, for example, the dose level of the drug. Equivalence of these treatments can be claimed if the difference in average outcome is below a certain threshold over the covariate range.

In this talk we present a new test procedure for the assessment of equivalence, assuming that the efficacy and toxicity of the treatments are measured as binary outcome variables. We start by using a two-dimensional Gumbel model for describing these bivariate binary (correlated) responses. The equivalence test is based on a parametric bootstrap approach, which generates data under the constraint that the distance between the curves is equal to the pre-specified equivalence threshold. We will explain the new method in detail, investigate its operating characteristics and finally present a case study as an illustration.

[OC31.3]

Designing cluster randomised trials with unequal allocation of clusters or measurements

Andrew Copas

MRC Clinical Trials Unit, University College London, United Kingdom

Background: Researchers may allocate individuals unequally to intervention and control in trials for cost, ethical or logistical reasons, e.g. more to intervention to learn about its implementation or side-effects. Researchers may also reduce sample size by allocating more than half the individuals to one arm if outcomes are more variable in that arm. In cluster randomised trials researchers may consider unequal allocation of clusters and/or measurements, and further statistical reasons to do so include different cluster size (measurements per cluster) or intra-cluster correlation coefficient (ICC) between arms. Interventions can affect the variance and/or ICC, e.g. reductions when the intervention focuses on improvements in poor performing clusters. Previous authors derived the optimal cluster allocation ratio (minimising the number of measurements required) under different variance and/or ICC between arms, for given cluster sizes by arm.

Objectives and methods: We derive simple expressions for the optimal proportions of clusters and measurements for each arm, when the cluster size can be freely chosen e.g. when clusters are communities and we only sample a fraction for measurement. We investigate designs where the ICC or variance differ between arms, and also designs under constraints such as a fixed proportion of clusters or fixed cluster size in one arm.

Results: When the ICC differs between arms but the variance is equal, the optimal design allocates more than half the clusters to the arm with the higher ICC, but slightly less than half the measurements and hence a smaller cluster size. Unequal allocation is also optimal even when the ICC and variance are equal between arms if we fix the cluster size or proportion of clusters in one arm. In realistic scenarios we show savings in the number of measurements required of up to 30% compared to designs with equal allocations to arms.

Conclusions: When cluster sizes can be chosen by the researcher this permits greater flexibility in design and sometimes reductions in the number of measurements. In design researchers need to balance the numbers of measurements and clusters recognising resources and costs, and keep enough clusters in each arm for a robust analysis.

[OC31.4]

Dose-finding designs with right censored endpoints. Evaluation through benchmark.

Anaïs Andrillon, Sylvie Chevret, Lucie Biard

Université de Paris - ECSTRRA Team (Epidemiology and Clinical Statistics for Tumor, Respiratory, and Resuscitation Assessments) INSERM U-1153 - CRESS (Centre of Research in Epidemiology and Statistics), France

Background: To obtain optimal benchmark performances of dose-finding algorithms, several methods based on the simulation of complete information on patient outcomes at all dose levels have been proposed. They have been developed for binary and continuous toxicity outcomes, and possibly efficacy outcomes. They use conventional monotonic dose-toxicity relationships, with binary outcomes mostly measured once (e.g., dose-limiting toxicity or not), after a fixed number of cycles. With new anti-cancer agents such as molecularly-targeted therapies and immunotherapies, late-onset toxicities are usual. Specific trial designs with prolonged observation windows and censored endpoints, analyzed using survival models, appear particularly suited to these settings.

Objectives: (i) To propose a dose-finding designs using survival models for censored endpoints allowing the outcomes to be delayed, and (ii) To develop a benchmark approach for evaluation.

Methods: Complete trial data is generated using a tolerance profile across dose levels: a uniform random variable is drawn for each patient to determine the toxicity outcome at each dose level by inverse-transform sampling on the cumulative incidence of toxicity. Non-parametric estimator is used to obtain estimates of the cumulative incidences of toxicity by dose and eventually select the best dose according to the dose-finding objective of choice. We compared the optimal benchmark performances to those of real trial design data selected from the complete data, and dose finding algorithm of a Continual Reassessment Method (CRM)-like design with censored toxicity endpoint, the Time-to-event (TITE)-CRM and the TITE-Bayesian Optimal Interval Design (BOIN) design. CRM and BOIN designs were adapted to censored endpoints by using one-parameter exponential working models.

Results: We evaluated the proposed methods in a simulation study, notably in terms of correct dose selection. Different scenarios of dose-toxicity relationship were considered, including monotone increasing dose-toxicity relationship and plateau. We found that the benchmark method provides a direct evaluation of the degree of complexity of scenarios (dose-toxicity curve, toxicity hazard shape, and sample size).

Conclusion: We propose a new framework for early dose finding trials with a censored toxicity endpoint. The proposed benchmark method is useful tool in planning trials with late-onset toxicity: it provides an optimal reference performance to compare designs

[OC31.5]

Estimation of the response rate in Simon's two stage design with early termination

Antonios Daletzakis[1], Rutger M. van den Bor[2], Harm van Tinteren[1], Kit Cb Roes[3]

[1] *Biometrics, the Netherlands Cancer Institute, Netherlands*
[2] *Biostatistics, Utrecht Univercity Medical Centre, Netherlands*
[3] *Biostatistics, Radboud University Medical Center, Netherlands*

Background: In early phase clinical studies in oncology, Simon's two-stage designs are widely used. The primary purpose of the design is to decide whether the experimental treatment is promising enough to be subjected to more elaborate evaluation. This design is also used in basket or umbrella designs, that may include many cohorts of patients. Estimation of the response rate in Simon's two stage designs is extensively addressed in the literature. The trial design could be made more efficient by stopping early in the second stage when it is clear that the required number of responses is reached, or it has become clear this cannot be reached any more. This efficiency is relevant in case of multiple cohorts, where reduction in sample size may prevent giving ineffective treatment to patients and resources are saved. Early stopping during the second stage will affect proper estimation of the response rate.

Objective: To derive an unbiased efficient estimator in Simon's two stage designs that allow stopping early in the second stage. We propose an unbiased estimator for the response rate, providing appropriate confidence intervals as well as evaluation of properties.

Method: Classically, response rates are modelled for the first and second stages using a binomial distribution. We used for second stage a „stopped" negative binomial distribution. The estimator is derived analytically using the Rao-Blackwell theorem, conditioning on the unbiased estimator of response for the completed first stage. We evaluated the estimator's properties in terms of bias and MSE, both analytically and by simulations. We derive appropriate confidence intervals based on stochastic orderings. Finally, we investigated expected sample size reduction.

Results: Analytical proof and simulation results shown that the proposed estimator is unbiased and performs good in terms of MSE, compared to alternative estimators. The expected sample size is at least smaller, by 2 patients on average, than the initial Simon's design's under the null hypothesis. In master protocols with many cohorts(100 or more) the savings are therefore highly relevant.

Conclusions: Our method provides a solution for estimating response rates in case of early stopping using a Simon's two-stage design.

OC32: Causal Inference 3

Chair: **Vanessa Didelez**

Leibniz Insitute for Prevention Research and Epidemiology – BIPS, Germany

[OC32.1]

Inference on Time-to-Event Data after Propensity Score Matching

Zhiyue Huang

Beigen, China

Propensity score matching (PSM) is widely used in observational analysis by clinicians because it is intuitive. However, establishing correct asymptotical results for a matched sample is challenging, especially for the analysis of time-to-event data due to censoring. In this work, we aim to provide warnings on current inference methods and propose a new class of test statistics for inference on time-to-event data after PSM. Large sample results are presented and examined through simulation. From our simulation study, our new method provides correct type I error under the null hypothesis, while other commonly used PSM inference methods have inflated type I error.

[OC32.2]

Data-adaptive methods for high-dimensional mediation analysis: Application to a tuberculosis vaccine trial

Margarita Moreno-Betancur[1], Nicole L. Messina[2], Kaya Gardiner[3], Nigel Curtis[4], Stijn Vansteelandt[5]

[1] Clinical Epidemiology and Biostatistics Unit, University of Melbourne and Murdoch Children's Research Institute, Australia
[2] Infectious Diseases Group, University of Melbourne and Murdoch Children's Research Institute, Australia
[3] Infectious Diseases Group, Murdoch Children's Research Institute, Australia
[4] Infectious Diseases Group, University of Melbourne, Murdoch Children's Research Institute and The Royal Children's Hospital Melbourne, Australia
[5] Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium; London School of Hygiene and Tropical Medicine, United Kingdom

Background: Statistical methods for causal mediation analysis are useful for understanding the pathways by which a certain treatment or exposure impacts health outcomes. While there have been many methodological developments in the past decades, there is still a scarcity of data-adaptive methods for mediation analysis with respect to high-dimensional mediators (e.g., biomarkers) and confounders. Existing methods necessitate modelling of the distribution of the mediators, which quickly becomes infeasible when mediators are high-dimensional. This work was motivated by the Melbourne Infant Study: BCG for Allergy and Infection Reduction (MIS BAIR), a randomised controlled trial investigating the effect of neonatal Bacillus Calmette–Guérin (BCG) (tuberculosis) vaccination on clinical allergy and infection outcomes, and its mechanisms of action.

Objective: To propose novel data-adaptive methods for mediation analysis that avoid such high-dimensional modelling, with focus on the context of a randomised treatment.

Methods: We propose novel methods for estimating the indirect effect of a randomised treatment that acts via a pathway represented by a high-dimensional set of measurements. The methods are shown to be doubly robust, which allows us to achieve (uniformly) valid statistical inference, even when machine learning algorithms are used for the two required models.

Results: We confirm adequate performance of the methods in an extensive simulation study and illustrate their application in the context of the MIS BAIR study. The hypothesis of this trial was that the heterologous effects of BCG on innate immunity have beneficial effects on the developing immune system, resulting in improved outcomes. We used the proposed methods to investigate the mediating role of immune pathways, which were represented by a high-dimensional vector of cytokine responses under various stimulants.

Conclusions: The proposed methods provide a feasible and flexible analytic strategy for incorporating high-dimensional mediators in mediation analyses in the context of randomised controlled trials.

[OC32.3]

Estimating causal effects from the large observational, with an application to multiple sclerosis

Camille Nevoret[1], Sandrine Katsahian[2], Agathe Guilloux[1]

[1] *Université Paris-Saclay, CNRS, Univ Evry, Laboratoire de Mathématiques et Modélisation d'Evry, 91037, Evry, France, France*
[2] *Inserm, UMRS 1138, team 22, Centre de Recherche des Cordeliers,, Université de Paris, France*

Background: With the increased availability of large databases of electronic health records (EHRs) comes the chance of evaluating treatment effect on new data sources. The French national health insurance information system (SNIIRAM) is a large observational database (LOD) built around medical reimbursements of more than 65 million individuals – over 98% of the French population.

Multiple sclerosis (MS) is a chronic degenerative inflammatory disease of the central nervous system of autoimmune origin and the leading cause of severe non-traumatic disability in young adults. Different treatments are used to reduce the frequency of relapses and slow the progression of the disease. All first-line treatment initiations were identified in the SNIIRAM to evaluate their efficacy on relapses.

Objective: Estimating the causal effect of a treatment from a LOD is a methodological challenge, because of the lack of randomization. Specific methods based on continuous covariates were developed to perform causal analysis on these data (see in particular Gran et al. 2017 method). However the SNIIRAM database contains only binary data with time stamps, which can be aggregated to counts on time intervals. In addition, their method is restricted to the at most one event situation. We propose a novel algorithm adapted to these longitudinal count data and repeated events.

Methods: We evaluate the average treatment effect on the treated (ATT) by simulating the counterfactual trajectories of patients covariates after treatment initiation using INGARCH models (Liboschik et al. 2015), the parameters of these models having been estimated on the trajectories observed before treatment initiation. In addition, we propose a new ATT estimation algorithm that takes into account estimation errors in the simulation of the counterfactual trajectories.

Results: A simulation study is performed and shows our method to be efficient to evaluate the ATT. We illustrate our method on MS data. Codes and simulations are available on Git lab.

Conclusion: We propose a novel algorithm that allows estimating a causal effect from LOD. We demonstrate how it estimates a causal link between first-line treatment initiations and frequency of relapses in MS.

[OC32.4]

Long-term effect of first-line multiple sclerosis treatments using time-dependent propensity score matching

Mathilde Lefort[1], Yohann Foucher[2], Rémi Lenain[2], Sandra Vukusic[3], Gilles Edan[4], Emmanuelle Leray[1]

[1] *EA 7449 REPERES, EHESP, France*
[2] *INSERM UMR 1246 SPHERE, Nantes University, France*
[3] *Department of Neurology, Lyon University Hospital, France*
[4] *Department of Neurology, Rennes University hospital, France*

Background: Studies of treatment effectiveness need to control indication bias as much as possible. Causal inference methods have evolved, in particular through the development of propensity score (PS) methods or marginal structural models. Their applications can be complex. In particular, using PS, the time of the therapeutic decision is difficult to define when controls are not treated and the exposure is time-dependent. Multiple Sclerosis (MS) is a chronic disease leading to irreversible disability. Beta-interferon (IFN) and glatiramer acetate (GA) are the two oldest first-line treatments. Their long-term effect on MS disability progression has resulted in controversial results, probably due to a lack of control of indication bias.

Objective: To apply a time-dependent PS matching method originally devised by Lu in 2005 to assess the long-term effect of IFN and GA in MS disability.

Methods: This retrospective observational study was based on a series of patients from the MS expert center in Rennes, France. The time-dependent PS was defined as the linear predictor of a Cox model estimating the hazard of being treated at each time from MS onset. The matching procedure resulted in two groups: patients matched as treated or as "not yet treated". The restricted mean times (RMST) to reach a moderate level of disability or a worsening of the disability were compared between the two groups in an intention-to-treat analysis. Bootstrapping was used to handle the randomness of the matching and to obtain confidence intervals.

Results: Of the 2,383 patients included in the study, 556 were matched as treated. The matching procedure provided a good balance of both the time-fixed and the time-dependent covariates. A slight difference was observed for the time to reach a moderate level of disability, in favor of the "not yet treated" group (difference in the RMST: -0.62 [-0.91; -0.33]) while no difference was found in terms of worsening of the disability (-0.03 [-0.24; 0.33]).

Conclusion: This unexpected result is probably due to unmeasured confounders. However, the time-dependent PS is a useful tool that warrants consideration in treatment effectiveness studies in a chronic disease context as it easy to apply and it provides intuitive results.

[OC32.5]

Propensity score diagnostics: assessing the accuracy of a propensity-adjusted effect estimate

Emily Granger[1], Jamie Sergeant[2], Mark Lunt[3]

[1] *Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom*
[2] *Centre for Biostatistics, The University of Manchester, United Kingdom*
[3] *Centre for Epidemiology Versus Arthritis, The University of Manchester, United Kingdom*

Background: Propensity scores are commonly used to deal with confounding bias in observational studies, by balancing covariate distributions between exposure groups. Poorly estimated propensity scores may not achieve adequate balance, leading to biased effect estimates. Consequently, it is essential to assess propensity scores using diagnostics. Unfortunately, there is currently no consensus on the best way to do this.

Objectives:

- Compare the performances of propensity score diagnostics;
- Produce guidelines on how best to build and assess propensity score models.

Methods: Diagnostics were categorised as either (1) individual or (2) overall. Individual diagnostics assess balance in covariates individually, whereas overall diagnostics assess overall balance achieved by propensity scores. Simulation studies were used to compare diagnostics.

Results: (1) Individual diagnostics were compared in terms of their ability to identify different types of model misspecification. Results indicated that diagnostics which work by comparing covariate means (e.g. standardised mean differences) can fail to identify when non-linear terms are misspecified in the propensity score model. Diagnostics which compare entire distributions (e.g. Kolmogorov-Smirnov statistic), performed worst in small samples sizes. The best performing individual diagnostics are new, and involve comparing the number of exposed subjects at each covariate value to that predicted by the propensity score (i.e. cumulative prevalence diagnostics).

(2) Two types of overall diagnostics were considered: weighted averages of balance (different balance metrics and weighting schemes were considered) and the standardised mean difference in prognostic scores (i.e. predicted outcomes under the control condition). Both types include information about the covariate-outcome associations, so that balance on strong predictors of the outcome can be prioritised. Overall diagnostics were compared in terms of their correlation with bias in the propensity-adjusted effect estimate. Results indicated that the standardised mean difference in prognostic scores performed best.

Conclusions: Based on the results above, we developed guidelines for building and assessing propensity score models. Researchers are recommended to use cumulative prevalence diagnostics to check model specification when building propensity score models. Once propensity scores have been estimated and used to balance covariates, prognostic scores can be used to check overall balance in the propensity-adjusted samples.

OC33: Survival Analysis 4

Chair: **Per Kragh Andersen**

Biostatistics, University of Copenhagen, Denmark

[OC33.1]

Statistical Intervals for the MRL Function via Empirical Likelihood Using Right Censored Length-biased Data

Ali Shariati, Hassan Doosti, Justin Wishart

Department of Mathematics and Statistics, Macquarie University

To study the natural history of a disease, survival data collected in a cohort of prevalent cases may be used to make statistical inference on survival functions. In such studies, we only observe subjects who have already been diagnosed with a condition or disease (e.g. HIV, cancer or dementia) but are yet to experience the failure event (e.g. AIDS, or death). Since data collected under this circumstance include a nonrepresentative sample of the target population of interest, statistical procedures and methodology for assessment of interventions or treatments based on such data often involve different bias, the most important of which are right censorship and length-bias.

The mean residual life (MRL) function is one of the most common functions that can be used to study survival data. In this talk, an empirical likelihood procedure is used to draw statistical inference for the MRL function using length-biased right censored data. The limiting distribution of the empirical log-likelihood statistic is derived and used to obtain confidence intervals for the MRL function of the target population. A simulation study is carried out to inspect the finite sample performance of the proposed method. Finally, a set of real data for elderly residents of a retirement care is analysed using the procedures.

[OC33.2]

An Analytic Approach on the Cumulative Mean Function for Recurrent and Terminal Events

Lars Bugiera, Jerome Lambert, Martin Wolkewitz

Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Germany

Background: The cumulative mean function (CMF) corresponds to the mean expected number of events in a population with recurrent events and a competing terminal event. This function can be particularly useful in the setting of hospital-acquired infections (HAI) where recurrent infections may happen.

Objective: We want to illustrate the effect of the terminal hazard on the CMF. Additionally, the number of possibles infection will play a key role to show an approximation for the CMF which can easily be obtained from the data.

Method: Using a simplified homogeneous time-continuous Markov chain with only two different transition functions (hazard rates) we derive an analytic approximation of the CMF. Using simulated data we will show how the hazard rates can affect the shape of the CMF.

Results: We will derive an analytic expression for the CMF which will lead to the approximated form. In that case we will focus on a high number of possible infections.

Additionally, we can derive an explicit closed form for the supremum of the CMF. This value corresponds to the mean number of infections after a long observation time. Using this plateau value we want to emphasize how two opposing situations can lead to similar results, e.g. a covariate can lower the mean number of HAI either directly by lowering the infection hazard or indirectly by increasing the discharge hazard.

Results: As a marginal measure, the effect of a covariate on the mean number of infections can reflect either a direct effect on the risk of getting an infection or an indirect effect on the risk of discharge. To have a clear insight on the process it is useful to take a closer look at the analytical expressions involved. This effect can not only be observed in simulated data and should raise awareness in the interpretation of the CMF. Simplifying the CMF suitably leads to an easily calculated and reproducible value of interest. A particular field of interest lies in clinical epidemiology with a focus on HAIs.

[OC33.3]

Stratifying event-history data by a binary time-dependent covariate without conditioning on the future

Reinhard Meister[1], Tobias Bluhmki[2], Evelin Beck[3]

[1] *FB2 Mathematics, Physics,Chemistrie, Beuth Hochschule Berlin, Germany*
[2] *Metronomia Clinical Research GmbH, Germany*
[3] *Pharmakovigilanz- und Beratungszentrum für Embryonaltoxikologie, Charité, Germany*

Context: For observational studies on drug safety, the relation of drug exposure time-pattern and the outcome is of interest. For a time-dependent covariate, many authors e.g. [1] have described erroneous inference by inappropriate selection of the individuals under risk. We recently demonstrated that length of exposure to a drug in pregnancy is related to the cumulative incidence of spontaneous abortion (SAB) [2] avoiding immortal time bias. We were using multi-state modeling and conditioning on landmarks, as in dynamic prediction. This landmark analysis for subgroups defined by different lengths of exposure intervals, however, could not provide results on the total over the whole length per group. We, therefore, present a way to get such results, without any additional assumptions in a general framework. We condition on so-called waiting-time-landmarks, allowing stratification without conditioning on the future.

Method: We are inspecting the sample space of the bivariate process of exposure state and survival state, both binary, in time. Switching to a representation of this space in the two-dimensional space of waiting time (for a change in exposure) and survival time, we explain how stratification on categorized waiting times generates a partition. The partition defines strata-specific time intervals, i.e. strata-specific risk-sets.

Results: Using those risk-sets for stratified analysis we are avoiding immortal time bias. Two examples from surgery and obstetrics [2] demonstrate applicability and generality of the approach.

Conclusions: Stratification by waiting-time-landmarks of a time-dependent binary covariate is an approach, substantially different from landmarking on survival. In dynamic prediction, landmarks help in understanding a time-dependent influence of covariates, observed either at baseline or at a survival landmark. Our stratification approach is suited for the assessment of overall survival, or characteristics thereof. Applying appropriate time-horizons for waiting-time defined strata, we do not need any other condition to be satisfied. The two approaches complement each other.

References:

1. *S Suissa (2007), Immortal time bias in observational studies of drug effects. Pharmacoepidemiol Drug Saf. 16(3):241-9.*
2. *T Bluhmki et al (2019), Multistate methodology improves risk assessment under time-varying drug intake–a new view on pregnancy outcomes following coumarin exposure. Pharmacoepidemiol Drug Saf. 28(5):616-24*

[OC33.4]

Discordant sampling designs from a longitudinal cohort: efficiency comparison

Francesca Graziano, Paola Rebora, Maria Grazia Valsecchi

Centre of Biostatistics for Clinical Epidemiology, University of Milano - Bicocca, Italy

Background: The availability of large epidemiological or clinical data storing biological samples allow to study the prognostic value of novel biomarkers, but efficient designs are needed for parsimonious and economical reasons. For example, an Italian randomised trial on childhood Acute Lymphoblastic Leukemia (AIEOP-ALL2000) stored biological samples of patients at diagnosis to allow the retrospective evaluation of the prognostic value of novel biomarkers.

Objective: The aim is to develop an efficient design strategy for subsampling from a clinical cohort in order to measure additional variables, such as new biomarkers that could serve as prognostic factors. In particular, we will focus on discordant, concordant and traditional sampling designs to optimally choose the sub-cohort for the measurement of the biomarker.

Method: We adopted a two-phase approach, considering the clinical trial cohort as the first phase and the subsample on which to measure additional variables, such as the biomarker, as the second phase. The performance of discordant sampling, defined as selecting early cases (i.e. relapses) with long surviving controls, concordant sampling, defined as selecting early (late) cases with early (late) censored controls, were investigated through simulations and compared with more traditional sampling, such as the nested case-control design. We investigated various scenarios, e.g. different censoring rates and control:case ratios. To assess the influence of the biomarker on relapse, we applied a Cox model weighted by the inverse of the empirical inclusion probability.

Results: Preliminary simulation results showed the advantage of carefully planning the sub-cohort design. Matching concordant subjects (e.g., early cases with early controls) did not seem to be efficient. The discordant sampling design demonstrated good efficiency and power.

Conclusions: Evaluating all available data before the analysis might be helpful to improve efficiency. Discordant sampling designs seems to be advantageous in term of efficiency.

[OC33.5]

A robust alternative for comparing time-event survival curves: initial simulations and cancer data analysis

Lubomir Stepanek

Department of Biomedical Statistics, Institute of Biophysics and Informatics, First Faculty of Medicine, Charles University, Czech Republic

Background: A situation of comparing two or more time-event survival curves is very common in applied biostatistics. Therefore, several well-established methods could be used. Regarding how many groups are supposed to be compared, either a log-rank test, a score-rank test, a Cox proportional hazards model, or even a Wilcoxon rank-sum test might be performed. However, each one of the described methods has its limitations, and its application is determined by meeting relatively rigorous statistical assumptions.

Objective(s): In this work, we introduce a new robust method for comparing two or more time-event survival curves. Furthermore, we test some of the properties using both high-throughput simulations on time-event data generated in R, and analysis of real survival data of stomach cancer provenience.

Method(s): The robustness of the method is in a nonparametric approach on how to model individual time-event curves, and also in considering the lengths of time intervals between every two consecutive survival events of interest. The next piece of our contribution is in connection with the robust statistical model, and combinatorial modeling of two time-event survival curves and numerical estimating of a surface of an area bounded by these two curves plotted onto a plane chart. The combinatorial modeling of time-event curves helps to avoid the usual hypergeometric distribution of a hazard function for the time-event curves, which makes the test comparing the curves more robust.

Results: The results show that the robust version of a log-rank and score-rank test have similar outputs as the traditional tests, but due to the robust approach behind the method are legit to be applied on small data. Consequently, the robust variant has got less statistical power. The area surface between two time-event corves seems to have interesting analytical properties regarding a p-value of the test.

Conclusions: Based on the simulations and preliminary analytical derivations, the robust version of the log-rank test could be a promising alternative on how to compare time-event curves regardless of any assumptions are met. Besides, the method and theory behind it could also be a topic for the development of a new R package.

OC34: Personalised Medicine 3

Chair: **Shaun Seaman**

University of Cambridge, United Kingdom

[OC34.1]

Predicting real world effectiveness of interventions using randomized and non-randomized data

Michael Seo, Orestis Efthimiou

Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland

Background: There has been growing interest in using data from non-randomized studies (NRS) to complement evidence from randomized controlled trials (RCTs) in medical decision-making. This is because, although RCTs are the best source of evidence regarding relative treatment effects, they often employ strict experimental settings, which may hamper their ability to predict outcomes in 'real-world' clinical settings. Currently, there is a gap in methods for combining individual patient data (IPD) from RCTs and NRS, when aiming to make patient-specific predictions about the real-world effects of medical interventions.

Objectives: Our objectives are (1) to describe a general framework for developing a prediction model that combines IPD from multiple randomized and non-randomized studies, regarding a range of alternative interventions, and (2) to illustrate how to implement this framework in practice, using a real example in rheumatoid arthritis (RA).

Methods: We used IPD from three RCTs and two NRSs involving three treatments for RA. The outcome of interest was Disease Activity Score in 6 months. We developed three generic approaches for building prediction models that combine IPD from studies of variable design: an approach that borrows elements from network meta-regression but does not account for differences in study design (approach I); an extension of approach I, where we also employ shrinkage methods for the regression coefficients related to treatment-covariate interactions (approach II); an approach that uses a differential weighting scheme for studies of different design (approach III). Within each approach, we developed a range of competing prediction models, utilizing different shrinkage methods on the effect of the covariates and exploring various weighting methods. We assessed the performance of the all models by comparing their predictions versus the actual observations, following an internal and an internal-external cross validation approach.

Results and Conclusions: Analyses showed that for the RA example, approach III led to models with better predictive performance. Models that employed shrinkage on the effects of the covariates performed generally better than simpler models. The proposed framework can be used to guide the development and validation of a prediction model that combines data from multiple sources.

[OC34.2]

An evaluation of the c-statistic for benefit

Thomas Debray[1], Orestis Efthimiou[2]

[1] *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Netherlands*
[2] *Institute of Social and Preventive Medicine, University of Bern, Switzerland*

Background: Clinical prediction models aim to facilitate treatment decisions, by providing estimates of absolute risk utilizing individual patient characteristics (such as age, gender and treatments received. Prediction models are usually evaluated for their ability to predict the target outcome. However, widely used measures of performance regarding discrimination and calibration do not quantify a model's ability to predict treatment benefit, i.e. the absolute difference in risk arising due to treatment. For this reason, the concordance-statistic for benefit ('c-for-benefit') has been recently proposed. This employs a procedure where individuals are matched according to their predicted benefit. The statistical properties of c-for-benefit are currently unclear, thus hampering its implementation and interpretation.

Objectives: We aim to assess the statistical properties of c-for-benefit and to propose alternative measures for quantifying a model's ability to predict treatment benefit.

Methods: We explore the potential advantages and limitations of c-for-benefit using theoretical arguments. We then perform a series of simulations, aiming to demonstrate these properties in practice. Hereto, we generate datasets with a binary outcome where the treatment effect is either constant or modified by patient-level covariates. These datasets are then used to develop prediction models that account for the received treatment. Subsequently, we assess the performance of all developed models using the c-for-benefit, and illustrate its properties in real clinical datasets.

Results: In accordance with our theoretical arguments, in many simulated scenarios the estimated c-for-benefit provided a distorted picture of the model's predictive performance. Simulations showed that estimates for c-for-benefit are often close to 0.5 and tend to be inaccurate, even for correctly specified models. Large values for c-for-benefit were only obtained when the interaction effect is very strong and depends on continuous covariates.

Conclusions: The c-for-benefit may be problematic for validating predictions of treatment benefit, unless there are strong interactions between treatment and continuous covariates. For this reason, we recommend alternative measures to quantify model performance.

[OC34.3]

Estimating the common distribution of two right-censored potential treatment responses given a biomarker

Michael Lauseker, Rüdiger Laubender, Ulrich Mansmann

Institute for Medical Information Processing, Biometry, and Epidemiology, LMU, Germany

Background: Laubender et al (BiomJ 2020) introduce the LML model (a trinormal model for two potential treatment responses in parallel group RCTs using a baseline biomarker measurement) and reconstruct the invisible correlation between both responses. Their main assumption is a linear dependence between biomarker and the sum of the potential outcomes. We apply this approach to right-censored log-normal event data by log-transforming the event times and combining the LML model with the EM-algorithm.

Objectives: (1) To assess potential bias in the relevant model parameters; (2) determine their confidence intervals; (3) application to real RCT data; (4) the relevance of log-normal event data in specific clinical settings.

Method: We performed an extensive simulation study. The data was produced by the LML model combined with an independent censoring mechanism. The EM algorithm provided a log-likelihood estimation on the simulated data set based on imputed (multiple) right-censored survival data. Bias and standard errors for the parameters between the full informative (no censoring) and the censored data were calculated.

Results: Sample size (5), correlation between responses and biomarkers (8), percentage of censoring (3) were varied in 120 simulation scenarios. Bias in parameters was given if in both treatment groups correlations between biomarker and event times were similar, by low (<100) group size, or relevant censoring in small groups. Relative efficiency (Varcensored / Varfull) is also influenced by sample size and censoring, but is acceptable in the setting of a typical RCT. Large group size and a high censoring rate is typical for RCTs with chronic myeloid leukaemia patients. Thus, we present an example from this field where age is a relevant biomarker and survival under bone marrow transplantation versus medical therapy (Interferon) are the outcomes. The examples shows several drawbacks of the censored log-LML model caused by small correlation coefficients between age and log-transformed survival.

Conclusions: The simulation studies show good performance of the censored log-LML model in typical RCT settings if normality assumptions are met. The clinical example teaches the need to check these assumptions carefully before applying the proposed model. The Doornik-Hansen test seems to be appropriate for these assessments.

[OC34.4]

Confidence interval estimation for the changepoint of treatment stratification

Bernhard Haller[1], Ulrich Mansmann[2], Dennis Dobler[3], Kurt Ulm[1], Alexander Hapfelmeier[1]

[1] Institute of Medical Informatics, Statistics and Epidemiology, Technical University of Munich, Germany
[2] Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Germany
[3] Department of Mathematics, Vrije Universiteit Amsterdam, Netherlands

Background: The major goal in stratified medicine is to administer the best treatment to each patient. Treatment stratification relies on identification of a covariate-by-treatment interaction, which is often assessed by inclusion of an interaction term between treatment and the covariate of interest in a regression model. For a continuous covariate, the cutpoint of the estimated treatment effect with the line of equivalence between both treatments can be derived from the regression coefficients under common linearity assumptions. We call this cutpoint the estimate for the changepoint of treatment stratification.

Objectives: We investigated different statistical approaches for estimation of a confidence interval for this changepoint. Analytical methods (delta method, Fieller's theorem) as well as resampling methods (normal, percentile-based, wild bootstrap) were compared with regard to confidence interval coverage and width.

Methods: A simulation study considering a time-to-event outcome was performed. Relevant aspects as the number of included patients, the proportion of censored observations, and the location of the true changepoint were varied. Under each scenario, 2000 datasets were generated. Proportions of estimated confidence intervals that covered the true changepoint and distributions of confidence interval widths were derived for each method.

Results: Observed coverage was close to 95% for all methods when the number of events was large. For small and moderate event numbers, coverage of the delta method varied relevantly depending on the location of the true changepoint. When Fieller's theorem was applied, coverage proportions were close to 95% for all scenarios, but a relevant proportion of infinitely wide confidence intervals was obtained when the number of observed events was small. While the percentile-based bootstrap was slightly conservative for most scenarios, the normal bootstrap did not provide acceptable coverage. The wild bootstrap performed similar to the delta method. All methods were also applied to data of the randomized SPACE trial.

Conclusions: After identification of a qualitative covariate-by-treatment interaction for a continuous covariate, the estimated changepoint should be presented with a corresponding confidence interval. In our simulation study, the estimator based on Fieller's theorem provided the most reliable results, irrespective of the number of observed events and the location of the true changepoint.

[OC34.5]

Statistical methods to assess dynamic treatment regimens in observational settings: a scoping review

Myra McGuinness[1], Robert Maher[1], Bibhas Chakraborty[2], John Carlin[1], Maarten I. Jzerman[3], Julie Simpson[1]

[1] *Centre for Epidemiology and Biostatistics, University of Melbourne, Australia*
[2] *Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore*
[3] *Centre for Health Policy, University of Melbourne, Australia*

Background: Repeated treatment decisions are required for patients with chronic diseases and often depend on the patient's response to current treatment. Optimal treatment pathways may be investigated through analysis of observational data from cohort studies, electronic health records and clinical registries. Marginal structural models, Q-learning, the parametric G-formula and G-estimation are examples of statistical methods available to investigate the outcomes of multistage decision problems which are known as dynamic treatment regimens (DTRs). However, to date, it has been unclear to what extent these methods have been applied to inform clinical decision making.

Objectives: To perform a scoping review with narrative synthesis to investigate the statistical methodology employed to assess DTRs in observational settings, and to investigate the application of these methods to inform clinical decisions.

Methods: Inclusion criteria included multistage human studies of non-randomised interventions published in English. Abstracts and titles from studies identified through a search of PubMed on 11 April 2019 were screened and potential papers were read in full and information on clinical area, data source, statistical methodology and clinical relevance was extracted. Two additional authors independently extracted data from a subset of papers to confirm eligibility and completeness.

Results and conclusions: From 165 abstracts initially identified, 33 full-text papers published between 2001 and 2019 were included in this review. The majority of papers assessed time-to-event outcomes (n=19, 58%) and for estimation of the DTR effect the most common method used was inverse probability weighting (n=17, 51.5%). Only a third included the statistical computing code used to perform their analyses (n=11). Over a third of studies used registry data (n=12, 36.4%) and a third applied the methods to HIV care-related decisions (n=11, 33.3%). However, only 2 papers were aimed primarily at informing clinical practice; the remaining papers used observational data solely to illustrate statistical methodology. These findings suggest that existing clinical databases and registries of patients with chronic conditions have not been used to full advantage. Promotion of the relatively recent and sophisticated statistical methods for investigation of DTRs has the potential to improve patient care by contributing to evidence-based recommendations for repeated treatment decisions.

OC35: Missing Data 2

Chair: **Havi Murad**

Gertner Institute for Epidemiology and Health Research Policy, Israel

[OC35.1]

Methods to deal with values under limit of detection or quantification

Marie-Cecile Fournier, Clémence le Gal, Corinne Maudet, Sébastien Leuillet, Florence Gillaizeau

44, Biofortis Mérieux NutriSciences, France

Background: In clinical trials, quantitative data (as biological or environmental data for example) can be reported as not detected or not quantified (Limit Of Detection - LOD - and Limit of Quantification - LOQ - respectively). These data which are left-censored rise difficulties in statistical analysis and require special consideration. For sake of simplicity we take the example of LOQ in the following abstract but all this work can be applied to LOD in the same way.

Objective: The goal of this work is to compare different statistical methods of handling values reported as not quantified in clinical trials using real and simulated datasets.

Method: Data came from two clinical trials previously analyzed by our company (BIOFORTIS Mérieux NutriSciences) in the presence of left-censored data. The censorship rates studied were 9%, 24% and 60%. The statistical methods tested were: simple imputation, multiple imputation, maximum likelihood modeling, the Kaplan-Meier method and the regression method on order statistics. A simulation study by resampling was also conducted from real datasets. Methods were compared in terms of α risk and statistical power (1-β).

Results: This methodological work showed that results varied according to the method used. The simulation study showed that results from maximum likelihood and substitution methods by LOQ/2 and LOQ/square root(2) were closed and generally the most efficient. When the censorship rate was high, no statistical method demonstrated efficacy.

Conclusion: The presence of values reported as not quantified requires a special attention. In view of results, the LOQ/2 substitution method, currently used in our routine analyses seems appropriate when the censorship rate is weak (lower than 25%).

[OC35.2] **StCA Winner**

Performance of Inverse Probability Weighting Compared to Multiple Imputation for Missing Binary Outcome in CRT

Sabbir Ahmed Hemo, Anower Hossain

Department of Applied Statistics, Institute of Statistical Research and Training, University of Dhaka, Bangladesh

Background: Missing outcomes are a very common threat in a cluster-randomized (CRTs) trial which may cause biased and inefficient inference if handled inappropriately or ignored. Two common approaches to handle missing outcomes in CRTs are multiple imputation (MI) and inverse probability weighting (IPW). Recent studies recommended IPW as an alternative to MI for handling missing outcomes when outcomes are missing depending on baseline covariate values. However, the comparison was made assuming that the intervention groups have the same missingness mechanisms and the same covariate effects, which might be a very strong assumption to hold in practice.

Objective: This study investigates the performance of IPW through a simulation study when missingness mechanisms and/or covariate effects are different between the intervention groups and then compare with the performance of MI.

Method: We considered individual-level analysis using generalized estimating equations (GEE) and random effects logistic regression (RELR).

Results: The simulation study shows that IPW for GEE gave small bias when the number of clusters is small and missingness mechanisms and covariates effects are different between the intervention groups. IPW for RELR gives biased estimates and lower coverage when missingness mechanism and covariates effects are different between the intervention groups. This study also reveals that IPW and MI perform similarly in GEE regardless of missingness mechanism and the covariate effects are different or same between the intervention groups.

Conclusion: However, RELR with MI performs better compare to RELR with IPW when the covariate effects are different between the intervention groups.

[OC35.3]

Missing covariates for competing risks models: an evaluation of available imputation methods

Edouard Bonneville[1], Matthieu Resche-Rigon[2], Liesbeth de Wreede[1], Hein Putter[1]

[1] *Medical Statistics, Leiden University Medical Center, Netherlands*
[2] *Biostatistics and medical information, Paris Diderot University, France*

Background: Multiple imputation by chained equations (MICE) has been widely used to impute missing values. Bias in the estimation of key parameters can be avoided under the missing-at-random assumption, with the additional condition that the imputation model used is compatible with the analysis model. In the case of Cox proportional hazards models, White and Royston (2009) showed that the imputation models for each covariate must include the event indicator, and the cumulative hazard as estimated by the Nelson-Aalen estimator. Since this is an approximation of the true model, its performance is not always optimal.

Objectives: The first aim of the study was to extend the work of White and Royston for cause-specific competing risks Cox models. In this framework, the second aim was to compare the performance of MICE with a variant thereof called 'substantive model compatible fully conditional specification' (smcfcs, Bartlett and Taylor 2016) - both in terms of bias in the estimated parameters, and in the predicted cumulative incidences for reference patients.

Methods: A large-scale simulation study was performed, with scenarios based on observational data describing outcomes after hematopoietic stem cell transplantation, where the common interest is in the effect of covariates on the competing events relapse and non-relapse mortality. The methods compared were complete-case analysis, smcfcs, and MICE with different imputation models. These were compared under different missingness mechanisms, including outcome-dependent missingness.

Results: In line with previous findings, substantial bias towards the null was observed when using MICE, particularly in the presence of large covariate effects. However, the magnitude of this bias depended on both the proportion of missing values, and the strength of the missingness mechanism. Furthermore, with competing risks, this bias also effects the estimates of the completely observed covariates, in possibly different ways for each cause of failure. Comparatively, the use of smcfcs yielded estimates with reduced bias.

Conclusions: The results of the simulation study represent a call for caution regarding the use of MICE to impute missing covariate values in competing risks analyses. Even under reasonable assumptions, MICE may still yield biased estimates. In light of this, the smcfcs approach may offer an improved alternative.

[OC35.4]

Clustering with missing values: what about congeniality?

Vincent Audigier[1], Niang Ndeye[1], Matthieu Resche-Rigon[2]

[1] *Laboratoire Cedric MSDMA, CNAM, Paris, France*
[2] *Université de Paris, INSERM U1153, ECSTRRA Team, France*

Background: Cluster analysis is commonly applied in medical research for exploration of large observational data sets. However, missing values are often a challenging task for such an analysis. Many multiple imputation methods have been proposed to tackle efficiently the missing data issue in a supervised framework for parametric models. Rubin's rules can be used for achieving this goal. However, missing values remain a recent field of research in the unsupervised framework for distance based cluster analysis. Previous works on clustering with missing values mainly discussed how to pool several partitions obtained from imputed tables notably using consensus method (Faucheux, 2019; Bruckers, 2017), but do not focus on the main issue of congeniality (Meng, 1994).

Objective: We developed new approaches aiming to handle congeniality issues and conducted a large simulation study to highlight how the clustering method is sensitive to the imputation models used. Data sets are simulated according to a mixture model and vary by the number of variables, the number of individuals, the number of classes and the correlation between variables.

Methods: Two clustering methods (k-means, hierarchical clustering) and several imputation models are investigated: 1) two new approaches handling the clustering structure at the imputation step according to congeniality arguments: multiple imputation by latent class models, fully conditional specification with a class indicator determined within the chained equation process 2) three existing approaches: multiple imputation ignoring the clustering structure as benchmark (Schafer, 1997), multiple imputation by k nearest neighbours (Kowarik, 2016), multiple imputation by principal component analysis (V. Audigier, 2016). Partitions obtained from each imputed table are pooled by consensus method (Strehl, 2002). Finally, pooled partitions are assessed by the misclassification rate. An illustration is given using data from an observational cohort data of 120 patients with Epstein-Barr Virus (EBV) reactivations after allograft of hematopoietic cells.

Results: Our study highlights that ignoring the cluster structure in the imputation model is uncongenial with cluster analysis. We provide guidelines for the imputation model choice according to the data structure and the clustering method wished.

OC36: Dynamic Prediction

Chair: **Eleni-Rosalina Andrinopoulou**
Erasmus MC, Netherlands

[OC36.1]

Minimizing the Burden of Invasive Procedures via Personalized Scheduling

Dimitris Rizopoulos[1], Anirudh Tomer[1], Ewout Steyerberg[2]

[1] *Biostatistics, Erasmus University Medical Center, Netherlands*
[2] *Biomedical Data Sciences, Leiden University Medical Center, Netherlands*

Background: Invasive procedures are typically used for confirming disease progression in early-stage chronic non-communicable disease patients (e.g., cancer, cardiovascular). Our work is motivated by prostate cancer patients with low-grade tumors who are routinely checked for progression via biopsies in active surveillance programs. The current standards require all patients to be monitored at exactly the same intervals. However, this imposes patients who progress slowly to high burden resulting in low compliance rates.

Objective: We aim to minimize patient burden an improve compliance by personalizing the schedule of such invasive procedures.

Methods: We use the framework of joint models for longitudinal and survival data to investigate and quantify the association between longitudinally measured biomarkers and the risk of progression. From the fitted joint model, we derive time-dynamic individualized risk predictions of progression. We propose to schedule an invasive procedure when the individualized risk of a patient exceeds a particular patient- and time-specific threshold value. To facilitate the choice of the dynamic threshold in a shared decision-making manner we translating its value into clinically relevant outcomes, namely, the expected number of future invasive procedures, and the expected delay in finding progression.

Results: We use data from the world's largest prostate cancer surveillance study PRIAS. Based on the joint model fitted to PRIAS, we have compared our method of personalized schedules versus the current standards of annual and 3-year interval schedules. Our approach results in much fewer biopsies for patients who progress slowly without compromising the effectiveness of surveillance in patients who progress fast. A shiny application implements our approach in practice.

[OC36.2]

Analysis of dynamic restricted mean survival time for right-censored data

Zijing Yang[1], Yawen Hou[2], Zheng Chen[1]

[1] *Public Health, Southern Medical University, China*
[2] *Statistics, Jinan University, China*

Background: For a study with a right-censored time-to-event outcome, the restricted mean survival time (RMST) has great potential as a meaningful and sensitive outcome measure. There are some advantages of RMST over the hazard ratio (HR), such as simple and straightforward interpretation and requires no particular structural assumptions (e.g. the proportional hazards assumption). However, the RMST is the mean survival time of all subjects in the study population followed up to a given time horizon (w), and cannot reflect how the prognosis of patients changes over time.

Objective: In this article we consider the use of dynamic or conditional RMST as an index to reflect the restricted mean survival time of a patient for further w years, given that he/she has already survived s years.

Methods: The proposed dynamic RMST is calculated based on pseudo-observations. Furthermore, hypothesis testing for the difference between two groups which is extended to regression analysis as well as the landmark prediction model are also developed Monte-Carlo simulation studies were conducted to examine the performance of the proposed methods with different combinations.

Results: Simulation results show that the proposed test performs well even under relatively heavy censoring and is robust in different scenarios (that is, the type I errors are all close to 0.05 as expected). These methods are applied to analyze several data sets, and the results show the proposed methods can effectively offer more scientific prognostic information for patients who have survived initial s years.

Conclusions: This developed dynamic RMST enables intuitive interpretations and can be used to make better individualized treatment decisions based on a dynamic assessment of the patient's prognosis.

[OC36.3]

Proposal of Clinical Dynamic Prediction Models to Predict Prognosis of Post-Treatment Prostate Cancer Patients

Harry Parr[1], Nuria Porta[1], Alison Tree[2], David Dearnaley[3], Emma Hall[1]

[1] *The Institute of Cancer Research, Clinical Trials and Statistics Unit (ICR-CTSU), United Kingdom*
[2] *Royal Marsden NHS Foundation Trust, The Institute of Cancer Research, United Kingdom*
[3] *The Institute of Cancer Research, Royal Marsden NHS Foundation Trust, United Kingdom*

Background: Prostate Cancer (PCa) is prevalent; it is estimated 1-in-6 men will be diagnosed during their lifetime and it has become the most common cancer in the UK. At diagnosis, patients present high values of prostate-specific antigen (PSA), a protein secreted by the prostate. Localised PCa is typically treated with hormone therapy (HT) neoadjuvantly and concurrently with radiotherapy (RT); this reduces PSA concentrations, inhibits testosterone, and oblates the tumoural tissue. A sharp-rise in post-treatment PSA indicates recurrence of PCa. Monitoring PSA trajectories over time provides vital information to inform clinicians of when a reappearance of cancer is likely to occur. Risk of recurrence is usually predicted from known prognostic factors (such as Gleason grade, tumour stage, risk group, and age) at diagnosis. Incorporating PSA trajectories over time will improve predictions and aid clinicians in creating a personalised post-treatment follow-up plan.

Objectives: We therefore propose a clinical dynamic prediction model to characterise prognosis of localised prostate cancer patients after their initial therapy.

Methods: Dynamic predictions are obtained with data from the CHHiP trial (N=3,071). The proposed shared-parameter joint models (JMs) predict participants' prognosis, using both a mixed-effects model to capture the nonlinear dynamics of the longitudinal PSA concentrations, and a relative-risk model for the risk of recurrence. The nonlinearity of PSA is captured by fitting cubic splines with pairwise interactions, the JM parameters are estimated using a Bayesian framework. Several association structures between the longitudinal biomarker and endpoint are fitted. Deaths from non-PCa causes are treated as competing risks. Internal validation procedures are used to evaluate the dynamic predictions and optimism.

Results: For risk of recurrence, the best association structures were a linear combination of the value-gradient-area, time-varying, and the shared-random effect structure with 4 internal knots respectively. Including follow-up PSAs improve dynamic predictive discrimination and calibration, each ½-year additional landmark corresponding between a 5-11% increase in discrimination at five-years.

Conclusions: With promising preliminary results, further external validation is required from other clinical trials, to show the model is suitable for use within current clinical treatment pathways with the intention of confirming those patients at risk and direct salvation care.

[OC36.4]

Joint modeling vs discriminant analysis for dynamic prediction based on longitudinal data: A simulation study

Rana Dandis, Joanna IntHout, Steven Teerenstra

Radboud Institute For Health Science, Radboud University Medical Center, Netherlands

Background: In literature, much emphasis has been placed on new statistical methods for dynamic risk prediction using a biomarker's longitudinal history, however few studies have compared their performance for predicting a long-term fixed dichotomous outcome. Therefore, our objective is to compare the dynamic predictive performance of two common methods namely: Longitudinal discriminant analysis (LoDA) and joint modeling for longitudinal and binary data (JMB), and to explore in which situations each of the two approaches is expected to give the best prediction.

Methods: We conducted an extensive simulation study by using a combination of different event rates, biomarker's within- and /or between-subjects variability informed by a real-world dataset. We also varied the amount of variability across the two diagnostic groups, i.e. the between- and/or within- subjects variability is larger in one of the groups compared to the other. Time-dependent predictive measures (area under ROC curve, mean squared error of prediction and misclassification rate) were used to evaluate the dynamic prediction performance.

Results: Our results showed that LoDA performed better than JMB in most of the simulation scenarios. In general, increasing the biomarker's between-subject variability reduced the predictive accuracy of both approaches to the same extent. Surprisingly, the increase in the number of events did not improve prediction accuracy of both methods. The predictive performance of LoDA was especially better than JMB when the biomarker's (within- and/or between-subject) variability was different between the groups.

Conclusion: Based on the obtained results we recommend using LoDA over JMB in situations when the between subject variability is high or when the variability of the biomarker differs among the groups.

[OC36.5]

A statistical methodology for data-driven partitioning of infectious disease incidence into age-groups

Rami Yaari[1], Amit Huppert[1], Itai Dattner[2]

[1] *Biostatistics and Biomathematics, Gertner institute, Israel*
[2] *Statistics, University of Haifa, Israel*

Understanding age-group dynamics of infectious diseases is a fundamental issue for both scientific study and policymaking. Age-structure epidemic models were developed in order to study and improve our understanding of these dynamics. By fitting the models to incidence data of real outbreaks one can infer estimates of key epidemiological parameters. However, estimation of the transmission in an age-structured populations requires first to define the age-groups of interest. Misspecification in representing the heterogeneity in the age-dependent transmission rates can potentially lead to biased estimation of parameters. We develop the first statistical, data-driven methodology for deciding on the best partition of incidence data into age-groups. The method employs a top-down hierarchical partitioning algorithm, with a metric distance built for maximizing mathematical identifiability of the transmission matrix, and a stopping criteria based on significance testing. The methodology is tested using simulations showing good statistical properties. The methodology is then applied to influenza incidence data of 14 seasons in order to extract the significant age-group clusters in each season.

OC37: Clinical Trials Methodology 6

Chair: **Andrew Copas**

MRC Clinical Trials Unit at University College London, United Kingdom

[OC37.1]

The impact of misclassification in a clinical trial with discrete longitudinal outcomes

Isabelle Smith[1], Linda Sharples[2], Jane Nixon[1]

[1] *Clinical Trials Research Unit, University of Leeds, United Kingdom*
[2] *London School of Hygiene and Tropical Medicine, United Kingdom*

Background: Pressure ulcers (PUs) are painful and debilitating for patients and use substantial healthcare resources. As they develop, skin goes through a series of 'discrete' states from normal skin to severe PU. Prevention studies monitor development of PUs over time, resulting in panel data representing transitions between discrete states. A review of PU prevention research showed that primary endpoints focussed on incidence or time to incidence of a PU, which fails to make use of longitudinal data. Simulation studies demonstrated that multi-state models lead to increased power compared to binary or time to event analysis methods and it was possible to increase the assessment intervals and/or reduce length of follow-up, whilst maintaining adequate power by analysing the longitudinal data. However, this assumed no misclassification of skin state. If we use non-specialist assessors rather than experts, we may introduce misclassification due to the subjective nature of assessment but could compensate by increasing frequency of assessments.

Objectives: To quantify the impact of misclassification on the power, coverage and treatment effect estimates from a clinical trial.

Methods: Data were simulated from a 4-state model of PU development, with misclassification informed by data from previous studies. Parameters that varied were; the assessment interval, length of follow-up, and extent of misclassification. In each case, a Markov model was fitted to the 'true' data, a hidden Markov model assuming misclassification was fitted to the 'observed' data and a Markov model was fitted to the 'observed' data, assuming no misclassification.

Results: Ignoring misclassification reduced power by up to 10%, coverage was as low as 8.8% and led to biased treatment effect estimates. In one scenario, the model ignoring misclassification concluded a harmful treatment effect despite the true treatment effect being beneficial. In the scenarios considered, Hidden Markov models accounting for misclassification had reasonable power, led to unbiased treatment effect estimates and coverage was maintained at 95%.

Conclusions: Non-specialist staff may be used in PU prevention trials providing misclassification is accommodated in design and analysis. These methods are applicable to any progressive disease where data are collected longitudinally using a discrete outcome measure, that is subject to misclassification.

[OC37.2]

Adaptive design with interim decision for single arm trial with external controls or randomized trial

Heiko Goette[1], Marietta Kirchner[2], Johannes Krisam[2], Arthur Allignol[1], Francois-Xavier Lamy[1], Armin Schueler[1], Meinhard Kieser[2]

[1] *GBEM, Merck Healthcare KGaA, Germany*
[2] *Institute of Medical Biometry and Informatics, University of Heidelberg, Germany*

Background: In early clinical development, where only few drug candidates have the potential for market approval, the gold standard, a randomized controlled trial (RCT), might not be the most efficient option for ethical and economic reasons. The required trial sample size could be reduced as the control treatment was already previously investigated. Therefore, a single arm trial with external controls (SATwEC) could be a pragmatic alternative. Using individual patient data, propensity score methods could provide unbiased comparisons between the experimental drug and the external control if all relevant confounders are observable. However, there is a practical challenge: the propensity score methods are applied at the end of the trial. If the resulting propensity score distributions are very different, the SATwEC might lead to inappropriate decisions for further development or require a subsequent RCT before going into late development.

Objectives: We aim to develop and assess through simulations an adaptive design that reduces the risk of unreliable decision making at the end of a SATwEC.

Methods: In stage I of the adaptive design, subjects are only assigned to the experimental arm. If the propensity score distributions at interim are comparable based on the preference score (Walker et al. 2013), further subjects in stage II will also be assigned to the experimental arm. If not, a randomized stage II will be conducted. Futility stop at interim is an additional design option.

In the simulation study, data is generated using a time-to-event model with observable as well as unobservable confounders. The confounder space will be modified to investigate the impact on treatment effect estimates and decision making.

Results: With comparable confounder spaces, the SATwEC is the best approach in balancing correct decisions and sample size. With strong deviations, the randomized design is preferable. The proposed adaptive design provides a compromise as being less prone to wrong conclusion in those "strong deviations"-settings than the SATwEC design while requiring less investment than a RCT.

Conclusions: The proposed adaptive design is a viable and pragmatic option for early clinical development if external control data are available for an interim analysis.

[OC37.3]

Extending the updating algorithm for the Graphical Approach

Cornelia Ursula Kunz

Global Methodology Statistics, Boehringer Ingelheim Pharma GmbH & Co.KG, Germany

Background: Due to increasingly complex trials, multiple testing problems are more and more often encountered in clinical trials. If not accounted for, multiple testing leads to inflation of the type I error rate. Several approaches exists to address this problem. Among them are the Graphical Approach and the Truncated Holm procedure.

Objectives: To extend the existing Graphical Approach allowing more flexible updating algorithms for the weights of the directed edges between the vertices.

Methods: Based on the truncated Holm procedure we show that the updating algorithm of the Graphical Approach is not uniquely defined and other updating algorithms for the weights can be found as long as certain criteria are fulfilled.

Results: We show that the truncated Holm procedure can be represented by a graph. However, the updating of the weights for the edges differs from the algorithm proposed by Bretz et al. [1]. When applied to a setting with two families of hypothesis, we show that the Graphical approach does not just define how much of the significance level is transferred from one hypothesis to the hypothesis connected via the edge but also how much of the significance level is kept within the primary family and how much is transferred to the secondary family. Comparing the results to the truncated Holm procedure, the Graphical Approach allocates more alpha to the secondary family once a hypothesis of the primary family is rejected as compared to the truncated Holm procedure. We also show that other updating algorithms can be defined allowing the specification of the fraction of the alpha level that is passed form family one to family two for each step of the algorithm.

Conclusion: While the Graphical Approach offers a neat solution to multiple testing problems, it can be extended so that more flexible updating algorithms can be defined allowing further tailoring the multiple testing procedures to the specific trial objective.

Reference:

1. *Bretz F, Maurer W, Brannath W, Posch M (2009): A graphical approach to sequentially rejective multiple test procedures. Statistics in Medicine, 28:586-604.*

[OC37.4]

EPAD: The European prevention of Alzheimer's dementia platform trial

Philip Hougaard

Biometrics, Lundbeck, Denmark

Background: Drug development trials have demanded more and more resources over time, in terms of patients, investigator time, calendar time and money. This is particularly a problem in Alzheimer's dementia, where prevention studies take many years and require many patients in order to demonstrate a delay in the disease development. Within oncology, one has started using more complex trial designs, also covered by the term master protocols. Time is now ripe for expanding these designs into other disease areas.

Objective: To create a proof-of-concept platform trial within Alzheimer's dementia.

Methods: Complex trials challenge the classical viewpoint that a clinical trial studies a single drug in a single disease. One such type of design is platform trials, also known as a trial machine, where multiple drugs are tested in the same disease, but in such a way that treatments enter when they become available and exit when enough knowledge has gathered, showing either futility or efficacy. Besides the operational efficiency, a platform trial allows for sharing the placebo group, which reduces the number of patients receiving an inactive treatment.

Results and Conclusions: This will be illustrated with the EPAD project. EPAD is a public-private partnership running two trials. The first trial studies untreated healthy participants. This study serves as feed-in to the second study, which is a proof-of-concept platform trial for treatments that are supposed to prevent the development of Alzheimer's disease. The trial as such is ready to start, but no new investigational drugs have entered yet, possibly due to the many failed projects in this indication.

[OC37.5]

Phase II trial with a trinomial efficacy-toxicity endpoint

Guillaume Mulier, Lucie Biard

2 rue d Alesia 75014 Paris, Inserm DR P5, France

Background: Immunotherapy, a new class of therapeutics in cancer, is associated with late-onset effects, both in toxicity and efficacy scales, compared to conventional cytotoxic chemotherapies. New designs combining efficacy and toxicity assessments are gaining popularity in the early evaluation of these innovating therapeutics. To that aim, a single-arm time-to-event Bayesian phase II optimal design was recently developed, weighting the likelihood on the available follow-ups at each interim analysis [1]. This design is particularly adapted to prolonged observation windows, since it can handle incomplete observations at each interim analysis, therefore limiting interruptions in patient accrual and gaining efficiency in the trial management.

This design also allows multiple endpoints using Dirichlet modeling. Specifically, the design proposes to combine response and toxicity outcomes in a 4-category endpoint, as follows: "response without toxicity", "response and toxicity", "neither response nor toxicity" and "toxicity without response". Nevertheless, in some clinical settings, for instance with frail cancer patients, the occurrence of toxicity directly rules out the candidate treatment for the patient, who usually discontinues the trial; the "response and toxicity" cannot be considered (either not observed or considered a failure).

Objective: To evaluate the operating characteristics of a modified phase II Bayesian optimal design with right-censored observations, applied to a trinomial endpoint ("response without toxicity", "toxicity with or without response", "neither response nor toxicity").

Methods: The design was adapted to the trinomial endpoint, modeled by a Dirichlet distribution. Interim stopping boundaries for futility and toxicity are defined using Bayesian inference and the posterior distribution of model-parameters. We performed a simulation study to assess the operating characteristics of the modified approach, compared to the original design.

Results: The modified design with the trinomial endpoint showed similar type I and II error rates compared to the original design, in different scenarios, varying the true endpoint distribution, the duration of the observation windows, the desired thresholds for decision rules.

Conclusion: The modified design has acceptable operating characteristics and can be implemented for phase II oncology trials where a trinomial patient outcome is preferred, with toxicity strictly defining a treatment failure.

Reference:

1. *Lin et al. J Natl Cancer Inst. 2020;112(1):38-45*

OC38: Causal Inference 4

Chair: **Bianca L. De Stavola**

University College London, United Kingdom

[OC38.1]

G-computation: a robust alternative for drawing causal inference when the positivity assumption does not hold

Maxime Léger[1], Arthur Chatton[2], Florent le Borgne[3], Yohann Foucher[4]

[1] *Anesthesia and Intensive Care, Angers University Hospital, France*
[2] *INSERM UMR 1246 - SPHERE, Nantes University, France*
[3] *A2COM-IDBC, France*
[4] *Nantes University, France*

Background: G-computation (GC), Inverse Probability Weighting (IPW), and Targeted Maximum Likelihood Estimator (TMLE) are methods for drawing a causal inference. Among the different assumptions, the adherence to the positivity assumption is essential for their valid use. Each subject should have a non-negligible probability to be exposed or unexposed.

Objective: We aimed to compare the robustness of GC, IPW, and TMLE when positivity does not hold, by using simulations and two applications on real-life datasets associated with a violation of the positivity.

Methods: We simulated different scenarios with a progressively expanding lack of positivity. Because the GC is based on the prediction of the outcome, it is theoretically robust to the positivity violation when the extrapolations for the corresponding patients are well-calibrated. Therefore, we performed additional scenarios with extrapolation issues. In all the comparisons, the estimand of interest was the average treatment effect in the entire population on a binary outcome.

Results: The mean absolute biases tended to increase for IPW and TMLE with the lack of positivity. However, estimates remained acceptable for the IPW and TMLE. The biases were lower for GC in the situations where the extrapolation issue was missing or moderate. The lowest variance estimation bias was observed for GC, except for the most caricatural scenarios where the extrapolation issues were very important, even non-realistic.

For the two applications, we computed the previous methods on the entire samples (where the positivity does not hold) and the corresponding subsamples of patients with no positivity issue. The GC was the most robust method, i.e., with the lowest differences between the results obtained in the entire samples and the corresponding subsamples. Additionally, GC displayed the lowest variance.

Conclusion: GC appeared to be more robust when positivity does not hold, even for a moderate extrapolation issue.

[OC38.2]

Correcting for selection bias when comparing separate treated and untreated cohorts

Rik van Eekelen[1], Nan van Geloven[2]

[1] *Centre for Reproductive Medicine, Amsterdam UMC, Academic Medical Centre, Netherlands*
[2] *Department of Biomedical Data Sciences, Leiden University Medical Center, Netherlands*

Background: we study the situation when separate cohorts of treated and untreated patients are compared and where the follow up period of untreated patients precedes the follow up period of those who were treated. For instance, this occurs when patients are untreated right after diagnosis and start treatment after waiting time. This comparison is very prone to selection bias since only the 'survivors', the patients that not yet experienced the event of interest during the untreated period, end up being treated. Even after correcting for all confounders, i.e., factors related to both the treatment decision and the outcome, selection bias will remain due to differences in prognosis between patients that clinicians are unaware of (unobserved heterogeneity), for example genes. These factors do not directly influence treatment decisions, but do become associated with treatment as time passes.

Objective: to compare different ways of dealing with this selection bias.

Methods: comparisons included: 1) including the time between diagnosis and start of treatment as a covariate, 2) including the time-updated untreated prognosis as a covariate or 3) including follow up of controls only from the average moment of treatment start. We compared these methods first in a simulation study and second in a prospective cohort of 1896 subfertile couples who were followed for time to pregnancy, first without treatment and later during insemination treatment. We restructured the data to mimic separate treatment and control datasets and compared results to those found in the whole cohort.

Results: preliminary results showed that when introducing heterogeneity by drawing individual outcome probabilities from a beta distribution, the option utilising time-updated untreated prognoses yielded least bias. When the heterogeneity was introduced by simulating an unobserved binary prognostic factor, including restricted cubic splines for time between diagnosis and treatment yielded least bias.

In the application to clinical data, both correcting using including time-updated predictions and a linear term for waiting time reduced the selection bias by approximately 50% on log-hazard scale.

Conclusion: our proposed methods reduce selection bias in treatment effects estimated from separate treatment and control databases. In our application, approximately half of the selection bias was eliminated.

[OC38.3]

Quantitative bias analysis for a misclassified confounder: marginal structural models vs conditional models

Linda Nab[1], Rolf H. H. Groenwold[1], Maarten van Smeden[1], Ruth H. Keogh[2]

[1] *Clinical Epidemiology, Leiden University Medical Center, Netherlands*
[2] *Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom*

Background: Observational data are increasingly used to estimate causal effects of treatments, through careful control for confounding by indication. The average treatment effect (ATE) estimator may be biased if a confounder is imperfectly measured.

Objective: We quantify the bias due to classification error in a dichotomous confounder in analyses that aim to estimate the ATE using a marginal structural model estimated using inverse probability weighting (MSM-IPW) and compare this with the bias in a conditional regression model. We focus on a point-treatment study with a continuous outcome.

Methods: Using the potential outcome framework, expressions were derived for the bias in the ATE estimator from MSM-IPW and a conditional regression model. Based on these expressions, we propose a quantitative bias analysis that enables researchers to study the impact of bias under different scenarios. A simulation study was conducted to study the finite sample performance of MSM-IPW and conditional models if a confounder is misclassified. The methods are illustrated through a quantitative bias analysis in a study of blood pressure lowering therapy.

Results: Compared to bias in the ATE estimator from a conditional model, the bias in MSM-IPW can be different in magnitude, but we show that the bias will always be equal in sign. Simulation results indicate that confidence intervals of the treatment effect obtained from MSM-IPW are generally wider and coverage of the true treatment effect is higher compared to a conditional model, ranging from over-coverage if there is no classification error to under-coverage when there is classification error. An online tool was developed to support the use of the bias expressions: https://lindanab.shinyapps.io/SensitivityAnalysis.

Conclusion: It is important to consider the potential impact of classification error in confounders in observational studies of causal effects. Quantitative bias analysis, using the derived expressions for the bias due to confounder misclassification, could inform studies where classification error may play a role.

[OC38.4]

Treatment effects in the principal stratum of patients who would comply, if treated with a specific treatment

Klaus Groes Larsen

Statistics, H. Lundbeck A/S, Denmark

Background: The draft ICH E9 (R1) addendum by the International Conference on Harmonisation working group opens for the use of a principal stratum in the analysis of study data for regulatory purpose, if a relevant estimand can be justified.

Objective: Inspired by the so-called complier average causal effect and work within this framework, we propose a novel estimator – Extrapolation based on propensity to comply – that estimates the effect of an active treatment relative to placebo, in the principal stratum of patients who would comply, if they were treated with the active treatment.

Results: Sensitivity of the approach to the number of covariates and their ability to predict principal stratum membership is investigated based on data from a placebo-controlled study of an antipsychotic in schizophrenia. The performance of the estimator is investigated in a simulation study and supports that the proposed estimator has a negligible bias even with small sample sizes, except when the covariate predicting compliance is very weak. Not surprisingly, precision of the estimate increases substantially with stronger predictors of compliance.

Conclusions: While the proposed methodology is technically easy to implement, choosing predictors for modelling compliance is a key issue in practice. The methodology is readily generalizable to many other types of post-baseline event (intercurrent events), e.g. to quantify the effect in patients who would develop anti-bodies as a response to immunizations.

[OC38.5]

Outcomes truncated by death in RCTs: a simulation study on the survivor average causal effect

Stefanie von Felten, Chiara Vanetta, Leonhard Held

University of Zurich, Department of Biostatistics, Switzerland

Functional outcome measurements truncated by death present a challenge in RCTs, especially if mortality differs between treatment arms. One way to deal with outcomes truncated by death is to estimate the survivor average causal effect (SACE). However, the SACE cannot be identified without non-testable assumptions. We are involved in an ongoing RCT to evaluate the effect of high-dose recombinant human erythropoietin (Epo) on neurocognitive outcomes of very preterm infants with intraventricular hemorrhage. The primary outcome is IQ at 5 years of age, but a mortality of at least 15% is expected in this highly vulnerable population. Inspired by this trial we designed a simulation study to compare the SACE (using the estimator developed by Hayden et al. 2005) with complete case analysis (of survivors) and multiple imputation of the missing outcomes (inappropriate here but potentially less biased than complete case analysis). We chose 9 scenarios combining positive, negative and no treatment effect on both outcome and survival. For each scenario we simulated 1300 data sets from a trial similar to that described above, with 500 patients. Each data set contained treatment (Epo vs. Placebo), a neurocognitive outcome and survival at 2 years of age, gestational age, head circumference at birth and the 5 min Apgar score as covariates, and was analyzed by all methods. The different methods were compared with regard to the corresponding estimands in terms of their estimates, standard errors, bias, MSE and coverage. "True values" of estimands were obtained from two simulated data sets of 1'000'000 patients. In scenarios with a positive treatment effect on survival, complete case analysis consistently estimated smaller positive or larger negative treatment effects (and even small negative effects in case of no treatment effect on outcome) than the other two methods. With a negative treatment effect on survival, complete case analysis estimated larger positive and smaller negative treatment effects. When survival was affected by treatment, estimates from complete case analysis were clearly biased regarding the SACE. With our work we hope to promote awareness of this problem and methodological knowledge of how it could be dealt with.

OC39: AFT Modelling& Survival

Chair: **Roch Giorgi**

Aix-Marseille Univ. Inserm IRD / SESSTIM, France

[OC39.1]

Effects of covariates on alternately occurring recurrent events in Accelerated Failure Time models

Moumita Chatterjee[1], Sugata Sen Roy[2]

[1] *Mathematics and Statistics, Aliah University, India*
[2] *Statistics, University of Calcutta, India*

Background: Alternating recurrent events occur frequently in real life. In clinical studies, successive times to relapse of a disease are followed by subsequent times to cure. Previous studies were primarily based on Cox Proportional Hazard model considering interdependence both between the events as also between recurrences.

Objective: To propose an alternative method for measuring the impact of covariates directly on the event times using an Accelerated Failure Time model, from which Cox PH model can be obtained as a special case.

Method: Suppose we have started with n individuals and (X_{1i}, Y_{1i},...,X_{mii}, Y_{mii}) be the sequence of alternating recurrent events for i=1,2,..n. The logarithm of the event times are modelled using a regression type model with error components which are likely to be correlated. This article develops a dependence structure within the error components of the two alternating recurrent event times. An extensive copula based study has been done considering the three possible bivariate error distributions: Bivariate Extreme Value, Bivariate Normal and Bivariate Logistic. In order to add an extra layer of dependence between the subsequent events, an AR type component has been employed. We restricted ourselves within Gaussian and Archimedean copula because of their well constructed dependence structure. Based on the above, the likelihood function is constructed and the parameter estimates are obtained considering several alternative methods.

Results: For illustration purpose we took the cystic fibrosis data (Fuchs et al.,1994). The empirical evidences (as discussed in the full paper) took us to believe that the errors follow Extreme value distribution and the joint error distribution has been constructed using Archimedean copula belonging to the Gumbel-Hougaard family. The hazard curves for treatment lie below that of the placebo.

Conclusions: The estimates of the regression parameters indicate that the effects of both the covariates are significant for the times to infection over all three cycles. Also rhDNase significantly delays the time to infection in all three cycles; both the time to infection and the time to cure depend on the past stages and the chances of infection are less under the treatment.

[OC39.2]

Embedded Likelihood based Estimation under Accelerated Failure Time Models for Clustered Censored Data

Dip Das[1], Hasinur Rahaman Khan[2]

[1] *Dhaka, University of Dhaka, Bangladesh*
[2] *Institute of Statistical Research and Training, University of Dhaka, Bangladesh*

Background: Likewise any statistical models, survival analysis depends on stochastic independence across subjects for proper inference. Observations within group are usually correlated because of many common features including genetic traits or shared environmental factors. This type of survival data is now availablle in many fiellds{genomics, proteomics and biostatistics. Traditional parameter estimation technique may still bring valid results, but estimates of standard errors and associated confidence intervals are generally not. One way to model the correlation directly through the shared frailty model that follows the latter path by explicitly modeling data clustering via random effects.

Objective: We develop an unified estimation technique using score estimating equations for censored cluster data under the accelerated failure time (AFT) models by extending the estimating equation based procedure as proposed in Ning et al. (2014). This model allows to describe the effect of covariates in latency and incidence and incorporates frailty parameter to explain cluster effects.

Methodology: The proposed method uses shared frailty AFT model to analyze clustered censored data. Estimating equations for regression parameters of AFT model are derived from the embedded likelihood function. The solutions for estimating equations are obtained by minimizing the Euclidean norm of the estimating equations. Then, frailties are derived using penalized partial likelihood method. For a fixed parametric value of frailty distribution, penalized partial likelihood is maximized and thus frailty distribution is estimated. Other properties of such estimates have been investigated.

Results: A number of simulated studies has been conducted under variety of settings including different number of sample sizes, censoring proportions and error distributions. The proposed model has also been implemented to a practical clustered censored data set. The performances of the proposed technique have been compared with several semi-parametric frailty models. Results from simulation studies suggest that the proposed technique performs well for different error distributions in terms of bias and efficiency.

Conclusion: The proposed semi-parametric model for clustered censored survival data can be used as a straight alternative to the existing techniques in literature for the analysis of clustered censored survival data.

[OC39.3]

Bayesian estimation of three-state accelerated failure time models for interval censored surveillance outcomes

Thomas Klausch, Eddymurphy Akwiwu, Johannes Berkhof, Veerle Coupé

Department of Epidemiology and Biostatistics, Amsterdam UMC, Netherlands

Background: In surveillance data disease states of patients typically are observed in time intervals. We consider the case with three interval-censored states (healthy, state A, state B) where reaching A leads to censoring of the surveillance series (e.g. because the condition is treated). As a consequence, the transition from A to B is never observed and state B is only observed if A and B occur in the same interval.

Objective: We suggest a simultaneous accelerated failure time (AFT) model for the time (denoted X) from baseline to state A, and time (S) from A to B. The model allows including covariates and specifying time distributions (e.g. Weibull, log-logistic, log-normal). An illustration is given using data from the POBASCAM cervical screening trial (with states A: HPV-infection; B: CIN2+ lesion) and colorectal cancer (CRC) surveillance in the Norwegian Cancer Register (A: advanced adenoma; B: CRC).

Method: We develop a Bayesian data augmentation approach treating X and S as latent variables. For each patient we observe the most recent interval I, the state Y, and covariates Z. The full conditional distributions of X and S are truncated. Although the truncation rules are complex (depending on X, S, I, Y, Z), the distributions are easily sampled. A Gibbs sampler then iterates between data augmentation and two Metropolis steps to sample from the posteriors of the model parameters. A so-called marginal augmentation approach accelerates MCMC convergence.

Results: We observed 20,896 healthy women's transition to either an HPV-positive state (n=627) or CIN2+ (n=82). We included the covariates 'age' and 'geographical region' and used information criteria to select the best time distribution. The estimated mean time from HPV to CIN2+ was 16.0 years but half of the population progressed in less than 10 years. Younger women in urbanized regions had shorter time to HPV and also to CIN2+. Similar illustrations are presented for the CRC surveillance data.

Conclusion: The three-state interval-censored AFT model has great practical relevance, because it allows flexible estimation of regressions for the time from state A to B which is often not directly observable in surveillance data due censoring after state A.

[OC39.4]

Flexible Modeling of Time-dependent and Non-linear covariate effects in Accelerated Failure Time model

Menglan Pang[1], Robert W. Platt[1, 2, 3], Tibor Schuster[3], Michal Abrahamowicz[1, 3, 4]

[1] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada
[2] Department of Pediatrics, McGill University, Montreal, Quebec, Canada
[3] Department of Family Medicine, McGill University, Montreal, Quebec, Canada
[4] The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada

Background: The accelerated failure time (AFT) model is considered an useful alternative to the Cox proportional hazards (PH) model in survival analysis. However, the validity of conclusions regarding the associations of prognostic factors with event times depends on whether the underlying modeling assumptions are met. Flexible methods for relaxing the PH and linearity assumptions in the Cox model have been extensively studied. In contrast, formal investigation of the corresponding assumptions of constant-over-time time ratio and linearity in the AFT model has been limited. Yet, many prognostic factors may have time-dependent and/or non-linear effects. Furthermore, parametric AFT models require correct specification of the baseline hazard function, which is treated as a nuisance parameter in the Cox PH model, and is rarely known in practice.

Objective: To develop a flexible multivariable AFT model that relaxes the inherent assumptions imposed in the AFT model.

Method: We propose a flexible AFT model where un-penalized regression B-splines are used to model (i) the baseline hazard function of arbitrary shape, (ii) the time-dependent covariate effects, and (iii) non-linear effects for continuous covariates. Maximum likelihood estimates of all functions are obtained through an iterative alternative conditional estimation algorithm. The accuracy of the estimated functions in multivariable settings is evaluated by simulation studies. To illustrate potential insights that offered by the proposed model, we apply it to re-assess the effects of prognostic factors on mortality after septic shock.

Results: Simulation results show that the proposed flexible AFT model yielded reasonably accurate estimates of various plausible shapes of both the NL and TD curves in multivariable analyses. The survival curves conditional on specific covariate vectors, can be accurately estimated, even in the presence of complex relationships between the covariates and the hazard. Application of the flexible AFT model suggests important non-linear and time-dependent effects for the association between several prognostic factors and mortality after septic shock.

Conclusion: The proposed flexible AFT model may offer potential new insights into the role of prognostic factors in clinical studies and may help to disseminate and encourage use of AFT modeling in time-to-event analyses.

OC40: Big Data

Chair: **Jörg Rahnenführer**

Technische Universität Dortmund, Germany

[OC40.1]

Comparing predictive models for continous outcomes with proper scoring rules: a simulation study.

Hugo Terrisse, Jean-Luc Bosson, Sophie Lambert-Lacroix

TIMC-IMAG, University Grenoble Alpes, France

Background: Some clinically relevant outcomes one might wish to predict are continuous. For example, predicting the length of stay can be useful for intensive care units. Many tools are available and used to predict continuous outcomes, such as generalised linear models or machine learning algorithms. Comparisons between predictive models for continuous outcomes often rely on the Mean Squared Error (MSE) or the Mean Absolute Error (MAE). However, these measures lack theoretical justification for model selection in general. Proper scoring rules, such as the Continuous Rank Probability Score (CRPS), could be useful alternatives because the expected value of a proper scoring rule is optimal for the true model that generated the data [1]. Implementations of the CRPS are already available in widely used statistical software [2].

Objective: To compare the behaviour in model comparison of the MSE, the MAE and the CRPS.

Methods: We simulated datasets according to two different simple models. The first one was a linear model. The second one was a generalised linear model with the same mean but with a gamma distribution. We split both datasets in training (n = 1,000) and test sets (n = 1,000). We adjusted a linear model and a generalised linear model on both datasets. We compared predictions of both models on the test sets, according to the NRMSE and the CRPS. The whole experiment was repeated 10,000 times.

Results: For the datasets simulated according to the linear model, the MSE selected the linear model in 54.2% (95%CI = [53.2–55.1]) of the simulations, the MAE in 53.7% (95%CI = [52.7–54.7]) of the simulations and the CRPS selected the linear model in 99.5% (95%CI = [99.3–99.6]) of the simulations. Similar results were obtained for selecting the generalised linear model when the data were generated using a gamma distribution.

Conclusions: Even in this simple simulation setting, the MSE and MAE seem inappropriate. The CRPS is a useful alternative to compare predictive models for continuous outcomes.

References

1. *Gneiting T, Raftery AE. Strictly Proper Scoring Rules, Prediction, and Estimation. J Am Stat Assoc. 2007 Mar;102(477):359–78.*
2. *Jordan A, Krüger F, Lerch S. Evaluating probabilistic forecasts with scoringRules. J Stat Softw. 2019;90.*

[OC40.2]

Deep Learning for predicting hospitalizations in elderly population using Electronic Health Records

Veronica Sciannameo[1], Daniele Jahier Pagliari[2], Corrado Lanera[1], Giuseppe Costa[3], Paola Berchialla[3]

[1] Department of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padova, Italy
[2] Department of Control and Computer Engineering, Politecnico di Torino, Italy
[3] Department of Clinical and Biological Sciences, University of Torino, Italy

Background: In the elderly population, defined as people aged 65 and over, a rising prevalence of polypharmacy, the use of multiple medications by the same individual, has been observed. Polypharmacy is closely related to multi-morbidity and it has been associated with poor adherence and potentially preventable hospital admissions.

Objective: The aim of this study is to explore the relationship between polypharmacy and medical history using medication purchases and hospitalization diagnoses from elderly north-western Italian population using Deep Learning (DL) on administrative Electronic Health Records (EHRs).

Method: The Bidirectional Encoder Representations from Transformers (BERT), a Natural Language Processing DL approach originally developed by Google, was applied to learn history of polypharmacy and hospital admissions using medication purchases and hospitalization diagnoses from administrative EHR in 2015-2016. Medications were represented with 4-digit ATC codes and diagnoses were collected from ICD-IX codes in hospital discharges records.

Masked language model (MLM) and next sentence prediction (NSP) were used to pre-train BERT in a self-supervised way, and results are reported in terms of accuracy, sensibility and specificity.

To evaluate the goodness of the medications embedding, a principal component analysis was performed. The DL was tested on the prediction of hospitalizations at 3 months.

Results: In the pre-training phase of BERT the MLM accuracy is 90.05% and the NSP accuracy is 99.60%. The embedding results of the top 10 occurring medications and their nearest neighbors in terms of cosine similarity will be shown. The embedding clusters of medications most likely bought together potentially reflect multi-morbidity conditions.

The pre-trained model was used in a preliminary experiment on a subset of 500,000 out of 24 million records to predict the occurrence of hospitalizations at 3 months. The accuracy was 80%, sensitivity 74% and specificity 81%. F1 measure was 70%.

Conclusions: Preliminary results on a small subset of the available EHRs suggest BERT is able to embed medication purchasing records, collected for administrative purposes, into patients' medical history for predicting future hospitalizations, providing a tool that could help to plan the allocation of healthcare resources.

[OC40.3]

Forward to the Past: The analysis of designed experiments and lessons for big data

Stephen Senn

Statistical Consultant, United Kingdom

The Rothamsted approach to the analysis of designed experiments observed Fisher's dictum that „that which is eliminated in the field must be eliminated in the laboratory". Analysis was matched to design and the choice of design was guided by the intended analysis. This particular philosophy reached its apotheosis in 1965 with John Nelder's theory of general balance. This, in turn, was incorporated in Genstat and is fundamental to the way it handles the analysis of designed experiments.

However, it does not seem to have been generally influential and we now are much more used to thinking of the analysis of data in terms of modelling, ironically a field to which Nelder himself made many contributions.

I shall illustrate the use of Nelder's general balance approach to handling components of variation in complex data-sets and show that it has uses in understanding observational data-sets also and hence implications for causal analysis, ‚big data' and the use of historical data for control in one-armed clinical trials. I shall also show that when applied to Lord's paradox it reveals directly that a commonly proposed solution depends on a strong unverifiable assumption.

I shall conclude that statistics seems to have lost something that once was regarded as fundamental and that we need to look back to make progress.

[OC40.4]

Construction and comparison of feature selection methods using a new measure of separability: the gamma-metric

Nicolas Ngo[1], Pierre Michel[2], Jean-François Pons[3], Stephane Delliaux[4], Roch Giorgi[5]

[1] *Aix Marseille Univ, INSERM, IRD, SESSTIM, France*
[2] *Aix Marseille Univ, CNRS, EHESS, Centrale Marseille, AMSE, France*
[3] *WitMonkiSAS, France*
[4] *Aix Marseille Univ, APHM, INSERM, INRAE, C2VN, Hop Nord, Service des explorations fonctionnelles respiratoires, France*
[5] *Aix Marseille Univ, APHM, INSERM, IRD, SESSTIM, Hop Timone, BioSTIC, France*

Background: Atrial fibrillation (AF) is one of the most frequent heart rhythm disorder. It is characterized by an irregular and often rapid heart rate and is associated with a five-fold increased risk of ischaemic strokes. In order to discriminate AF and normal sinus rate (NSR) numerous features are derived from ECG signals. Using classification models that include as many features, could lead to overfitting, thus non-generalizable predictors. To address this issue, feature selection methods use evaluation functions to assess the most relevant features or feature subsets for the prediction. The γ-metric, a separability measure, could then be seen as an evaluation function for feature selection.

Objective(s): To propose the γ-metric as an evaluation function designed for feature selection in binary classification models and evaluate the resulting models on various independent datasets.

Method(s): We first used PhysioNet database dedicated to the detection of AF to compute more than a hundred features. Then, we implemented fourteen methods, univariate and multivariate, for feature selection. Among them, four were implemented using the γ-metric, nine used classical methods and we also consider a full model, without feature selection. We used support vector machine as a classifier on each subset of features selected by the thirteen methods. Sensitivity, specificity and matthews correlation coefficient were estimated in the training and validation sets. External validation was performed, with datas provided by the Heart Rhythm Department, University Hospital, Marseille. Estimation of the performance indicators were computed by bootstrap over each dataset.

Results: In terms of feature selection, the γ-metric based methods selected less than five features whereas other methods selected up to sixty features. In terms of classification performances, all the models had a sensitivity and specificity over 99%, including the full one, on the training and validation datasets. However, in the external validation, the models with the most features had very poor performances, while the ones with the γ-metric selected features have kept performances over 90% for all indicators.

Conclusions: In this application the γ-metric fulfil its main objective, considerably reduce the dimension of the data and maintaining good predictive capacities on an independent dataset.

[OC40.5]

Propensity score matching and stratification using multiple sources without pooling individual level data

Jixian Wang

Celgene, Switzerland

Propensity score matching (PSM) and stratification are powerful approaches to eliminating or reducing confounding bias for the comparison of treatments where there is no randomization. However, applying these approaches to multiple data sources, e.g. registries and electronic health record systems for treatment comparison can often be challenging because of data privacy constraints. Our research is motivated by studies to compare a new treatment in a clinical trial with a control treatment using multiple data sources. For confidentiality reasons, individual patient data (IPD) could not be pooled. In this situation, usual PSM and stratification approaches may not be feasible because 1) a model for propensity scores (PS), e.g., a logistic model, cannot be fitted without pooling covariate data; 2) even if the PS model can be fitted by secure-multiparty computation, handling a large number of potential confounders is even more challenging to the need of secure real-time data transfers; 3) pooled outcome data may be needed to estimate the SE of treatment effect estimates (e.g., for 1:1 matching). We propose a novel approach to PSM and stratification and analysis based on them without the need for pooling IPD. For matching and stratification, we propose to use a quasi-propensity score (QPS) which is a monotone function of the propensity score and is computationally distributable, i.e., it can be calculated with aggregated data from each party. Possible choices for the QPS include linear discriminants and score transformation. When the number of potential confounders is large, compared with the sample size, sparse linear discriminants estimated by l_1 minimization is used. For analyses based on 1:1 matching we adapt a secure computation method to calculate within-pair SE. When QPS values themselves are confidential, we propose an algorithm to approximately stratify subjects in, e.g., quintiles based on histograms of QPS. This approach only requires the sharing of aggregated data. For illustration, the approach is applied to a study comparing Tamoxifen and comparators among cancer patients with many potential clinical and genomic confounders. We will also examine the performance of our approach with sparse linear discriminants when some confounders are categorical, based on a small simulation study.

OC41: Machine Learning

Chair: **Stanislav Katina**

Masaryk University, Czech Republic

[OC41.1]

Multi Objective Semi-Supervised Clustering with a right censored end point and in presence of missing data

Lilith Faucheux[1], Vassili Soumelis[2], Sylvie Chevret[1]

[1] *Paris, Université de Paris, Sorbonne Paris Cité, Center of research in Epidemiology and Statistics, INSERM UMR-1153, ECSTRRA Team, 1 avenue Claude Vellefaux F-75010, France*
[2] *Université de Paris, Sorbonne Paris Cité, INSERM U976, 1 avenue Claude Vellefaux F-75010, France*

Background: To deal with missing data in cluster analysis, consensus-based clustering algorithms have been proposed, based on multiple imputations. In addition to the similarity information used by unsupervised clustering, in many cases a small amount of knowledge is available concerning outcome information. Therefore, it should be interesting to use predictive information towards the outcome of interest when partitioning the data. This could be handled by a new semi-supervised clustering approach using multiobjective optimization, based on both internal and external validity indexes (Khorshidi, Health Inf Sci Syst, 2019). Briefly, the internal validity index captures the unsupervised property (K-means, for instance) while the external validity index (such as cross-validation error) captures the supervised information.

This work was motivated by a clinical study aiming to identify different patterns of cytokine concentrations secreted by the tumor microenvironment among 395 breast cancer patients, with the relapse-free survival as outcome of interest.

Objectives: We propose a multi-objective clustering consensus-based algorithm with a right censored end point in presence of missing data. More specifically, we assessed: (a) Which cross-validated error of prediction of the Cox model where clusters are predictors, should be used? (b) Should the final partition be selected via the positive ideal vector or obtained using a consensus? and (c) Should the pool of cluster centers be initialized by clustering algorithms?

Methods: To handle the missing data and time-to-event framework, we first modified the Khorshidi's method, by reproducing Rubin's approach and using the cross-validation error of prediction for Cox model as the external index. An alternative method based on a consensus of all the pareto optimal solutions, was secondly proposed.

We performed a simulation study to empirically assess and compare their performances, in presence of varying proportions of missing data and different strengths of association between the data structure and the outcome.

Results: A comparison of the proposed approaches is presented. Then, the proposed methods are applied to the real dataset of cytokine concentrations and breast cancer relapse-free survival.

Conclusion: Semi-supervised consensus clustering using multiobjective optimization can provide a useful framework to simultaneously handle missing data and right censored data.

[OC41.2]

Discriminating between healthy and malaria-diseased cytoskeleton red blood cells by fast-AFM and deep learning

Ido Azuri[1], Irit Rosenhek-Goldian[2], Elya Dekel[3], Dana Yaffe[3], Yael Ohana[3], Shimrit Malihi[3], Georg E. Fantner[4], Michal Sharon[3], Sidney R. Cohen[2], Neta Regev-Rudzki[3]

[1] *Life Science Core Facilities, Weizmann Institute of Science, Israel*
[2] *Chemical Research Support, Weizmann Institute of Science, Israel*
[3] *Biomolecular Sciences, Weizmann Institute of Science, Israel*
[4] *Laboratory for Bio- and Nano- Instrumentation, Ecole Polytechnique Fédérale de Lausanne, Switzerland*

The cytoskeleton is an underlying protein network supporting the cell membrane. Cytoskeleton integrity is crucial for the Red Blood Cells (RBCs) deformability, which in turn mediates the function and viability of the cell. In the particular context of infectious models, the extent of deformability of the RBCs can be linked to the likelihood of invasion, as in the case of the human malaria parasite. These deadly parasites secrete Extracellular Vesicles (EVs), while growing inside the RBCs. However, the mechanistic role of parasite-derived EVs on the RBC host membrane is not understand yet.

Developing a method with the ability to quantitatively recognizing if cell is healthy or diseased is crucial for investigating the role of parasite-derived EVs in malaria infection mechanism in addition to supporting and improving expert-human diagnostics.

We have applied atomic force microscopy (AFM) and deep learning to study the mechanical changes occurring in cells treated with malaria-derived EVs, as well as morphological transformations in the cellular cytoskeleton. High-resolution images of dried cells with exposed cytoskeleton show distinct morphological features associated with the breakdown and softening of the cell structure. In order to acquire the large volume of images necessary for deep learning and to obtain sufficient statistics, the images were obtained using a newly designed fast-scanning AFM system.

A deep learning model (Convolutional Neural Network) was developed and applied in order to verify and quantify the differences between images of healthy and diseased cytoskeleton. The model succeed in differentiating between healthy and diseased cells with accuracy of $> 90\%$. We also used the model to extract features that discriminate between the healthy and diseased cells, supporting it with statistical tests.

The model also provided an independent test of the efficacy of drug treatment for prevention of EV-induced damage. Here, the affected cells were treated by a proteasome inhibitor, meant to prevent the damage. The same model applied to this set showed that $> 90\%$ of the images are of typical healthy cells

[OC41.3]

Making time-series deep learning models more interpretable: a study in patients with cardiovascular disease

Ilaria Gandin[1], Arjuna Scagnetto[2], Giulia Barbati[2]

[1] *Research Unit, Area Science Park, Italy*
[2] *Department of Medical Sciences, University of Trieste, Italy*

Interpretability is fundamental in healthcare problems and the lack of it in deep learning models is currently the major barrier in the usage of such powerful algorithms in this context.

This study describes the implementation of an attention layer for Recurrent Neural Networks (RNNs) that provides a useful picture on the influence of the several input variables included in the model.

The study uses the MIMIC III dataset, an openly available database of electronic health records including all patients admitted to an ICU at Boston's Medical Center from 2001 to 2012. A cohort of 10,639 patients were selected based on ICD9-CM related to cardiovascular diseases. For each patient, at the start of the hospitalization, we considered a 10-length sequence 1-hour windows in which 43 clinical parameters have been extracted. The outcome of interest was death in the next 7 days (disregarding the specific time of death). In order to obtain a predictive model for the binary outcome (dead/survivors), we implemented a RNN with Long Short-Term Memory cells incorporating an attention mechanism inspired on the findings of Deepak et al. 2019: instead of time-dimension, the activation function works on the feature-dimension and so weights are directly referred to features rather than time-steps.

The performance of the RNN model, measured in terms of AUC, was 0.805, 95% CI [0.767,0.843]. Regard our primary objective, i.e. model interpretability, the attention mechanism provided a set of values (one for each feature) within the interval [0,1] that showed properties in line with their interpretation as "feature-importance" weights: 1) using the common strategy of permuting and predict one-by-one variable, we observed a clear relationship between the magnitude of the weights and decreasing performance of the model; 2) we observe correlation (r=0.663, 95% CI [0.452,0.803]) between weights and the estimated coefficient of a LASSO-based logistic regression model calculated on time-aggregated features.

Despite some limitations, mainly related to the assessment of single-variable importance that can be misleading in the case of strong dependence among features, this study shows the potential of attention mechanisms as strategy to make deep learning model more interpretable.

[OC41.4] **CASc Winner**

Iterative Least-squares Regression with Censored Data: A Survival Ensemble of Learning Machine

Nayma Hossain

Institute of Statistical Research and Training (ISRT), University of Dhaka, Bangladesh

Background: Dealing with modeling for high-dimensional censored data is challenging because of the complexities in the data structure. Many variable selection methods have been proposed for high-dimensional censored survival data for accelerated failure time (AFT) model.

Objective: This study attempts to focus on extending a variable selection procedure for censored high-dimensional data with the AFT models using the survival ensemble of popular machine learning techniques.

Methodology: The proposed method modifies the iterative least squares estimation as proposed by Jin et al. (2006) for AFT models by survival ensemble of random forest and boosting machine learning techniques for the purpose of precise estimation and variable selection. The implementation of these machine learning tools has been developed in light with recent works by Khan and Shaw (2016).

Results: The performance of the proposed methods has been demonstrated with high-dimensional censored data through a number of simulation examples and then has been compared with two similar methods in literature known as the modified resampling-based Buckley–James method and Buckley–James Dantzig selector as developed by Khan and Shaw (2016). A microarray data known as Diffuse Large-B-cell Lymphoma (DLBCL) has been used to select the genes that are linked with the survival time of DLBCL patients. The simulation studies demonstrate very satisfactory variable selection performances for the proposed methods. Particularly, both proposed methods boosting and random forest outperform the existing methods for many cases. The DLBCL data analysis also suggests that both the boosting method and random forest method are capable to find the important genes that are related to survival of patients and also can predict the survival time of future patients with small prediction error.

Conclusion: The proposed methods are easy to understand and they perform estimation and variable selection simultaneously. Hence they can be used as parallel or alternative to the existing techniques in literature for analysis high-dimensional survival data.

[OC41.5]

Covariate adjustment in randomised trials: when and how?

Clemence Leyrat, Karla Diaz Ordaz, Elizabeth Williamson

Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom

Background: Although unadjusted analyses of randomised controlled trials (RCTs) lead to unbiased treatment effect estimates, adjustment on baseline covariates may increase power. Adjustment is particularly beneficial for RCTs with moderate sample sizes, prone to chance imbalance. However, adjustment in trials is often criticized because of concerns regarding model misspecification, missing covariate data and lack of transparency in variable selection. Current guidelines recommend adjustment on a small number of pre-specified variables and not to include interaction terms with treatment. Therefore, these guidelines might prevent the use of promising new data-adaptive methodologies such as targeted maximum likelihood estimator (TMLE).

Objectives: We aim to evaluate the performance of adjustment methods for RCTs and develop guidance for a transparent yet flexible planning of these analyses.

Methods: We performed a simulation study comparing the performance of multivariable regression, g-computation, inverse-probability-of-treatment weighting, covariate-balancing propensity scores and TMLE, with and without use of the Super Learner (SL), an ensemble machine learning algorithm, which finds in a data-adaptive manner the functional form for the models. We considered binary and continuous outcomes and we generated complex associations between the outcome and the covariates and extreme covariate distributions to evaluate the impact of misspecification of the analysis model. We also implemented these methods for the analysis of the ViDiAs trial, a RCT investigating the benefits of vitamin D3 supplementation to prevent asthma exacerbations.

Results: All methods lead to the estimation of unbiased treatment effect estimates, but g-computation with bootstrapped standard errors and TMLE with SL were associated with larger gains in power. Similar results were observed in the re-analysis of the ViDiAs trial: the difference in peak expiratory flow rate between arms was 0.359 (-27.41,28.12) without adjustment. After adjustment, mean differences were similar, with narrower: -3.87 (-18.33,10.59) with multivariable regression and -4.27 (-16.94,8.41) with TMLE-SL.

Conclusions: Covariate adjustment can increase power without introducing bias. Data-adaptive methods are more efficient, and do not require the pre-specification of the functional form of the analysis model. While this could be seen as problematic, we argue that specifying a priori a data-adaptive strategy helps transparency, and prevents introducing biases due to model misspecification.

OC42: Meta-analysis 3

Chair: **Gerta Rücker**

Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Germany

[OC42.1]

Federated regression modeling for selecting biomarkers under data protection constraints

Daniela Zöller[1], Daniel Amsel[2], Stefan Lenz[1], Patrick Fischer[3], Henning Schneider[3], Hildegard Dohmen[2], Till Acker[2], Harald Binder[1]

[1] *Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Germany*
[2] *Institute of Neuropathology, Justus-Liebig-University Gießen, Medical Faculty, Gießen, Germany*
[3] *Institute of Medical Informatics, Justus-Liebig-University Gießen - Medicine Faculty, Gießen, Germany*

Background: With the advent of high-throughput screening methods, genetic biomarkers become increasingly important for clinical management of patients. However, the discovery of new biomarkers requires large cohorts. In many countries, data protection constraints forbid exchange of individual-level data between different research institutes, but researchers would like to share the contained information. To circumvent this problem, only non-disclosive aggregated data is exchanged, which is often done manually and requires explicit permission before transfer. The framework DataSHIELD enables automatic exchange in iterative calls, but methods for performing more complex tasks such as variable selection are missing.

Objective: We propose a multivariable regression modeling approach for identifying biomarkers by automatic variable selection solely based on non-disclosive aggregated data from different institutions in iterative calls. The approach should be applicable in a setting with high-dimensional data with complex correlation structures in consortia. This also implies that the amount of transferred data and the number of data calls should be limited to enable manual confirmation of compliance with data protection constraints.

Methods: We propose a regularized regression approach based on componentwise likelihood-based boosting, only requiring univariate effect estimates obtained from a linear regression for the endpoint of interest and the covariance matrix of the covariates. Additionally, we present a heuristic version of the approach aiming at reducing number of data calls.

Results: Assuming globally standardized data, the analysis is mathematically equal to an analogue individual-level analysis. In a simulation study, the information loss introduced by a local standardization is seen to be minimal. In a typical scenario, the heuristic decreases the number of data calls from more than 10 to 3, rendering manual data releases feasible. Furthermore, we demonstrate the approach to grant protected access to a single site in an application with DNA methylation biomarker data.

Conclusion: Gradient-based methods can be adapted easily to a federated setting under data protection constraints. The here presented method can be used in this setting to perform automatic variable selection and can thus aid in the process of identifying new biomarkers. We provide an implementation of the heuristic version in the DataSHIELD framework.

[OC42.2]

Test of informative cluster size with survival data

Alessandra Meddis[1], Aurélien Latouche[2]

[1] *INSERM, U900, Saint Cloud, Institut Curie PSL Research University, France*
[2] *Paris, Conservatoire National des Arts et Métiers,, France*

Background: Clustered survival data often arise in biomedical research and the assumption of non informative cluster size is commonly used even though it may be not true in some situations. The cluster size is said to be informative when the outcome depends on the size of the cluster conditional on a set of covariates. Moreover, in this context, two marginal analyses are of interest. The first refers to a typical member observed in the overall population (AOM), and the second refers to a typical member of a typical cluster (TOM).

Objectives: Under non informative cluster size (NICS) these two marginal analyses coincide. When informative cluster size (ICS) is detected, they differ in general, and more care is needed in the interpretations of results. Moreover, under ICS, the obtained results depend on the study design to collect the data and it might be difficult to generalize them to others populations. Informative cluster size is a challenging problem and its presence has an impact on the choice of the correct methodology. Several approaches have been described to take into account information carried by cluster sample size. Howerver, no method has been introduced to check for ICS in survival analysis. We want to fill this gap proposing a test for ICS with clustered survival data.

Methods: The null hypothesis of NICS is translated with the equality of the two marginal analyses. The test statistic is based on the difference of the Nelson-Aalen estimator for the cumulative hazard obtained from TOM and AOM analyses. The asymptotic distribution under the null hypothesis is provided. The performance of the test has been assessed by simulation for different settings. An application is illustrated by metastatic cancer data where the response to treatment is of interest and several metastasis sites are explored. Observations of the same patient are correlated and the number of sites may be linked to the outcome.

Results: The test statistic under the null is asymptotically equivalent to a gaussian process and the simulation study shows good power for both highly clustered data and data with a few number of big clusters. No significant impact of right censoring is highlighted.

[OC42.3]

Use of real world data in bridging disconnected networks of first and second lines of therapies

Sylwia Bujkiewicz[1], Janharpreet Singh[1], Lorna Wheaton[1], David Jenkins[1], Reynaldo Martina[1], Keith Abrams[1], Kimme Hyrich[2]

[1] *Health Sciences, University of Leicester, United Kingdom*
[2] *Faculty of Biology, Medicine and Health, University of Manchester, United Kingdom*

Background: Evidence base for health care decision making traditionally consisted of data from randomised controlled trials (RCTs), which were considered gold standard in evaluation of health technologies. More recently, a lot of research has focussed on inclusion of real world data (RWD), from observational studies, with the aim of overcoming some limitations of RCT data.

Objectives: We aim to develop methodology for incorporating RWD into decision-making using a case study in rheumatoid arthritis (RA). We explore how RWD can be used to optimise an evidence base by using evidence on first-line therapies to inform second-line effectiveness estimates in the evidence synthesis of biologic therapies in RA.

Methods: We make use of data from the British Society for Rheumatology Biologics Register in Rheumatoid Arthritis (BSRBR-RA) to supplement RCT evidence. We do so by (i) deriving prior distribution for the correlation between treatment effects of first and second-line biologic therapies (ii) using treatment effects derived from the registry data directly. We use these estimates to inform a bivariate network meta-analysis (NMA) model of first and second-line treatments. Estimates from registry data are used to bridge disconnected networks for the two lines of therapy. Disease activity score (DAS-28) and American College of Rheumatology response criteria (ACR20) were used as outcome measures.

Results: Data were obtained from 44 trials of biologic therapies including 5 for second-line treatment. The bivariate second-line NMA results showed a decrease in uncertainty for individual treatments when compared to a second-line only univariate NMA; for example OR= 2.7 (0.97, 5.35) compared to 3.35 (0.66, 10.72) for ACR20 response to Golimumab versus traditional DMARD. The bivariate NMA approach allowed for predictions of treatment effects that had not been evaluated in trials in a second-line setting.

Conclusions: The bivariate NMA provides effectiveness estimates for all treatments in first- and second-line, by using the correlation between them to predict estimates in second-line where these estimates did not exist when data were synthesised in the univariate approach. This novel methodology may have further applications, for example for bridging networks of trials in children and adults.

[OC42.4]

jarbes: An R Package for Combining Randomized and Observational Studies in Meta-Analysis

Pablo Verde

Coordination Center for Clinical Trials, University Düsseldorf, Germany

Comparative Effectiveness Research (CER) often involves the meta-analysis of results coming from studies of different qualities and types (e.g. observational and randomized studies). In addition, these studies could report aggregated results or individual participant data (IPD). The main hurdle in this kind of meta-analysis is that we are not only combining results of interest, but we are also combining their multiple biases. Therefore, commonly applied meta-analysis methods may lead to misleading conclusions. The implementation of Bayesian meta-analysis methods, which account for the CER complexity, could be challenging for most of the public health practitioners in CER.

In this work, we present the R package jarbes, standing for "Just a rather Bayesian Evidence synthesis". This R package aims to facilitate the use of complex Bayesian models in meta-analysis.

jarbes implements multi-parameters Bayesian approaches for meta-analysis, which generalize the standard random-effects model. These models allow to adjust the potential external and internal validity of the studies included in a systematic review.

The package contains a series of innovative statistical techniques including meta-analysis models with a mixture of distributions, where one of the mixture components represents hidden biased results. We illustrate the use of the package with real world examples, such as the effectiveness of stem cells treatment in heart disease patients, or the effectiveness of adjuvant treatments to heal foot lesions in diabetic patients.

[OC42.5]

Quantifying replicability and consistency in systematic reviews

Iman Jaljuli[1], Yoav Benjamini[1], Liat Shenhav[2], Orestis A. Panagiotou[3], Ruth Heller[1]

[1] *Department of Statistics and Operations Research, Tel-Aviv University, Israel*
[2] *Department of Computer Science, University of California Los Angeles, United States*
[3] *Department of Health Services, Policy & Practice, Brown University, United States*

Systematic reviews are important tools for synthesizing evidence from multiple studies. They serve to increase power and improve precision, in the same way that larger studies can do, but also to establish the consistency of effects and replicability of results across studies which are not identical. In this work we propose to incorporate tools to quantify replicability of treatment effects and assess the consistency of findings.

We suggest that these tools accompany the fixed-effect or random-effects meta-analysis, and we show that they convey important additional information for the assessment of the intervention under investigation.

We motivate and demonstrate our approach and its implications by examples from systematic reviews from the Cochrane library, and offer a way to incorporate our suggestions in their standard reporting system.

Mini – Symposia

Thursday, 27 August 2020

MS1: Statistical methods in pharmacoepidemiology and pharmacoeconomics

Organiser: Robert Platt

Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Canada

[MS1.1] Overview lecture – state of the art and current problems

Robert Platt

Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Canada

This lecture will provide an overview of some of the methodological challenges in pharmacoepidemiologic research. Pharmacoepidemiology is the study of safety and effectiveness of medications in practice, usually conducted using real-world data which was not collected for research purposes.

In practice, unlike in controlled trials, medications are not assigned at random (i.e., confounding is present), measurement of baseline conditions and of outcomes is not done under controlled conditions, and follow-up times are not standardized. Each of these problems can introduce bias.

Confounding is a particular problem in pharmacoepidemiology, because the data are not collected for research purposes. In practice, some key confounders are measured; however, some are not and it is hypothesized that other variables may serve as useful proxies for these confounders. The high-dimensional propensity score (hdPS) and its counterpart the high-dimensional disease risk score (hdDRS) are tools for semi-automated variable selection in real-world data; it has good operating characteristics in practice, but its statistical properties are not well-explored. I will describe the hdPS and hdDRS and some of their properties, and introduce some methods developed to address their limitations. I will also discuss the potential that measurement error may affect our ability to control for confounding and to estimate the effects of drugs. Finally, I will give an introduction to the problem of non-random visit patterns. It is very plausible in real-world settings that patients' visits to their physicians are associated with the patients' health states, and that this association may bias the effect of medications on outcomes. I will describe some examples of this problem and some solutions.

The lecture will conclude with a discussion of some ongoing and future problems in this area, including the introduction of electronic health record and other real-time data, privacy-protecting analyses, and other open problems.

[MS1.2] Machine learning in pharmacoepidemiology

Jason Roy

Department of Biostatistics and Epidemiology, Rutgers School of Public Health, United States

Data adaptive methods can be used in conjunction with causal inference methods to better account for measured confounding in pharmacoepidemiologic studies. We briefly describe several methods for doing this, including g-computation, ensemble learning, and mixture of experts' models. We then focus on heterogeneity of treatment effects, and show how modifications of Bayesian additive regression trees can be used to estimate the effects of interest while flexibly accounting for confounding. Our methods are demonstrated with simulation studies and an application to dataset involving adults with HIV/Hepatitis C coinfection who newly initiate antiretroviral therapy.

[MS1.3] **The importance of informative visit patterns**

Thomas Debray

University Medical Center Utrecht, The Netherlands

Non-randomized routine care data offers many opportunities to study the effectiveness of therapeutic interventions in less controlled environments. However, beyond the well-known concerns about exposure-related bias, the analysis of non-randomized routine care data may be prone to bias due to informative missingness of relevant patient outcomes. This situation can, for instance, arise when outcomes are assessed at irregular visit schedules.

In this talk, I will describe and illustrate (a selection of) common methods for estimating comparative treatment effects from routine care registry data when outcomes are assessed at irregular visit schedules. These methods transform the patient visits into equally spaced observations and adopt (simple) imputation methods to account for the presence of missing outcome data. Subsequently, I discuss how multilevel models can be used to estimate the individual patients' disease trajectories, and to alleviate the need for generating equally spaced observations.

Each method is illustrated using real world data from a multi-center registry of patients that suffer from multiple sclerosis (MS). MS is a chronic progressive disorder that affects approximately 2.3 million people worldwide. Although the efficacy of disease-modifying therapies has previously been studied in multiple randomize trials, real-world evidence can provide valuable insight into the effectiveness in routine medical practice, outside the structured clinical trial settings.

Subsequently, all methods are evaluated in an extensive simulation study, where we mimic routine care data from patients suffering from multiple sclerosis (MS). Individual disease trajectories were generated for MS patients from multiple practices. Patient outcomes were expressed as Expanded Disability Status Scale (EDSS), a standard reference scale to assess progression of MS disease, and generated for distinct months using multilevel normal distributions correlated over time. To mimic the irregular visit times in clinical practice, patient outcomes were censored according to an informative missingness procedure. For each simulated dataset, treatment effect estimates were estimated in terms of time to confirmed EDSS progression at 6 months, as defined in clinical practice.

This work was supported by Biogen (Cambridge, MA, USA).

[MS1.4] **Targeted learning in pharmacoepidemiology**

Menglan Pang[1], Tibor Schuster[2], Kristian Filion[1], Mireille Schnitzer[3], Maria Eberg[4], Robert Platt[1]

[1] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Canada
[2] Department of Family Medicine, McGill University, Canada
[3] Faculté de pharmacie, Université de Montréal, Canada
[4] Centre For Clinical Epidemiology, Lady Davis Research Institute, Jewish General Hospital, Canada

Background: Targeted Maximum Likelihood Estimation (TMLE) has been proposed for estimating causal effects, allowing specification of both the treatment and outcome models. This approach has been shown to be robust against misspecification of either model (double robustness). High-dimensional covariate spaces are commonly encountered in pharmaco-epidemiologic research. While flexible modeling methods are implemented in both the outcome and propensity score models in TMLE, the performance of data-adaptive variable selection in propensity score estimation has not been fully evaluated.

Objectives: To demonstrate the application of TMLE in high-dimensional covariate settings, to incorporate the use of high-dimensional propensity score (HDPS) into TMLE, and to compare the performance of TMLE to that of inverse probability weighting (IPW) estimators by simulations.

Methods: We implemented the TMLE procedure in a real-world pharmaco-epidemiological study: statins and the 1-year risk of all-cause mortality post myocardial infarction using large administrative database. A range of known potential confounders was considered, and empirical covariates were selected using the HDPS algorithm. The treatment effect was estimated using TMLE and IPW estimator with a variety of covariate selection strategies. We then used simulations to evaluate the performance of TMLE and IPW estimator in high-dimensional covariate settings based on the study of statins and mortality.

Results: In the case study, TMLE, adjusting for the pre-specified confounders and the HDPS covariates in either the treatment model or the outcome model, provided comparable estimates to those reported in the previous meta-analysis. The simulation showed that the performance of TMLE and IPW estimator differed when a large number of covariates were included in modeling the treatment mechanism. We found large standard errors and occasionally numerical problems by TMLE with correctly specified HDPS treatment model. The performance can be improved by truncation of the propensity score.

Conclusions: HDPS can be used in TMLE to account for potential proxies for unmeasured confounders. Although TMLE is doubly robust in general, including high-dimensional covariates are preferable in the outcome model than the treatment model, when near positivity violation is detected.

MS2: STRATOS initiative - more on guidance for analysis of observational studies

Organisers: Georg Heinze
Medical University of Vienna, Austria

Willi Sauerbrei
University of Freiburg, Germany

[MS2.1] Outstanding issues in selection of variables and functional forms in multivariable analysis

Willi Sauerbrei[1], Aris Perperoglou[2], Georg Heinze[3]

[1] Institute for Medical Biometry and Statistics, Faculty of Medicine and Medical Center University of Freiburg, Germany
[2] Center for Medical Statistics, Informatics and Intelligent Systems, Section for Clinical Biometrics, Austria
[3] Data Science and Artificial Intelligence AstraZeneca, United Kingdom

Background: When creating a descriptive multivariable model, variable selection and identification of functional forms for continuous variables are key concerns. Ad hoc 'traditional' approaches to variable selection have been developed at least 50 years ago and are still in use. Methods for determining functional forms for continuous variables were also first suggested many years ago but are still underused. More recently, many alternative approaches to address these two challenges have been proposed, but knowledge of their properties is scarce, probably because of a lack of meaningful comparisons between them. Therefore, there are many outstanding issues in multivariable modelling that prevent us to define a state of the art and to provide evidence-supported guidance.

Objective: Our main aims are to identify and illustrate such gaps in the literature.

Methods: We briefly discuss general issues in building descriptive regression models, strategies for variable selection, different ways of choosing and modeling functional forms for continuous variables and methods for combining the selection of variables and functions.

Results: Our overview revealed that there is not yet enough evidence on which to base recommendations for the selection of variables and functional forms in multivariable analysis. Such evidence may come from comparisons between alternative methods. In particular, we highlight seven important topics that require further investigation. We discuss issues in variable selection in more detail.

Conclusions: Selection of variables and of functional forms are important topics in multivariable analysis. To define a state of the art and to provide evidence-supported guidance further comparative research is required. Simulation studies and comparative analyses of real data sets will play a key role.

References:

1. Sauerbrei et al. Diagnostic and Prognostic Research (2020) 4:3. https://doi.org/10.1186/s41512-020-00074-3
2. Heinze et al. Biometrical Journal (2018) 60(3). https://doi.org/10.1002/bimj.201700067

[MS2.2] Calibration of risk prediction models: making decisions with the lights on or off?

Ben Van Calster[1], **Ewout W. Steyerberg**[2]

[1] *KU Leuven, Belgium*
[2] *Leiden University Medical Centre, the Netherlands*

Clinical risk prediction models are being published in increasing numbers, and this year's pandemic has not been an exception to this rule. The ultimate aim of predicting a clinical event is to assist clinicians in making decisions for their patients. Good decisions are based on the ability of models to distinguish between patients who have (diagnosis) or will develop (prognosis) the event. However, clinical usefulness also requires risk predictions to be accurate or 'calibrated'. Calibration ensures that risk estimates as well as risk thresholds to select high risk patients make sense. We will discuss statistical issues related to different levels of calibration and different types of predicted outcomes (e.g. binary, time-to-event, multicategory). Then we address practical issues related to calibration such as heterogeneity between centers, which make calibration 'the Achilles heel of predictive analytics'. Finally, we discuss three myths about choosing risk thresholds for prediction models.

References:

1. *Wynants L, van Smeden M, McLernon DJ, Timmerman D, Steyerberg EW, Van Calster B. Three myths about risk thresholds for prediction models. BMC Medicine 2019;17:192.*
2. *Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Medicine 2019;17:230.*

[MS2.3] Measurement error and misclassification of variables in observational epidemiology - an overview

Veronika Deffner, Helmut Kuechenhoff

Department of Statistics, Ludwig-Maximilians-University Munich, Germany

Measurement error and misclassification of variables (where observed measurements differ from what is wished to be observed) frequently occur in epidemiology and involve variables important to public health. Their presence can impact strongly on results of statistical analyses involving such variables. However, investigators commonly fail to pay attention to biases resulting from such mismeasurement (Shaw et al. (2018)).

The STRATOS topic group 4 (TG4) has prepared state-of-the art review papers on methods for handling and modelling mismeasurement (Part 1: Keogh et al (2020); Part 2: Shaw at al. (2020); Introductory paper: Wallace (2020)). In this talk, we present some of the main aspects relevant for epidemiological practice.

Methods and Results: In the first part, we give an overview of methods for measurement error in regression with a focus on the method of moments and Bayesian methods. We illustrate ways to address measurement errors by the example of modeling the association between individual exposure to air pollution and human health. We also discuss how these methods can be used for regression models with longitudinal data and more complex measurement error structures (Deffner et al. (2018)). The second part deals with handling misclassification in regression and prevalence estimation. The matrix method and likelihood approaches with extensions are presented. As an example, we show data and models from a study on age-related macular degeneration (AMD), see Günther et al. (2019). There, we use data resulting from machine-learning based disease classification. Since data of this type could occur more often in future applications, adequate handling of possible classification errors is a highly relevant methodological issue.

Conclusions: TG4 of the STRATOS initiative increases awareness of the problems caused by measurement error and misclassification in statistical analyses and removes barriers to use statistical methods that deal with such problems by providing guidance for this common topic in observational epidemiology.

References:

1. *Shaw PA, et al., Keogh,RH et al.: STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology. (Part 1 and 2) 2020. Statistics in Medicine*
2. *Wallace M (2020), Analysis in an imperfect world. Significance, 17: 14-19. doi:10.1111/j.1740-9713.2020.01353.x*

[MS2.4] Analysis of time-to-event for observational studies: Guidance to the use of intensity models

Maja Pohar Perme on behalf of the TG8 survival group

Institute of Biostatistics and Medical Informatics, Medical Faculty, University of Ljubljana, Slovenia

In this presentation, we give an overview of the STRATOS initiative paper providing guidance on the conduct of time-to-event analysis in observational studies based on intensity (hazard) models. Discussions of basic concepts like time axis, event definition and censoring are given. Hazard models are introduced, with special emphasis on the Cox proportional hazards regression model. We provide check lists that may be useful both when fitting the model and assessing its goodness of fit and when interpreting the results. Special attention is paid to how to avoid problems with immortal time bias, we discuss predictions based on hazard models and the issues in drawing causal conclusions from such models. A series of examples is presented to illustrate the methods and exemplify the check lists.

[MS2.5] Framework for the Treatment And Reporting of Missing data in Observational Studies: The TARMOS framework

James Carpenter[1], Katherine Lee[2], Kate Tilling[3], Rosie Cornish[3]

[1] Department of Medical Statistics, London School of Hygiene & Tropical Medicine, United Kingdom
[2] Murdoch Children's Research Institute Clinical Epidemiology and Biostatistics Unit, Australia
[3] MRC Integrative Epidemiology Unit, University of Bristol, United Kingdom

Missing data are ubiquitous in medical research. Although there is increasing guidance on how to handle missing data, practice is changing slowly and misapprehensions abound, particularly in observational research. Importantly, the lack of transparency around methodological decisions is threatening the validity and reproducibility of modern research. We present a practical framework for handling and reporting the analysis of incomplete data in observational studies, which we illustrate using a case study from the Avon Longitudinal Study of Parents and Children. The framework consists of three steps:

1) Develop an analysis plan specifying the analysis model and how missing data are going to be addressed. An important consideration is whether a complete records analysis is likely to be valid, whether multiple imputation or an alternative approach is likely to offer benefits, and whether a sensitivity analysis regarding the missingness mechanism is required.
2) Explore the data, checking the methods outlined in the analysis plan are appropriate, and conduct the pre-planned analysis.
3) Report the results, including a description of the missing data, details on how the missing data were addressed, and the results from all analyses, interpreted in light of the missing data and the clinical relevance. This framework seeks to support researchers in thinking systematically about missing data, and transparently reporting the potential effect on the study results, therefore increasing the confidence in and reproducibility of research findings.

[MS2.6] **A replication crisis in methodological research? On the design of comparison studies**

Anne-Laure Boulesteix

Institute of Medical Information Processing, Biometry and Epidemiology (IBE), Ludwig-Maximilians-University of Munich, Germany

Statisticians are often eager to point to the statistical aspects of the replication crisis in other scientific fields, but what about good practice issues in their own–methodological– research? When developing and evaluating statistical methods, do methodological statisticians adhere to the principles that they promote in application fields of statistics? We claim that many of the well-known issues related to the replication crisis affecting health research, such as publication bias, selective reporting/p-hacking, poor design and poor reporting, are also relevant to methodological statistical research.

In this worrying context, the statistical community needs more neutral studies, such as those intended within STRATOS papers, which are (i) not conducted with the aim of demonstrating the superiority of a particular (new) method; and (ii) authored by researchers who, as a collective, are approximately equally familiar with all considered methods. But how should they be conducted? The design of comparison studies of statistical methods, either based on simulated or real data, has hardly drawn attention so far, although the design of experiments is essentially a statistical topic.

In my talk, I will reflect those issues with special focus on the (simulation panel of the) STRATOS initiative.

References:

1. T.P. Morris, I.R. White, M.J. Crowther, 2019. Using simulation studies to evaluate statistical methods. Statistics in Medicine 38:2074-2102.

2. A.-L. Boulesteix, H. Binder, M. Abrahamowicz, W. Sauerbrei, 2018. On the necessity and design of studies comparing statistical methods. Biometrical Journal 60: 216-218.

3. A.-L. Boulesteix, R. Wilson, A. Hapfelmeier, 2017. Towards evidence-based computational statistics: lessons from clinical research on the role and design of real-data benchmark studies. BMC Medical Research Methodology 17:138.

MS3: Advanced statistical methods for the analysis of high-dimensional medical data

Organisers: Małgorzata Bogdan

Institute of Mathematics and Computer Science, Wroclaw University of Technology, Poland

Jarosław Harezlak

Department of Epidemiology and Biostatistics, Indiana University, United States

[MS3.1] Advanced technologies, large data and modern statistics in medicine

Małgorzata Bogdan

Institute of Mathematics and Computer Science, Wroclaw University of Technology, Poland

In recent years we have observed a rapid development of measurement and computer technologies, which allow us to gather and store large amounts of medical data. Efficient extraction of information from such large data bases requires the development of novel statistical methods, which often take advantage of the prior biological and medical knowledge. During this symposium 10 prominent speakers will discuss a variety of topics arising in the analysis of high dimensional medical data, including the methodology for dealing with missing data, regularization and model selection techniques in application for brain imaging and selection of genetic biomarkers, analysis of data from wearable accelerometers, Bayesian image analysis and modeling spread of disease across contact networks.

[MS3.2] Recycling genome-wide association studies with Mendelian Randomisation

Armando Teixeira-Pinto

School of Public Health, University of Sydney, Australia

Genome-wide association studies (GWAS) have identified thousands of variants associated with complex traits. The Mendelian Randomisation methodology has been developed to use these variants as instrumental variables and inferring causal associations, in the presence of unobserved confounding. This method has been around for a while but the availability of large repositories publicly available, with summaries of GWAS, has potentiated its use and the number of MR studies has been increasing rapidly.

In this talk, I will introduce the principles of MR and discuss some of the challenges with these studies. In particular, we will see the main assumptions that must be made about the genetic instrument(s) used in the MR analysis and how meta-analytic tools are used to summarise evidence from multiple instruments.

[MS3.3] Selecting predictive biomarkers from genomic data

Florian Frommlet

Medical University Vienna, Austria

Recently there have been tremendous efforts to develop statistical procedures which allow to determine subgroups of patients for which certain treatments are effective. This article focuses on the selection of prognostic and predictive genetic biomarkers based on a relatively large number of candidate Single Nucleotide Polymorphisms (SNPs). We consider models which include prognostic markers as main effects and predictive markers as interaction effects with treatment. We compare different high-dimensional selection approaches including adaptive lasso, Sorted L-One Penalized Estimator (SLOPE), Bayesian adaptive version of SLOPE (SLOBE) and modified version of the Bayesian Information Criterion. We compare them with classical multiple testing procedures for individual markers. Having identified predictive markers we consider several different approaches how to specify subgroups susceptible to treatment. Our main conclusion is that selection based on FDR controlling selection procedures has similar predictive performance as the adaptive lasso while including substantially fewer biomarkers.

[MS3.4] Modeling Dependence in Multivariate Time Series

Hernando Ombao[1], Marco Pinto[2]

[1] King Abdullah University of Science and Technology (KAUST), Saudi Arabia
[2] Oslo Metropolitan University, Norway

In this talk, we will explore some characterizations of dependence between components of a multivariate time series. This work is motivated by the current interest in neuroscience which is focused on modeling functional brain connectivity which is because functional brain networks is associated with cognitive function and mental and neurological diseases. There is no single measure of dependence that can capture all facets of brain connectivity. In this talk, we shall briefly discuss some of the well-known measures of connectivity such as partial coherence, partial correlation and partial directed coherence. The focus will be on current work on spectral causality and models for exploring potential non-linear cross-frequency interactions. This is joint work with Marco Pinto (KAUST and Oslo Metropolitan University).

[MS3.5] (Dual-Frequency)-dependent dynamic functional connectivity analysis of visual working memory capacity

Anna Dudek

AGH University of Science and Technology, Poland

We propose a novel methodology for studying the dynamic functional connectivity within the brain from EEG traces. Our observations consist of replicated realizations of spatio-temporal processes that are locally time-harmonizable. We propose a novel method to estimate both the spatial time-varying Loeve-spectrum and the spatial time-varying dual-frequency coherence functions under realistic modeling assumptions. We apply block bootstrap approach to construct confidence intervals for these parameters of interest. We illustrate the application of this methodology on a data set arising from an experiment designed to assess the visual working memory capacity. Our real data analysis pipeline starts with the clustering of our replicated time series obtained from toroidal mixture modeling of the corresponding response variables which describe the quality of memorization. Then, we estimate the spatial time-varying dual frequency coherence functions and the corresponding connectivity matrices within each cluster. This procedure allows us to potentially identify specific patterns in the dynamic functional connectivity characterizing each cluster. More specifically we reveal that better visual working memory performance is apparently associated to longer connectivity period within the prefrontal cortex between the alpha-beta frequency bands during the memorization task.

[MS3.6] **Treatment effect estimation with missing attributes**

Julie Josse

École Polytechnique, France

Inferring causal effects of a treatment or policy from observational data is central to many applications. However, state-of-the-art methods for causal inference suffer when covariates have missing values, which is ubiquitous in application.

This work is motivated by medical questions about different treatments based on a large prospective database.

Missing data greatly complicate causal analyses as they either require strong assumptions about the missing data generating mechanism or an adapted unconfoundedness hypothesis. In this talk, I will first provide a classification of existing methods according to the main underlying assumptions, which are based either on variants of the classical unconfoundedness assumption or relying on assumptions about the mechanism that generates the missing values. Then, I will present two recent contributions on this topic: (1) an extension of doubly robust estimators that allows handling of missing attributes, and (2) an approach to causal inference based on variational autoencoders adapted to incomplete data.

I will illustrate the topic an observational medical database which has heterogeneous data and a multilevel structure to assess the impact of the administration of a treatment on survival.

[MS3.7] **Regularization methods for multimodal brain imaging**

Jaroslaw Harezlak

Department of Epidemiology and Biostatistics, Indiana University, United States

A problem frequently occurring in brain imaging research is a principled incorporation of information from different imaging modalities in regression models. Often data from each modality are considered separately resulting in a loss of information. We propose a novel regularization method incorporating information from structural brain imaging, structural connectivity and functional connectivity in a regression setting. In our work, the penalty term is defined by the structural and functional connectivity modularity information. We address both theoretical and computational issues and show that our method adapts to the incomplete or mis-specified brain connectivity information. Our regularization method is evaluated via extensive simulation studies and it is applied in a study of HIV+ individuals' neurodegeneration.

[MS3.8] **Revealing the common structure of brain networks by combining the nuclear and L1 norms**

Damian Brzyski

Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Poland

In medical applications, the measurements very often are collected in a form of multidimensional arrays. For example, it is of clinical interest to measure the electrical activity of different brain regions over time, three-dimensional white-matter structure data collected from diffusion tensor imaging (DTI) for patients with multiple sclerosis (MS) or brain's metabolic activity data collected from three-dimensional positron emission tomography (PET) imaging. In my talk, I will present the method for finding the common pattern of patient-specific data appearing in a form of matrices. Specifically, I will focus on the problem of revealing the brain network structure based on the measurements indicating the strength of the connection for any pair of considered brain regions. The approach I will introduce was designed to extract the common structure from the collection of connectivity matrices and identify the clusters of strongly connected brain regions. To achieve this goal, the combination of two norms, nuclear and L1, is used under the specific penalized optimization. Here, the nuclear norm plays a role of a convex relaxation of matrix rank and can emphasize its „background" built by the dense clusters while L1 induces the sparsity outside. The method was implemented in MATLAB based on the alternating direction method of multipliers (ADMM) and was applied to recover the common brain connectivity structure for HIV-infected subjects.

[MS3.9] Novel approach for precise walking cadence estimation from high-density tri-axial accelerometry data

Marta Karas[1], Jonas Dorn[2], Jacek Urbanek[3]

[1] Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, United States
[2] Novartis International AG, Basel, Switzerland
[3] Department of Medicine, Center on Aging and Health, Division of Geriatric Medicine and Gerontology, Johns Hopkins University, Baltimore, United States

Quantifying gait parameters in free-living and monitoring of changes in these parameters have become increasingly important in epidemiological studies and clinical settings. Here, we propose a novel approach for precise walking identification and daily walking cadence estimation from high-density tri-axial accelerometry measurements collected with a wearable sensor worn at a wrist.

The problem of precise walking segmentation remains difficult due to the complexity and variability of walking signals across individuals, high-dimensionality and high-density of the data, and a lack of a gold-standard labelled data sets needed to validate and compare stride segmentation methods.

We apply our method to accelerometry data signals collected in the study of 30 arthritis patients and 15 healthy individual patients who wore a physical activity sensor at the wrist for 4 weeks. We further show that the resulted daily walking cadence estimates are associated with quality-of-life scores obtained with SF-36 survey.

[MS3.10] Bayesian image analysis in transformed spaces (BITS) and the BIFS/BIWS Python packages

John Kornak, Karl Young, Ross Boylan

Department of Epidemiology and Biostatistics, University of California, San Francisco, United States

Bayesian image analysis can improve image quality by balancing a priori expectations of image characteristics with a model for the noise process. We will give a reformulation of the conventional image space Bayesian image analysis paradigm into Fourier and wavelet spaces. By specifying the Bayesian model in a transformed space, spatially correlated priors, that are relatively difficult to model and compute in conventional image space, can be efficiently modeled as a set of independent processes in an appropriately transformed space. The originally inter-correlated and high-dimensional problem in image space is thereby broken down into a series of (trivially parallelizable) independent one-dimensional problems. We will describe and show examples of the Bayesian image analysis in transformed space (BITS) modeling approach for both Fourier and wavelet space using both parametric and data-driven priors. In the process, we will showcase our Python package(s): BIFS/BIWS that can allow easy and fast implementation of BITS.

[MS3.11] Dynamical Survival Analysis for COVID-19 Predictions in Ohio

Grzegorz Rempala

Department of Mathematics, Ohio State University, United States

Over the last several weeks many mathematicians, statisticians and data scientists have found themselves involved with various efforts in response to the public health crisis caused by the COVID-19 pandemic.

Did predictive modeling really help with COVID preparedness and decision making? Following up on my earlier lectures on the topic over the summer, I will try to give a perspective of how various mathematical methods turned out to work (or not) in practical settings of the daily predictions of the pandemic size in Ohio. In particular, I ill briefly outline some new ideas and possible improvements in the methodology of „dynamic survival analysis" developed by the OSU COVID response team to help predict COVID hospital burden.

Early Career Biostatisticians' Day

Thursday, 27 August 2020

Emily Karahalios

Senior Research Fellow (biostatistics) in the Melbourne School of Population and Global Health, The University of Melbourne and in the School of Public Health and Preventive Medicine, Monash University. She has held appointments as Clinical Biostatistician at Western Health (a health centre comprised of three teachings hospital in the western suburbs of Melbourne, Australia), and as a Research Fellow at the Cancer Epidemiology Centre, Cancer Council Victoria.

Since completing her PhD in 2014, she has been involved in biostatistics teaching and research. Emily's research focuses on the statistical methods for systematic reviews (i.e. pairwise and network meta-analysis). However, she has collaborated on numerous research projects including research in malaria, multiple sclerosis, and cancer.

Havi Murad

Senior statistician at the Biostatistics & Biomathematics Unit in the Gertner Institute for Epidemiology and Health Policy Research, Sheba Medical Center, Israel. She earned an M.A. in Statistics, "summa cum laude", at the Hebrew University, Jerusalem and PhD in Mathematics & Statistics at the Bar-Ilan University, Ramat-Gan. Her research interests include Methods for imputing missing data, methods for correcting for measurement error (in particular estimating and testing interactions). Her article "Metformin Treatment and Cancer Risk: Cox Regression Analysis, with Time-Dependent Covariates, of 320,000 Individuals with Incident Diabetes Mellitus" has been selected as the American Journal of epidemiology 2019 Article of the Year.

Societies: Editor of the ***Biometric Bulletin,*** **International Biometric Society** newsletter (2016-2018); Secretary (2009-2011) and President (2013-2015) of ***Eastern Mediterranean Region*** of the IBS (EMR); Co-chair of the ***EMR2013*** conference in Tel-Aviv (2013). Chair of the ***Israeli Biostatistics Forum*** (IBF), a branch of EMR (2014 to date). She renewed its meetings since 2014. Active member in the ***Israeli Statistical Association***, serving as a member on its Organizing/ Scientific Committee for the annual conference (2014-2018). Serving on the local Organizing and Scientific Committees of the ***EMR2018*** conference, Jerusalem. Member of the Education Subcommittee of the International Society for Clinical Biostatistics (ISCB) (2003-2006)

Stephen Senn

Born Swiss, Stephen Senn was head of the Competence Center for Methodology and Statistics at the Luxembourg Institute of Health in Luxembourg, 2011-2018, Professor of Statistics at the University of Glasgow, 2003 to 2011, and Professor of Pharmaceutical and Health Statistics at University College London 1995-2003. He has also worked in the Swiss pharmaceutical industry, in higher education in Dundee and for the National Health Service in England. He is the author of the monographs *Cross-over Trials in Clinical Research* (1993, 2002), *Statistical Issues in Drug Development* (1997, 2007), *Dicing with Death* (2003) and over 300 scientific publications. In 2001 Stephen Senn was the first recipient of the George C Challis award for Biostatistics of the University of Florida, and in 2009 was awarded the Bradford Hill Medal of the Royal Statistical Society. In 2017 he gave the Fisher Memorial Lecture. He is a Fellow of the Royal Society of Edinburgh and an honorary life member of Statisticians in the Pharmaceutical Industry (PSI) and the International Society for Clinical Biostatistics. He retired in 2018 but is still researching and consulting in statistics and holds honorary professorships at the Universities of Sheffield and Edinburgh.

Early Career Biostatisticians' Day Programme	
12.00	Welcome address: **Myra McGuinness** (AU)
	Session 1 – Moderator: Michael Grayling (UK)
12.15	**Invited Speaker: Emily Karahalios** (AU) Working as a biostatistician – effectively planning, organizing and monitoring the progress of a project
13.00	Rushani Wijesuriya (AU) Bachelors to PhD: taking the leap
13.15	Monsurul Hoq (AU) A biostatistician's journey from aid work to academia
13.30	**Discussion**
13.40	Break
	Session 2 – Moderator: Camila Olarte Parra (BE)
14.00	**Invited Speaker: Havi Murad** (IL) Working in public health versus academia or pharmaceutical industry
14.45	Mohammad Fayaz (IR) From wrong analysis to the wrong decisions; the lift chart and the others
15.00	Anna Schritz (LU) Possible pitfalls when performing sample size calculation with online tools
15.15	**Discussion**
15.30	Break
	Session 3 – Moderator: Laure Wynants (BE)
16.00	Phillip Awodutire (NG) Survival analyst in Nigeria: my experience
16.15	**Invited Speaker: Stephen Senn** (UK) The seven habits of highly effective statisticians
17.00 – 17.15	**Discussion**

Working as a Biostatistician – Effectively Planning, Organizing and Monitoring the Progress of a Project

Emily Karahalios

The University of Melbourne and Monash University, Australia

Working as a biostatistician, either in a consulting or academic position, can be challenging. Biostatisticians are in high demand, which leads to a multitude of projects that require the use of different statistical techniques. Further, other researchers often wait until the last minute to engage with a biostatistician, which leads to tight timeframes for project completion. To be successful in your role as a biostatistician, you not only need to be technically savvy, but you also need skills to ensure that you are organized and are able to keep track of multiple projects and deadlines. In this talk, Emily will share scenarios from her personal experiences that challenged her organizational and planning skills. She will then share her tips and tricks of how to plan a project, organize herself and project teams and monitor the progress of a project.

Bachelors to Phd: Taking the Leap

Rushani Wijesuriya

Department of Paediatrics, University of Melbourne, Australia

Diving headfirst into a PhD after a bachelors can be quite a challenge. Transitioning from a naive undergraduate student to a full-time researcher in a short space of time can throw many curve balls your way, making your PhD experience a truly unique journey. This talk covers my experience of progressing straight from an honours degree to pursuing a PhD, overcoming the twists and the turns, and evolving through it with a lot of lessons learnt.

Biography: Rushani is a second year PhD student at the Clinical Epidemiology and Biostatistics Unit, Department of Paediatrics, The University of Melbourne. Her main area of statistical research is multiple imputation methods for handling missing values in clustered data structures with multiple levels of clustering.

A Biostatistician's Journey from Aid Work to Academia

Monsurul Hoq

Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Australia

Fresh from university after completing an MSc in Statistics, I started my professional career at an international non-government organisation (INGO) in Bangladesh. My responsibilities included preparing data, analysing data, writing reports, and creating dashboards. The INGO had a health intervention and analysing the health data drew my interest to public health interventions. Therefore, after analysing data for two years, I pursued a Master's in Public Health and gained some skills in Medical Anthropology, Epidemiology, and Qualitative studies. Immediately after obtaining my MPH degree, I flew to South Sudan, which was awaiting a referendum to decide its fate and fighting internally. In South Sudan, I was responsible for data analysis, but with more focus on data collection and ensuring quality. I realised that we were collecting and analysing data that did not necessarily address the bigger question, i.e., whether the intervention was effective. The INGO had built infrastructure and provided access to health services, but there was no control group for comparison. Literature in support of the intervention was also lacking. After spending two and half years in East Africa, I came back to Bangladesh, this time with the responsibility of setting up a monitoring and evaluation (M&E) system for a health project at another INGO. I designed the M&E system effectively so that it was useful in answering the bigger question. Together with other colleagues, I published five papers using the secondary data collected as part of the M&E system. However, this opportunity highlighted my weakness in academic research. Hence, I moved to Australia and started working in an academic research institute as a Biostatistician. After a few years I started studying for a PhD in Biostatistics. Now I provide statistical advice to different clinical and public health research projects and communicate the findings of the statistical analysis. With the experience of working in different sectors and multiple projects I have develop skills for advocating to involve a Biostatistician in every steps of a research project. From aid work to academia, I have realised that the objective of the research (or the bigger question) should be well defined for a Biostatistician to make a meaningful contribution.

Working in public health versus academia or pharmaceutical industry

Havi Murad

Biostatistics and Biomathematics, Gertner Institute, Sheba Medical Center, Tel-Hashomer, Israel

From Wrong Analysis to the Wrong Decisions; the Lift Chart and the Others

Mohammad Fayaz

Department of Biostatistics, Shahid Beheshti University of Medical Sciences, Islamic Republic of Iran

Background: There are some topics in the statistics which are not addressed routinely in the real-world problems at the working place. On the other hand, these subjects are taught in the universities. Therefore, students think they don't need these skills, because they don't use them at all. And even if they use these methods to improve the decisions from the data, these topics are not well-understood by their colleagues and they may think that these methods are adding complexity and waste of time.

Objectives: I'd like to share some of my experience in using Confounding, Missing values and Bayesian analysis. The lift chart analysis for campaign promotions was improved the targeting population by stratifying on a confounding variable. Comparing the drug effect estimates in the presence of missing values in the clinical trial dataset with Generalized Estimation Equation (GEE) and Generalized Linear Model (GLM) methods were given us a substantial confidence about the hypothesis conclusion. The trend of COVID-19 confirmed cases were modeled with Bayesian Penalized Splines and were calculated the pointwise 95% credible sets. In this regard, we check the assumptions for using functional data analysis methods and show the appropriateness of the Bayesian approach.

Methods: Some books and related documents were provided, the SAS and R code of the case studies were publically available on the GitHub profile and Shiny. The dataset was simulated with a logistic regression model and the lift charts were estimated and compared without/ with confounding. The next one is simulating the dataset with GLM and consider %5 to 30% random missing in the dataset and using GLM and GEE. Final is testing same distribution of functional observation, COVID-19 confirmed cases, assumptions with Kolmogorov–Smirnov test, clustering, and visualize the result with heat maps. The Functional Principal Component Analysis and Bayesian Penalized splines were fitted.

Results: Considering the confounding variable improves the reachability of the target population. The GEE works well with missing values. The Bayesian estimates are appropriate.

Conclusions: Using advanced statistical methods will help to make better decisions, even if they are not seemed to be at the first.

Possible Pitfalls When Performing Sample Size Calculation With Online Tools

Anna Schritz

Competence Center for Methodology and Statistics, Luxembourg Institute of Health, Luxembourg

Background: I am a member of the Animal Welfare Structure (AWS) at my institution, which reviews study protocols for animal experiments. In the protocols, a sample size justification needs to be included as per regulations to limit the number of wasted animals. Using online tools for sample size calculations is often a popular and easy way for non-statisticians to do this. Unfortunately, many of these online tools are not validated and/or not well described. This can lead to wrong calculations and implications for the sample size of the experiment.

Objective: To summarize the pitfalls I encountered when reviewing sample size calculations of mostly non-statisticians, especially when online tools were used, and to compare different available tools.

Method: The internet is screened to find online tools that are freely available to perform standard calculations of sample size or power. The tools are reviewed and compared against each other as well as compared against sample size software such as GPower and nQuery. Possible pitfalls and misunderstandings that often occur are shown and discussed. For different

example values and distributions, sample sizes are calculated and compared.

Results: Many online tools for sample size calculations use the normal distribution without giving a clear definition. This leads to an underpowered study when other distributions, like the t-distribution, are applied for the final statistical analysis. Additionally, the definition of the input variables is often not well described, nor are examples provided to facilitate the application. In general, only standard calculators are provided for simple study designs not taking into account multiple testing.

Conclusions: Even though we are happy to see an increase in sample size justifications and calculations performed in the first versions of ethical protocols, there is still a need to further educate, especially non-statisticians, on how to use different tools and what to be aware of for adequate statistical calculations.

Survival Analyst in Nigeria: My Experience

Phillip Awodutire

Mathematical Sciences, University of Africa, Toru Orua, Nigeria

Statisticians in Nigeria have little interest in survival analysis as a field in statistics. Most researchers in Nigeria are interested in distribution theory, econometrics, and experimental design, to mention but a few. Whilst most research using survival analysis focuses more on descriptive analyses relating to survival times. This presentation will therefore highlight my experience as a survival analyst in Nigeria during my master's programme, starting from project supervision, to data collection and analysis, and getting the project properly written as a thesis and subsequently as a journal article.

The Seven Habits of Highly Effective Statisticians

Stephen Senn

The University of Edinburgh, United Kingdom

If you know why the title of this talk is extremely stupid, then you clearly know something about control, data and reasoning: in short, you have most of what it takes to be a statistician. If you have studied statistics then you will also know that a large amount of anything, and this includes successful careers, is luck.

In this talk I shall try share some of my experiences of being a statistician in the hope that it will help you make the most of whatever luck life throws you, In so doing, I shall try my best to overcome the distorting influence of that easiest of sciences hindsight. Without giving too much away, I shall be recommending that you read, listen, think, calculate, understand, communicate, and do. I shall give you some example of what I think works and what I think doesn't

In all of this you should never forget the power of negativity and also the joy of being able to wake up every day and say to yourself 'I love the small of data in the morning'.

Posters Overview

PO1: Clinical Trials: Methods & Applications

Tuesday, 25 August 2020, 17.40 - 19.10

Chairs: **Kinga Sałapa**
Astra Zeneca Pharma Poland

Alexia Iasonos
New York, United States

[PO1.01]

Nonparametric Limits of Agreement for small to moderate sample sizes - a simulation study

Maria Elisabeth Frey[1], Hans Christian Petersen[2], Oke Gerke[3]

[1] *Department of Toxicology, Charles River Laboratories Copenhagen A/S, Denmark*
[2] *Department of Mathematics and Computer Science, University of Southern Denmark, Denmark*
[3] *Department of Nuclear Medicine, Odense University Hospital, Denmark*

Background: The assessment of agreement in method comparison and observer variability analysis on quantitative measurements is often done with Bland-Altman Limits of Agreement for which the paired differences are implicitly assumed to follow a Normal distribution. Whenever this assumption does not hold, the respective 2.5% and 97.5% percentiles are often assessed by simple quantile estimation.

Objective: Sample, subsampling, and Kernel quantile estimators as well as other methods for quantile estimation have been proposed in the literature and were compared in this simulation study.

Methods: Given sample sizes between 30 and 150 and different distributions of the paired differences (Normal; Normal with 1%, 2%, and 5% outliers; Exponential; Lognormal), the performance of nonparametric quantile estimators in generating prediction intervals for one newly generated observation was evaluated by their respective coverage probability. Fourteen nonparametric quantile estimators were chosen, three of which were sample quantile estimators, four were subsampling quantile estimators, two were Kernel quantile estimators, and five were other quantile estimators.

Results: For n=30, the most simple sample quantile estimator (smallest and largest observation as estimates for the 2.5% and 97.5% percentiles) outperformed all other estimators. For sample sizes of n=50, 80, 100, and 150, only one other sample quantile estimator (a weighted average of two order statistics) complied with the nominal 95% level in all distributional scenarios. The Harrell-Davis subsampling estimator and estimators of the Sfakianakis-Verginis type achieved at least 95% coverage for all investigated distributions for sample sizes of at least n=80 apart from the Exponential distribution (at least 94%).

Conclusions: Simple sample quantile estimators based on one and two order statistics can be used for deriving nonparametric Limits of Agreement. For sample sizes exceeding 80 observations, more advanced quantile estimators of the Harrell-Davis and Sfakianakis-Verginis types that make use of all observed differences are equally applicable, but may be considered intuitively more appealing than simple sample quantile estimators that are based on only two observations per quantile.

[PO1.02]

Association rule mining to find correlations between adverse events and drugs in clinical trials

Milan Geybels[1], Frederik Kreiner[2], Jannik Nielsen[3], Kajsa Kvist[4], Jens Tarp[5], Benjamin Wolthers[6]

[1] *Biostatistics, Novo Nordisk A/S, Denmark*
[2] *Medical Writing, Novo Nordisk A/S, Denmark*
[3] *Safety Surveillance, Novo Nordisk A/S, Denmark*
[4] *Real World Data and Science, Novo Nordisk A/S, Denmark*
[5] *Data Analytics, Novo Nordisk A/S, Denmark*
[6] *Medical Affairs, Novo Nordisk A/S, Denmark*

Background: The data mining technique 'association rule mining' (ARM) uses rule-based, unbiased machine learning to find potential correlations between variables in high-dimensional datasets and to rank the interestingness (e.g. confidence) of derived rules. ARM, used independently or to guide traditional methods, may be valuable in the safety evaluation of clinical trials to ultimately improve the speed and quality of drug development.

Objective: To use ARM to elucidate potential associations between cardiovascular (CV) and other adverse events (AEs) and treatment in a large clinical trial.

Methods: We used data from the LEADER trial (9430 patients with diabetes and high CV risk; median follow-up 3.8 years), which confirmed a significant CV benefit of liraglutide (vs placebo). The dataset was dichotomized: patients with ≥1 AE (30,688 AEs in 7004 patients) or no AE during follow-up. Using the Apriori algorithm, we derived rules between common AEs (left-hand side [LHS]) and treatment (right-hand side [RHS]; liraglutide or placebo). A two-stage procedure was used: first, the first-order rule for a specific common (or composite) CV AE was identified; second, all higher-order rules with the first-order rule as a component were identified. The rule importance was evaluated using the lift criterion; to focus on rules positively treatment-associated, rules with a >1 lift were selected. Rules found for at least 10 subjects were considered.

Results: For the prevailing CV AE (acute myocardial infarction [AMI]; 357 patients), the following first-order rule was derived: {AMI} → {placebo} (lift = 1.14; p-value = 0.0114). Dependent on this first-order rule, 14 higher-order rules were identified (lift: 1.14 to 1.66). Most (10) of the additional event categories in the higher-order rules represented cardiac disorders, CV-related procedures, or renal disease. Similar rules were derived using composite variables (top 3 or 5 most frequent CV AEs) on the LHS, and several were replicated in another large trial (SUSTAIN-6).

Conclusions: ARM-derived association rules aligned with and corroborated the previously reported CV benefits of liraglutide confirmed with traditional methodologies. The results support that ARM may facilitate efficient and unbiased analysis of safety data from clinical trials; however, method validation using additional and heterogenous datasets is needed.

[PO1.03]

Using electronic health records to inform the development of clinical trials in an uncommon skin disease

Monica Sofia Maria Persson[1], Karen E. Harman[1], Kim S. Thomas[1], Joanne Chalmers[1], Yana Vinogradova[2], Sinead M. Langan[3], Julia Hippisley-Cox[4], Sonia Gran[1]

[1] *Centre of Evidence Based Dermatology, University of Nottingham, United Kingdom*
[2] *Division of Primary Care, University of Nottingham, United Kingdom*
[3] *Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, United Kingdom*
[4] *Nuffield Department of Primary Care Health Sciences, University of Oxford, United Kingdom*

Background: Bullous pemphigoid (BP), an uncommon blistering skin disease of older people, is commonly treated with oral corticosteroids. Although effective, corticosteroids have multiple side-effects. In order to plan clinical trials of safer alternatives it is important to understand the number of people with BP.

Objective: To estimate the number people in England diagnosed with new-onset BP that could be eligible for a trial involving oral corticosteroids.

Method: We identified adults with incident BP from the Clinical Practice Research Datalink and Hospital Episode Statistics Admitted Patient Care (January 2015 to December 2017). We determined the proportion affected by: (a) cautions to oral corticosteroid use listed in the British National Formulary, and (b) conditions determined by clinical expertise as „very likely" to cause exclusion from a trial involving oral corticosteroids. The proportion affected by dementia were estimated. Rates were extrapolated to the population of England.

Results: Of 22 categories leading to cautions for the use of oral corticosteroids, eight were deemed "very likely" to cause trial exclusion. The electronic records for 237 adults with incident BP were reviewed. The most common cautions at BP diagnosis were hypertension (n=163, 68.8%), diabetes mellitus (n=55, 23.2%), heart failure (n=44, 18.6%), and osteoporosis (n=39, 16.5%). Overall, 196 (82.7%) of those with incident BP had at least one caution and 117 (49.4%) had at least one caution considered "very likely" to cause exclusion.

Extrapolated to the English population, one would expect approximately 10,700 new diagnoses of BP over a three-year recruitment period. Of these, only 1,800 would have no caution to corticosteroid use. The pool of eligible participants could increase to 5,400 if more relaxed eligibility criteria were employed (only excluding those with conditions "very likely" to cause exclusion). Up to 20% of participants may require alternative consent processes due to dementia.

Conclusions: Electronic health records may be used to guide the development of clinical trials in uncommon diseases by estimating the potential pool of eligible participants. Findings from the present study will allow researchers to determine if a trial is feasible, whether a pragmatic or explanatory approach is taken, and the effect of applying exclusion criteria.

[PO1.04]

Development and preliminary validation of a novel risk and life expectancy calculator

Jacek Kopec[1], Eric Sayre[2], Benajir Shams[2], Linda Li[3], Hui Xie[4], Lynne Feehan[3], John Esdaile[5]

[1] *School of Population and Public Health, University of British Columbia, Canada*
[2] *Arthritis Research Canada, Canada*
[3] *Department of Physical Therapy, University of British Columbia, Canada*
[4] *Faculty of Health Sciences, Simon Fraser University, Canada*
[5] *Department of Medicine, University of British Columbia, Canada*

Background: A large number of disease risk calculators have been developed, but most have significant limitations. They are usually limited to one or a few conditions and timeframes, based on data from a single cohort, and do not provide the results in terms of life expectancy (LE).

Objective: To develop a new risk and life expectancy calculator that will address these limitations.

Methods: We used data from the Global Burden of Disease (GBD) Study for Canada to obtain age/sex-specific deaths rates for 270 diseases, population means for 59 exposures, and relative risk functions for 15,930 disease-exposure pairs. These data were used to create multiple-decrement period life-tables for males and females. We assumed independent effects of the risk factors, with options for multiplicative and additive interaction models. To incorporate a delay in the effects of changes in exposure on risk, we assumed an exponential decay model, varying by exposure and disease. The calculator was programmed as an SQL database with stored procedures, and VB Script web connectivity for online use. The results include LE at any age and probability of dying from each disease over various timeframes. To verify that the results are internally consistent and plausible, we computed the effects of selected risk factors (e.g., smoking, obesity) and compared the results with data from the Big Life calculator based on a Canadian cohort.

Results and Conclusions: Life expectancy (LE60) for a 60-year old Canadian male smoker (20 cigarettes a day since age 20, keeping other variables at population means) was 80.5 years, compared with 85.9 years for a non-smoker. For an obese (BMI 35) male non-smoker LE60 was 83.2 compared with 87.2 for a normal-weight (BMI 25) non-smoker. Among females, LE60 was 80.6 for a smoker and 88.0 for a non-smoker. LE60 was 85.5 for an obese and 89.2 for a normal-weight female non-smoker. Data from the Big Life calculator were generally similar. For LE60, the only difference greater than 2 years was for extremely obese female non-smokers. These results suggest that the model provides plausible results and, after further validation, could be a useful tool for health practitioners.

[PO1.05]

A practical approach to blinded sample size re-estimation in a randomized trial with a survival endpoint

Markus Pfirrmann[1], Anette-Gabriele Ziegler[2]

[1] *Institut fuer Medizinische Informationsverarbeitung, Biometrie und Epidemiologie (IBE), Ludwig-Maximilians-Universitaet Munich, Germany*
[2] *Forschergruppe Diabetes, Medical Faculty, Technical University Munich, Germany*

Background: In the Fr1da trial (ClinicalTrials.gov identifier: NCT04039945), trial numbers were based on oral insulin halving the exponentially distributed hazard of the events dysglycemia and diabetes as compared with placebo.

Objective: After recruitment of 163 children, a blinded sample size re-estimation was planned to review the assumption that 220 children randomized 1:1 provide 86.6% power to observe a significant difference at alpha level 0.033 if 2-year event-free survival was 70% in the placebo arm, recruitment 55 months, additional follow-up 36 months, and drop-out 13%.

Methods: Data on event-free survival, drop-outs, and a randomization ratio of 81:82 was available. Group allocation was unknown. Now, 1,000 samples were created, all by randomly drawing, without replacement, 82 children for a hypothetical placebo arm. Assuming exponential distribution, the event hazard was estimated in both resulting treatment groups. Sample size re-estimation was based on these estimates and the original assumptions. The method of Schoenfeld and Richter was used.

Results: Median observation time for the 163 children was 23 months. With 81.4% at 2 years, event-free survival probabilities were higher than expected. For 19 of 163 children, a drop-out was observed; 2-year probability was 9.2%. Supposing that results for the unknown real allocation into the two arms would follow the original assumption of twice the hazard for placebo, in dependence on the estimated hazards, re-calculated sample sizes ranged between 254 and 266 (n=10); and between 214 and 224, if power was reduced to 80.0%. Changes remained considerable if the hazard ratio was smaller (sample sizes from 620 to 250 for ratios from 1.5 to 1.9) or larger (sample sizes from 196 to 132 for ratios from 2.1 to 2.5).

Conclusions: The observation of higher event-free survival probabilities i.e. fewer events (n=28) than anticipated resulted in higher re-estimated samples sizes. Ten samples with hazard ratio 2.0 confirmed that a final sample size of 220 would be reasonable if all assumptions and the estimated current overall hazard were correct and a power reduction to 80.0% was acceptable. Considerable variation in sample sizes with small changes in the hazard ratio indicated restricted information provided by the current data.

[PO1.06]

Testing treatment effects for Split-plot design for Clinical trials

Lamis Alamoudi

London, MS, United Kingdom

Validating statistical tests to prevent any sort of biasness in experiments is crucial in clinical trials and related applications. This is primarily carried out using different randomisation techniques which have been extensively used in such studies. When a restricted randomisation takes place or occurs in a planned experiment, it is known as a split plot design. Linear mixed models (LMM) or analysis of variance (ANOVA) are often used as default methods to analyse split-plot designs. It is well known that these parametric methods require the assumptions of normal distribution for all random parts (errors) and equal variances. However, these assumptions are often violated or are difficult to check in practice and the traditional approaches may not be suitable under certain circumstances. The nonparametric permutation test is considered to be distribution free and the preferred alternative choice for statistical inference and hypothesis testing. In complex designs of experiments such as split plot designs, adopting one method may be unable to answer the study question. However, by combining various methods, such designs are analysed more effectively. The purpose of this study is to compare randomisation-based methods, permutation tests and a parametric method of the standard t-test for testing treatment effects to analyse different split-plot designs, in order to find out the most appropriate method statistically and practically in clinical trials. Also, a permutation test is required to determine how small the number of whole plots has to be before the standard analysis is not good enough. Next, we extend our framework to unbalanced and non-orthogonal designs. Intensive number of simulations are carried out, to generate datasets with different distributions of errors taking into account the type 1 error rate under the null and the power of these tests. The most obvious finding to emerge from the analysis is that when the permutation test based on the reduced residuals was compared with the randomisation based approach, this tended to provide good control of type 1 error in small split-plot designs with 6 and 8 whole plots.

[PO1.07]

Allocation bias in randomized clinical trials with binary response

Martin Reugels, Ralf-Dieter Hilgers, Nicole Heussen

Department of Medical Statistics, RWTH Aachen University, Germany

Background: The presence of allocation bias in clinical trials prevents a valid comparison of the treatment effects between two study groups. Allocation bias may occur even in single- or double-blinded randomized clinical trials if past treatment assignments are unmasked. Models for allocation bias in clinical trials with a continuous endpoint or with a time-to-event outcome have been recently established. However, there exist no suitable model for the case of binary endpoints to meaningfully investigate the influence of allocation bias on the test decision. Especially in view of the fact that binary endpoints are frequently used and evaluated in the context of drug or treatment effects, this is an important missing element for an accurate analysis.

Objective: The objective is to develop a model that can be used to investigate the influence of allocation bias on the test decision for binary endpoints, based on the methodology proposed in the ICH Guideline E9 and the research results for continuous or time-to-event endpoints to date.

Methods: We consider a randomized single center clinical trial with a two-arm parallel group design with no adaption aiming a target 1:1 allocation ratio. Based on the idea of the implementation of a biasing policy, possible ways to model allocation bias in the binary endpoint case are developed considering theoretical aspects and analysed via a simulation study.

Results and Conclusions: We propose and evaluate different models for biasing policies in the case of binary endpoints, which are used in further analyses to investigate the influence of allocation bias on the test decision. Since the presented biasing policies depend on the random allocation of patients and therefore are directly connected to the implemented randomization procedure, the results can further be used to compare different randomization procedures with respect to their susceptibility to allocation bias.

[PO1.08]

What happens when you switch the analysis scale in a non-inferiority trial?

Nan van Geloven[1], Matteo Quartagno[2], Stefan Böhringer[1]

[1] *Biomedical Data Sciences, Section Medical Statistics, Leiden University Medical Center, Netherlands*
[2] *MRC Clinical Trials Unit, UCL, United Kingdom*

Background: It is well known that the size of the non-inferiority margin strongly influences the required sample size of non-inferiority trials. What is largely ignored however, is that for trials with binary outcomes the scale of the margin - absolute risk difference (ARD), risk ratio (RR) or odds ratio (OR) – also has a major impact on the power and interpretation of the trial. Switching the scale of the margin could for instance be desirable when the event percentage in the control group turns out much lower than expected. Currently, many non-inferiority trials are designed and analysed without taking this impact of scale into account.

Objective: Our aim is to provide an overview of the consequences of switching between analysis scales in non-inferiority trials with binary outcomes.

Methods: We study two ways of switching between ARD, RR and OR margins: either using the anticipated or the observed event percentage in the control group. We provide sample sizes needed in a broad range of designs when using the first approach. Next, we compare the two approaches in a simulation study focussing on both type I and type II error. To separate effects induced by mapping the non-inferiority margin from those induced by switching the inference scale, we also study a reference approach of mapping the limits of the confidence intervals produced on one scale to another scale directly.

Results: Switching the scale of the non-inferiority margin based on anticipated control event rates can strongly influence power, with sample size requirements differing up to a factor two. We point out for which cases which analysis scale requires lowest sample size. Differences were most pronounced in case of large non-inferiority margins (e.g., >10% ARD) and designs with unequal anticipated event rates. When switching between scales based on observed control event rates, power differences were smaller, but this impacted type I error.

Conclusions: Our results can be used by trialists when choosing the non-inferiority scale in the design phase of a trial and can inform them on the consequences of performing an analysis on a different scale than planned in design phase.

[PO1.09]

Design and analysis of trials including subgroups defined by a biomarker

Linda Sharples[1], Kate Edgar[1], Dan Jackson[2], Kirsty Rhodes[2], Carl-Fredrik Burman[3]

[1] *Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom*
[2] *Statistical Innovation, Astra Zeneca, United Kingdom*
[3] *Biostatistics, Astra Zeneca, Sweden*

Introduction: The target population for a randomised controlled trial may comprise two distinct subgroups, with biomarker positive (Bpos) patients more likely to respond than biomarker negative (Bneg) patients, although the Bneg patients may benefit from the treatment. A model that includes a treatment by subgroup interaction term is unlikely to be statistically significant at the 5% level due to low power to detect an interaction. Lack of statistical significance does not provide sufficient reassurance to recommend treatment in Bneg patients. In this case we may wish to impose additional conditions on Bneg results before recommending treatment.

Objectives: To consider conditions imposed on Bneg results before treatment is recommended for the subgroup.

Methods: We assume a parametric regression equation including fixed effects for treatment, subgroup and treatment-subgroup interaction. Estimands of interest are regression coefficients representing treatment effects in Bneg and Bpos subgroups, which may be independent or may be correlated due to inclusion of baseline characteristics in the analysis. We explore the conditional power of three criteria for Bneg results, given that there is a significant treatment effect in the full population. The criteria are based on either (1) a minimum treatment effect in Bneg patients, (2) Bneg data 'adds' to statistical significance or (3) lack of significance of the treatment-subgroup interaction for some Type I error. These frequentist criteria are compared with a Bayesian approach based on sharing of information across Bpos and Bneg subgroups.

Results: Rules based on requiring a minimum treatment effect in Bneg generally have higher (conditional) power than those based on increasing statistical significance. Stringent levels for statistical significance resulted in Bneg patients being recommended for treatment in few scenarios, whilst the opposite was true for interaction tests. Bayesian methods provided a richer range of estimands and allowed exploration of treatment effects across a range of assumptions about the Bpos/Bneg relationship.

Discussion: Minimum requirements for treatment recommendations across biomarker-defined subgroups can be derived in both frequentist and Bayesian paradigms and can inform trial design and analysis.

[PO1.10]

Recommendations for analysing multi-arm non-inferiority trials and controlling family-wise error rate

Jake Emmerson[1], Julia Brown[1], Sue Todd[2]

[1] *Leeds Institute of Clinical Trials Research, United Kingdom*
[2] *University of Reading, United Kingdom*

Background: Multi-arm non-inferiority (MANI) trials can be useful in situations where several viable treatments exist for a disease area or for testing different dose discontinuation schedules. There are issues that need to be considered, from both the multi-arm and the non-inferiority perspective, in order to maintain the statistical integrity of such trials.

Objectives: We investigated whether methodological considerations made for multi-arm superiority trials are appropriate in the non-inferiority setting and whether issues specific to MANI trials exist. Of particular interest is the use of adjustment for multiple testing, as non-inferiority trial results are typically presented as a confidence interval while most adjustment methods are designed for p-values. The effect of different trial adaptations for MANI trials was also considered.

Methods: A comprehensive literature review across multiple databases was conducted in order to identify and investigate all current literature regarding the statistical methods and design considerations of MANI trials. Separate searches were executed for practical examples of MANI trials in order to investigate whether and how they addressed the potential statistical issues found.

Results: We found 43 past and current MANI trials, of which two had planned potential adaptations (adding arms and altering the NI margin, neither made further statistical considerations for these) and two had an option to stop early for futility. Seventeen adjusted for multiple testing (five with the use of adjusted CIs).

From investigating the methodological literature, we found that considerations made for multi-arm superiority trials typically remained necessary for non-inferiority trials. Issues specific to MANI trials were found when controlling the trial type I error when carrying out adaptations e.g. altering the sample size. Since simple corresponding methods for stepwise p-value adjustment for multiple testing do not currently exist within confidence intervals, there was no clear conclusion as to the best way to present MANI trial results when adjustment is required.

Conclusions: Despite finding MANI-specific considerations that may need to be made, trials to date have rarely required them to be utilised, particularly as they rarely involve adaptations. Although trials often adjusted for multiple testing there is room for improvement in how their adjusted results are presented.

[PO1.11]

Using routine data to conduct a pragmatic RCT to reduce antibiotic prescriptions in Swiss primary care

Soheila Aghlamandi[1], Florian Halbeisen[1], Dominik Glinz[1], Ramon Saccilotto[2], Giusi Moffa[1], Andreas Kronenberg[3], Andreas Zeller[4], Andreas Widmer[5], Julia Bielicki[6], Heiner C. Bucher[1]

[1] *Basel Institute for Clinical Epidemiology & Biostatistics, University Hospital Basel, Switzerland*
[2] *Department of Clinical Research, University Hospital Basel, Switzerland*
[3] *Institute for Infectious Diseases, University Hospital Basel, Switzerland*
[4] *Centre for Primary Health Care, University Hospital Basel, Switzerland*
[5] *Division of Infectious Diseases & Hospital Hygiene, University Hospital Basel, Switzerland*
[6] *Infectious Diseases and Paediatric Pharmacology, University Hospital Basel, Switzerland*

Background: The antibiotic consumption is very high in primary care in Switzerland, and Nesting intervention trials into routinely collected registry data is an innovative approach to address clinical problems in need of system-wide interventions such as antibiotic overuse in primary care.

Objective: To reduce antibiotic use in primary care by providing personalized antibiotic prescription feedback to individual primary care physicians in Switzerland.

Methods: We conducted a nationwide pragmatic randomized intervention trial of routine antibiotic prescription feedback in general practitioners (GP). The goal was to reduce overall antibiotic prescriptions by 5% in the second intervention year (longer-term effect months 12-24 of the intervention (primary endpoint)). The secondary aims were to lower the use of quinolones and cephalosporins. We used routinely collected individual patient claim data from the three largest health insurers covering approximately 45% of the Swiss population to prepare interventional feedback and to assess endpoints. The target population consists of the top 75% antibiotics prescribers among all GPs who see at least 100 patients a year.

Results: We randomized 3426 Swiss GPs in a 1:1 ratio to intervention and control arms. The 1713 GPs on the intervention arm only once at the beginning of the trial received evidence-based guidelines for the management of acute respiratory and urinary tract infections. Then they received quarterly personalized antibiotic prescription feedback. The 1713 GPs in the control group were not actively notified about the study and received no guidelines and no prescription feedback. The two-year intervention phase started in January 2018 and ended on December 31st, 2019. The trial set-up guarantees the full anonymization of GPs and their served patients. During the intervention phase, 74 physicians actively withdrew from the trial, 106 closed their practices, and 4 died. Data collection by health insurers is still ongoing. The last data transfer from health insurers to the University Hospital Basel server is planned for May 2020.

Conclusion: Our trial demonstrates that routinely collected data can be used to conduct efficient clinical trials on system level, in particular, for the neglected and unresearched field of primary care research.

[PO1.12]

Comparison of penalized linear mixed models for high-dimensional data

Julien St-Pierre

Department of Biostatistics, McGill University, Canada

Background: Numerous small cohorts have been put together in order to conduct genomewide association studies (GWAS) for many diseases in order to overcome their lack in power due to smaller sample sizes. However, several analytical and statistical challenges arise from combining individual level data from different cohorts in mega-analyses. Heterogeneity between studies can be induced by different study designs, different ethnic groups or different environmental exposures. Not accounting for such sources of heterogeneity can lead to spurious associations and a loss in statistical power. Moreover, fitting multivariable models of GWAS data requires careful consideration because the number of predictors (SNPs) greatly exceeds the number of observations.

Objectives: The objective of this presentation is to compare the performance of existing methods when combining individual level data from multiple cohorts with respect to (1) association testing/ variable selection and (2) phenotype prediction in the presence of multiple sources of heterogeneity.

Methods: First, we compare the different methods via simulation studies. Secondly, we demonstrate their performance when trying to identify the genetic determinants of Temporomandibular disorders (TMD), combining the individual level data from each of four well-characterized TMD datasets: The Orofacial Pain Prospective Evaluation and Risk Assessment (OPPERA) Study waves 1 and 2, the Brazilian case-control study (SPB) and the Complex persistent PAIN conditions (CPPC) study.

Results: This work aims to elucidate some outstanding questions in the high-dimensional genetics literature. When analyzing data from multiple studies simultaneously, we need to account for both between subject and between study variation which requires multiple random effects terms. However, many of the current linear mixed models (LMMs) can only handle one random effect. Moreover, most of existing penalized regression methods do not currently allow for estimation of random variables in LMMs. Hence, there is a need for developing interaction and mixed-effects models for high-dimensional data that simultaneously perform variable selection while estimating their effects, in the presence of many potential sources of heterogeneity.

[PO1.13]

The evolution of Master Protocol Clinical Trial Designs - A Systematic Review

Elias Laurin Meyer[1], Peter Mesenbrink[2], Cornelia Dunger-Baldauf[3], Hans-Jürgen Fülle[3], Ekkehard Glimm[3], Yuhan Li[2], Martin Posch[1], Franz König[1]

[1] *Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria*
[2] *Novartis Pharmaceuticals Corporation, United States*
[3] *Novartis Pharma AG, Switzerland*

Background: The recent years have seen a change in the way that clinical trials are being conducted. There has been a rise of more flexible designs allowing the investigation of multiple sub-studies with possibly different objectives, interventions and subgroups conducted within an overall trial structure. These types of designs can be summarized by the term master protocol.

Objectives: The aim of this review is to structure and thereby shed more light on these recent developments.

Methods: We conducted a comprehensive systematic search to review current literature on master protocol trials, from a design and analysis perspective, with a focus on platform trials, but also considering basket and umbrella trials. Additionally, we investigated planned or conducted master protocol trials which are connected to the included literature.

Results: We have observed an exponential growth in publications in this domain over the last few years, which we assume has not yet reached its peak. We report the results of the literature search as well as interesting features of the identified planned and conducted trials, e.g. most of the trials were designed as single-arm (58%), phase II trials (64%) in oncology (84%) using frequentist decision rules (74%). We furthermore identified several statistical questions with respect to planning master protocol trials, especially more complex platform trials, which are still unanswered.

Conclusions: Master protocol trials and especially platform trials have the potential to revolutionize clinical drug development, if the methodological and operational challenges can be overcome.

[PO1.14]

Sample size estimation for cancer randomized trials in the presence of heterogeneous data

Derek Dinart[1], Carine Bellera[1], Virginie Rondeau[2]

[1] *UMR1219, Inserm CIC-EC1401/ Institut Bergonie, France*
[2] *UMR1219, Univ Bordeaux, Inserm, France*

Background: Most of randomized controlled trials using time-to-event criteria as the primary endpoint are designed, powered and analyzed based on an hypothetical hazard ratio (HR) corresponding to the targeted effect size between experimental and control arms. Usually, one assumes that populations are homogeneous within each treatment arm, that is, within each arm, (i) the baseline risk is identical for all patients, and (ii) the treatment effect is identical for all patients.

Objective: Provide statistical methods taking into consideration variations across clusters of the baseline hazard function and variations across clusters of the treatment effect for determining sample size.

Methods: We have developed sample size estimation methods for additive frailty models. These are proportional hazard models with two cluster-specific random effects that allow to describe the dependency of clustered event times. They include one random intercept that acts on the baseline hazard and one random treatment effect that measures the variation of treatment effect. Extensive simulations were conducted to validate the proposed method and to assess the impact of the presence of heterogeneity, the number of groups and their size on the power of the statistic.

Results: Failure to take account of these heterogeneities leads to an underestimation of the sample size required. The impact of the variation of baseline hazard is small compared to that of the treatment effect. The development of appropriate statistical methods is essential in order to guarantee the required statistical power, particularly in cancerology, where multicenter trials and the parallel evaluation of multiple therapeutic strategies are increasingly frequent.

[PO1.15]

The impact of primary outcomes measures with ceiling effects on non-inferiority designs

Dawn Teare[1], Nikki Totton[2], Ellen Lee[2]

[1] *Faculty of Medical Sciences, Newcastle University, United Kingdom*
[2] *ScHARR, University of Sheffield, United Kingdom*

The use of outcome measures with ceiling or floor effects is generally avoided due to their lack of sensitivity to quantify change. However, in some situations their use is unavoidable. When such outcome measures are used in non-inferiority designs this may mask larger differences in treatment effect size due to the censoring of the measurement either post randomization, pre randomization or both. We explore the impact of ceiling effects in the design and analysis of non inferiority trials using scenarios observed in existing trials.

PO2: Longitudinal Data / Biomarkers

Tuesday, 25 August 2020, 17.40 - 19.10

Chairs: **Aleksander Owczarek**
Medical University of Silesia, Sosnowiec, Poland

Orlagh Carroll
London School of Hygiene and Tropical Medicine, London, United Kingdom

[PO2.01]

Personalized Comparisons of Flare Rates after Hydroxychloroquine Tapering or Discontinuation in Systemic Lupus

Celline Almeida-Brasil[1], Christian Pineau[1], Evelyne Vinet[1], John Hanly[2], Christine Peschken[3], Ann Clarke[4], Paul Fortin[5], Michal Abrahamowicz[6], Sasha Bernatsky[1]

[1] *Department of Medicine, McGill University, Canada*
[2] *Department of Medicine and Pathology, Dalhousie University, Canada*
[3] *Internal Medicine and Community Health Sciences, University of Manitoba, Canada*
[4] *Department of Medicine, University of Calgary, Canada*
[5] *Department of Medicine, Université Laval, Canada*
[6] *Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Canada*

Background: Hydroxychloroquine (HCQ) is a cornerstone treatment for Systemic Lupus (SLE). Given unclear risk/benefits in long-term use, patients and physicians struggle with decisions about tapering/stopping HCQ.

Objectives: To identify factors associated with SLE flare risk after HCQ is tapered or discontinued, and to compare flare rates among high-risk subgroups.

Methods: We analyzed prospective data from 5 Canadian cohorts of SLE patients treated with HCQ, using annual study visits between 1999-2019. We assessed patients from date of HCQ tapering (cohort 1) or discontinuation (cohort 2) until disease flare, defined by therapy augmentation, increase of ≥4 points in the SLE Disease Activity Index-2000 (SLEDAI-2K) or hospitalization for SLE. Multivariable Cox regression was used to identify demographic and clinical factors associated with time to first flare. We compared unadjusted flare rates and 95%CIs stratified by the risk factors. A third cohort of patients remaining on HCQ was used as a control, to assess if the same factors influenced flare rates even when therapy was unchanged.

Results: In cohort 1 (N=398, 731.2 person-years), multivariable analysis suggested that non-Caucasians, patients with active SLE and those on prednisone at the time HCQ was tapered had a greater flare risk. Patients with >1 of these characteristics at the time of HCQ taper had a higher flare rate (4.2/1000 person-years, 95%CI 3.6, 4.9) than patients without any such characteristics (2.8, 95%CI 2.3-3.4). In cohort 2 (N=395, 788.9 person-years), flare risk after HCQ discontinuation was greater for non-Caucasians and those with SLE onset at age <25 years. Patients belonging to one or both of these subgroups had an increased flare rate after HCQ discontinuation (4.2/1000 person-years, 95%CI 3.5-5.0) versus those without these factors (2.2, 95%CI 1.8-2.6). In patients remaining on HCQ (N=621), there was a trend towards higher flare rates in patients with risk factors (2.8/1000 person-years, 95%CI 2.5-3.3) versus those without (1.9, 95%CI 1.4-2.6), though 95% CIs overlapped.

Conclusions: Non-Caucasians, patients with SLE onset at age <25, those with more active SLE and/or using prednisone have more flares after HCQ tapering or discontinuation. Being aware of these factors may help physicians and patients in personalizing decisions regarding HCQ de-escalation or maintenance.

[PO2.02]

Imputation techniques for longitudinal data - applications in clinical trials

Adam Korczyński

Institute of Statistics and Demography, Statistical Methods and Business Analytics Unit, Warsaw School of Economics, Poland

Clinical trials are aimed at assessment of various effects over time. One of the challenges in longitudinal studies is missing data in the form of either intermittent missing values or patients' attrition which introduces uncertainty into the conclusion about the study endpoints.

The aim of the paper is to describe and assess different approaches for missing data analysis taking the properties of the estimates provided by each method as key criteria. The assessed methods are multiple imputation and the expectation-maximization (EM) algorithm supplemented by Newton-Raphson algorithm.

The study contains a comparison of the estimates provided by the two independent imputation techniques. Furthermore, the empirical part presents the use these imputation techniques in application for analysis of clinical trial data.

Both multiple imputation and EM algorithm reduce bias caused by missing data. The EM algorithm is more efficient than multiple imputation in terms of the computational requirements. Both methods provide consistent estimates. The advantage of the EM algorithm over multiple imputation lies in its deterministic character. It does not require multiple draws from posterior distributions, which are built-in the multiple imputation technique. Multiple imputation adjusts the standard errors and provides confidence intervals corrected for the uncertainty due to missing data. In this respect multiple imputation surpasses the EM algorithm.

Both multiple imputation and the EM algorithm can be applied to wide range of estimation problems where samples are incomplete. However multiple imputation can be easily adjusted to test different assumptions for missing data such as Washout imputation, Multiple Imputation based on Retrieved Drop-outs (MI-RD) and Tipping-Point analysis.

[PO2.03]

Estimation and Construction of CIs for the Cutoffs of Continuous Biomarkers in Trichotomous Settings

Brian Mosier, Leonidas Bantis

Department of Biostatistics, University of Kansas Medical Center, United States

Pancreatic ductal adenocarcinoma (PDAC) is one of the most lethal forms of cancer with a 5-year survival rate between 5-7%. Thus, its early detection could be key to reducing PDAC mortality. Novel biomarkers are currently being studied for its early detection. The ROC based Youden index is a popular method for choosing the optimal cutoff values of biomarkers and was recently generalized to the three-class case. We provide new parametric and non-parametric approaches that outperform the Youden based cutoffs in terms of variance and thus narrower confidence intervals are achieved. We further generalize our approaches to address k-class problems. Our approach is illustrated using real data from pancreatic cancer patients from a study conducted at the MD Anderson Cancer Center.

[PO2.04]

Three zone diagnostic decisions for numerical data

Natasa Kejzar

Institute for Biostatistics and Medical Informatics, University of Ljubljana, Faculty of Medicine, Slovenia

Background: In diagnostic medicine, biomarkers can be used as surrogates for the diagnosis. They are used when gold standard is very expensive, invasive or difficult to perform. Biomarkers have to exhibit good replacement for the gold standard. This is usually quantified by ROC curves for numerical biomarkers. Detecting the optimal ROC cut point for separating positively from negatively diagnosed patients is problematic. The optimal cut point is most likely impossible to obtain, since surrogates are usually not perfect. Therefore, two cut points are sometimes used to separate the set of patients in the three zones: the likely positive, likely negative and the (middle) inconclusive.

Objective: Few different definitions of cut points for the three zones are known from the literature (Coste et al 2003, Cannesson et al 2011, Landsheer, 2016). We intend to explore their properties and usefulness from the perspective of users in the field of medicine.

Methods: Some properties of the three-zone cut points are shown theoretically. The usability of all the definitions is explored with different simulation settings: varying sample size, distribution of biomarker and prevalence.

Results and Conclusions: It turns out that population with its underlying prevalence of disease greatly impacts the selection and relevance of the defined three-zone cut points in clinical setting. Scenarios are discussed and described where cut point(s) cannot be found. Some recommendations about the use of cut points in medical decision-making are given and illustrated with real-world examples.

[PO2.05]

A variable selection approach for highly correlated predictors in high-dimensional settings

Wencan Zhu[1], Nils Ternès[2], Céline Lévy-Leduc[1]

[1] *MMIP, AgroParisTech, France*
[2] *Translational Medicine Statistics, SANOFI R&D, France*

Introduction: The remarkable development of new technologies has allowed us to collect omic data with unprecedented size, where biomarkers are likely to be highly correlated. In such high-dimensional settings, variable selection has become one of the essential steps in biomarker discovery. Regularized methods such as Lasso are popular approaches in this context. However, it is highly impacted by the correlations between the predictors which could make the method fail to select relevant biomarkers. Hence, new methods are needed to overcome this challenge.

Methods: We propose a new penalized method based on the de-correlation of the columns of the design matrix. Since the variable selection inconsistency of Lasso may come from the correlation between the columns of the design matrix, we alleviate this problem by right multiplying the design matrix by a transformation matrix which makes its columns uncorrelated. The generalized Lasso is used to fit the transformed model and obtain estimations. These estimations are then re-transformed to obtain the final estimations. In contrast to methods based on Lasso preconditioning (HOLP among others) which consist in left multiplying the design matrix and the response variable simultaneously, our method does not modify the error term. Indeed, a modification of this term can add extra noise in the model.

Results: Numerical simulations have been conducted to compare our method to existing approaches including classical Lasso, and up-to-date methods aiming to handle the correlation such as HOLP and precision Lasso. In high-dimensional simulated scenarios where the correlation between biomarkers is high, our method achieved smaller false positive rates than the other methods for fixed true positive rates. Furthermore, unlike precision Lasso, our method has similar performance as the classical Lasso in weak correlation frameworks. We also illustrate our method on gene expression data in the oncology field.

Conclusion: In high-dimensional settings with highly correlated predictors, our approach outperforms the classical Lasso and the up-to-date methods handling correlations.

[PO2.06]

Prediction of recovery in trauma patients using latent Markov models

Roos J. M. Havermans[1], Felix J. Clouth[2], Koen W. W. Lansink[3], Jeroen K. Vermunt[2], Mariska A. C. de Jongh[1], Leonie de Munter[4]

[1] *Brabant Trauma Registry, Network Emergency Care Brabant, Netherlands*
[2] *Department of Methodology and Statistics, Tilburg University, Netherlands*
[3] *Department of Surgery, ETZ Hospital Tilburg, Netherlands*
[4] *Department of Trauma TopCare, ETZ Hospital Tilburg, Netherlands*

Background: With improved mortality rates in trauma centers, the population of patients at risk of short- and long-term disabilities after a trauma is increasing. Patient Reported Outcome Measures (PROMs) are used to assess a patient's health status and/ or functional outcome after injury and previous research identified prognostic factors for poor recovery. Information about the trauma and expected recovery can prevent unrealistic expectations and lead to more satisfaction with the chosen treatment. Decision aids increased participants' knowledge and appeared to have a positive effect on patient-clinician communication. However, the information provided in these decision aids is most often generic and not applicable for an individual patient.

Objective(s): To develop a model for personalized predictions on functional outcomes after injury that can be used in patient decision aids.

Method(s): Data of a cohort of trauma patients were derived from the Brabant Injury Outcome Surveillance study. Adult patients (18 – 75 years) were included if they were admitted to an ICU or a ward in Noord-Brabant (the Netherlands) within 48 hours after injury and survived to hospital discharge between August 2015 and November 2016. PROMs were assessed with the EuroQol-5D (EQ-5D) 3-level version and the Health Utilities Index (HUI) at one week, and one, three, six, twelve and twenty-four months after injury. Additionally, all patients received a questionnaire about their pre-injury level of function and socio-demographic information. Clinical patient information was received from the Brabant Trauma Registry. Four latent Markov models were estimated to combine the indicators from the physical dimensions of the EQ-5D and HUI and multinomial logistic regression was used to determine the effects of selected predictors on the initial state and transition probabilities. 10-fold cross-validation was performed.

Results: In total, 1107 patients were included in this study. Preliminary results show that most patients transition out of the worst outcome state within one month and, except for pain, reach the best outcome state within twelve month after injury.

Conclusions: We developed four cross-validated prediction models for functional outcomes after injury that predict the trajectory of recovery at one week, and one, three, six, twelve and twenty-four months after injury.

[PO2.07]

Transition between living arrangements among elderly Norwegians 2011-2016. Use of the IPLOS register.

Tore Wentzel-Larsen

Norwegian Centre for Violence and Traumatic Stress Studies, Norway

Background: The IPLOS register includes a comprehensive follow-up of Norwegians receiving health and care services from the municipality, with consecutive periods with differences in services and living arrangements.

Objective: To study transitions between six stages of use of services and living arrangements, and to death.

Method: Data for all persons aged 80 and above, for the years 2011-2016, were retrieved from the IPLOS register and initially investigated for inconsistencies resulting in deletion of some persons. For the remaining persons a seven state variable for use of services and living arrangements was constructed, from living at home with no use of services to long-term stay at an institution, and eventually death. Date were set up for multistate analysis using the R package mstate and transition probabilities from each stage were computed. Analyses were repeated separately for persons initially living in small municipalities, less than 5000 inhabitants, medium and large municipalities, more than 20000 inhabitants.

Results: A total of 271 419 persons was reduced to 264 503 persons, 97.4%, without missing or inconsistent gender recordings, overlapping periods, age recording above 112 years or death after 2017. From most states, about half was estimated to die during follow-up. The most common long-term transition when alive was to living at home with some assistance. Those initially having a long-term stay at an institution were an exception, most of these persons eventually died, and most of those alive tended to be in the same state also after long time of follow-up. The differences between municipality sizes were generally small.

Conclusions: It is possible to prepare data from comprehensive national registers such as IPLOS, with a large number of consecutive periods for a large number of persons, for multistate analysis using mstate and obtain interpretable information. The resulting information may be useful in planning of services for the elderly.

[PO2.08]

Coronary Heart Disease Risk Map according to Mediterranean Diet Score and Dietary variety

Alice Bonomi[1], Manuela Amato[1], Monica Giroli[1], Federica Laguzzi[2], Jose Pablo Werba[1], Elena Tremoli[1], Fabrizio Veglia[1]

[1] *IRCCS Centro Cardiologico Monzino, Italy*
[2] *Karolinska Institutet, Sweden*

Background: In medical research, an important goal is an accurate risk assessment of patients. However, continuous variables used to estimate the risk are often categorized, with a resulting loss of precision.

Objective(s): The aim of this study is to quantify the risk of coronary heart disease (CHD) with greater accuracy, through the construction of a risk map with two continuous variables, a measure of adherence to Mediterranean Diet and a measure of food variety.

Method(s): 178 patients with CHD and 155 healthy controls (matched for age and gender) were administered the European Prospective Investigation into Cancer and Nutrition (EPIC) food frequency questionnaire. The adherence to Mediterranean Diet was estimated by the Mediterranean Diet Score (MDS), an index developed by Trichopoulou (2003). Dietary variety was computed as the count of single food items consumed at least once a month. We compared four different logistic regression models according to the AIC value:

1. stratifying the covariates above/below the median;

2. in tertiles;

3. in quartiles and

4. using continuous variables.

A risk map was drawn by using PROC GCONTOUR (SAS Institute Inc., 2009) applied to the parameters obtained with the best fitting logistic regression model.

Results: As expected, the best fitting model was the model 4 (AIC=439 vs AIC=446, 442 and 441 for models 1, 2 and 3, respectively). CHD risk ranges (from 0.05 to 0.60) were represented on the map with different shades of colors, varying from white (risk<0.05) to dark red (risk≥0.60). Therefore, one can visually estimate the probability of CHD for any given combination of MDS and dietary variety (the two Cartesian coordinates of the point on the map).

Conclusions: The risk map can be a useful tool for a quick and easy classification of patient's risk. It is easy to use for clinicians and can be constructed without losing the high predictive value of continuous variables.

[PO2.09]

Does having amblyopia affect school readiness and cognitive performance?

Lisanne Gitsels, Mario Cortina-Borja, Jugnoo Rahi

Population, Policy and Practice Research and Teaching Department, UCL Great Ormond Street Institute of Child Health, United Kingdom

Background: Amblyopia is a neurodevelopmental condition causing reduced vision, for which international programmes of whole population child vision screening exist. There is an ongoing debate about the value of screening due to the lack of evidence about meaningful functional impacts of amblyopia and the extent to which these can be mitigated by treatment.

Objective(s): To determine whether amblyopia is associated with school readiness and early cognitive performance.

Method(s): Data from the prospective Millennium Cohort Study of children born in the United Kingdom in 2000-01 and followed-up to age 7 years (n=13,967). Using parental self-report on eye conditions and treatment coded by clinical reviewers, participants were grouped into no eye conditions, strabismus alone, refractive amblyopia, or strabismic/mixed (refractive plus strabismic) amblyopia. The outcomes were poor school readiness using Bracken School Readiness Assessment <25th percentile (age 3); and cognitive tests and their age-related trajectories using British Ability Scale II Naming Vocabulary (ages 3/5) and Pattern Construction (ages 5/7).

Results: Multivariable analyses showed that compared to children without any eye conditions, those with strabismic/mixed amblyopia had increased risk of poor school readiness (OR=2.04, 95%CI 1.09-3.82) but neither those on treatment (p=0.45) nor with refractive amblyopia (p=0.85) or strabismus alone (p=0.91). The small differences in mean scores for Naming Vocabulary and Pattern Construction of children with amblyopia were not of clinical significance (>10 points) compared to those without any eye conditions, irrespective of whether the treatment had already started. The age-related cognitive trajectories of children with amblyopia did not differ from those without any eye conditions for Naming Vocabulary (p=0.62) and Pattern Construction (p=0.51).

Conclusions: Amblyopia does not appear to affect cognitive performance and trajectories in early schooling and there is no evidence that this is due to a mediating effect of treatment. Although amblyopia combined with strabismus is associated with poor school readiness, this is not translated to poor cognitive performance. These novel findings may explain the lack of associations between amblyopia and educational outcomes in adult life and suggest that the impact of amblyopia on education may not of itself be a justification for population screening aimed at detecting this disorder.

[PO2.10]

Using multilevel models to estimate weight percentiles in twin pregnancies

Rocio Aznar[1], Luis Mariano Esteban[2], Sergio Sabroso[3], Gerardo Sanz[4], Ricardo Saviron[5], Rafael del Hoyo[1]

[1] *ITAINNOVA, Spain*
[2] *Escuela Politécnica La Almunia, Universidad de Zaragoza, Spain*
[3] *Instituto de Investigación Sanitaria de Aragón, Spain*
[4] *Métodos Estocásticos, Universidad de Zaragoza, Spain*
[5] *Obstreticia y Ginecologia, Hospital General de Villalba, Spain*

Background: Fetal growth models were built depending on gestational age (GA). Log transformation of weights was adjusted by gestational week using a polynomial dependence. Hadlock (1991) analysed variability in fetal weight by week and concluded that the coefficient of variation is constant among gestational week. Taking this into account, percentile weight by GA were estimated. Recently, some authors proposed the use of multilevel models in order to provide a more accurate estimation of variance in fetal weights by GA.

Objective: To estimate percentile weights in a Spanish twin population using multilevel models.

Methods: The study was performed for 1036 twins delivered between 2012 and 2017 in the Miguel Servet University Hospital, Zaragoza, Spain. Different multilevel mixed models were fitted for both dichorionic diamniotic (DC) and monochorionic diamniotic (MCDA) twin pregnancies. To provide estimated percentile weight (EPW), the variance was estimated for every gestational week from the random effects of the multilevel models, allowing calculation of percentiles. Besides, to estimate the performance of twin growth standards, we calculated the percentage of Small for Gestational Age (SGA), EPW under 10%, and Large for Gestational Age (LGA), EPW over 90%, from the DC and MCDA standards in our cohort. Comparison were performed with European and American growth models.

Results and Conclusions: The final models considered as fixed effect the GA using restricted cubic splines with 5 knots and included random slopes at the pregnancy level (between pregnancy and a cubic polynomial structure of GA) and random intercepts for each pregnancy and each fetus. For the use of the model we developed the R package PTwins freely available in CRAN repository https://cran.r-project.org/web/packages/PTwins/index.html.

For DC cases, our standard (9.8-8.2) showed good calibrations for the 10th and 90th percentiles while the other compared standards underestimated or overestimated them. For MCDA cases, both our standard (10.2-8.5) and that of Shivkumar (11.4-6.8) had the most suitable calibration. The correlation analysis between SGA and LGA cases provided by standards, showed clear differences among them. The twin growth standards depend on the population characteristics and model structure, hence a validation procedure is recommended before its use.

[PO2.11]

Investigating the use of blood metabolites as biomarkers of early cognitive changes relevant to dementia

Becki Green[1], Jodie Lord[1], Jin Xu[1], Jane Maddock[2], Min Kim[1], Richard Dobson[1], Christina Legido-Quigley[1], Andrew Wong[2], Marcus Richards[2], Petra Proitsi[1]

[1] *King's College London, United Kingdom*
[2] *University College London, United Kingdom*

Background: Decline in cognitive function in late midlife is indicative of dementia. Identifying early markers in this key phase is critical for identifying those at risk and guiding effective interventional strategies. As blood metabolites closely reflect the physiological status of an individual, they show promise as minimally invasive markers for population-level implementation.

Objectives: Using the MRC 1946 British Birth Cohort, we aim to:

1) Identify blood metabolites associated with early cognitive changes relevant to dementia
2) Tease apart mediating life course factors to address key aetiological questions
3) Employ a network approach, weighted coexpression network analysis (WGCNA), to glean insights into underlying biology.

Methods: At age 60-64, levels of 1019 metabolites were quantified using liquid chromatography-mass spectrometry (N=1740). At the same age and 5-9 years later, four aspects of cognitive function were assessed: short-term and delayed verbal memory, processing speed, and the Addenbrook's Cognitive Examination-IIII (ACE-III). Using WGCNA, 14 modules of highly connected metabolites were identified. Linear regression models evaluated associations between metabolites/metabolite modules and cognitive outcomes, sequentially adjusting for sex, childhood cognitive ability, education, socioeconomic status (SES), lipid medication and lifestyle factors. To infer biological function, pathway analyses were conducted to identify enriched pathways within metabolite modules.

Results: After correcting for multiple testing, 154 metabolites were associated with cognitive outcomes. Many associations attenuated after adjusting for early life cognition and SES, with seven metabolites belonging to lipid, amino acid and xenobiotic families remaining. WGCNA revealed two modules, enriched in vitamin A and sphingolipid metabolism, to be positively associated with cognitive outcomes. However, after adjusting for education and childhood cognition, these associations were largely reduced. Two modules, enriched in amino acid and fatty acid metabolism, were negatively associated with cognitive outcomes, with associations of the fatty acid metabolism module remaining after adjusting for life course factors.

Conclusions: These results provide support for a role for metabolites in reflecting early changes relevant to dementia, revealing insights into underlying biological mechanisms and lifelong interplay. It is hoped that these insights can be utilised to inform early diagnosis and the development of disease-modifying treatments and interventions.

[PO2.12]

Assessing the Correlation and Agreement of Repeated Measures: Application to Bis Data

Zeynep Yavuz[1], Beyza Doğanay Erdoğan[1], Atilla Halil Elhan[1], Başak Ceyda Meco[2]

[1] *Department of Biostatistics, ANKARA university, Turkey*
[2] *Department of Anaesthesiology & Intensive Care, ANKARA university, Turkey*

Background: Measures of correlation and agreement for continuous data are widely used statistical methods. Pearson product moment correlation, intraclass correlation coefficient (ICC) and Lin's concordance correlation coefficient (CCC) are the most frequently used methods when the observations are independent. However when the assumption of independence is violated, as in the case of repeated measures, statistical methods that takes this dependency into consideration and provides global measures of agreement or correlation should be preferred.

Objective(s): The aim of this study is to introduce the measures of correlation and agreement for repeated measurements and present the results for the Bispectral index (BIS) data.

Method(s): Bispectral index (BIS), the most common method to monitor anaesthesia depth, values from the sensors placed on the forehead which is the standard position and across the nasal dorsum as an alternative position of 62 patients scheduled for neurosurgical operations were used in this study. The correlation and agreement between the BIS values were obtained using the R packages "rmcorr" and "cccrm".

Results: The results indicated that the frontal (standard position) and the nasal position (alternative) BIS measurements were highly correlated (rmcorr:0.901, 95% CI: 0.889-0.911, $p<0.001$) while the agreement level was found to be moderate (repeated ccc: 0.645, 95% CI: 0.581-0.701).

Conclusions: Raising the awareness of the clinicians on the coefficients that offers overall measures of correlation and agreement and accounts for the dependency between observations can be a way to increase correct usage of the appropriate statistical methods especially in the fields like anaesthesiology where the repeated measurements are constantly taken.

References:

1. *Bakdash JZ and Marusich LR (2017) Repeated Measures Correlation. Front. Psychol. 8:456. doi: 10.3389/fpsyg.2017.00456.*
2. *Carrasco JL, Phillips BR, Puig-Martinez J, King TS, Chinchilli VM (2013) Estimation of the concordance correlation coefficient for repeated measures using SAS and R. Comput Methods Programs Biomed.109(3):293-304.*

[PO2.13]

Incorporating baseline outcome data in individual participant data meta-analysis of non-randomized studies

Lamprini Syrogiannouli[1], Lea Wildisen[1], Douglas C. Bauer[2], Anne R. Cappola[3], Jacobijn Gussekloo[4], Stella Trompet[4], Rudi G. J. Westendorp[5], J. Wouter Jukema[6], Luigi Ferrucci[7], Graziano Ceresini[8], Bjørn O. Åsvold[9], Christiaan L. Meuwese[10], Wendy P. J. den Elzen[11], Orestis Efthimiou[12], Nicolas Rodondi[1], Cinzia Del Giovane[1]

[1] *Institute of Primary Health Care (BIHAM), University of Bern, Switzerland*
[2] *Departments of Medicine and Epidemiology & Biostatistics, University of California, United States*
[3] *Department of Medicine, University of Pennsylvania , United States*
[4] *Department of Internal Medicine, Leiden University Medical Center, Netherlands*
[5] *Public Health and Center for Healthy Aging, University of Copenhagen, Denmark*
[6] *Department of Cardiology, Leiden University Medical Center, Netherlands*
[7] *National Institute on Aging, United States*
[8] *Clinical and Experimental Medicine, University Hospital of Parma, Italy*
[9] *Department of Public Health and Nursing, Norwegian University of Science and Technology, Norway*
[10] *Department of Intensive Care Medicine, University Medical Centre Utrecht, Netherlands*
[11] *Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Center, Netherlands*
[12] *Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland*

Background: In non-randomized studies (NRSs) where a continuous outcome is assessed at baseline and follow-up, it is likely to observe baseline imbalance between the treatment/exposure and control group. Unless adjusted for, this imbalance may confound the estimated effects in each study and consequently bias the results of a meta-analysis (MA). Appling different methods to analyze the same continuous outcomes leads to different study estimates, which may influence the meta-analytical estimates. Analysis of individual participant data (IPD) allows to standardize the methods across studies.

Objectives: 1) To identify methods used in published IPD-MA of NRSs for continuous outcomes, 2) To compare different methods to account for baseline outcome data in IPD-MA of NRSs using six studies on patients with hyperthyroidism.

Methods: For the first objective, we searched EMBASE, MEDLINE, and Cochrane (until December 2019) to identify IPD-MA of NRSs of continuous outcomes that incorporated baseline outcome in the analysis. For the second objective, we considered the following methods: A) using only follow-up values, i.e. ignoring baseline; B) ANCOVA, assuming linearity; C) using change from baseline as outcome; D) using propensity score, based on the conditional probability of having subclinical hyperthyroidism, given the baseline outcome; E) a variation of ANCOVA that modelled the effect of baseline using splines. We applied each method to IPD from 24'524 patients with subclinical hyperthyroidism with the outcome being depressive symptoms.

Results: Our bibliographic search identified 8 published reviews. Change from baseline was the most widely used approach (7/8). In our hyperthyroidism example, baseline outcome data were balanced across groups. Study-specific point estimates varied substantially according to the statistical method used. Overall, no method found strong evidence for an association between depressive symptoms and subclinical hyperthyroidism. We will present the updated results for objective 1 and results from an application of the 4 methods to an additional example with baseline imbalance between groups at the Conference.

Conclusions: Despite Cochrane recommends using ANCOVA, change from baseline remains commonly used. Different methods used to deal baseline outcome data in the analysis of continuous outcomes in IPD-MA of NRSs may lead to different and potentially inappropriate summary estimates.

[PO2.14]

Structured reporting to improve transparency of analyses in prognostic biomarker studies

Willi Sauerbrei[1], Tim Haeussler[1], James Balmford[1], Marianne Huebner[2]

[1] *Institute for Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg, Germany*
[2] *Department of Statistics and Probability, Michigan State University, East Lansing, MI, United States*

Background: Reporting guidelines for prognostic tumor biomarker studies and a corresponding 'explanation and elaboration' paper have been available for many years (Mc Shane et al, JNCI 2005, Altman et al, PLoS Med 2012). However, a recent review showed that even basic information about study populations and relevant details of statistical analyses are often not provided (Sekula et al PLoS ONE 2017). In a systematic review Kempf et al (2018, BJC) showed that overinterpretation of findings of prognostic biomarker assessment is frequent in high impact journals. Clearly, poorly reported single studies also hinder the conduct of meaningful meta-analyses of prognostic biomarkers (Sauerbrei and Haeussler, BJC2018).

Objective: The two-part REMARK profile, a structured display summarizing key aspects of a study, especially the derivation of the sample and information about the analyses performed, has been proposed to improve completeness and transparency of reporting, specifically of statistical analyses (Altman et al PLoS Med 2012). Created prospectively, it helps authors develop the statistical analysis plan and increases the transparency of the analyses conducted (Winzer et al, PLoS ONE 2016).

Methods: We created REMARK profiles for three published biomarker studies with a time-to-event outcome from each of five cancer research journals (BCRT, Cancer, EJC, IJC, JCO). We summarized the analysis steps performed and whether sufficient details of each analysis were provided.

Results: We found that the reporting of analyses was insufficient in nearly all of the studies we reviewed. Concerning the patient population, information about exclusion of patients was incomplete in over half of the studies. Even for the primary outcome, the number of events (the effective sample size) was often not reported, nor was this mentioned for many subgroup analyses.

Conclusions: We argue that the REMARK profile is a suitable instrument to improve the transparency of analyses of prognostic studies. It can also help to reduce the common problem of 'fishing for significance' if a statistical analysis plan is registered. These principles can be transferred to many other types of studies.

[PO2.15]

Perturbation of proteomic biomarker data for sharing: gaining privacy while preserving utility?

Jeppe Christensen[1], Georg Heinze[1], Harald Mischak[2]

[1] *Medical University of Vienna, Austria*
[2] *Mosaiques Diagnostics and Therapeutics AG, Germany*

Background: Release of medical data is important in the scientific world, but it compromises patient privacy, which is a major concern.

Objective: Releasing perturbed versions of the original data sets might preserve some degree of patient privacy, but more privacy leads to less utility. With proteomic biomarker data, perturbation is complicated by zero-inflated and non-symmetric distributions. We investigated to which extent perturbation methods can preserve utility while guaranteeing a certain level of privacy.

Methods: The original proteomics data set is split in training- and test sets multiple times to control for uncertainty in splitting. Each training set is perturbed and LASSO regression models are trained on the original and privacy preserving sets, which are then applied to the test set. The utility of the sets is assessed by the average difference in AUROC and Brier scores and average agreement on selected biomarkers expressed by true and false positives. Privacy is assessed by assessing each real patient's proximity to the closest perturbed-data subject, expressed in terms of metrics based on the Euclidian distance.

Approximately symmetric distributions of the variables are achieved by log-transformation. Perturbation is performed by transforming the data to principal components, permuting some of them and then transforming back.

Results and Conclusions: The models trained on the privacy preserving data on average yield AUROC almost as high as those trained on the original data. They also chose the same variables as in the original models fairly often. However, the models resulting from training on the privacy preserving data seem more sensitive to the splitting, as the AUROC and the variable selections are more volatile with more permuted principal components.

Satisfying AUROC and somewhat satisfying biomarker selection can on average be achieved, while identifiability of specific patients becomes infeasible. Limitations are that we only tested utility of the data regarding prediction modelling and only quantified resilience towards attacks that make use of distance. We will continue this research by considering further applications and simulations, and by considering other measures of privacy.

PO3: Survival analysis / Machine Learning

Tuesday, 25 August 2020, 17.40 - 19.10

Chairs: **Janie Coulombe, Steve Ferreira Guerra**
McGill University, Montreal, Canada

[PO3.01]

Pre-existing autoimmune disease and immune-related adverse events with checkpoint inhibitors in melanoma

Marina Machado[1], Cristiano Moura[2], Kelvin Chan[3], Jeffrey Curtis[4], Marie Hudson[5], Michal Abrahamowicz[6], Ines Colmegna[1], Rahima Jamal[7], Louise Pilote[1], An-Wen Chan[3], Sasha Bernatsky[1]

[1] *Department of Medicine, McGill University, Canada*
[2] *Research Institute of the McGill University Health Centre, Canada*
[3] *Department of Medicine, University of Toronto, Canada*
[4] *Department of Medicine, University of Alabama at Birmingham, United States*
[5] *Lady David Research Institute for Medical Research, Canada*
[6] *Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Canada*
[7] *Centre Hospitalier de l'Universite de Montreal, Canada*

Background: Metastatic melanoma patients with pre-existing autoimmune disease were excluded from clinical trials investigating checkpoint inhibitors because of concerns of immune-related adverse events (irAE).

Objective: We used real-world data to evaluate risk of irAE in metastatic melanoma patients with and without pre-existing autoimmune diseases initiating checkpoint inhibitors.

Methods: We assembled a cohort of adults with metastatic melanoma using US MarketScan® Databases. All patients were assigned a time-zero as the date of first exposure to ipilimumab (IPI), pembrolizumab (PEM), nivolumab (NIV), or NIV/IPI over Jan 2012-July 2017. We required health/drug plan coverage for 12 months before time-zero and identified in this period all autoimmune disease diagnostic codes from physician claims/hospitalizations. Outcomes (incident irAE) following time-zero were identified from physician claims/hospitalizations primary diagnostic codes. Patients were censored due to in-hospital death, loss of health/drug plan, end of study period (Dec 31, 2017), gap between doses ≥ 60 days, or therapy switch. We calculated irAEs rates stratified by pre-existing autoimmune disease and sex. We used Cox model to estimate hazard ratios (HR) with 95% confidence intervals (CI) adjusted for age, sex, calendar year, comorbidities, past health care use, past and current cancer therapy, and pre-existing autoimmune disease.

Results: We studied 2315 patients initiating IPI (62.2%), PEM (16.8%), NIV/IPI (11.8%), and NIV (9.3%). The median age was 60 years (interquartile range 52-67), 62.1% were male, and 27.7% had pre-existing autoimmune disease including hypothyroidism (16.4%), myositis (3.3%), and interstitial lung disease (2.5%). The rate (per 100 person-years) of irAEs in patients with pre-existing autoimmune diseases was 7.3 (95%CI 4.2-12.6) versus 4.1 (95%CI 2.5-6.6) in those without autoimmune disease. The rate was 5.4 (95%CI 3.5-8.3) in men and 4.4 (95%CI 2.4-8.3) in women. In multivariate analyses, presence of pre-existing autoimmunity was associated with higher irAE risk vs. absence (HR 2.17, 95%CI 1.01-4.66); irAE risk was lower with IPI vs. NIV/IPI (HR 0.19 95%CI 0.04-0.86); we did not to see difference in women vs. men (HR 0.77, 95%CI 0.35-1.70).

Conclusions: In this study, we observed higher irAE risk in patients with pre-existing autoimmunity and a decreased irAE risk in IPI, although 95% CIs around estimates were wide.

[PO3.02]

Mixture and Non-Mixture Cure Rate Models Using the Beta Type I Generalized Half Logistic Distribution

Phillip Awodutire

Department of Mathematical Science, University Of Africa,Toru Orua, Nigeria

In this work, we studied the Beta Type I Generalized Half Logistic distribution in the presence of cure fractions. The estimates of the parameters were derived using the maximum likelihood estimation method. Simulation studies under mixture and non-mixture cure rate models were carried out to assess the performance of their estimates. We applied these models to data from patients with gastric adenocarcinoma in which the necessary inferences were drawn.

[PO3.03]

A New Method to Evaluate Post-Discharge Intervention on Risk of Adverse Events with Observational Data

Liping Tong

Advocate Aurora Health, United States

Background: Evaluations for the effect of intervention programs on reducing adverse events are very important topics in medical practice. When the adverse events are dichotomous, people tend to directly compare rates of adverse events between the with- and without-intervention groups. This can lead to misleading conclusions when the intervention is a post-discharge action. As an alternative, the time-dependent Cox proportional hazard model in survival analysis provides an appropriate way of evaluation. However, the complexity of the survival model has prevented it from widely used and accepted in practice. Our goal is to develop a new method that is easy to understand and interpret and meantime has a comparable efficiency to the survival model.

Methods: Using the example of an intervention (follow-up visit) to reduce 30-day readmission rate, we propose a new probabilistic model for the dynamic process of intervention and readmission. We build regression models based on it and show how to perform the corresponding hypothesis tests. More importantly, we identify a direct connection between the model coefficients and daily conditional readmission probabilities. With such a connection, further calculations besides significance tests on intervention can be easily performed. This new method is named Probabilistic Model Based Method (PMBM).

Results: Results from simulation studies show that the proposed PMBM can recover the dynamic process for intervention and readmission very well. Comparison on parameter estimations with true values further validates the estimation process. Power analysis based on 1000 simulations indicates that our new method is essentially equivalent to the time-dependent survival model.

Conclusions: Our studies have clearly show that a testing or evaluation method ignoring the time dependent feature of a post-discharge intervention may be severely biased and misleading. After incorporating the time-dependent feature, our new method is as efficient as the time-dependent survival analysis but much easier to interpret. More importantly, our fitted model provides rich information for health care authorities to evaluate their intervention procedure and explore best intervention strategies. Furthermore, our models can be easily extended for other post-discharge intervention effect evaluation problems, even with time-dependent covariates.

[PO3.04]

Estimation of the transition probabilities conditional on repeated measures in Multi-state models

Gustavo Soutinho[1], Luís Meira-Machado[2]

[1] *EPIUnit, ISPUP - Instituto de Saúde Pública do Porto, Portugal*
[2] *Department of Mathematics, University of Minho, Portugal*

Background: The topic of joint modelling of longitudinal and survival data has received remarkable attention in recent years. In cancer studies for example, these models can be used to assess the impact that a longitudinal marker has on the time to death or relapse. Analyses of such studies, in which individuals may experience several events, can be successfully performed by multi-state models. In these models the estimation of the transition probabilities is of particular interest, since they allow for long-term predictions of the process.

Objectives: The goal of this work is to introduce feasible estimation methods for the transition probabilities conditionally on covariates observed with repeated measures.

Methods: A joint multi-state model can be decomposed into a linear mixed sub-model for the longitudinal data and a multi-state model for the event history data both linked by a function of the shared random effects from the longitudinal sub-model. The use of specific samples of data, consisting of subjects occupying a given state at a particular time (also known as landmarking), allows the adaptation of existing methods for joint modelling of longitudinal and survival data to estimate the transition probabilities for each individual while taking into account the trend of the repeated measures.

Results: Results of the simulation studies confirm the superiority of the proposed estimator when compared to methods that do not take in consideration the effect of the covariate on the estimated transition probabilities or do not assume all the existence of repeated measures (Breslow estimator). In contrasts to alternative methods, the proposed methods show the effect of the longitudinal marker in each transition taking in consideration its trend and its values.

Conclusions: Results obtained from simulation studies and in the real data application confirmed the good performance of the JMLM estimator, providing accurate estimated transition probabilities. The proposed method also demonstrated to have more sensibility to reflect the evolution of the longitudinal measures when comparing to the Breslow's based method which only makes use of a single value of the covariate.

[PO3.05]

Summarising a Global Interaction Test, Quantitative Interaction Terms and Qualitative Interaction Tests Summary

Marion Procter, Christine Campbell, John Townend

Frontier Science Scotland (Ltd), United Kingdom

Background: In clinical trials, stratification factors expected to impact on prognosis are often used in randomization to ensure balance between treatment groups. When stratification factors are used, they should be accounted for in the statistical analysis. Potentially, there may be statistically significant and clinically meaningful stratification factor by treatment interaction effects. In such a scenario, as well as describing the hazard ratio from the intent-to-treat population, it may be of interest to describe the hazard ratio at particular levels of a stratification factor, taking into account the significant interaction terms.

Objectives: Present a table shell which can be used to summarise a global interaction test, quantitative interaction terms and qualitative interaction tests.

Methods: The statistical analysis illustrated considers a Cox proportional hazard model (Cox 1972). Here, the covariates are the stratification factors and are categorical. A Cox model with treatment group, all covariates, and all 2-way covariate-by-treatment interaction terms is compared to a model which excludes the interaction terms. A global interaction test is applied based on the change in model deviance. If the global interaction test is significant, the cause and type of the interaction(s) will be investigated. Stepwise backwards selection will be performed on the full model. To help clinically interpret any significant interactions, for each of these in turn: a model is fitted including the treatment group, all covariates and the covariate-by-treatment interaction in question.

Hazard ratios at each level of the covariate in question are then estimated from this model. Additionally, for each of the significant interaction terms a test will be carried out for qualitative interaction (Gail and Simon 1985).

Results: We developed a table shell which can be used to summarise this approach.

Conclusions: When there are statistically significant and clinically meaningful stratification factor by treatment interactions, it may be of interest to describe the hazard ratio at a particular level of a stratification factor. We considered categorical stratification factors as covariates in a Cox proportional hazards model. We developed a table shell which can be used to summarise a global interaction test, quantitative interaction terms and qualitative interaction tests.

[PO3.06]

Application of machine learning algorithms for recurrent musculoskeletal pain prediction

Anna Giczewska[1], Steven George[1], Nrupen Bhavsar[2], Brooke Alhanti[1], Charles Thigpen[3]

[1] *Duke Clinical Research Institute, United States*
[2] *Duke University School of Medicine, United States*
[3] *ATI Physical Therapy, United States*

Background: Electronic Health Record (EHR) data are increasingly being used in biostatistical research to improve patient care. EHR data are often used to create predictive models to inform both patient care and outcomes, particularly in applications with recurring interactions with the healthcare system. Musculoskeletal (MSK) pain is highly prevalent, associated with frequent healthcare utilization, and can lead to a decreased range of motion or disability. As MSK pain conditions are often recurrent, accurate prediction of multiple episodes of care could allow for more effective care management. Therefore, there is value in the accurate prediction of recurrence of MSK pain episodes. EHR data has the potential to identify high-risk patients in this population, however best prediction practices still remain unknown.

Objective: The main goal of this analysis is to compare machine learning (ML) algorithms to identify the best performing method for the prediction of recurrent MSK pain using EHR data.

Methods: Data was obtained from the ATI Patient Outcomes Registry. Patients who were at least 14 years old, indicated MSK pain, and had at least three separate episodes of care (defined as a group of related visits with <90 days between each encounter) were included in the analysis. Candidate predictor variables were identified using clinical expertise and univariable logistic regression models. The LASSO and random forest ML algorithms were compared. Performance statistics (discrimination and calibration) were calculated using a 10-fold cross validation method. All analyses were conducted using R (R Foundation for Statistical Computing; Vienna, Austria).

Results: The analytic dataset had 574,351 patients where 86% of patients had a single episode of care and 14% had multiple care episodes. LASSO and random forest performed similarly (AUC=0.59), but LASSO exhibited better calibration. Key predictors from the LASSO models included insurance type, arthritis and primary body part. Random forest model included body mass index, insurance type and age category as important factors.

Conclusion: The LASSO method was a better prediction method than the random forest in this application. LASSO and random forest methods selected different variables as key predictors. Guidelines on the selection optimal ML algorithms are needed in this rapidly growing field.

[PO3.07]

Missing Data: Imputation methods for survival analysis

Anna Lecka

Parexel, Poland

Missings are often found among longitudinal data because some subjects may abandon the research or be unable to be taken some measures. Considerable amount of solutions regarding missing data is based on imputation, however most frequently used method is Multiple Imputation. In SAS there are two procedures that use Multiple Imputation: MI and MIANALYZE. The latter one can be considered as an extension to MI procedure. After using MI procedure in SAS, it is possible to invoke dataset with imputed data. Procedure MIANALYZE can create statistical results with estimation of new parameters from imputed dataset. In created model missing data is imputed for three independent variables: size of tumor in breast, number of progesterone receptors, number of estrogen receptors. Purpose of the study is to show how to deal with missing data in survival analysis and present two procedures available in SAS that impute data using Multiple Imputation method.

[PO3.08]

On the use of neural networks for survival models with censored data

Elvire Roblin[1], Paul-Henry Cournède[2], Stefan Michiels[3]

[1] Villejuif, Department of Biostatistics and Epidemiology, Gustave Roussy & Oncostat U1018, Inserm, University Paris-Saclay, labeled Ligue Contre le Cancer & MICS - Laboratory of Mathematics and Computer Science, CentraleSupélec, France
[2] University of Paris-Saclay, MICS - Laboratory of Mathematics and Computer Science, CentraleSupélec, France
[3] Villejuif, Department of Biostatistics and Epidemiology, Gustave Roussy & Oncostat U1018, Inserm, University Paris-Saclay, labeled Ligue Contre le Cancer, France

Background: Recently Artificial Neural Networks have been increasingly used to model complex patterns and prediction problems. As they are flexible non-linear models, they are particularly relevant when a high number of candidate covariates and complex interactions are to be evaluated. In survival analysis, handling censoring is a key point.

Objective(s): The objective of this work is to compare neural network models with time-to-event data, using specific ways to handle censored observations such as pseudo-observations and loss functions, and to study their operating characteristics in a simulation study.

Method(s): We compared survival models based on neural networks with different loss functions: Cox-MLP (Kvamme et al., 2019) uses a special loss based on a case-control approximation. Lee et al. (2018) proposed DeepHit, a model that estimates the probability mass function and combines a log-likelihood with a ranking loss. DNNSurv (Zhao et al., 2019) circumvent the problem of censoring by using pseudo observations. These 3 models allow non-proportional effects. We used random survival forests by Yshwaran et al. (2008) as a benchmark. We investigated the prediction ability of these models using data simulated from the AFT model proposed by Friedman et al. (2001), with 3 different censoring rates (20%, 40% and 60%). For a given rate, we simulated 100 datasets of 1,000 samples and 20 variables each, with pairwise interactions and non-linear effects of random subsets of these variables. Models were compared using the concordance index, dynamic AUC and the Brier Score. We applied the methods to the METABRIC breast cancer data set, including 1,960 patients (divided into 1,470 training and 490 test samples), 6 clinicopathological covariates and the expression of 203 genes.

Results: Detailed results from the simulation study will be shown at the conference. The results on the METABRIC test set showed comparable performances in terms of 5-year and 10-year concordance indices.

Conclusions: We present different models based on neural networks to perform survival analysis. Pseudo-observations may be more appropriate in the context of high censoring.

[PO3.09]

Enhancing the optimal cut-point identification of continuous covariates for predicting survival outcomes

Paola M. V. Rancoita

University Centre of Statistics in the Biomedical Sciences (CUSSB), Vita-Salute San Raffaele University, Italy

Background: In the clinical practice, often decision rules (e.g. for treatment decision) are based on categorized clinical or biological continuous variables. The choice of the optimal cut-point is fundamental to guarantee the best prediction of the patient outcome. For survival outcome, few methods exist in the literature, such as the survival tree and the method of Contal and O'Quigley (1999), which is based on the log-rank statistic. However, they may be prone to overfitting in some situations, since they are based on an exhaustive search of the optimal cut-point among the observed data.

Objective: A new methodology is proposed for optimal cut-point identification in case of survival outcome prediction, overcoming the limitations of current methods.

Method: Assuming a proportional hazards model, the proposed method estimates the cut-point as a parameter of the Cox's regression model. The numerical maximization of the partial likelihood is not straightforward, since it may present many local maxima. Moreover, if the global maximum is not found, some local maxima could be higher than others that are instead closer to the global maximum. Thus, several maximization procedures are defined with a grid of initial values of the parameters and different criteria for selecting the maximum.

Results: The best version of the proposed method was chosen using an extensive simulation study, mimicking several challenging aspects that often characterize real data (e.g. by varying the sample size, the distribution of the continuous covariate, the percentage of censored data and the distance among the survival curves of the two groups derived from the categorized covariate). When comparing with other methods, the one of Contal and O'Quigley usually achieved the worst performance. The proposed method performed better than the survival tree in several scenarios especially the challenging ones (e.g unbalanced groups with high percentage of censored data or asymmetrically distributed covariate). Even using real training-test sets of breast and other cancers, the proposed methodology outperformed the survival tree providing cut-points giving better results on the corresponding test set.

Conclusions: Overall, the proposed methodology provides a better estimation of the optimal cut-point in the survival setting than other methods in the literature.

[PO3.10]

Using multiple imputation to impute missing cancer stage in the estimation of relative survival

Catherine Welch[1], John Broggio[2], Mark Rutherford[1], Paul Lambert[1], Michael Sweeting[1]

[1] Department of Health Sciences, University of Leicester, United Kingdom
[2] Cancer Analysis, Public Health England, United Kingdom

Background: Cancer registry data are commonly utilised to produce national cancer survival statistics for different cancer sites. Survival estimates are often stratified by cancer stage, an important prognostic variable for cancer survival. A proportion of patients missing stage information are commonly excluded from stage-specific survival estimates. Analysing only patients with complete records can bias inferences if the complete sample is not representative of the whole dataset. Multiple Imputation (MI) can be preferable to impute missing values because MI can achieve unbiased inferences if applied correctly.

Objectives: We investigate different algorithmic and statistical approaches to handle missing cancer stage for patients from the English National Cancer Registry and Analysis Service (NCRAS) to estimate relative survival and compare estimated 1 and 5-year net survival estimates under each approach.

Methods: We extracted data on patients in NCRAS diagnosed with colorectal or breast cancer from 2012 to 2017 and followed-up until 2018. Relative survival methods commonly used in registry studies estimate excess mortality in the cancer patients compared with population mortality rates.

We consider three different approaches to handle missing cancer stage when the relative survival model is non-parametric (Pohar-Perme) or flexible parametric (Royston-Parmar):

1. Complete records analysis – only analyse patients with no missing cancer stage values;
2. "Missing is late" - replace missing cancer stage with stage 4 because most of the missing values are believed to be stage 4;
3. MI –impute missing cancer stage using a multinomial logistic regression imputation model conditioning on all variables in the substantive model plus the Nelson-Aalen estimate of the all-cause cumulative hazard and all-cause event indicator.

We compare net survival estimates when including background mortality rates in the imputation model and adding auxiliary information on patient, tumour and treatment characteristics recorded in the registry plus previous hospital admissions information from linked databases, e.g. Hospital Episode Statistics.

Results: We compare net survival estimates to assess the robustness of including more auxiliary variables in the imputation model. Adding auxiliary variables can reduce bias and make MI assumptions more plausible.

Conclusion: We will advise on sensible approaches to MI in a relative survival context.

[PO3.11]

Multilevel nonlinear joint model to characterize the tumor response variability under immunotherapy

Marion Kerioui[1], Solène Desmée[2], François Mercier[3], Jin Jin[4], René Bruno[5], Jérémie Guedj[1]

[1] Université de Paris, INSERM IAME, Paris, France
[2] Université de Tours, Université de Nantes, INSERM SPHERE UMR 1246, Tours, France
[3] Biostatistics, Roche Innovation Center Basel, Switzerland
[4] Clinical Pharmacology, Genentech Inc., South San Francisco, CA, United States
[5] Clinical Pharmacology, Roche/Genentech, Marseille, France

Joint models are increasingly used to characterize the relationship between biomarker kinetics and a time-to-event. Mechanistic models relying on nonlinear mixed effects model allow to describe complex tumor kinetics and provide biologically interpretable parameters. In previous work in metastatic bladder cancer patients, nonlinear joint modelling allowed to characterize the association between tumor size and survival, and to develop efficient tools to predict individual survival under immunotherapy treatment. Nevertheless, studies report higher variability of tumor responses under immunotherapy than chemotherapy. Thus, considering the kinetics of each target lesions of one patient and integrating the impact of the location of each lesion to capture the tumor kinetics in different organs could improve the survival predictive ability of the model. Here we aimed to develop a multilevel nonlinear joint model to describe the lesions kinetics depending of their location and their association with patient survival, depending on the treatment. We investigated data from a phase 3 clinical trial (IMVigor211) of 900 advanced urothelial carcinoma patients randomized between immunotherapy treatment (Atezolizumab) and chemotherapy control arms. We built a multilevel nonlinear joint model, exploring effect from the treatment arms and the location sites on patient response. An additional level of random effects was implemented to capture correlation in the kinetics of patient's lesions and estimate the intra-patients variability. The hazard function depended on the tumor dynamics in each organs, and we assessed the gain in term of survival predictive ability, comparing to the original model. The data analysis demonstrated a stronger decrease of the tumor size due to treatment effect under chemotherapy but a more durable response under immunotherapy. We found significantly different tumor dynamics from one organ to another. Notably, we notice a higher growth of the lesions in the liver, and a stronger treatment effect in the lymph nodes. The association between tumor kinetics and survival turns out to be sensitive to treatment arm and location of the lesion. This work opens the way for the development of more complex joint models allowing to capture variability of treatment responses and provide more precise survival predictions.

[PO3.12]

The Influence of Repeat Pregnancies on the Generalizability and Accuracy of Risk Prediction Models for Time-to

Sonia Grandi[1], Kristian Filion[2], Jennifer Hutcheon[3], Graeme Smith[4], Robert Platt[1]

[1] *Bethesda, NICHD, United States*
[2] *Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Canada*
[3] *University of British Columbia, Canada*
[4] *Queen's University, Canada*

Background: In perinatal epidemiology, the choice of development cohort for the creation of risk prediction models is complicated by the potential contribution of more than one pregnancy per woman, the correlation between pregnancy outcomes, and the variation in predictor-outcome associations by parity. The manner in which repeat pregnancies influences the predictive accuracy of models for long-term outcomes including obstetrical history is unclear.

Methods: Four analytical cohorts including 1) first deliveries, 2) a random sample of deliveries, 3) all eligible deliveries per woman, and 4) all eligible deliveries and censoring of follow-up at subsequent pregnancies were created using the Clinical Practice Research Datalink to assess the influence of parity on the association between predictors and the outcome, as well as the influence of ignoring correlations between pregnancies. Using plasmode simulations, we varied the predictor-outcome association across the cohorts. Weibull models were used to estimate the parameters of the model, and predictive accuracy was assessed via bootstrap resampling. Robust variance was used to account for clustering of pregnancies within women.

Results: We found important differences in the relative contribution of predictors to the overall prediction of survival times and the discriminative accuracy of models developed in a cohort of first deliveries versus a random sample of deliveries (C-statistic: 0.70 versus 0.62; Nagelkerkle's R2: 0.08 versus 0.03). Minimal differences were found when comparing the random sample of deliveries to the all deliveries cohort. Accounting for clustering and censoring upon subsequent pregnancies had negligible influence on model performance.

Conclusions: This study suggests that a model including first deliveries has the best predictive accuracy but is not generalizable to women of varying parities. Moreover, the inclusion of repeat pregnancies may not improve the predictive accuracy of the models. Multiple models may be needed in order to improve the generalizability and predictive accuracy of models for the prediction of cardiovascular risk.

[PO3.13]

Dietary patterns as markers for lifestyle - prediction of long-term mortality

Nele Taba[1], Tõnu Esko[2], Andres Metspalu[2], Krista Fischer[3]

[1] *Institute of Molecular and Cell Biology, University of Tartu, Estonia*
[2] *Estonian Genome Center, Institute of Genomics, University of Tartu, Estonia*
[3] *Institute of Mathematics and Statistics, University of Tartu, Estonia*

Dietary patterns are known to have an effect on individuals' health-related outcomes and mortality. Nevertheless, studies investigating long-term effects have been limited either by sample size or follow-up time. Thus the long-term effect of dietary pattern on mortality and health-outcomes (such as Type II Diabetes (T2D), Coronary Artery Disease etc.) in a large cohort-based sample are yet to be investigated. To do that, one first needs to answer two questions: 1) how to create summaries of dietary patterns based on a food-frequency questionnaire (FFQ); 2) how to assess the effect of dietary patterns on health-related outcomes; and fulfill the requirement of large cohort-based sample with long follow-up.

Since single dietary items lack interpretation without context, and principal components are hard to interpret, we decided to use k-means algorithm to cluster individuals based on their answers to 17 items on the FFQ. We chose the number of clusters (k) based on how much the prediction of mortality improves by increasing the number of clusters from k to k+1. For assessing the effect of dietary patterns on health-related outcomes we used Cox proportional hazards model using age as time-scale and adjusting for gender, smoking status and education level. And lastly, for the requirement set, the Estonian Biobank suits well with >10 years of follow-up for >49 000 individuals who reported their dietary habits in the baseline FFQ.

We found that the dietary clusters differ clearly by their risk of mortality (hazard ratio (HR) ranging up to 1.42) and other health outcomes (HR up to 1.8 for T2D); but not only: there are clear discrepancies in the phenotypic background of the clusters that were formed solely based on dietary data. The latter indicates that aside affecting directly one's health, dietary pattern can be viewed as a marker for general lifestyle of an individual.

[PO3.14]

Comparison of Semiparametric and Parametric Survival Models with the Use of Time-Dependent ROC Curve

Jadwiga Borucka

Statistical Programming, Parexel, Poland

Background: In the analysis of cancer diseases, overall or recurrence-free survival is often the primary endpoint and Cox model is one of the most commonly used methods to analyze simultaneous effect of several prognostic factors on the risk of the event. Parametric models are chosen relatively rarely in medical data analysis. In order to consider parametric models as an alternative to the Cox model some measures need to be defined to compare semiparametric and parametric models.

Objectives: The current paper presents an attempt to use time-dependent ROC curve proposed by Heagerty, Lumley and Pepe to compare discriminative power of Cox model and selected parametric survival models. The aim of the paper is to answer the question whether it is possible to indicate a model which has considerably higher discriminative power for some pre-defined timepoints on the basis of time-dependent ROC curve.

Methods: The outcome of interest is the overall survival measured as time in months since the disease diagnosis for breast cancer patients. Subjects who are alive at the last contact date are censored at the study cutoff date. For the overall survival, Cox model and selected parametric models are estimated. For each model time-dependent ROC-curve is plotted for several pre-defined timepoints and AUC corresponding to each ROC is calculated. Data used for analysis comes from SEER database. Calculations are performed in SAS Base 9.3 and R 3.0.3.

Results: Obtained results show that considered models have similar discriminative power for the overall survival in breast cancer patients, however it is possible to indicate some of them which perform slightly better: log-logistic and gamma models. Cox, exponential and Weibull models represent a little worse discriminative power in this case.

Conclusions: It can be concluded that for the analyzed sample gamma model is the best choice in terms of discriminative power for 5-, 10- and 20-year survival prediction while Cox, exponential and Weibull models represent slightly worse properties. Time-dependent ROC curve and AUC are useful tools to compare survival models of two different types: semiparametric and parametric since they have the same interpretation, regardless which model was chosen.

[PO3.15]

Variable importance metric using modern machine learning techniques for clustered data

Mohammad Kaviul Anam Khan[1], Rafal Kustra[2]

[1] *Dalla Lana School of Public Health/Biostatistics, University of Toronto, Canada*
[2] *Dalla Lana School of Public Health, University of Toronto, Canada*

In modern biomedical research, a large number of patient characteristics and history can be obtained to understand disease progression. Using the novel machine learning techniques, the patient prognosis or the risk of diseases can often be predicted more accurately than with standard statistical methods. Their applications in biomedical domain can be hampered, due to the need to understand the predictive and causal effects of input variables. Breiman (2001) proposed Variable Importance Metric (VIM) for Random Forests model to identify important variables for prediction. The VIM provides the ranking of the covariates based on their impact on changes in prediction error. It can be shown however, that VIM approach can be well understood when applied to the conventional linear models, generalized linear models and generalized additive models which offers a basis for their intepretation, including in causal framework. For example, Fisher et.al. (2018) showed that the VIM can also be represented in terms of expected conditional average treatments effect for a simple binary treatment. However, the causal interpretation of VIM is still not available in a more complex treatment setting.

The objective of this is study is demonstrate the decomposition the Variable Importance Metric (VIM) proposed by Breiman (2001) into a causal parameter of interest for a treatment variable with more than two categories and also for continuous treatments.

We extend the framework provided by Fisher et.al. (2018) for binary treatment to treatments with more than two categories and also for continuous treatment, which allows us, for example to study cluster-specific causal effects. The decomposition is also an extension of the decomposition of variance of potential outcomes shown by Chen et.al (2019). The VIM can be presented as a multiplication of treatment variance and the expected conditional average treatment effect squared.

We apply our VIM decomposition to long term care facilities data to estimate facility specific causal effect on cognitive decline in elderly. The usefulness of such decomposition is that the predictions from complex machine learning algorithm. The important aspect of the decomposition is that it is method agnostic.

PO4: Big Data / Machine Learning

Wednesday, 26 August 2020, 10.20 - 11.50

Chairs: **Małgorzata Ćwiklińska-Jurkowska**
Nicolaus Copernicus University, Torun, Poland

Lara Lusa
University of Primorska, Koper/Capodistria, Slovenia

[PO4.01]

Comparing survival functions with interval-censored data in the presence of an intermediate clinical event

Jinheum Kim[1], Chung Mo Nam[2], Sohee Kim[3]

[1] *Department of Applied Statistics, University of Suwon, Republic of Korea*
[2] *Department of Preventive Medicine, Yonsei University College of Medicine, Republic of Korea*
[3] *Department of Clinical Statistics, Yuhan Corporation, Republic of Korea*

In the presence of an intermediate clinical event, the analysis of time-to-event survival data by conventional approaches, such as the log-rank test, can result in biased results due to the length-biased characteristics. In this talk, we extend the studies of Finkelstein and Nam & Zelen to propose new methods for handling interval-censored data with an intermediate clinical event using multiple imputation. The proposed methods consider two types of weights in multiple imputation: 1) uniform weight and 2) the weighted weight methods. Extensive simulation studies were performed to compare the proposed tests with existing methods regarding type I error and power. Our simulation results demonstrate that for all scenarios, our proposed methods exhibit a superior performance compared with the stratified log-rank and the log-rank tests. In the absence of intensive iterations, our proposed methods show a superior performance compared with the stratified log-rank and the log-rank test regarding type I error and power. Finally, data from a randomized clinical study to test the efficacy of sorafenib/sunitinib vs. sunitinib/sorafenib to treat metastatic renal cell carcinoma were analyzed under the proposed methods to illustrate their performance on real data.

[PO4.02]

Fetal health risk assessment around the onset of labour: the role of an hour of cardiotocography monitoring

Antoniya Georgieva

Nuffield Dep Women's and Reproductive Health, University of Oxford, United Kingdom

Background: Reliable, objective monitoring of the fetal health during childbirth remains a massive unmet clinical need. Continuous cardiotocography (CTG) which displays the fetal heart rate and uterine contractions and is visually evaluated by clinicians, remains the maystray tool worldwide. The subjective assessment of a short CTG trace at 'admission' to the labour ward has not shown benefits in clinical trials and is debated.

Objective: Using a large dataset of routinely collected maternity and CTG data, to develop and evaluate statistical prediction models to assess objectively the risk (just before or soon after the onset of labour) for severe fetal compromise.

Methods: A total of 11835 births were analysed within a nested case-control design: 11481 healthy infants, delivered spontaneously with arterial cord pH>7.15, and not requiring resuscitation or admission to the neonatal intensive care unit (NICU); and 354 severely compromised infants (stillbirth or neonatal death; seizures; neonatal encephalopathy; resuscitation followed by NICU admission for >48hrs).

Three logistical regression models (multivariable fractional polynomials, implemented in R) were proposed: (1) Basic Clinical Model (BaseClinM) including: gestation, maternal temperature, maternal age (continuous variables); nulliparity, preeclampsia, and established labour (binary variables); (2) Joint Model (JointM) including the same clinical risk factors as BaseClinM, but also computer-based CTG characteristics of the first available hour of CTG monitoring: baseline, short term variability, maximal decelerative capacity, number and amplitude of accelerations; and number of decelerations; (3) Joint Model to include also the presence of thick meconium (JointM_Mec).

The sensitivity (Se) of the models only for specificity (Sp) of 99% is reported here.

Results: We obtained Se of: 10.7% for BaseClinM; 13.6% for JointM; and 18.7% for JointM_Mec. About half of the CTGs were taken before established labour; 70% in < 2hrs from the onset of established labour, and nearly 90% < 5.5hrs.

Conclusions: We developed for the first time an objective data-driven risk assessment around the time of onset of labour. We conclude that a substantial proportion of severely compromised infants could potentially be detectable around the onset of labour with a very low false positive rate (1%), allowing targeted interventions and/or appropriate consultation with the parents.

[PO4.03]

An Innovative Study to evaluate the treatment efficiency of Percutaneous Nephrolithotomy by DEA

Adnan Karaibrahimoglu[1], Sefa Alperen Ozturk[2], Ayhan Golcukcu[3]

[1] *Department of Biostatistics & Medical Informatics, Turkey*
[2] *Department of Urology, Turkey*
[3] *Department of Statistical Research, Turkey*

Background: Urinary stone disease is a common problem and its prevalence is an increasing problem in recent years. Depending on the location and size of the stone, according to the EAU guidelines, the treatment of renal stone 2 cm or more is surgical and percutaneous nephrolithotomy (PNL).

Objective(s): The aim of this study was to measure the treatment efficiency of PNL patients by Data Envelopment Analysis (DEA), as a novel method with minimized output model.

Method(s): This is a retrospective study in which 152 patients with kidney stones over 2 cm were enrolled. Some biochemical perioperative and postoperative characteristics of patients (Sodium (NA), potassium (K), creatinine (CRE) and hemoglobin (HB)) were analyzed. Contrary to classical DEA theory, a novel output minimization model with uncountable output was proposed, and then Malmquist analysis was used to find out the difference between patients according to gender and age.

Results and Conclusions: The decomposition of index revealed that no remarkable difference within decision making units clearly (within groups $IE^{Female\text{-}Male} = 1.069$ and $IE^{<40\ -\ 40+} = 0.961$), however between age and gender groups the results were quite different. The Malmquist results showed that the females' response to treatment was extremely different than males', and the patients' respond over 40 years was also slightly different to the patients' under 40 years (between groups $IF^{Female\text{-}Male} = 21255.09$ and $IF^{<40\ -\ 40+} = 0.13$). This is a new research concept, and it can be suggested to develop alternative treatment modalities for PNL by gender, age or various groups.

[PO4.04]

Sample size planning and optimal design for estimating regression-based reference values

Francesco Innocenti, Frans Tan, Math Candel, Gerard van Breukelen

Department of Methodology and Statistics, Maastricht University, Netherlands

Background: Normative studies are needed to obtain reference values (norms) for comparing patients with the reference population in terms of relevant clinical measures, such as scores on mental tests or physiological variables (e.g. blood pressure), taking into account possible age or gender differences. To prevent mistakes in patients' assessment, the stability of norms is important. Stability can be achieved by carefully planning both the size and the design of the sample from which the norms are derived (normative sample).

Objective(s): First, sampling variance formulae for Z-scores and percentile rank scores, two norm statistics that are typically used in normative studies, are derived. Second, the optimal joint distribution of predictors in the regression-based norming model (i.e. the optimal design) is derived for five regression models with a quantitative (e.g. age) and a qualitative predictor (e.g. sex). Third, a procedure to determine the required size of the normative sample for the optimal design is proposed.

Method(s): Sampling variance formulae for Z-scores and percentile rank scores are derived using the delta method. Their bias and that of their estimators are investigated through simulation studies. These variance formulae are subsequently used to derive the optimal design. Since the optimal design depends on the assumed regression model, competing designs are compared in terms of efficiency relative to the optimal design to find the design that yields the highest worst-case relative efficiency across all considered models.

Results: For efficient estimation, the normative sample need not be representative of the reference population with respect to the distribution of the predictors. For instance, designs including ten or more different age levels, typically used in normative studies, can be very inefficient compared with the optimal design (e.g. requiring 60% more persons than the optimal design). The required sample size depends on the norm statistic of interest, because Z-scores require smaller sample sizes than percentile rank scores.

Conclusions: The stability of norms can be improved by drawing normative samples as prescribed by the optimal design under the chosen regression-based norming model, or by the design with the highest worst-case relative efficiency in case of uncertainty about the underlying model.

[PO4.05]

Estimating Diagnostic Test Accuracy Whilst Adjusting for Imperfect Interrater Agreement in the Reference Stand

Enzo Cerullo[1], Alex Sutton[2], Nicola Cooper[2], Terence Quinn[3]

[1] Department of Health Sciences, University of Leicester, United Kingdom
[2] Department of Health Sciences, Complex Reviews Support Unit, University of Leicester, United Kingdom
[3] Institute of Cardiovascular and Medical Sciences, Complex Reviews Support Unit, University of Glasgow, United Kingdom

Background: Diagnostic accuracy studies typically assume that the sensitivity and specificity of the reference test is perfect. Latent class models (LCMs) allow estimation of the true accuracy of a reference test, and comparisons between the performance of the reference test and the index test(s). A potential source of error is imperfect interrater agreement (IRA).

Objectives: (1) To develop a method to obtain imperfect reference standard-adjusted accuracy estimates for the meta-analyses of diagnostic accuracy studies, and (2) to apply the methods developed to the area of Clinical Dementia diagnosis.

Methods: A Cochrane meta-analysis on the Mini-Mental State Examination (MMSE) was reanalysed using Bayesian LCMs, with prior distributions constructed based on relationships between the observed prevalence, misclassification probabilities, and IRA (kappa-statistic, k). To determine the degree of IRA, a systematic review and meta-analysis was carried out for the reference tests used in the diagnosis of clinical dementia.

Results: Most studies reported k between 0.4 and 0.8. Using the novel Bayesian methods, it was found that for the 24 point MMSE cut-off, the estimated sensitivity was between 18.1% to 5.6% lower than the original study estimate for k=0.4 and 0.8 respectively, and 8.4% lower for k=0.65. However, little bias was found for the specificity. The reference tests were found to be between 4.0% and 1.1% less sensitive than the MMSE for k=0.4 and 0.5, respectively. For the 25 cut-off analysis, very little evidence of bias was found, and the reference tests were slightly more sensitive than the MMSE. For both cut-off points, the reference tests were consistently more specific than the MMSE.

Conclusions: Imperfection in the reference tests due to imperfect IRA can result in biased estimates of diagnostic accuracy, and adjusting for IRA provides better estimation of diagnostic test accuracy.

[PO4.06]

Predicting patient engagement in IAPT services: A statistical analysis of electronic patient records

Alice Davis

Department of Mathematical Sciences, University of Bath, United Kingdom

Background: The Mental Health Foundation found that two-thirds of people say they have experienced a mental health problem and that collective mental health is deteriorating. Across England, 12% of all Improving Access to Psychological Therapy (IAPT) appointments are missed, and on average around 40% of first appointments are not attended, varying significantly around the country. In order to intervene effectively, it is important to target the patients who are most likely to miss their appointments.

Objective: This research aims to develop and test a generalised mixed effect model to predict whether an IAPT patient will attend their first appointment.

Methods: Data from 19 adult IAPT services were analysed in this research. A multiple logistic regression was used at an individual IAPT service level to identify which patient, appointment and referral characteristics are associated with attendance. These variables were then used in a generalised linear mixed effects model (GLMM). We allowed for random effects in the GLMM for variables where we observed high service to service heterogeneity in the estimated effects from service specific logistic regressions.

Results: We find that patients who self-refer are more likely to attend their appointments. The older a patient is, the fewer the number of previous referrals and consenting to receiving a reminder short message service are also found to increase the likelihood of attendance.

Conclusions: Our model is expected to help IAPT services identify which patients are not likely to attend their appointments by highlighting key characteristics that are associated with attendance. This analysis will help to identify methods IAPT services could use to increase their attendance rates.

[PO4.07]

Adaptive lasso approaches for pharmacovigilance signal detection: new penalty weights for variable selection

Emeline Courtois, Ismaïl Ahmed, Pascale Tubert-Bitter

U1018, Inserm, France

Background: Methods used to generate safety signals from pharmacovigilance spontaneous reporting databases classically rely on disproportionality analysis of counts aggregating patients' reports for each drug-adverse event pair. In recent years, alternative methods have been proposed to analyze individual spontaneous reports relying on penalized multiple logistic regression, typically lasso regression (LR), or propensity score in high dimension (PSHD).

One difficulty with LR is to optimally choose the regularization parameter in a variable selection context. In particular, cross-validation leads to the selection of too many covariates. On the other hand, PSHD-based approaches are computationally very intensive.

Objectives: We present a new automated signal detection strategy based on the adaptive lasso, which aims at improving the guidance of the variable selection operated by the lasso through adaptive penalty weights (PWs) specific to each covariate.

Methods: We propose two new PWs derived from (i) a LR for which the regularization parameter is chosen using the Bayesian Information Criterion (BIC) (ii) the Class Imbalanced Subsampling Lasso (CISL) algorithm, an extension of the stability selection adapted to signal detection in pharmacovigilance. These PWs are then incorporated in a LR with the BIC for choosing the regularization parameter.

We compare the different versions of our approach to (i) more classical implementations of adaptive lasso, (ii) LRs considering cross-validation, BIC or permutations for choosing the regularization parameter, (iii) CISL, (iv) several PSHD-based approaches. This comparison is performed through extensive simulations and an empirical study conducted on the French pharmacovigilance database using a large reference signal set pertaining to drug-induced liver injuries. For both studies, we evaluate the methods in terms of false discoveries and sensitivity in a lesser extent.

Results: In the simulations and the application, both proposed PWs show equivalent or better performances than the other competitors, with an advantage for the CISL-based PWs. CISL and LR using BIC are solid alternatives.

Conclusions: Our adaptive approaches are promising methods for signal detection in pharmacovigilance. They are easy to implement and less computationally intensive than PSHD methods. Although we cannot rely on test theory, our approaches show a low and stable False Discovery Rate in all simulation settings.

[PO4.08]

Markovchart: an R package for cost-optimal patient monitoring and treatment

Balazs Dobi, Andras Zempleni

Faculty of Informatics, University of Debrecen, Hungary

Background: The cost-effectiveness of the therapeutic regimens and monitoring of chronic diseases is an important factor for every party involved. To be able to tune such a setup properly, one needs to model a complex process involving disease progression, treatment effects and the costs emerged. Similar problems also arise in industry, where cost effectiveness and monitoring have well developed tools and theory.

Objective: To solve the above problem we set out to generalise control charts used in industrial statistics in a way that allowed their application to patient monitoring. This required several generalisations, as contrary to assumptions in traditional control chart theory, here, we needed to allow random shift sizes (i.e. disease progression), random repair (i.e. treatment effect) and random sampling times (i.e. patient non-compliance).

Methods: The basis of our work were simple X-bar charts with a sample size of 1 (representing the monitoring of a single patient at a time). The task required the modelling of the patient's state (i.e. disease level), which was carried out using a Markov chain-based model. This allowed us to calculate the stationary distribution of the Markov chain which became the basis of cost estimation. Disease progression, treatment effects, patient non-compliance and the costs involved are modelled using different distributions and functions which can be fine-tuned to the problem at hand. The methods were implemented in the R programming language as the Markovchart package, which was tested using different parameter setups, simulations and also on real-world data involving patients with diabetes.

Results: Testing showed that the model is capable of showing non-trivial relationships between parameters and the resulting average costs and cost standard deviations. Applications showed that the method can accurately model a complex system involving different therapies and can help in designing cost-efficient treatment regimens.

Conclusions: The creation of cost-optimal treatment and monitoring setups of patients can be facilitated using industrial statistics. Mathematical theory concerning a Markov chain-based cost-optimal control chart model was established and implemented in the R programming language as the Markovchart package. The model was successfully applied to real-world data of diabetic patients.

[PO4.09]

Modeling and Mapping Low Birth-Weights: Role of Interactions among Determinants and Spatial Adjustment

Rubaiya, Mohaimen Mansur

Institute of Statistical Research and Training, University of Dhaka, Bangladesh

Background: Low birth weight is a long-standing public health problem in Bangladesh. It is linked to under-nutrition and consequent health problems in children at later ages, even to premature deaths. This paper develops a statistical model which is able to identify important health-related and socio-economic risk factors of low birth weight in Bangladesh and provide reliable district-level estimates of prevalence of low birth weight.

Objective: The objective of the paper is two-fold. The first is to identify vulnerable groups of mothers who are at high risk of giving birth to under-weight children. The second aim is to generate and map district-level estimates of prevalence of low birth weight in Bangladesh in order to facilitate formulation of regional interventions and policies.

Methods: Data from a nationally representative health survey is used for the empirical analysis. A machine learning method known as regression tree is combined with classical linear regression to build a model with interactions among possible risk factors of low birth weight. Variants of the model are then developed by making two types of adjustments: a within-district correlation and a between neighboring district association among individuals. Finally, individual-level models are compared for prediction accuracy via cross-validation and the best model is used to generate district-level estimates of low birth weight.

Results: The Regression tree identifies interactions among risk factors of low birth weight that appear statistically significant in the regression model. In particular, a baby which is not born through C section is on average heavier than a cesarean baby, and babies who are from Rangpur division and born in poor or middle-class households have lower birth weight, on average. Adjusting for district-level clustering effects improve predictive accuracy, but allowing for more structured spatial correlation deteriorates model performance. The district-level prevalence estimates from the best fitted mixed effects model show that Rangpur, Kurigram, Nilphamari and Gaibandha are the four districts most vulnerable to low birth-weights.

Conclusion: Findings will be useful for identifying demographic and socio-economic risk factors that are closely associated with low birth-weight in Bangladesh. Regional estimates will be useful to policy makers for designing area-specific intervention programs to improve birth weight.

[PO4.10]

A dynamic treatment regime model to enhance adherence of patients with type 2 diabetes with Q-learning

Shotaro Maedera[1], Tomomi Takeshima[2], Ayano Chida[2], Kosuke Iwasaki[2], Nobuhiro Shojima[3], Tomohide Yamada[4]

[1] *Graduate School of Pharmaceutical Sciences, The University of Tokyo, Japan*
[2] *Milliman Inc., Japan*
[3] *Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, University of Tokyo, Japan*
[4] *King's College London, University Institute for Population Health, United Kingdom*

Background: Type 2 diabetes (T2D) is one of the most common and serious diseases. As the treatment options are many and the efficacy of a treatment depends on each individual and his/her disease status, it is difficult for patients and doctors to choose the optimal treatment. Currently the choice heavily relies on the doctors' experience and implicit knowledge. By making it explicit, and sharing and accumulating the know-how, the improvement of the quality of the treatment is expected.

Objective: To discover the optimal DTR to enhance adherence of patients with T2D with Q-learning.

Methods: We extracted 312,399 records of antidiabetic drug prescriptions and diagnostic information concerning 50,000 patients with T2D from a US claims database (MarketScan commercial). We defined rewards reflecting adherence as the number of prescriptions. Q-functions were modeled with information of the top 10 diseases most frequently diagnosed and antidiabetic drug prescription. As Q-function models, seven models were employed: linear, quadratic and cubic polynomial regression models, and 1 layer, 2 layer and two types of 3 layer Q-network models. To train the models, we performed temporal difference learning with Nesterov's accelerated gradient method. Additionally, we evaluated the learning models with the expectation of total rewards calculated with a simulation analysis based on the observational data used in this study. The learning and simulation steps were performed with TensorFlow on Python.

Results and Conclusions: In the learning process, the Q-network models exhibited better convergence than that of polynomial models, and especially 1 layer Q-network model showed the best convergence among all the models. Besides, in the simulation analysis, DTRs with all the Q-functions except for that of the linear model earned higher rewards than the conventional treatment regimen virtually reconstructed from the data did. In particular, 1 layer Q-network showed the best performance and the earned rewards was 48.19, while the conventional treatment regimen gained 9.67. This result suggests the potential of Q-learning for constructing a DTR enhancing adherence of patients with T2D. We would like to conduct further analysis to examine its effectiveness in real clinical settings.

[PO4.11]

Selecting small sets of diagnosis codes with high prediction performance in large electronic medical datasets

Thomas Cowling[1], David Cromwell[1], Linda Sharples[2], Jan van der Meulen[1]

[1] *Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, United Kingdom*
[2] *Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom*

Background: Large routine datasets of electronic medical records often contain a wide range of diagnosis codes representing clinical information. Many studies use these codes to model patient multimorbidity or comorbidity. Models incorporating more codes may predict patient outcomes better, but these models will also be more difficult to interpret, present, and apply in future studies or clinical practice.

Objective: To examine an approach to produce models that include relatively few diagnosis codes and predict outcomes almost as well as models that include large, comprehensive sets of codes.

Methods: Modelling study of national administrative hospital data (Hospital Episode Statistics) linked to mortality data in England. The populations were patients admitted for acute myocardial infarction (n=200,119), hip fracture (n=169,646), or major surgery for colorectal cancer (n=56,515) from 2015 to 2017. The outcome was death within 365 days of the date of admission or procedure. Logistic regression was used to estimate associations with all ICD-10 codes recorded for at least 0.5% of patients in the preceding year (full models). A model approximation method proposed by Harrell was used to develop models with fewer codes that explained at least 95% of variation in full model predictions (reduced models).

Results: One-year mortality was 17.2% (34,520) after myocardial infarction, 27.2% (46,115) after hip fracture, and 9.3% (5,273) after colorectal surgery. Full models included 202, 257, and 209 ICD-10 codes in these populations, respectively. Optimism-adjusted c-statistics for these models were 0.884 (95% CI: 0.882–0.886), 0.798 (0.795–0.800), and 0.810 (0.804–0.817). Reduced models included 18, 33, and 41 codes (same order) and had c-statistics of 0.874 (0.872–0.876), 0.791 (0.788–0.793), and 0.807 (0.801–0.813). The performance of the full and reduced models was also similar when measured using Brier scores. All models were well calibrated.

Conclusions: Our approach produced relatively small sets of diagnosis codes that performed comparably to large, comprehensive sets of codes, in three clinical populations. This may be useful to the many studies that measure patient morbidity using electronic medical records. Such measures will be increasingly important in the global context of population ageing and greater burdens of non-communicable disease.

[PO4.12]

A posteriori maternal dietary patterns and human milk: an application of the principal component analysis

Matteo Di Maso[1], Francesca Bravi[1], Simone Eussen[2], Stefano Sottemano[3], Enrico Bertino[3], Guido Moro[4], Alessandra Mazzocchi[5], Carlo Agostoni[5], Elisabetta Vacca[2], Adriano Decarli[1], Bernd Stahl[2], Monica Ferraroni[1]

[1] *Department of Clinical Sciences and Community Health, Branch of Medical Statistics, Biometry and Epidemiology "G.A. Maccacaro", Università degli Studi di Milano, Italy*
[2] *Danone Nutricia Research, Netherlands*
[3] *Department of Public Health and Pediatrics, Neonatal Intensive Care Unit, Università degli Studi di Torino, Italy*
[4] *Italian Association of Human Milk Banks (AIBLUD), Italy*
[5] *Department of Clinical Sciences and Community Health, Pediatric Intermediate Care Unit, Fondazione IRCCS Ospedale Cà Granda-Ospedale Maggiore Policlinico, Università degli Studi di Milano, Italy*

Background: Human milk from healthy and well-nourished mothers offers the optimal nutrition for infants. During lactation, maternal diet can modify milk composition through several metabolic pathways. For instance, maternal intake of fats reflects the composition of the milk fatty acid (FA) profile. Conversely, the intake of other nutrients is less related to the corresponding contents in milk.

Objective: To further explore the relationship between maternal diet and milk composition, we identified a posteriori dietary patterns from maternal diet and assessed their correlation with milk composition.

Method: The MEDIDIET study included 300 Italian healthy breastfeeding mothers who donated a fresh sample of breast milk. Maternal diet was assessed by a food frequency questionnaire and the corresponding nutrient intakes were computed using an Italian food composition database. Milk analyses included macronutrients (lactose, dry mass, proteins, fats, and energy) and FA profile. We performed a principal component analysis with a varimax rotation on the correlation matrix of 31 nutrients from maternal diet to identify underlying factors. Factors were selected according to scree-plot examination, factor eigenvalue >1, and interpretability. We labelled dietary patterns using nutrients reporting factor loadings ≥0.63 in absolute value which can be considered "dominant nutrients" for such pattern.

Results: We identified five a posteriori dietary patterns: 1) Vitamins, minerals and fibres, 2) Proteins and FAs with legs (i.e. deriving from animal sources), 3) FAs with fins (from fish), 4) FAs with leaves (from vegetables), and 5) Starch and vegetable proteins. As common in nutrition, observed correlations between dietary patterns and milk composition were weak. In particular, the Proteins and FAs with legs pattern correlated with saturated FAs and polyunsaturated ω-6 FAs in milk (r=0.121 and 0.119, respectively); the FAs with fins correlated with polyunsaturated ω-3 FAs (r=0.226), and the FAs with leaves correlated with monounsaturated FAs (r=0.170). Vitamins, minerals and fibres and Starch and vegetable proteins patterns did not correlate to milk macronutrients or FA profile.

Conclusion: Although correlations were weak, dietary patterns based on fats seem to be related to different milk FA profile. This consolidates existing evidence of the relationship between maternal fat intake and milk FA profile.

[PO4.13]

Bayesian within-host modelling of red blood cell dynamics and primaquine induced haemolysis in G6PD deficiency

Parinaz Mehdipour[1], James Watson[2], Saber Dini[1], Sophie Zaloumis[1], Robert Commons[3], Julie Simpson[1]

[1] *Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Australia*
[2] *Mahidol Oxford Research Unit, Faculty of Tropical Medicine, Mahidol University, Thailand*
[3] *Menzies School of Health Research, Charles Darwin University, Australia*

Background: Almost half of the malaria cases in Asia and South America are due to Plasmodium vivax (P. vivax). Primaquine is the only widely available drug that targets dormant P. vivax parasites in the liver, thereby preventing relapsing vivax malaria. The current recommended dose of primaquine can cause severe haemolysis in glucose-6-phosphate dehydrogenase (G6PD) deficient individuals.

Objectives: In this study, a compartmental within-host model for Red Blood Cell (RBC) dynamics was developed to explore the safety of primaquine regimens with escalating doses for patients with G6PD deficiency.

Method: The model captured deviations from the normal process of RBC production and destruction by modifying two main parameters, RBC lifespan and the release time of reticulocytes into the circulation. The within-host RBC model was fitted to longitudinal haemoglobin and reticulocyte measurements from 75 G6PD deficient patients using a Bayesian hierarchical framework, implemented in the software Stan. Posterior-predictive simulations of the model were used to compare the reticulocyte and haemoglobin profiles of hypothetical patients administered the current and alternative dosing regimens.

Results: Posterior predictive simulations demonstrated that a stepwise increase in daily administered primaquine dose would be relatively safe for G6PD deficient individuals. To facilitate the dissemination of the proposed stepwise dosing schemes to policy makers, a R Shiny app was created which allows users to evaluate the impact of different dosing schemes on haemoglobin and reticulocyte profiles by using slider bars to change the centre of the sampling distributions of each model parameter.

Conclusion: The results of this study suggest an alternative ascending primaquine dosing regimen to the current 0.75 mg/kg weekly dosing scheme will reduce the risk of primaquine-induced anaemia.

[PO4.14]

Shift the mean or the study variance to detect outliers in network meta-analysis

Maria Petropoulou

University of Ioannina, Greece

Background: When considering data from many trials, it is likely that some of them provide a markedly different intervention effect or an inflated (shifted) effect variance. Outlying or influential studies in a meta-analysis can affect the intervention effect or can contribute to the heterogeneity existence. Several outlier detections measures, such as deletion measures, allow a shift in the mean (study result). An alternative approach based on shift variance. Network meta-analysis (NMA) has become an increasingly popular methodology allowing multiple treatment comparisons. Despite the rapid methodological development of NMA, the issue of outlier detection in NMA provides a methodological gap.

Objectives: To provide the methodological development for detecting outliers with several measures considered study deletion, to extend the method with shift variance in NMA evidence structures and, to disseminate the developed R package NMAoutlier.

Methods: Outlier deletion measures were actually considered a study deletion in NMA. Several well-known deletion measures were extended from pairwise to NMA such as Cook distance, the ratio of variance-covariance matrix, standardized and studentized deleted residuals. Methodological extension by shift the random variance for one included study each time was provided in NMA structures. An estimate of the shift in the error variance associated with that study. A large shift may indicate a possible outlier and, if desired, can be downweighed. Methodological challenges to extend the proposed methods in a network of interventions with the presence of multi-arm studies were provided and the technicalities of model fitting were discussed. NMAoutlier R package was developed to provide the replicability of the proposed methods to meta-analysts who are interested in applying the methodology to their data. Applications with outlier detection methods were conducted to real NMA data sets.

Results: Examples in real NMA datasets with the NMAoutlier package are provided and potential outliers are identified. Downweighing potential outlying or influential studies through sensitivity analyses indicated more promise results compared to primary analyses. Identification of studies responsible for heterogeneity or inconsistency overcame the corresponding problems.

Conclusions: Considering the advantages of outlying detection, the proposed methodologies can be useful tools for outlier detection in NMA overcoming heterogeneity and inconsistency problems.

PO5: Causal Inference

Wednesday, 26 August 2020, 10.20 - 11.50

Chairs: **Agnieszka Pac**

Jagiellonian University Medical College, Chair of Epidemiology and Preventive Medicine, Krakow, Poland

Myra McGuinness

University of Melbourne, Melbourne, Australia

[PO5.01]

How to deal with multidimensional mediators? Overview and application with health literacy

Nadia Dardenne[1], Marine Paridans[2], Michèle Guillaume[2], Anne-Françoise Donneau[1]

[1] *Public Health, Biostatistics, University of Liège, Belgium*
[2] *Public Health, Nutrition, Environment and Health, University of Liège, Belgium*

Background: To evaluate the mediation effect of a multidimensional score, common methods analyse either each dimension separately, which is statistically not recommended, or study an overall score. However, this last approach assumes that such a score is validated and/or has clinical meaning.

Objectives: This study reviews various statistical methods that can be used to consider multidimensional mediators. An application to Health literacy (HL) questionnaire was realized to highlight potential different results between those methods.

Methods: A total of 328 adolescents were recruited from schools based in Liège, Belgium. Health literacy was determined using "KidsHealth KidsPoll of Health Literacy". Three dimensions of HL were evaluated: understanding, applications and motivation/interest. Mediation effects of multidimensional HL on the relationship between the school socio-economic index and two health indicators, the body mass index (BMI) and perceived health, were tested in several ways: (1) construction of a latent HL with Structural Equation Modeling (SEM) using the R package Lavaan, and (2) fitting natural effect models (NM) with all HL dimensions, with an overall but not validated HL score and with each dimension separately by means the R package medflex. Analyses were adjusted for gender and age.

Results: The relationship between the school socio-economic index and BMI was totally mediated by multidimensional HL ($p = 0.030$), the overall score ($p = 0.0081$) and for each HL dimension ($p < 0.05$) while SEM concluded to a partial mediation ($p = 0.0080$). Results from SEM and NM also differed with perceived health where multidimensional and overall HL were partial mediators ($p = 0.011$ and $p = 0.032$ respectively) with NM but totally mediator with SEM ($p = 0.001$). Only the dimension application was a partial mediator ($p = 0.014$).

Conclusion: Even if NM has the advantage of evaluating a mediation effect that goes through several dimensions while SEM is based on the construct of a latent variable, all analyses concluded to mediation effects with only slight differences between partial and totally mediator. Nevertheless, investigations should be made with other multidimensional mediators to conclude to significant differences between methods.

[PO5.02]

Causal Inference Concepts Applied to Three Observational Studies in the Context of Vaccine Development

Phoebe Nakanwagi[1], Emilia Gvozdenovic[2], Lucio Malvisi[3], Emmanuel Aris[1], Dominique Rosillon[1], Elisa Cinconze[3]

[1] *GSK Vaccines, Belgium*
[2] *Galapagos Pharma, Belgium*
[3] *GSK Vaccines, Italy*

Background: Randomized controlled trials are considered the gold standard to evaluate causal associations. Conversely, assessing causality in observational studies is more challenging.

Objective: We assessed the applicability of causal inference methods to 3 observational studies.

Methods: We evaluated the potentially causal relationship between the exposure and the outcome using the classical (Hill's criteria) and more recent methods (i.e. counterfactual reasoning) in the following publications: 1. a retrospective matched cohort study to determine the association between type 2 diabetes and herpes zoster disease, 2. a retrospective cohort study to assess the effect of the AS04-HPV-16/18 vaccine (Cervarix, GSK) on the risk of autoimmune diseases, and 3. a prospective matched case-control study to evaluate the effectiveness of a rotavirus vaccine (Rotarix, GSK) in preventing hospital admissions due to rotavirus gastroenteritis. In addition, directed acyclic graphs (DAGs) were drawn to illustrate and support the previous assessments.

Results: In the 1st and 2nd studies, only two Hill's criteria were met (i.e. temporality and plausibility). In the 3rd study, 8 out of the 9 Hill's criteria (i.e. strength, consistency, specificity, temporality, plausibility, coherence, analogy and experiment) were met.

Regarding counterfactual reasoning, exchangeability (the most critical assumption) could not be demonstrated. However, some evidence supporting conditional exchangeability between the exposed groups was found in the 3rd study. Positivity was found in studies 1 and 3, and consistency, while met for the 2nd and 3rd studies, could not be ascertained for the 1st study. This assessment enabled the determination of a certainty level on causality for each study, from very likely in the 3rd study to less likely or unlikely in the 1st and 2nd studies.

Furthermore, DAGs provided the most accurate scenario in support of causal inference assessment, through visual structures that identified confounding and bias.

Conclusions: Although a causal relationship could not be established in any of the three selected observational studies, causal inference approaches are valuable in determining the magnitude of evidence for concluding on the likeliness of a causal relationship. Application of an integrated causal inference framework should be considered when designing and interpreting observational studies.

Funding: GlaxoSmithKline Biologicals SA

[PO5.03]

Agreement among raters: revisiting Hubert's kappa

Antonio Martín Andrés[1], Maria Alvarez Hernandez[2]

[1] *Department of Statistics and Operations Research, University of Granada, Spain*
[2] *Defense University Center, Naval Academy, Spain*

In many fields of science, there is a frequent need to measure the degree of agreement between R observers that independently classify n subjects within K nominal or ordinal categories. For the case of two raters, the most common method is to eliminate the chance effect by Cohen's kappa coefficient, which was initially defined for nominal categories. When the categories are ordinal or when a weight is assigned to the nominal case, the Cohen's weighted kappa coefficient is used. Assuming quadratic weights in the ordinal case, Cohen's kappa has the advantage of coinciding with the intraclass and concordance correlation coefficients. When R>2, there are more disparities because the definition of the kappa coefficient depends on how the phrase "an agreement has occurred" is understood. The interpretation of Hubert, „an agreement occurs if and only if all raters agree on the categorization of an object", is taking into account and Hubert's (nominal) and Schuster-Smith's (ordinal) kappa coefficients are obtained. Expressions for the approximate large-sample variances are given for the latter coefficients, leading to different ways of carrying out inferences (tests and confidence intervals). In addition, it is shown a demonstration that the Cohen's weighted kappa coefficient and Schuster-Smith's kappa coefficient are equal to the concordance correlation coefficient when quadratic weights are assigned.

[PO5.04]

Comparison of weight-for-height and BMI-for-age for estimating under-five thinness burden: Policy implications

L. Naga Rajeev[1], Monika Saini[1], Ashish Kumar[1], Shikha Sinha[2], Clive Osmond[3], Harshpal Singh Sachdev[4]

[1] *Department of Mathematics and Statistics, Manipal University Jaipur, India*
[2] *Sitaram Bhartia Institute of Science and Research, India*
[3] *MRC Lifecourse Epidemiology Unit, University of Southampton, India*
[4] *Pediatrics and Clinical Epidemiology, Sitaram Bhartia Institute of Science and Research, India*

Background: Thinness below five-years of age, also known as wasting, is used to assess the nutritional status of populations for programmatic purposes, and is also a Sustainable Development Goal (SDG). Thinness may be defined, either when weight-for-height or when Body-Mass-Index-for-age (BMI-for-age) lie more than two standard deviations (SD) below the respective WHO standard.

Objective: To compare these two definitions for quantifying the burden of thinness.

Methods: The effect of ignoring age on thinness cut-offs was first estimated theoretically. For different height Z-scores, we compared the age- and sex-specific values of weight that would define thinness using weight-for-height and BMI-for-age. For varying height Z-scores, we also compared the actual WHO BMI-for-age cut-offs with BMI-for-age cut-offs derived from weight-for-height at that particular age. Estimates of thinness from these two indices were made in simulated populations (short, intermediate and tall), and using individual records from the National Family Health Survey–4 (NFHS-4) in India.

Results: From 6-59 months age, BMI-for-age cut-offs were lower than weight-for-height cut-offs in shorter children, but higher in taller subjects. In simulated data-sets, using weight-for-height definition, the burden of thinness was higher in short, but lower in tall populations. In NFHS–4 data (mean, SD for height-for-age Z: -1.56, 1.64), prevalence of thinness was greater with weight-for-height in comparison to BMI-for-age cut-offs (19.2% vs 17.0%). The corresponding figures for children with height-for-age below -2SD and above 0SD were 16.4% vs 10.6%, and 32.9% vs 37.2%, respectively. The overall differences were more prominent in boys. A reverse pattern was evident before six months age.

Conclusion: The two definitions produce cut-offs, and hence estimates of thinness, that differ with the age, sex and height of children. In stunted children, aged from 6-59 months, wasting estimates are substantially higher with weight-for-height in comparison to BMI-for-age definition, which influences the programmatic input for supplementary feeding. For estimating thinness in under-five children, the relative invariance of the BMI-for-age definition with age, sex and height favours its use as the preferred index for policy and programmatic purposes, and for evaluating the progress of nations in achieving the related SDG.

[PO5.05]

Assessing the performance of population adjustment methods for indirect comparisons: a simulation study

David Phillippo[1], Sofia Dias[2], Tony Ades[1], Nicky Welton[1]

[1] *Bristol Medical School, University of Bristol, United Kingdom*
[2] *Centre for Reviews and Dissemination, University of York, United Kingdom*

Background: Standard network meta-analysis and indirect comparisons combine aggregate data from multiple studies on treatments of interest, assuming that any factors that interact with treatment effects (effect modifiers; EMs) are balanced across populations. Population adjustment methods including multilevel network meta-regression (ML-NMR), matching-adjusted indirect comparison (MAIC), and simulated treatment comparison (STC) relax this assumption using individual patient data from one or more studies, and are becoming increasingly prevalent in health technology appraisals and the applied literature.

Objectives: Motivated by two recent reviews of applications, we aimed to assess the performance of these methods in a range of realistic scenarios under various failures of assumptions.

Methods: We undertook an extensive simulation study, investigating the impact of varying sample size, missing EMs, strength of effect modification and validity of the shared EM assumption, validity of extrapolation and varying between-study overlap, and different covariate distributions and correlations. We assessed bias, standard error, and coverage for MAIC, STC, and ML-NMR, alongside standard indirect comparisons.

Results: ML-NMR and STC performed similarly throughout, eliminating bias and estimating standard errors well when the requisite assumptions were met. MAIC performed poorly in almost all simulation scenarios, in some cases increasing bias compared with a standard indirect comparison. MAIC required full overlap between populations otherwise estimates were biased and unstable, especially when sample size was small. All methods incurred bias when EMs were missing from the model.

Conclusions: Serious questions are raised about the suitability of MAIC, currently the most popular approach, which is only valid in scenarios where there may be little benefit over a standard indirect comparison. ML-NMR and STC are robust methods for population adjustment, but careful selection of potential EMs prior to analysis is essential to avoid bias. ML-NMR offers additional advantages over MAIC and STC, including synthesising larger treatment networks and producing estimates in any target population, making this an attractive choice in many scenarios.

[PO5.06]

Non-Specific Effects of Oral Polio Vaccine Campaigns on Child Mortality? Triangulation of Epidemiological Data

Sebastian Nielsen[1], Christine Stabell Benn[1], Sören Möller[1], Peter Aaby[2]

[1] *OPEN, Department of Clinical Research, University of Southern Denmark, Denmark*
[2] *Bandim Health Project, University of Southern Denmark, Guinea-Bissau*

Background: Increasing evidence supports that vaccines have non-specific effects (NSEs) on child health, altering susceptibility to non-targeted diseases.

Randomised Controlled Trials (RCTs) are the gold-standard to examine casual relationships, however they are expensive and cause ethical challenges for vaccines already implemented, as RCTs would often imply delaying or omitting recommended vaccines. Results from observational data and secondary/exploratory analyses of RCTs are often dismissed due to potentially uncontrolled bias or confounding. Triangulation, combining different epidemiological studies using different methods and data sources, offers a different approach to establish causality.

Objective: Child mortality decreased 2-3-fold in most low-income countries in the last two decades. Numerous campaigns with Oral Polio Vaccine (C-OPV) were conducted in this period. We used triangulation to assess whether C-OPV alter child mortality patterns and could explain the decline in child mortality.

Methods: We investigated the impact of C-OPV by addressing the following research questions:

Was C-OPV associated with child mortality in observational data?

Did C-OPV alter effect estimates of other interventions obtained in RCTs?

Did C-OPV alter the effect of known child mortality risk factors in observational data?

Results: In observational data from Guinea-Bissau child mortality was 25% (95% CI: 15-33%) lower after C-OPV among children under 3-years-of-age between 2002-2014.

In an RCT comparing the effect of an extra early measles vaccine (MV) vs. the usual single MV the hazard ratio for early vs. no early MV was 1.39 (0.97-1.99) for children that had received C-OPV before enrolment, but 0.79 (0.62-1.00) for children that had not. Divergent results in three RCTs (Guinea-Bissau, 2003-2019) could be explained by frequency of C-OPV.

In-house crowding is a known risk factor for child mortality. The risk of dying per each additional child under 3-years-of-age increased by 9% (3-16%) before C-OPV but was 18% (11-26%) after receiving C-OPV.

Conclusions: RCTs may be the gold-standard; however, not all epidemiological questions can be investigated in RCTs.

Triangulation of the evidence revealed effects of C-OPV on child mortality shows effects of C-OPV which cannot be explained by its specific effects against polio, and thus supports that C-OPV has major beneficial NSEs.

[PO5.07]

Evaluation of substantive-model-compatible multiple imputation approaches for incomplete three-level data

Rushani Wijesuriya[1], Margarita Moreno-Betancur[1], John B. Carlin[1], Anurika De Silva[2], Katherine J. Lee[1]

[1] *Department of Paediatrics, University of Melbourne, Australia*
[2] *School of Population and Global Health, University of Melbourne, Australia*

Background: Three-level data structures arising from repeated measures on individuals who are clustered within larger units are common in health research studies. Missing data are prominent in such studies and are often handled via multiple imputation (MI). In analyses with interactions or non-linear terms, it has been shown that conventional MI approaches can produce biased parameter estimates. Substantive-model-compatible (SMC) MI has shown great promise for tackling these issues in the context of single-level data. While there have been SMC extensions of MI in the context of multilevel data, to date only one approach that explicitly handles incomplete three-level data is available. Therefore, imputation of incomplete three-level data structures compatibly with complex analysis models can be a challenge in applications and it is unclear how these specialised routines compare to extensions of single and two-level imputation models.

Aims: In this study we aimed to extend and evaluate pragmatic adaptations to the single-level and two-level MI methods using dummy indicators and/or a 'just another variable' approach and compare their performance to that from the only available implementation of three-level SMC MI.

Methods: We considered two analysis models that are of interest to many researchers in the longitudinal data setting - a multilevel model with an interaction between a time-varying exposure and time, and a multilevel model with a non-linear effect of the time-varying exposure. The various MI methods were evaluated via simulations and illustrated using empirical data based on a case study from a longitudinal cohort estimating the effect of early depressive symptoms on the academic performance of students over time, clustered by school.

Results: Results showed that the three-level SMC MI approach and adaptations of single-and two-level SMC MI approaches performed well in terms of bias and precision when the target analysis involved an interaction with time, while the three-level SMC MI approach resulted in better performance in the presence of non-linear terms.

Conclusions: Researchers may use extensions to standard single- and two-level MI approaches to adequately handle incomplete interactions with time in three-level data, while for non-linear terms, MI approaches that explicitly model three-level variation may be more appropriate.

[PO5.08]

Comparison of statistical methods for estimating age-specific paediatric reference intervals

Monsurul Hoq[1], Susan Donath[1], Paul Monagle[2], John Carlin[1]

[1] *Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Australia*
[2] *Royal Children's Hospital, Australia*

Background: Reference intervals (RIs) are commonly used in laboratory medicine for identifying patients who may need further investigation. A standard statistical approach is to identify measurements that fall outside the $\alpha/2$ and $(100 - \alpha/2)$ percentiles where $0 < \alpha < 100$. Studies have demonstrated changes in RIs with age in children by establishing RIs for different age groups, but grouping creates artificial "step changes". Addressing this, RIs varying continuously with age have been developed using a range of curve fitting approaches. The choice of statistical method may be important as there may be differences in RIs due to the underlying assumptions of these methods. Hence, we developed a simulation study with the aim of comparing the behaviour of a range of statistical methods for estimating age-specific paediatric RIs.

Objective: To evaluate and compare statistical method(s) for estimating age-specific paediatric RIs under different scenarios and for different sample sizes.

Methods: We compared five methods of estimating age-specific paediatric RIs. These were Cole's LMS, Generalised Additive Model for Location Scale and Shape (GAMLSS), Royston's fractional polynomial regression, and a new method based on quantile regression using power variables in age selected by fractional polynomial regression for the mean. Data were generated using hypothetical true curves based on five analytes with varying complexity of association with age i.e. linear or nonlinear association with age, consistent or inconsistent variation across age, and for four different sample sizes (100, 200, 400 and 1000). Mean square error (MSE) was used as the primary performance measure for comparing the methods.

Results and conclusions: There was no one method that performed better in all scenarios. Our new method performed consistently across all scenarios for sample sizes of at least 400, and had smallest average MSE in a scenario with inconsistent variation over age. This method was applied to obtain age-specific RIs for 30 biochemistry analytes using data from a study of normative children.

We recommend that our proposed method should be used for estimating age-specific paediatric RIs regardless of the complexity of the association between analytes and age, for sample sizes of minimum 400.

[PO5.09]

Martingale residual based approach for Cox modeling from high-dimensional data

Anastassia Kolde[1], Meelis Käärik[1], Tõnu Esko[2], Peter Joshi[3], Krista Fischer[1]

[1] *Institute of Mathematics and Statistics, University of Tartu, Estonia*
[2] *Estonian Genome Center, University of Tartu, Estonia*
[3] *Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, United Kingdom*

Background: Survival modelling is a natural approach to study genetic associations to many phenotypes and it has been shown to be more sensitive than traditional regression based methods [1]. Still, it is not commonly used in Genome Wide Association Studies (GWAS) due to prohibitive computational cost.

Objective: The Cox proportional hazard model can be approximated with linear regression by converting the survival trait to martingale residuals, thus enabling the use of common GWAS tools. This approach has been taken in many papers [2], however, it is known that, theoretically, the approximation only works within certain bounds [3].

Method: The goal of this work is to explore the bounds theoretically and experimentally through simulations to determine if and when the approximation is usable within GWAS and other omics data analysis settings.

References:

1. *Hughey, J.J., Rhoades, S.D., Fu, D.Y., Bastarache, L., Denny, J.C. and Chen, Q., 2019. Cox regression increases power to detect genotype-phenotype associations in genomic studies using the electronic health record. BMC genomics, 20(1), pp.1-7.*
2. *Joshi, P.K., Fischer, K., Schraut, K.E., Campbell, H., Esko, T. and Wilson, J.F., 2016. Variants near CHRNA3/5 and APOE have age-and sex-related effects on human lifespan. Nature communications, 7(1), pp.1-7.*
3. *Therneau, T.M., Grambsch, P.M. and Fleming, T.R., 1990. Martingale-based residuals for survival models. Biometrika, 77(1), pp.147-160.*

[PO5.10]

Prediction versus transferability of betas: a PCA informed strategy

Katri Parna, Davide Marnetto, Luca Pagani

Institute of Genomics, University of Tartu, Estonia

Genome-wide association studies (GWASs) are a common method to detect single nucleotide polymorphisms (SNPs) associated with a trait of interest, whereas all these SNPs are combined together into the polygenic risk score (PRS) representing each person's genetic susceptibility for a specific disease. Currently, GWASs summary statistics obtained from European populations are hardly transferrable, and hence not beneficial, across other worldwide populations, or even across different European cohorts, due to differences in allele frequencies, linkage disequilibrium, disease prevalence and genetic architecture. Existing methods based on Principal Component (PCs) corrections may help mitigating these transferability issues, and we here put forward the hypothesis that obtaining PCs from a reference dataset and projecting case/controls over it may minimize cohort-specific biases. The main aims of the current study are (1) to test the transferability of polygenic risk score computed based on GWASs adjusted for different populations principal components (PCs) obtained through projection on a reference PC space, and (2) to assess predictability of these PRSs in understudied populations.

Data from Estonian Biobank (n=52,000) (EstBB), 1000Genomes Project (n=2504) and UK Biobank (n=500,000) were used. We computed PCs for the EstBB samples (PCEst) and projected the same EstBB samples onto PC spaces from principal component analyses conducted on world-wide (PCworld) or European (PCEur) populations of 1000Genomes reference panel. Four different GWASs for type 2 diabetes were performed in EstBB subset (n=5959), where SNP effect sizes were: (1) adjusted for PCworld, (2) adjusted for PCEur, (3) adjusted for PCEst, and (4) no adjustment. Ongoing work concerns the calculation of PRSs based on SNPs adjusted for these different PCs and the testing of their predictability and transferability in EstBB and UK Biobank. The same analysis will be conducted for continuous traits as body mass index and height.

Preliminary results of the current study show that there are perturbations in SNP effect sizes obtained from GWASs in EstBB when adjusted for PCs compared to no adjustment or to adjustment for other population PCs, but further analysis are needed to test PRS predictivity and transferability to other cohorts.

[PO5.11]

Multiple mediation modalities for social class disparities in cardiovascular patients

Nicola Fitz-Simon, Martin O'Donnell

National University of Ireland, Galway, Ireland

Background: There are disparities between social classes in cardiovascular outcomes, which could potentially be reduced if appropriate targets for intervention are identified. Recent developments in causal mediation analysis may help clarify suitable targets for intervention.

Objective: We aim to identify potential interventions to reduce social class disparities in cardiovascular event outcomes, and explore practical challenges in a real-life application of interventional mediation analysis.

Methods: We use data from a large national cardiovascular registry of patients with follow-up to six months. We consider complex hypothetical interventions via three multivariate mediating pathways (1) age, (2) severity (3) in-hospital treatment. Under an interventional mediation framework, we counterfactually shift the distributions of each pathway separately so that the lowest social class has the same distribution of mediators as the highest. Differences between survival probabilities under different interventions on the distributions of mediators allows us to estimate interventional indirect effects via each pathway. We estimate effects via Monte Carlo simulation and bootstrap for confidence intervals.

Results: There is a large disparity in six-month survival between the highest (26%) and lowest (20%) social class. The lowest social class tends to experience events younger, which confers an immediate survival advantage, but their events are more severe, which decreases survival. Allowing for these, if we shift the distribution of in-hospital treatments in the lowest social class to that in the highest, six-month survival could be improved by 2.3% (95% CI 0.6-4.4%).

Conclusions: An interventional mediation analysis helps to clarify appropriate targets for intervention to reduce social class disparities in patients who have experienced a cardiovascular event. We discuss practical and interpretational challenges of real-life application of this methodology.

[PO5.12]

Using independent cross-sectional survey data to approximate post-migration health trajectories among refugees

Stella Preussler[1], Louise Biddle[2], Meinhard Kieser[1], Kayvan Bozorgmehr[3]

[1] *Institute of Medical Biometry and Informatics, University of Heidelberg, Germany*
[2] *Department of General Practice and Health Services Research, University Hospital Heidelberg, Germany*
[3] *Department of Population Medicine and Health Services Research, School of Public Health, Bielefeld University, Department of General Practice and Health Services Research, University Hospital Heidelberg, Germany*

Background: Collection of longitudinal data is resource-intensive, whereas cross-sectional data is usually less laborious to obtain. Therefore, estimation of health trajectories from repeated independent cross-sections, i.e., data collected at different time points (T0, T1) among independent samples from, e.g., real world data provided by routine data collection systems would be desirable.

Objective(s): Our aim is to extrapolate a pseudo-panel of independent cross-sectional data (e.g. data of T0 and T1) to approximate a longitudinal health trajectory (T0-T1). Methods will be illustrated by the example of studying contextual effects (e.g. remoteness of refugee accommodation) on health (e.g. low/high self-rated general health) among refugees by calculating transition probabilities with associated variances.

Method(s): Following the post-migration trajectory, two large-scale cross-sectional health surveys among randomly selected refugee samples in reception centers (T0) and accommodation centers (T1) located in Baden-Württemberg were done as part of the RESPOND study. Self-reported measures of physical and mental health, health-related quality of life, health care access, and unmet medical needs of 560 refugees were collected. Missing data was imputed by multiple imputations based on Fully Conditional Specification and the Predictive Mean Matching method. For each imputed data set transition probabilities were calculated based on (i) Probabilistic Discrete Event Systems with Moore-Penrose generalized inverse matrix method, and (ii) Propensity score matching. By application of sampling approaches, exploiting the fact that status membership is multinomially distributed, results of both methods were pooled by Rubin's Rule, accounting for within and between imputation variance.

Results: The results of most of the analyzed estimates of the transition probabilities and their variances are comparable between both methods. However, it seems that they handle sparse cells differently: either assigning an average value for the transition probability for all states with high certainty (i), or assigning a more extreme value for the transition probability with large variance estimate (ii).

Conclusions: Further research on the potential to extrapolate the results of cross-sectional data is needed. To inform and advise future studies, the results of this analysis will be compared to the results of a prospective natural experiment study with longitudinal data collection of contextual and individual factors.

[PO5.13]

A comparison of methods for causal inference with a rare binary outcome

Mariella Gregorich, Georg Heinze

Section for Clinical Biometrics, Medical University of Vienna, Austria

Background: Causal inference from observational studies can be challenging with a rare outcome event and many potential confounding variables. The probability of an individual to receive the treatment given the patient's information, known as the propensity score, can be used in the process of matching or weighting the observational data to combat the inherent imbalance of a patient's baseline characteristics in observational studies. Alternative proposals to tackle the shortcomings of observational data are based on traditional outcome regression models, also in combination with inverse probability weighting. However, a major strength of the propensity score methods is supposed to be their handling of a large number of confounders and few observed outcomes due to the focus on the exposure-covariates associations as an intermediary step.

Objectives: We compared the performance of four well-known causal inference approaches (Propensity score [PS] matching, Inverse Probability of Treatment Weighting [IPTW], G-computation and Double Robust G-computation) in the situation of many confounders and few observed events.

Methods: A study to estimate the marginal causal treatment effect of a computer tomography scan examination of patients undergoing coronary artery bypass surgery (CABG) on the postoperative stroke risk served both as a real-data example and as a motivation for a comparative simulation study. The target estimand in the study was the average treatment effect, and performance was evaluated by comparing the root mean squared error.

Results and conclusion: Even though PS matching and IPTW circumvent problems of estimating a multivariable outcome model with too many variables and few events, they both did not result in more precise estimates of the average treatment effect, neither in the simulation nor in the case study. All chosen approaches suffered in performance due to the small number of events, but the results illustrate that propensity score analyses cannot necessarily improve over other proper causal inference techniques when events are rare.

PO6: Epidemiology: Methods and Applications / Electronic Health Records

Wednesday, 26 August 2020, 10.20 - 11.50

Chairs: **Agnieszka Doryńska**

Institute of Public Health, Jagiellonian University Medical College, Krakow, Poland

Kristel Van Steen

[PO6.01]

Probabilistic quantification of bias - combine strengths of population-based register data and clinical cohort

Aleksandra Turkiewicz[1], Peter M. Nilsson[2], Ali Kiadaliri[1]

[1] *Clinical Sciences, Lund, Orthopedics, Clinical Epidemiology Unit, Lund University, Sweden*
[2] *Department of Clinical Sciences, Internal Medicine Research Group, Lund University, Sweden*

Background: Probabilistic quantification of bias is an efficient method to correct statistical analysis of observational data and account for both random and systematic error.

Objective: We propose to combine population-based register data, with limited selection bias, with a nested clinical cohort to correct misclassification and unmeasured confounding in the register data through probabilistic quantification of bias. We illustrate the method in analysis of mortality associated with osteoarthritis.

Methods: Using the Swedish Population Register we included all persons aged 56 to 84 and resident in the Skíne region in 2008. We classified persons as having knee osteoarthritis if had been diagnosed between years 1998 and 2008, based on data from Skíne Healthcare Register. We studied mortality until year 2017 and estimated hazard ratios (HR) of death using adjusted Cox proportional hazards model. We used data from Malmö Osteoarthritis Study (MOA), a clinical cohort from Skíne (nested in the population based sample), to derive bias parameters and their distributions for probabilistic quantification of bias, to correct the HR estimate for differential misclassification of osteoarthritis diagnosis and confounding from unmeasured obesity. We used beta distributions to parametrize sensitivity, specificity and prevalence of osteoarthritis and obesity in four strata defined as combinations of exposure and outcome.

Results: We included 292,000 persons in the Skíne population, with mean (SD) age of 69 (8) years. Osteoarthritis was diagnosed in 23,595 persons. The MOA study consisted of 1491 persons, whereof 11% had osteoarthritis according to gold standard definition and 15% were obese. The adjusted association of osteoarthritis with all-cause mortality in the MOA sample was (HR [95% confidence interval]) 1.13 [0.81, 1.56] and thus inconclusive. The association in the Skíne general population was 0.95 (0.93, 0.98), while the bias-corrected estimate was 1.00 (0.82, 1.23), suggesting no relevant excess mortality.

Conclusions: The population-based sample yielded overly precise and slightly biased estimate. The clinical cohort yielded biased results with wide, inconclusive confidence intervals. Combining population-based register data with clinical cohorts can lead to more valid and precise results, than using either data source separately.

[PO6.02]

Creation of WHO indicators of Infant and Young Child Development (IYCD): metadata synthesis and item mapping

Gillian Lancaster[1], Gareth McCray[1], Patricia Kariger[2], Melissa Gladstone[3], Magdalena Janus[4], Vanessa Cavallera[5], Tarun Dua[5], Claudia Alves[6], Muneera Rasheed[7], Limbika Maliwichi[8], Andrew Titman[9], Jaya Chandra[3]

[1] *School of Primary, Community and Social Care, Keele University, United Kingdom*
[2] *School of Public Health, University of California Berkeley, United States*
[3] *Institute of Translational Medicine, University of Liverpool, United Kingdom*
[4] *Department of Psychiatry and Behavioural Neurosciences, McMaster University, Canada*
[5] *Neurological Diseases and Neuroscience, World Health Organisation, Switzerland*
[6] *Faculty of Medicine, Federal University of Minas Gervais, Brazil*
[7] *Department of Paediatrics, Aka Khan University, Pakistan*
[8] *Chancellor College, University of Malawi, Malawi*
[9] *Department of Mathematics and Statistics, Lancaster University, United Kingdom*

Introduction: The World Health Organisation (WHO), UNICEF and World Bank have all called for increased investment in early childhood programs that promote holistic development, and are evidenced to lead to improved health outcomes and socio-economic status in later life. For the first time the United Nations Sustainable Development Goals (SDGs) include targets to ensure equity in access to quality early development and pre-primary enrichment programs. Valid and reliable methods of measurement are crucial for monitoring achievement of SDG targets across countries and for determining the success of interventional programs. Currently, there is a dearth of low-cost, culturally-validated, simple-to-implement tools that can be used in low and middle income (LMIC) settings for the most sensitive 0-3 year age range. In a WHO-led initiative we aimed to develop a tool to address this gap.

Methods: Through meta-data synthesis and innovative item mapping using 14 datasets comprising 21,083 children from 10 LMIC countries assessed by 7 different tools, and an expert consensus meeting, the first IYCD prototype was created. Subsequent field testing (in Brazil, Malawi, Pakistan) refined the prototype to a 100 item-tool that assesses motor, language-cognitive, socio-emotional skills, and general behaviour, as a caregiver-interview that can be administered via tablet, providing audio-visual aids especially useful in low-literacy settings.

Results: Data visualisation aided expert judgement and item adaptation, and statistical analyses of field test data demonstrated item performance, reliability and validity of the tool, including creation of a Development for Age z-score (DAZ) via item-response and GAMLSS modelling.

Discussion: We present the methodological approach and challenges encountered working in multicultural settings to design the study and create the WHO IYCD tool.

Reference:

Lancaster G.A., McCray G., Kariger P., Dua T., et al. Creation of the WHO Indicators of Infant and Young Child Development (IYCD): meta-data synthesis across ten countries. BMJ Global Health 2018,3:e000747.

[PO6.03]

Time until the diagnosis of pulmonary tuberculosis in Portugal: spatiotemporal clustering

Patricia Soares, Carla Nunes

NOVA National School of Public Health, Public Health Research Centre, Universidade NOVA de Lisboa, Portugal

Background: A delayed diagnosis of tuberculosis (TB) is a threat for TB control since a higher delay leads to an increased infectious period, hence higher transmission. In recent years, several EU countries have registered an increase in the time to diagnosis TB.

Objectives: This study aims to identify spatiotemporal clusters of high and low delay in Portugal and to characterise and compare the clusters profiles at the individual and contextual level.

Methods: All the cases of pulmonary TB, notified between 2008 and 2017 and diagnosed by passive screening (symptoms) in Portugal were considered. SaTScan was used to identify spatiotemporal clusters with a weighted normal distribution and at county level (308 counties).

Results: During the study period, 15046 cases were notified. Of those, 2092 (13.9%) did not have information about symptoms onset and were discarded from the analysis. Two clusters were found, one high cluster included 48 counties from Alentejo, Algarve and Lisboa, between 2016 and 2017, and a low cluster including 37 counties from North and Centre regions, between 2008 and 2009.

The high cluster had a delay mean inside the cluster of 110 days, whereas the low cluster had a delay mean inside the cluster of 71 days. At the individual level, the high cluster had 361 patients, and a significantly higher proportion of females (33.5% vs 26.9%), of foreign-born (25.8% vs 2.31), comorbidities (30% vs 12.7%), a significantly lower proportion of long-term unemployment (9.4% vs 18.6%), and significantly older patients (mean age: 49 vs mean age: 43), compared with the low cluster. At an ecological level, the high cluster was characterised by a significantly lower number of beneficiaries of the social income per 1000 individuals (28.3 vs 65), lower number of doctors in health centres per 10.000 habitants (6.2 vs 7), and lower TB incidence rate (14.7 vs 27.3), compared with the low cluster.

Conclusions: The time to diagnosis is increasing temporally and in areas of low incidence and with a smaller proportion of doctors. Interventions should be planned to improve the health literacy of the population and healthcare professionals.

[PO6.04]

Risk of severe infections and mortality in patients with newly diagnosis of systemic lupus erythematosus

Kai Zhao, Hui Xie, Lingyi Li, John Esdaile, J. Antonio Aviña-Zubieta

Arthritis Research Canada, Canada

Background: Infections are serious complications in systemic lupus erythematosus (SLE) and often associated with premature mortality. Studies on the risk of infection in SLE patients often suffer from small sample sizes. We conducted a large population-based study to: 1) estimate the differences in the risk of severe infection requiring hospitalization in terms of the time from diagnosis of SLE to onset of infection, the rate of infection recurrence and mortality rate due to infection in SLE and non-SLE cohorts; and 2) identify risk factors for severe infections among SLE patients.

Methods: Using administrative data and a validated case definition, we assembled a cohort of all incident SLE patients between 1997 and 2015. The case inclusion criteria were at least two physician visits, at least two months apart, within two years, with ICD indicating SLE. The non-SLE cohort was a random sample from the general population and 1:5 matched to the SLE patients on age, sex and SLE index year. We used: 1) Kaplan-Meier estimates with log-rank test and multivariable Cox proportional hazard models to compare time to the onset of infection post SLE diagnosis and to estimate hazard ratios; 2) Poisson regression to compute the ratios for infection recurrence rate; and, 3) Competing risk models accounting for other causes of death to test the difference in infection-caused mortality rates in the SLE and non-SLE cohorts.

Results: We identified 5,169 SLE patients with newly diagnosed SLE and 25,845 non-SLE controls (86.17% females, mean age 46.9), yielding 955 and 1,986 severe onset infections during 48,367 and 260,712 person-years, respectively. The crude incidence rate ratios were 2.59 (95% CI; 2.40-2.80) and 2.10 (95% CI; 1.28-3.45) and the adjusted hazard ratios were 1.66 (95% CI; 1.52-1.82) and 1.00 (95% CI; 0.55-1.80), respectively for incident infections and mortality due to infections. The rate ratio was 3.24 (95% CI; 3.06-3.43) and adjusted rate ratio 1.93 (95% CI; 1.81-2.06) for recurrent infections.

Conclusion: We found SLE was associated with increased risk of hospitalization for infection onset, infection recurrence and mortality when compared to the non-SLE cohort. This highlights the higher severity of infection in patients with SLE.

[PO6.05]

Classification trees applied to identify predictors of autism spectrum disorder diagnosed in early childhood

Simon Grøntved[1], Nicolai Søndergård Schjøtt[1], Anne Marie Svane[2], Marlene Briciet Lauritsen[3], María Rodrigo-Domingo[1]

[1] *Unit for Psychiatric Research, Region North Psychiatry, Denmark*
[2] *Department of Mathematical Sciences, Aalborg University, Denmark*
[3] *Research Unit for Child and Adolescent Psychiatry, Region North Psychiatry, Denmark*

Introduction: Machine learning techniques, e.g. classification trees, have become very popular in many research fields, including psychiatry, within the last decade. These techniques are best suited to analyze large amounts of high quality data, like the ones available in the Danish national registers.

Objectives: To show how classification trees can be used in the field of psychiatry, and to investigate whether known risk factors are good predictors of a registered diagnosis of autism spectrum disorder (ASD) in early childhood.

Methods: We use classification trees to predict a diagnosis of ASD before the age of 3 as per the Danish national registers between 1997 and 2012. Information is available on 1,133,703 individuals, and 0.1% of them received a diagnosis. Potential predictors include known familial, pregnancy, birth, and early life risk factors for ASD from the literature. The trees are visually inspected and their predictive ability assessed via cross-validated AUC.

Results: The highest cross-validated AUC among all trees was 0.65 (95% CI: [0.64, 0.67]), which is considered poor.

Conclusions: In light of our results, we believe that further research into the causes of autism is necessary, as known risk factors could not reliably predict a registered diagnosis of ASD.

[PO6.06]

A Review on the Application of Adaptive Designs in Oncology Clinical Trials

Jerissa Samuel Pezhumkattil[1], Smruti Mokal[2], Sadhana Kannan[1], Anjali Tiwari[1], Akshay Patil[2]

[1] *Clinical Research Secretariat, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), India*
[2] *Clinical Research Secretariat, Tata Memorial Hospital, India*

Background: Optimal treatment approach is important for success of a trial. In traditional design the rules set in the beginning are maintained till the last. However adaptive clinical trials are designed to take advantage of response of accumulating data and make forethought modification to study design. This has helped the researchers and the statisticians to design a "smarter" trial by reducing the chance of false failure, minimise the use of resources such as time and human subjects. Although the concept of adaptive design is being practiced for last many years there is still deficiency of understanding the statistical significance of adaptive design in clinical trials.

Objective: The aim of our study is to understand the potential advantage of using adaptive design in oncology clinical trials over past 10 years.

Methodology: Registered and published articles from 2011 to 2019 with an adaptive design in clinical trials were identified from PubMed Database. Phase 2, phase 3 and seamless phase2/3 oncology randomized control trials were included. Phase 1 clinical trial and Review articles were excluded. Basic properties of the study and important variables such as use of interim analyses, type of adaptive design used, objective for using particular design and result of attainment of the objective were extracted from the selected articles. The results are summarised using descriptive statistics.

Result: The search retrieved 88 results from major oncology journals in PubMed data base from which 22 were eligible for analysis. Adaptive designs were more commonly applied in phase 2 setting (59%). Majority of trials had planned an interim analysis (72.7%). The most commonly used adaptive design was bayesian adaptive randomization (31%) and adaptive group sequential design (31%), followed by response adaptive randomization (27%) and others (11%). The major reason for choosing an adaptive design was to ensure randomization probabilities were in favour of the better performing treatment (63%). Majority of the study has met their objective of using adaptive designing over the traditional design (95%).

Conclusion: This review suggests the use of adaptive design over the traditional approach has potential advantage and will benefit in conducting an efficient trial with faster progress, limited resource usage and provide statistical efficiency.

[PO6.07]

Internal bootstrap validation of logistic predictive models: Stata command bsvalidation

Borja Manuel Fernandez-Felix[1], Alfonso Muriel[1], Teresa Pérez[2], Javier Zamora[1, 3]

[1] *Clinical Biostatistics Unit, Hospital Universitario Ramón y Cajal (IRYCIS). CIBER Epidemiology and Public Health (CIBERESP), Spain*
[2] *Department of Statistics and Data Science, Complutense University of Madrid, Spain*
[3] *Queen Mary University of London, United Kingdom*

Background: The TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) standards for predictive model reporting includes internal model validation as a fundamental step. Bootstrap techniques are the most appropriate procedures for internal model validation, since it uses all the data used in model development and it permits to quantify the optimism.

Objective: To provide researchers with a Stata post-estimation command to perform internal bootstrap validation of a logistic regression model.

Methods: The validation method follows the following algorithm:

To adjust the final predictive model using logistic or logit Stata commands.

To generate `b´ bootstrap samples with replacement.

In each bootstrap sample, either to fit the model using the same variables selected for the final model or to allow for a variable selection strategy used in the development of the final model.

To determine bootstrap and test performance.

To calculate the measures of optimism.

To obtain estimates of model performance and to adjust model coefficients by optimism.

Results: To assess the overall performance the command reports the Brier score. For Discrimination, it shows the C-Statistic and Somer's D. For calibration, the command presents the calibration slope, calibration-in-the-large and the ratio of expected to observed events. In addition, it produces a calibration plot.

Conclusions: This tool makes the internal validation procedure more accessible to researchers, and allows to better reporting of predictive models according to the TRIPOD standards.

[PO6.08]

Hospital readmissions and death within 30 days and 1 year for patients with hip fracture – a multi-state model

Doris Tove Kristoffersen[1], Anette Hylen Ranhof[2]

[1] *Cluster for Health Services Research, Norwegian Institute of Public Health, Norway*
[2] *Norwegian Institute of Public Health, Norway*

Background: Hospital performance is commonly measured by quality indicators. One such indicator is the 30-day mortality for elderly patients with hip fracture. Overall, about 9% die in- or out-of hospital within 30 days of admission compared to about 25% after one year. While many of these patients are frail due to aging and chronic conditions, the rather high 1-year mortality may be affected by the follow-up care provided by primary care. When attributed to hospitals, both 30-day mortality and readmission vary significantly between hospitals.

Objective: To use a multi-state model to study time to death and readmissions for patients with hip fracture.

Methods: Patients were identified from administrative data from all Norwegian hospitals for the period 2015-2017, linked with date of death obtained from the National Registry. The multi-stage model described jointly the length of hospital stay, risk of readmissions and mortality.

Results and Conclusions: About 8500 elderly persons are annually hospitalized due to hip fracture, and the majority (75%) are females. The mortality of men was much higher for men compared to women; 1.3 % for females versus 3.1% men for patients aged 65-69 years. For patients more than 90 years of age, 11.7% of the females and 22.2% of the males died within 30 days. Overall 10% of the patients are readmitted within 30 days. The 1-year mortality has not decreased over the period.

[PO6.09]

Predictive discrimination models to diagnose malabsorption from routine clinical diagnostics tests in Africa

Frissiano Honwana[1], Maia Lesosky[1], Sabelo Hlatshwayo[2], Mashiko Setshedi[2], Sandie Thomson[2]

[1] *School of Public Health and Family Medicine, University of Cape Town, South Africa*
[2] *Division of Gastroenterology, Department of Medicine, University of Cape Town, Groote Schuur Hospital, South Africa*

Malabsorption disorders are conditions that result from impaired absorption of one or more dietary nutrients in the gastrointestinal tract. The clinical manifestations of malabsorption are highly variable, but left undiagnosed and untreated can lead to severe and irreversible complications. Malabsorption symptoms are often subtle and may overlap with common, but benign, conditions like irritable bowel syndrome. The clinically silent nature of the condition often leads to broad generalization of its causes although some patients may present well recognized symptoms. With the scarcity of data from sub-Saharan Africa, burden disease discussions of commonest cause of malabsorption like coeliac often becomes guesswork. Discriminant analysis models have the potential of addressing the diagnostic yield of malabsorption without the need for invasive and expensive tests.

This research aimed to use partial least squares discriminant analysis (PLSDA) and penalized regression models to identify patients at high risk for malabsorption using data collected during routine work up including endoscopic duodenal biopsies. A comparison of statistical models was evaluated against a clinically determined diagnosis.

The study used electronic histology data from the South Africa's National Health Laboratory Services of duodenal biopsies in n=3291 patients investigated for malabsorption syndrome. Ridge logistic regression and PLSDA models were fitted. K-folds cross validation were used respectively, for the regularization of the ridge model and choosing optimal PLSDA model. Prediction accuracy was compared against clinical diagnoses.

In the NHLS duodenal biopsies for patients investigated for malabsorption syndrome, 658 (20.0%) had suspected coeliac. Penalized regression showed improved prediction of the yield of malabsorption through coeliac with an accuracy of 0.79 (95%CI: 0.76; 0.82) as compared to standard logistic regression. PLSDA model showed similar estimates with an AUC of 0.77. Structures approaches using routine data may augment clinical decision making in this context.

[PO6.10]

Applications of Regularised Structural Equation Modelling to Psychometrics

Isobel Ridler, Cedric Ginestet, Silia Vitoratou, Daniel Stahl

Department of Biostatistics and Health Informatics; Institute of Psychiatry, Psychology and Neuroscience, King's College London, United Kingdom

Background: A novel method has been proposed that extends the use of regularisation to structural equation modelling (SEM) by Jacobucci, Grimm, and McArdle (2016). Regularisation is especially useful when a model contains a large number of variables (P) relative to sample size (N), as is common in psychiatric research. Regularised SEM has the potential to perform automatic variable selection, thereby producing more parsimonious latent factor solutions in a single-step procedure, using modern statistical learning methodology. This study will investigate the benefits of this method applied to psychometrics data with latent structural models.

Objectives:

1. Generate several simulation datasets to assess the performance of regularised SEM for various ratios of P:N and compare with traditional SEM,
2. Apply regularised SEM to a longitudinal dataset, extending the measurement model to see if we can optimise a psychometric scale based on prediction of later outcome, and assess the differential predictive power of individual items over and above the common factor.

Method: First, maximum likelihood estimation (MLE) and regularised SEM (following the method developed by Jacobucci (2017)) will be applied with Monte Carlo simulations of psychometric models. Second, MLE and regularised SEM will be applied to a suitable psychometric scale (large, relatively new and under-validated as a whole, or contains some items which have uncertainty surrounding their inclusion in the scale) and related outcome in a longitudinal dataset.

Results: Initial results of simple (five items, one latent factor, one outcome) model simulation show that regularised SEM lasso can select final models which reach parameter estimates close to the true simulated value (loadings and regression coefficients), increasing in accuracy and precision with increasing N. Type I error, Type II error, relative bias and root mean square error for MLE SEM and regularised SEM lasso will be compared for all models.

Conclusions: Preliminary results suggest regularised SEM can be utilised to investigate psychometric scales, with outcome prediction models aiding the selection of a subset of items in a scale.

References:

Jacobucci R, Grimm KJ, McArdle JJ. Regularized structural equation modeling. Structural equation modeling: a multidisciplinary journal. 2016 Jul 3;23(4):555-66.

Jacobucci R. regsem: Regularized structural equation modeling. arXiv preprint arXiv:1703.08489. 2017 Mar 24.

[PO6.11]

Heart Rate Asymmetry assessment during head up tilt table test for healthy women

Rafał Pawłowski[1], Katarzyna Buszko[1], Paweł Zalewski[2]

1 Department of Biostatistics and Biomedical Systems Theory, Ludwik Rydygier Collegium Medicum in Bydgoszcz Nicolaus Copernicus University in Toruń, Poland
[2] Department of Hygiene, Epidemiology, Ergonomics and Postgraduate Education, Ludwik Rydygier Collegium Medicum in Bydgoszcz Nicolaus Copernicus University in Toruń, Poland

Background: Heart Rate Asymmetry (HRA) is recently discovered attribute of human heart. It is a phenomenon in which contribution of accelerations to heart rate variability is unequal to those of decelerations [1]. It is interesting to examine its appearance depending on the physiological and external circumstances. The occurrence of HRA in majority of healthy human population was an inspiration for us to assess HRA in healthy women's response to the head up tilt test (HUTT).

Objectives: The purpose of the study is the assessment of HRA changes in response to the HUTT for group of 30 healthy women. HRA was calculated during HUTT in supine position and after the tilt.

Methods: The tilt tests were performed using Task Force Monitor device. We analyzed RR intervals extracted from high resolution ECG signal recorded during the test. The analysis of HRA is based on Poincaré plot. The asymmetry affects the relative position of the points on the plot [2]. Pairs with possible non-sinus origin were removed from the comparison using quotient filtering. Short-term HRA descriptors SD1a and SD1d were calculated and compared in supine position and tilt using Sign Test on significance level $\alpha = 0.05$.

Results: We observed statistically significant differences between supine and tilt for short-term HRA ($p = 0.022$). The appearance of short-term HRA (SD1d > SD1a) was observed for 9 women in supine and for 20 women after the tilt. Differences between long-term as well as total HRA were not statistically significant.

Conclusions: The study shows that head-up tilt table test influences the appearance of short-term HRA for healthy women. We suppose that short-term HRA can be modified in response to the HUTT and can be used as a parameter describing autonomic nervous system response to the tilt.

References:

1. *Piskorski J and Guzik P, Geometry of the Poincaré plot of RR intervals and its asymmetry in healthy adults, Physiol. Meas. 28 287–300, 2007*
2. *Guzik et al, Heart rate asymmetry by Poincaré plots of RR intervals, Biomed. Technol. 51 272–5, 2006*

[PO6.12]

Efficacy of treatments for patients with recurrent glioblastoma: A systematic review and network meta-analysis

Susanne Schmitz[1], Anna Schritz[1], Nassera Aouali[1], Aurelie Fischer[1], Coralie Dessenne[1], Guy Berchem[2], Laetitia Huiart[1]

[1] Department of Population Health, Luxembourg Institute of Health, Luxembourg
[2] Luxembourg Institute of Health, Luxembourg

Background: Glioblastoma is the most aggressive brain tumour in adult patients, with an average survival of 15 months following maximal surgical resection. Treatment options for patients with recurrent glioblastoma are limited with poor outcomes. While many treatment regimens have been explored in randomised clinical trials, there is little evidence on the relative efficacy between these options.

Objective: The objective of this study is to provide an overview of existing treatment strategies for recurrent glioblastoma and to estimate the relative efficacy of these regimens in terms of progression free survival and overall survival.

Method: We conducted a systematic review according to PRISMA guidelines to identify randomised control trials investigating any treatment regimens in adult patients suffering from recurrent glioblastoma. Studies reporting at least one of our primary outcomes (median progression free survival and median overall survival) were included in a network meta-analysis. The analysis was conducted in WinBUGs. We report estimates of relative efficacy as well as a ranking of treatment options.

Results: Thirty-nine randomised controlled trials fulfilled our inclusion criteria evaluating the efficacy of thirty-eight drugs as monotherapy or combination therapy. Results of the network meta-analysis show poor outcomes for patients (median progression free survival ranging from 1 to 6 months; median overall survival ranging from 3 to 19 months) with little differences between treatments.

Conclusion: This systematic review provides a comprehensive overview of existing treatment options for recurrent glioblastoma. The large network meta-analysis provides relative effects of efficacy for many of these treatment regimens, allowing for a comparison between regimens, which have not been directly compared. Overall, outcomes for patients with recurrent glioblastoma remains poor across all treatment options, making this an area of high need for innovative treatment options in the future.

[PO6.13]

Doubly robust inference procedure for relative survival ratio in population-based cancer registry data

Sho Komukai, Satoshi Hattori

Graduate School of Medicine, Osaka University, Japan

Population-based cancer registry data are utilized for various research questions in cancer epidemiology. For example, the CONCORD study series have reported worldwide surveillances about the impacts of cancer death for some nation and some period. One of the important features of cancer registries is that the information about the cause of death isn't available. To address this, a net survival, which is defined as the hypothetical survival probability if a patient would not die due to reasons other than cancer, is used as the measure for cancer prognosis in the analyses of the cancer registry data. To estimate the net survival, the Pohar-Perme (PP) estimator, which is the nonparametric estimator under the independent censoring, is proposed and tends to use as the standard estimator in recent years. The presence of the covariate-dependent censoring in practice was pointed out, and to address it the inverse probability of censoring weighted (IPW) estimator and the doubly robust estimator (Komukai and Hattori, 2017) was proposed.

As an alternative to the net survival, a relative survival ratio is also used. It is defined as the ratio between the survival probability of the cancer population and that of the general population. In this research, we propose a doubly robust estimator for the relative survival ratio in the presence of covariate-dependent censoring which we often encounter in practice. Although the inference for the net survival by Komukai and Hattori (2017) relies on untestable assumptions on dependence among potential time-to-events and censoring, it is possible to verify the underlying assumption in the inference for the relative survival ratio. We also propose a doubly robust test to assess the underlying assumption for the inference of relative survival ratio. We demonstrate the results of simulation studies to examine the finite-sample performance and also compare the behaviors of estimators among the proposed methods and the other existing methods. We illustrate our proposed method by applying to a cancer registry data for gastric cancer patients in Osaka, Japan.

[PO6.14]

Weight change and the incidence of cardiovascular diseases in healthy adults; emulating trials using EHR

Michail Katsoulis[1], Bianca DeStavola[2], Karla Diaz-Ordaz[3], Manuel Gomes[4], Alvina G. Lai[1], Pagona Lagiou[5], Goya Wannamethea[4], Kostas Tsilidis[6], Tom R. Lumbers[1], Spiros Denaxas[1], Amitava Benarjee[1], Constantinos A. Parisinos[1], Rachel Batterham[7], Claudia Langenberg[8], Harry Hemingway[1]

[1] *Institute of Health Informatics, UCL, United Kingdom*
[2] *GOS Institute of Child Health, UCL, United Kingdom*
[3] *Medical Statistics Department, LSHTM, United Kingdom*
[4] *Institute of Epidemiology & Health Care, UCL, United Kingdom*
[5] *School of Public Health, Harvard, United States*
[6] *Department of Epidemiology and Biostatistics, Imperial College London, United Kingdom*
[7] *Centre for Obesity Research, UCL, United Kingdom*
[8] *MRC Epidemiology Unit, University of Cambridge, United Kingdom*

Background: Medical practitioners encourage overweight or obese patients to lose weight, as high BMI is associated with increased risk for cardiovascular diseases (CVDs). Nevertheless, evidence in favour of weight loss is conflicting in observational studies and there are no randomised controlled trials (RCTs) conducted in otherwise healthy individuals. Emulation of target trials using electronic health records (EHRs) opens up new opportunities beyond what is feasible in RCTs.

Objectives: To estimate the effect of 2-year weight change interventions on CVD in individuals with normal weight (BMI:18.5-24.9kg/m^2), overweight (BMI:25-29.9 kg/m^2) and obesity (BMI:30-39.9 kg/m^2),

Methods: We identified 138.567 healthy individuals in England, aged 45-69 years old, at baseline during 1998-2016 and followed them for 7 years. Individuals were classified into one of the hypothetical interventions: i.weight loss (-20% to -3% per year or bariatric surgery), ii.weight maintenance (-3% to +3% per year), or iii.weight gain (3% to 20% per year). We used information on the first year of follow-up to "assign" individuals to these interventions and we performed pooled logistic regression to estimate per-protocol effects. We used inverse-probability weighting to adjust for baseline and time-varying variables, to account for deviation from the allocated intervention or censoring through follow-up. We used positive and negative control outcomes (diabetes and non-melanoma skin cancer respectively) to detect potential biases due to unmeasured confounding. We applied sensitivity analyses assuming that a set of chronic diseases occurred 1-3 years before the recorded date.

Results: Among normal weight individuals, both weight-loss and weight-gain were associated with increased rates of CVD [HR vs weight maintenance=1.53 (1.18-1.98) and 1.43 (1.19-1.71 respectively)]. Among overweight individuals, weight-loss and weight-gain had moderately higher rates of CVD [HR=1.20 (0.99–1.44) and 1.17 (0.99–1.38), respectively]. Among obese individuals, the weight-loss group was associated with moderately lower rates of CVD [HR =0.90 (0.72–1.13)] and lower rates of CHD [HR =0.66 (0.49–0.89)]. Sensitivity analyses estimates were attenuated among overweight individuals, and stronger for weight-loss among obese individuals.

Conclusion: Weight maintenance had lower rates of CVD among normal-weight and overweight individuals. In the obese, weight loss had moderately lower rates of CVD, which was more pronounced for CHD.

[PO6.15]

Age-Period-Cohort Analysis Of Colorectal Cancer Incidence In New Zealand 1994-2018

Malgorzata Hirsz[1], Lynne Chepulis[2], Michael Mayo[3], Lynette Hunt[1]

[1] *Department of Mathematics and Statistics, University of Waikato, New Zealand*
[2] *Medical Research Centre, University of Waikato, New Zealand*
[3] *Department of Computer Science, University of Waikato, New Zealand*

Background: The overall colorectal cancer (CRC) incidence in New Zealand (NZ) has been declining over the last two decades, but, curiously, it has been increasing in young adults and in the indigenous (Maori) population. Cohort and period components of trends can provide clues about factors which drive the trends but have not yet been examined in NZ.

Objectives: To disentangle the contribution of age, period and cohort effects to the temporal trends of CRC by anatomical sub-site and to investigate the contribution of those effects to the trends in Maori and non-Maori populations.

Methods: Data from CRC incident cases registered in the New Zealand Cancer Registry between 1994 and 2018 were analysed (n=65,530). We used age-period-cohort (APC) modelling to estimate average annual percentage change in incidence rates, incidence rates by 1-year age group (30–90 years) and incidence rate ratios by birth cohorts (non-overlapping 1-year birth cohorts 1904–1988).

Results: The overall decrease in CRC incidence rates between 1994 and 2018 by 1.25% per year was limited to individuals older than 45 years. Age-specific incidence in those younger than 45 years increased in successively younger birth cohorts, regardless of gender and ethnicity, similarly for proximal, distal and rectal tumours. The APC analysis revealed a strong cohort effect that could nearly entirely explain trends in CRC incidence, showing that generations born in the 1970s and 80s were being affected by the increased incidence rates. The cohort effects were very different in Maori and non-Maori populations born before 1955. Specifically, there was a sharp decrease in incidence in non-Maori born between 1939 and 1955, which Maori did not experience.

Conclusions: The APC analysis shows that in individuals born in the 1970s and 80s, the incidence is increasing sharply with later year of birth. The generation born in the year 1980, currently 40 years old, will in 20 years' time be 60 years old. At that age, CRC incidence is already high, even without the additional cohort effect. According to the fitted APC model, in the future, the combination of higher age and cohort effect will result in high CRC incidence rates in those generations.

Regular Posters

RP1: Applications

[RP1.01]

A cluster randomised controlled trial of lifestyle intervention for adolescents using 'SPRAT' program

Kazue Yamaoka[1], Junko Watanabe[2], Mariko Watanabe[3], Misa Adachi[4], Asuka Nemoto[1], Toshiro Tango[5]

[1] *Graduate School of Public Health, Teikyo University, Japan*
[2] *Department of Nutrition Management, Minami Kyushu University, Japan*
[3] *Showa Women's University, Japan*
[4] *Nutrition Support Network LLC, Japan*
[5] *Center for Medical Statistics, Japan*

Background: Severe subjective psychosomatic symptoms (SPS) in adolescents are a major public health concern, and lifestyle modification for reducing SPS is an important topic. Recently, we developed a school-based lifestyle education program involving parents for reducing SPS of adolescents (SPRAT) [1].

Objective: This study aimed to evaluate the effectiveness of SPRAT in reducing SPS among adolescents by a cluster randomised clinical trial (cRCT).

Methods: This was a 6-month, cRCT with two intervention arms (SPRAT vs. usual school education). SPRAT is an intervention to reinforce the role of parent participation in adolescents' healthy lifestyle modifications to reduce SPS. The study population was middle school students (aged 12–14 years) with their parents/guardians in Japan. The SPS score (with 9 items) was assessed according to the progression of SPRAT at baseline, 2, 4, and 6 months after. The primary effect size was the difference of the change from baseline SPS scores to those obtained after 6 months between the two groups. Crude and multivariate adjusted effect sizes were examined by sex using ANCOVA and linear mixed models. For handling missing data, we applied 1) likelihood-based ignorable analysis, 2) LOCF, and 3) multiple imputation under the MAR assumption. Lifestyle factors (11 items) were used as covariates. The sample size calculation was shown in elsewhere [1]. (Trial registration number: UMIN000026715)

Results: At baseline, 23 schools (n=2005) were included in this cRCT. The baseline SPS scores were almost similar among the groups. By the ANCOVA adjusted for covariates, the effect size was significant for boys (p=0.022) but not for girls (p=0.463). By the longitudinal analysis using 0, 2, 4, 6 months data, significant effects for the adjusted models were obtained among boys (p=0.009) and girls (p=0.004). The improvement of the lifestyle was related to the improvement of the SPS score.

Conclusion: The findings of this study will correspond to the critical need to develop effective and practical measures for minimizing SPS, and its potential influence among adolescents. The inclusion of parent-participation may help to obtain favorable effects.

Reference:

1. *Watanabe J, et al. BMJ Open 2018;8:2:e018938.*

[RP1.02]

Trends in incidence, mortality and survival of childhood cancer in the Czech Republic, 1994–2016

Denisa Krejci[1], Ivana Svobodova[2], Jaroslav Sterba[3], Jan Stary[4], Ladislav Dusek[2]

[1] *Faculty of Medicine, Masaryk University, Institute of Biostatistics and Analyses, Czech Republic*
[2] *Institute of Health Information and Statistics of the Czech Republic, Czech Republic*
[3] *University Hospital Brno and Faculty of Medicine, Masaryk University, Department of Paediatric Oncology, Czech Republic*
[4] *Department of Paediatric Haematology and Oncology, 2nd Faculty of Medicine, Charles University in Prague and Motol University Hospital, Czech Republic*

Background: Childhood cancers (or paediatric cancers) are classified among rare diseases as their incidence rates are orders of magnitude lower than cancer incidence rates in adults. About 270 new cases of childhood cancer are diagnosed in the Czech Republic each year. Although paediatric cancers are rare, they are the second leading cause of death among children (after injuries).

Objective: Our study aimed to evaluate incidence, mortality and survival trends for childhood cancer in the period 1994–2016 in the Czech Republic.

Methods: Data from a clinical database of childhood cancer patients were combined with retrospective data of the Czech National Cancer Registry and with data from death certificates. Cancers were classified into groups according to the International Classification of Childhood Cancer, 3rd Edition (ICCC-3)[1]. Data from death certificates were used to evaluate long-term trends in mortality. Incidence and mortality trends were assessed by the average annual percentage change (AAPC). The life tables method was used to calculate the overall age-standardised five-year survival. The period analysis [2] was used to calculate survival rates for three periods: 1999–2004, 2005–2010 and 2011–2016.

Results: The age-standardised incidence trend was stable (AAPC=0.2; 95% CI: –0.2, 0.6; p=0.343); however, there was a significant long-term decrease in mortality (AAPC=−5.1; 95% CI: −5.7, −4.4; p<0.001). The age-standardised five-year survival has markedly increased over the evaluated period. Regardless of the ICCC group, the five-year survival increased from 80% in the period 1999–2004 to 87% in the most recent period (2011–2016). Particularly large increases were observed for bone (+23%), neuroblastoma (+17%) and germ cell tumours, trophoblastic tumours and neoplasms of gonads (+11%).

Conclusions: This study provides reliable information on trends for childhood cancer in the Czech Republic.

References:

1. *Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P (2005): International Classification of Childhood Cancer, third edition, Cancer 103 (7): 1457–1467.*
2. *Brenner H, Gefeller O (1996): An alternative approach to monitoring cancer patient survival. Cancer, 78(9): 2004–2010.*

[RP1.03]

Absence of causal association of BMI with eGFR: One- and two-sample Mendelian randomization analyses

Minsu Park[1], Hyung-Jin Yoon[2]

[1] *Department of Statistics, Keimyung University, Republic of Korea*
[2] *Department of Biomedical Engineering, Seoul National University College of Medicine, Republic of Korea*

Background: Many studies observed that elevated body mass index (BMI) was associated with lower estimated glomerular filtration rate (eGFR), but their causal relationship remains uncertain.

Methods: To investigate the causal effect of BMI on eGFR levels, we performed Mendelian randomization analyses with a one-sample setting, using a two-stage least squares (2SLS) estimator on an Ansan-Ansung cohort of 7,829 participants, and a two-sample setting, using an inverse variance weighted (IVW) estimator, Mendelian randomization Egger regression (Egger), and weighted median estimator (WME) on the Genetic Investigation of Anthropometric Traits consortium and Nuclear Magnetic Resonance Genome-Wide Association Studies data. eGFR was calculated with the Chronic Kidney Disease Epidemiology equation.

Results: Although BMI was negatively associated with eGFR (coefficient = –0.377, p-value <0.001), the causal effect was not observed as a result of using 2SLS (coefficient = –0.19, p-value = 0.728) based on allele scores of 47 single nucleotide polymorphisms (SNPs) satisfying the conditions as instrumental variables among the 64,000 SNPs in the Ansan-Ansung cohort data. In a two-sample, Mendelian randomization setting, using four SNPs, which were significant in the Genetic Investigation of Anthropometric Traits consortium among SNPs analyzed in both consortia and one-sample setting, the causal estimates of IVW, Egger, and WME were not significant (IVW, coefficient = –0.236, p-value = 0.281; Egger, –0.192, 0.781; WME, –0.258, 0.242, respectively).

Conclusions: Our results do not support the causal effect of BMI on eGFR; the true nature of their association needs to be studied further.

[RP1.04]

Is there a difference in survival of patients with first primary and subsequent primary cancer?

Lucie Pehalova[1], Ivana Svobodova[2], Ladislav Dusek[2]

[1] *Faculty of Medcine, Masaryk University, Czech Republic*
[2] *Institute of Health Information and Statistics of the Czech Republic, Czech Republic*

Background: In the last decades, significant progress has been made in the prevention, diagnosis, and treatment of malignant neoplasms, leading to increasing life expectancy of patients with oncological diseases. At the same time, cancer survivors have an increased risk of developing subsequent primary malignancies due to a variety of unique factors, such as genetic predisposition, environmental exposures or late effects of anticancer treatment. Previous studies found significantly decreased survival rates of patients with subsequent primary cancer compared to first primary cancer [1].

Objective: The aim of this analysis was to assess the difference in survival of patients with first primary and subsequent primary cancer with respect to different oncological diagnoses and clinical stages.

Methods: We utilized the Czech National Cancer Registry to calculate overall and relative survival for the time interval 2013-2017 using the period method [2].

Results: Our results revealed significantly lower overall and relative survival of patients with subsequent primary cancer compared to first primary cancer for most of the analyzed diagnoses. The greatest differences were evident for endometrial cancer, prostate cancer, bladder cancer, Hodgkin lymphoma, and myelodysplastic syndrome. On the other hand, we found also significantly increased survival rates of patients with subsequent primary cancer compared to first primary cancer for few oncological diagnoses, namely liver cancer, lung cancer, cervix cancer, and chronic myeloid leukemia. In case of solid tumors, this unexpected increased survival was caused by higher proportion of early clinical stages in patients with subsequent primary cancer.

Conclusions: Significantly worse prognosis for subsequent primary cancer compared to first primary cancer for most of the analyzed diagnoses suggests that neoplasms diagnosed as subsequent primary should be considered as a different diagnostic group in terms of treatment.

References:

1. *Keegan T, Bleyer A, Rosenberg A, Li Q and Goldfarb M (2017): Second Primary Malignant Neoplasms and Survival in Adolescent and Young Adult Cancer Survivors. JAMA Oncol, 3(11):1554-1557.*

2. *Brenner H and Gefeller O (1996): An alternative approach to monitoring cancer patient survival. Cancer, 78(9):2004-2010.*

[RP1.05]

Competing risks analyses for evaluating autologous stem cell transplantations in elderly myeloma patients

Kaya Miah, Axel Benner

Division of Biostatistics, German Cancer Research Center (DKFZ), Germany

Multiple myeloma (MM) is a malign tumour disease of the blood building system. High-dose therapy followed by autologous stem cell transplantation (ASCT) is considered the standard of care for patients with newly diagnosed MM younger than 65 years. However, the median age at diagnosis is above 65 years and so far, there is only little evidence for the effectiveness and safety of ASCT in elderly patients. Since cost absorption of ASCT was only unconditionally granted for patients with diagnosed MM up to the age of 65 years by statutory health insurance, safety and risk assessment of ASCT for elderly patients is of great interest in particular for the Medical Service of German Statutory Health Insurance providers.

Our main aim focuses on the evaluation of the impact of ASCT in treating elderly patients with MM. To this end, prospective data of a randomised multicentre phase III trial comprising patients with newly diagnosed MM until 70 years of age were analysed with respect to a comparison of three predefined age groups (≤60 years, 61-65 years and 66-70 years). Following these findings, currently a registry-based study of elderly patients from 71 years of age throughout Germany is being initiated.

With regard to the statistical analysis, we have to face at least three major challenges. First, ASCT was not performed by randomisation. Second, risks and benefits of ASCT must be assessed against each other. Third, the number of events is expected to be small due to reduced sample sizes in subgroups. For investigating the impact of ASCT, the competing events of time to progression and non-relapse mortality as a potential consequence of ASCT have to be taken into account.

As a consequence, the comprehensive analyses imply that patients up to the age of 70 years may receive upfront ASCT if they are considered transplant-eligible. Based on these results, cost absorption of ASCT is now covered for patients with diagnosed MM to the age of 70 years by statutory health insurance. Beyond, cost transfer for transplanted patients from 71 years of age will be recommended within the context of the planned registry-based study.

[RP1.06]

Comparison of new diagnostic methods for determination of ethyl alcohol in forensic toxicology

Elżbieta Chełmecka[1], Marcin Tomsia[2], Joanna Nowicka[2]

[1] Department of Statistics, Medical University of Silesia, Poland
[2] Department of Forensic Medicine and Forensic Toxicology, Medical University of Silesia, Poland

Background: In some forensic autopsies blood is not available hence other matrices are sampled for toxicological analysis. The aims of the study were: to examine if ethanol can be detected in costal cartilage and to investigate whether analyses of different forms of costal cartilage can give useful information about alcohol concentrations in peripheral blood. This paper presents the results of a comparative study of alcohol concentration in post-mortem costal cartilage, blood, and urine samples collected during medico-legal autopsies.

Objectives: Diagnosis of the correlation between the concentration of ethanol in cartilage and the concentration of alcohol in autopsy blood. Comparison of two different methods of preparation/fragmentation of cartilage for toxicological analyzed. Assessment of the possibility of determining ethanol in rib cartilage.

Methods: The costal cartilage (CC) fragment was taken from the rib arch. The studied group included CCs taken from cadavers in which the presence of ethyl alcohol in blood and urine was noted, the control group consisted of CCs taken from cadavers with no alcohol detected. Each sample was divided into two sub-samples that were processed differently. For each cadaver (n=100) we obtained costal cartilage samples differing in the degree of fragmentation: (1) Unground Costal Cartilage and (2) Ground Costal Cartilage. Ethanol concentrations were determined in samples of UCC, GCC, femoral venous blood, and urine.

Based on the constructed ROC curves, the optimal cut-off point determining the sobriety (alcohol content in the cartilage) was calculated for both of analyzed methods. Two different cut-off points were considered, because according to the Polish law, the blood alcohol concentration >0.2 is defined as the 'after use' condition and the blood alcohol concentration >0.5 is referred to as the state of insobriety. To measure the accuracy of the new diagnostic tests, the area under a ROC curve (AUC) was analyzed.

Conclusions: Statistical analysis showed considerable utility of the CCs particle test in posthumous determination of ethanol concentration. In addition, it has been shown that higher ethanol concentrations were better determined in ground samples. Further investigations on ethanol post-mortem ethanol redistribution in costal cartilage are important for future diagnostic applications, especially in cases of advanced cadaver decomposition.

[RP1.07]

Disentangling causal relationships with Alzheimer's Disease using multivariable Mendelian Randomization

Jodie Lord[1], Rebbeca Green[2], Shing Wan Choi[3], Christopher Huebel[4], Marcus Richards[4], Pak Sham[5], Cristina Legido-Quigley[6], Ammar Al-Chalabi[7], Richard Dobson[8], Petroula Proitsi[1]

[1] *Institute of Psychology, Psychiatry & Neuroscience, Kings College London, United Kingdom*
[2] *NIHR Maudsley Biomedical Research Centre, South London and Maudsley, London, United Kingdom*
[3] *Icahn School of Medicine, Mount Sinai, New York City, United States*
[4] *Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden*
[5] *MRC Unit for Lifelong Health & Ageing, UCL, London, United Kingdom*
[6] *Department of Psychiatry, University of Hong Kong, Hong Kong*
[7] *Institute of Pharmaceutical Science, Kings College London, United Kingdom*
[8] *Farr Institute of Health Informatics Research, UCL Institute of Health Informatics, London, United Kingdom*
[9] *NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at South London and Maudsley NHS Foundation, London, United Kingdom*

No disease-modifying interventions currently exist for Alzheimer's Disease (AD), and whilst factors like schooling, intelligence and hypertension show consistent association, it remains unclear how these mediate biological processes to influence risk. Blood metabolites can provide vital clues to how such factors influence biological mechanisms and offer accessible sources of targeted intervention. Epidemiological studies have demonstrated AD-metabolite associations, however confounding and reverse causation have prevented causal conclusions. Here, we seek to disentangle metabolite-AD causality through a combination of polygenic risk scores (PRS) and Mendelian Randomization (MR). Then, using schooling, cognition, and hypertension, we investigate mediating relationships.

Summary statistics from the largest metabolomics genome-wide association study (GWAS) (Kettunen et al., 2016) were utilized to generate per-metabolite PRS models. These were applied to AD case/controls to assess genetic overlap (N=4725). Using the most predictive metabolites together with data from the largest AD GWAS (Kunkle et al., 2019), causality was then investigated through bi-directional MR. This included robust methods like MR-egger and weighted median, and sensitivity analyses such as leave-one-out. Cognition (Savage et al., 2018), schooling (Lee et al., 2018), and blood-pressure (Neale, 2017) data was then introduced, firstly to investigate univariable causal relationships with AD, then, via a multivariable framework, to interrogate mediating relationships. A significant difference between total (derived from univariable-MR) and direct (derived from multivariable-MR) effects were taken as evidence of mediation.

34 metabolites demonstrated suggestive genetic associations with AD. Of these, glutamine indicated a significant causal association. This was in the inverse direction, indicating a protective effect (β= -0.222, SE=0.069, p=0.001). Cognition and schooling both also demonstrated inverse causal associations. However, these did not mediate metabolite-AD relationships, nor did metabolites mediate causal relationships in the opposite direction. Blood-pressure demonstrated no causal association with AD; not directly nor via metabolites.

Using a multivariable framework together with the more conventional univariable-MR, results demonstrate that whilst glutamine and schooling/intelligence show a causally protective effect on AD, these are likely working via independent causal pathways. Intervening directly may thus go some way to preventing AD but our data suggests attempting to indirectly impact one via the other to influence disease risk may be ineffective.

[RP1.08]

Considerations in Choosing Methods of Randomization for Cluster-Randomized Studies

Jody Ciolino[1], Chen Yeh[1], Madeleine Organ[2], Alicia Diebold[3], Jessica Johnson[3], Kathryn Jackson[3], Stephen Persell[3], Darius Tandon[3]

[1] *Department of Preventive Medicine - Biostatistics, Northwestern University, United States*
[2] *Press Ganey, United States*
[3] *Northwestern University, United States*

Introduction: In cluster-randomized trials (CRTs), constraining randomization using baseline cluster-level variables is a technique used to control covariate imbalance across arms. Implementation of these methods varies depending upon complexities in study design and logistics. We use several CRTs as use cases coupled with simulations to illustrate a range of covariate-constrained randomization (CCR) algorithm(s) and explore their operating characteristics.

Methods: Complexities encountered in our applied examples include: staggered randomization due to logistical constraints, unequal allocation across more than two arms, mixtures of continuous and categorical covariates, cluster dropout, and cluster-level variables that are poor representations of participant-level covariates used in analyses (we call this "misspecification" for simplicity).

The general CCR logic involves:

1. Enumerating a large subset of possible allocation schemes.
2. Evaluating imbalance for each variable for each scheme.
3. If the imbalance is acceptable according to some pre-specified criterion, then save this scheme in a smaller subset for implementation.
4. Of those from step #3, randomly select one sequence for use in the current study.

When randomization was staggered, we employed this logic repeatedly within a trial. We explored multiple measures of imbalance (step #2 above) and criteria for acceptable levels of imbalance (step #3). We used simulation studies to explore performance of our imbalance metrics, especially in the presence of covariate "misspecification."

Results: We successfully implemented CCR in multiple unique studies in the United States -ranging from the primary care clinic setting to home visiting centers in low-income, underserved perinatal populations. Our cluster-level summary statistics illustrate comparable intervention arms; however, failing to account for misspecification in the trials where the unit of analysis is the individual participant decreases the efficiency of the algorithm.

Conclusion: With our novel, yet intuitive approaches to CCR methods, we achieved comparable baseline variable distributions across arms at the cluster level. While the efficiency decreases in the case of misspecification, simulations illustrate that the performance of CCR still protects against large chance imbalances of simple randomization. Taking our experiences and results together, we conclude with recommendations for researchers considering implementing CCR in cluster-randomized trials.

[RP1.09]

An application of marginal structural joint frailty models on arrhythmogenic cardiomyopathies

Caterina Gregorio, Marco Merlo, Davide Stolfo, Giulia Barbati

Department of Medical Sciences, University of Trieste, Italy

Background: Arrhythmogenic Right-Ventricular Cardiomyopathy/ Dysplasia (ARVC/D) is a complex life-threatening disease and patients in their lifetime experience severe arrythmias and can develop Heart Failure (HF). To prevent arrhythmic events and to slow HF progression, among other drugs, beta-blockers can be prescribed. However, the efficacy of this treatment is still debated. Scientific evidence on this subject is lacking because it is a rare and complex disease and the only data available comes from observational registries.

Objective: The aim is to provide a statistical modelling strategy to estimate the treatment effect of beta-blockers on the risk of recurrent arrythmias and on the risk of heart transplant or death in the contest of an observational study on ARVC/D.

Methods: Because of the observational nature of the study and the high number of possible confounders in the treatment-outcome causal relation, a propensity score method was used. Since treatment, and most of the confounders are time-dependent, a marginal structural approach was applied (Hernán, Brumback and Robins, 2000). The first step consisted in specifying a model for estimating time-dependent inverse probability of treatment weights for beta-blockers.

The effect of treatment on recurrent arrhythmias and the terminal event was estimated by a Joint Frailty model (JF) (Liu, Wolfe and Huang, 2004). The JF allows to jointly estimate the risk of recurrent events and the terminal event. It assumes a unique individual frailty term which models both the association within the recurrent events and between the latter and the terminal event.

Results: Beta-blockers at a dosage ≥ 50% of the target dose showed a protective effect on the risk of recurrent events (HR: 0.10; CI95%0.02-0.46). Moreover, regardless of the dose, they decreased also the risk of the terminal event (HR: 0.19; CI95%0.06-0.70). Finally, the variance of the frailty term indicated a high within-subject variability. Furthermore, the estimated model confirmed the increase of the risk of death due to recurrent events.

Conclusion: Marginal structural joint frailty models represent a valid modelling strategy to estimate the treatment effect on the risk of recurrent events and death in longitudinal and observational studies with dynamic treatments such as in the ARVC/D setting.

[RP1.10]

COPD readmissions prediction model

Jose Maria Quintana Lopez[1], Ane Anton-Ladislao[1], Cristobal Esteban[2], Amaia Aramburu[2], Miren Orive[1], Iratxe Lafuente[1], Susana Garcia-Gutierrez[1]

[1] *Research Unit, Hospital Galdakao, Spain*
[2] *Respiratory Service, Hospital Galdakao, Spain*

Background: Chronic Obstructive Pulmonary Disease (COPD) is one of the more common chronic conditions nowadays. Readmissions after an index admission by a COPD exacerbation (eCOPD) is very common and increase their probability of complication or death. Therefore, knowing predictive factors of short term readmissions could provide with tool to prevent them

Objective: To develop a predictive model of readmissions at 60 days after an index admission by eCOPD.

Methods: Observational prospective cohort study of patients who attended an emergency department (ED) at one of 16 hospitals and were admitted by an eCOPD. Information on possible predictor variables was gathered from the medical records and from the patient during the ED stay, during admission, at hospital discharge time and until 60 days after discharge. Readmission within 2 months of the index admission was the main outcome. Statistical analysis: Multilevel multivariable regression models were employed, using generalized linear mixed models (GLMMs), to predict readmission. To assess predictive ability of the model the Area Under the receiver operating characteristic (ROC) curve (AUC) was used, as well as Hosmer and Lemeshow test to evaluate the calibration of the final model.

Results: Our study sample was comprised of 1436 patients, having 414 (28.83%) patients a readmission within 2 months. Final multivariable model, adjusted by participant hospital, included the following predictors of readmissions: Having admissions in the previous year (Odds Ratio (OR): 2.405. (95% confidence interval (95%CI: 1.636 – 3.537))), moderate or severe stable COPD (OR: 1.770. 95%CI: 1.194 -2.626), presence of cardiac comorbidity (OR: 1.520. 95%CI: 1.128 – 2.048), low level of blood hemoglobin (OR:1.088. 95%CI: 1.011 – 1.171), not walking before discharge (OR:1.672. 95%CI: 1.102 – 2.538), prescription of non-invasive mechanical ventilation at home after discharge (OR:1.799. 95%CI: 1.168 – 2.770) and absence of social support at home (OR: 2.882. 95%CI: 1.954 – 4.251). AUC for the model was 0.710 (95%CI: 0.674 – 0.747). Hosmer-Lemeshow test was 0.9941.

Conclusions: This model includes predictors from the background of the patient, present exacerbation, status at discharge and social support. Some of them are preventable or modifiable parameters that could help to establish future interventions.

[RP1.11]

Socioeconomic status and survival after stroke – using mediation and sensitivity analyses to assess the effect

Anita Lindmark[1], Bo Norrving[2], Marie Eriksson[1]

[1] Department of Statistics, Umeå School of Business, Economics and Statistics, Umeå University, Sweden
[2] Department of Neurology, Lund University, Sweden

Background: It has been established that low socioeconomic status is linked to increased risk of adverse outcome after stroke but the mechanisms behind this link are still unclear.

Objective: In this study we aim to shed light on the relationship between income level and survival after stroke by investigating the extent to which differences in stroke severity account for differences in survival.

Methods: The study was based on patients registered in Riksstroke (the Swedish stroke register) with first time ischemic stroke (n = 51 159) or intracerebral hemorrhage (n = 6777) in 2009-2012. We used causal mediation analysis to decompose the effect of low income on 3-month case fatality into a direct effect and an indirect effect due to stroke severity. Since causal mediation analysis relies on strong assumptions regarding residual confounding of the relationships involved, recently developed methods for sensitivity analysis were used to assess the robustness of the results to unobserved confounding.

Results and Conclusions: After adjustment for observed confounders, patients in the lowest income tertile had a 3.2% (95% CI: 0.9-5.4%) increased absolute risk of 3-month case fatality after intracerebral hemorrhage compared to patients in the two highest tertiles. The corresponding increase for case fatality after ischemic stroke was 1% (0.4-1.5%). The indirect effect of low income, mediated by stroke severity, was 1.8% (0.7-2.9%) for intracerebral hemorrhage and 0.4% (0.2-0.6%) for ischemic stroke. Unobserved confounders affecting the risk of low income, more severe stroke and case fatality in the same directions could explain the indirect effect, but additional adjustment to observed confounders did not alter the conclusions. This study provides evidence that as much as half of income-related inequalities in stroke case fatality is mediated through differences in stroke severity. Targeting stroke severity could therefore lead to a substantial reduction in inequalities and should be prioritized. Sensitivity analysis suggests that additional adjustment for a confounder of greater impact than age would be required to considerably alter our conclusions.

[RP1.12]

Modelling time-varying effects on recurrent pneumonia in children under two years

Jordache Ramjith[1], Kit Roes[1], Heather Zar[2], Marianne Jonker[1]

[1] Department of Health Evidence, Radboudumc, Netherlands
[2] Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital and University of Cape Town, South Africa

Background: Pneumonia is one of the leading causes of morbidity and mortality in children under two years in low-to-middle income countries. To study risk factors for pneumonia and other recurrent diseases, a shared frailty Cox proportional hazards (CPH) model is often fitted. The assumption of proportional hazards is not always true in practice. Sex, for instance, is constant over time, but its effect on the hazard rate does not need to be.

Objectives: We aim to investigate the time-varying behaviour of the effect of risk factors on pneumonia incidence, in the presence of recurrent events.

Methods: The Drakenstein child health study (DCHS) was designed to investigate the incidence of and risk factors for childhood pneumonia. In total, 1137 children were followed from birth to two years or until loss to follow up/death. We fitted the data with a shared frailty model. The effects of risk factors were estimated with spline functions (as a function of the follow-up time) to study the time-varying behaviour of these effects on the incidence of pneumonia and episode recurrence in the DCHS data.

Results: Effects plots for the change in the hazard ratio over time showed time-varying hazard ratios. Specifically, male sex and HIV exposure were high when children were very young and then gradually decreased until a null effect. The large pneumonia hazard ratio between boys and girls found by the CPH model can be explained by much larger hazard differences between very young boys and girls and almost no differences in the hazard between older boys and girls. The effect of preterm birth showed a time-constant hazard ratio implying that preterm birth persists as a risk factor throughout all ages of the first two years of life.

Conclusion: Most time-to-event analyses assume constant effects over the risk period. However, for pneumonia incidence, the effect of some risk factors change as the children get older.

[RP1.13]

The use of recurrent models to assess the risk of subsequent PCI or death in ischemic heart disease patients

Michał Michalak[1], Daria Adamczak[2], Jarosław Hiczkiewicz[3], Stefan Grajek[2]

[1] Department of Computer Science and Statistics, Poznań University of Medical Sciences, Poland
[2] 1-st Department of Cardiology, Poznań University of Medical Sciences, Poland
[3] Department of Cardiology of Specialised Hospital, Nowa Sol, The Faculty of Medicine and Health Sciences of the University of Zielona Gora, Poland

Background: Many medical studies compare the effects of treatment with the incidence of adverse events. The Cox proportional hazard model has often been used to identify risk factors for death or adverse events in both cardiology and oncology. However, these publications tend to analyze to the first event, but ignore other events that are correlated with each other.

Objectives: The aim of this research was to identify the effect of selected factors on the risk of a complex endpoint–whether another hospitalization or death–based on the multiple event models available in the literature. The next step was to compare the results using the analyzed models in order to assess their effectiveness in detecting relevant factors, and also as discuss their suitability for analyzing this type of medical data.

Methods: The 3397 patients enrolled in the study had had at least one coronary angioplasty (PCI) in the period between 2009 and 2015. The data on risk factors and comorbidities were extracted from patient interviews, the course of the first PCI treatment, the electrocardiogram and the heart echo taken upon admission were taken into analysis. To identify the effect of selected factors on the risk of subsequent hospitalization or death in patients with ischemic heart disease, using the following multiple event models: The Andersen–Gill Model (AG), The Prentice–Williams–Peterson total time model (PWP-TT), The Prentice–Williams–Peterson time gap model (PWP-GT), The Lee–Wei–Amato marginal means and rates model (LWA), The Wei–Lin–Weissfeld model (WLW), The frailty model.

Results and Conclusions: Of the 60 analyzed factors, 2 to 38 significant risk factors for subsequent hospitalizations were demonstrated, depending on the model used. The most effective model proved to be WLW. The second most effective is the AG model. In third place come the PWP-TT, PWP-GT, LWA models, and the frailty model with the Gaussian inverse distribution of random effects. Worst was the frailty model with the gamma-distributed random effect.

[RP1.14]

A Systematic Review of Subgroup Analyses in Randomized Clinical Trials about the Cardiovascular Disease

Alexander Hapfelmeier[1], Korbinian Brand[2], Kurt Ulm[1], Bernhard Haller[1]

[1] Institute of Medical Informatics, Statistics and Epidemiology, Technical University of Munich, Germany
[2] School of Medicine, Technical University of Munich, Germany

Background: Subgroup analyses are frequently used to assess heterogeneity of treatment effects in randomized clinical trials (RCT). Unreliable, inconsistent, inappropriate and incomplete implementation, reporting and interpretation have been identified as ongoing challenges. Recently, recommendations and guidelines have been provided in order to improve the reporting in this regard.

Methods: This systematic review was based on a systematic literature search within the online databases of three selected medical journals, including The New England Journal of Medicine (NEJM), The Lancet and Circulation. We reviewed articles of RCTs in the domain of cardiovascular disease which were published in 2015 and 2016. We screened and evaluated the selected articles for the mode of implementation and reporting of subgroup analyses.

Results: We were able to identify a total of 130 eligible publications of cardiovascular RCTs. In 89/130 (68%) articles results of at least one subgroup analysis were presented. This was dependent on the considered journal ($p < 0.001$), the number of included patients ($p < 0.001$), and the lack of statistical significance of a trial's primary analysis ($p < 0.001$). The number of reported subgroup analyses ranged from one to 101 (median = 13). The specification time was deducible for 71/89 (80%) articles with 55/89 (62%) articles presenting exclusively pre-specified analyses. This information was not always provided by trial protocols and often did not include the pre-definition of cut-off values for the categorization of subgroups. The use of interaction tests was reported in 84/89 (94%) articles.

Conclusion: Compared to previous reviews in this context we observed improvements in the reporting of cardiovascular RCTs. Nonetheless, critical shortcomings, such as inconsistent reporting of implementation, insufficient pre-specification and non-addressed multiplicity issues, persist.

[RP1.15]

Association of red meat consumption and cardiovascular diseases in participants with and without obesity

Jufen Zhang[1], Karen Hayden[1], Ruth Jackson[1], Rudolph Schutte[2]

[1] School of Medicine, Anglia Ruskin University, United Kingdom
[2] School of Allied Health, Anglia Ruskin University, United Kingdom

Background: There is increasing evidence that red and processed meat (RPM) intake is associated with increased risk of cardiovascular (CV) disease. However, there is little literature evaluating the association among people with and without obesity.

Objectives: To evaluate the association between RPM intake and the risk of CV death, cerebrovascular, and ischemic heart diseases in participants with obesity and non-obesity using the UK Biobank data

Methods: In this large prospective population-based cohort study, the RPM consumption was assessed through the UK Biobank touch-screen questionnaire at baseline. The estimated hazards ratios (HRs) with 95% confidence intervals (CIs) were obtained from the Cox proportional hazard models to assess the association between RPM intake (including a total RPM consumption and terciles of RPM intake respectively) and the risk of CV diseases. The potential non-linear relationships were assessed by using restricted cubic spline models.

Results and Conclusions: Of 428,070 participants who had RPM consumption, 100,175 (23.4%) were obese with the mean age of 56 (SD: 7.9) years and 54% were female. Participants without obesity, the mean age was 56 (SD: 5.2) and 55% were female. The overall median follow-up was 7.2 (IQR: 6.5-7.8) years. RPM intake had increased risk of CV death (HR (95%CI):1.04 (1.01-1.08) per week serve for those who had obesity and 1.04 (1.02-1.07) for no obesity) after adjusted for age, sex, ethnicity, education, smoking and alcohol status and overall health. The positive association between RPM intake and ischemic heart disease was only observed in participants without obesity (HR (95%CI): 1.03 (1.00-1.05) for the highest versus lowest terciles of RPM intake). No association was found with cerebrovascular disease in the participants regardless of obesity. In conclusion, consumption of red and processed meat is associated with higher risk of CV death in both obesity and non-obesity groups. The risk of ischemic heart disease associated with red and processed meat consumption may be higher in participants without obesity. Further studies are needed to understand the full extent of the mechanism of the association.

[RP1.16]

Associations Between Commonly Prescribed Medications And Cancer Risk: A Series of Nested Case-Control Studies

Ronald McDowell[1], Carmel Hughes[2], Peter Murchie[3], Chris Cardwell[1]

[1] Centre for Public Health, Queen's University Belfast, United Kingdom
[2] School of Pharmacy, Queen's University Belfast, United Kingdom
[3] Division of Applied Health Sciences, University of Aberdeen, United Kingdom

Background: Screening studies have been used to identify medications which may be associated with an altered cancer risk. However none have taken place in the United Kingdom (UK) where prescribing patterns differ.

Objectives: To determine candidate medications associated with a higher or lower risk of cancer.

Methods: Nested case-control studies were undertaken using the Primary Care Clinical Informatics Unit database, a database of over two million Scottish patients. Each cancer case was matched with up to five controls based on age, gender, general practice and year of registration. The index date for each matched set was the first date of primary cancer diagnosis for the case. The exposure period, during which medication use was determined, was truncated within each matched set to ensure it was identical for all members. Prescriptions in the year prior to the index date were excluded as they may have been issued due to undiagnosed cancer. Associations between the most commonly-prescribed medications and cancer were estimated using conditional logistic regression, adjusting for Charlson conditions specific risk factors not known to be caused by medication. Medications deemed worthy of further study (i.e. "signals") were those with modest effect sizes, of statistical significance, and which exhibited a dose-response relationship with cancer risk.

Results: 338,599 patients were included in the analyses. There were 353 signals with at least 50 case users or 250 control users of medication. Results replicated well-known associations between cancer and increased risk (e.g. estrogen (hrt) and breast cancer: Odds Ratio (OR) (any use): 1.24, 95%CI (1.18,1.32), $p<0.001$) or decreased risk (e.g. aspirin and oesophageal cancer: OR (any use): 0.72, 95%CI (0.53,0.98), $p=0.03$). They also suggest other medications may be associated with an altered cancer risk, such as allopurinol and colorectal cancer (OR (≥6 items v<6 items): 1.35 (1.16,1.57), $p<0.001$), or clopidogrel and pancreatic cancer (OR (≥6 items v<6 items) 2.16, 95%CI (1.35,3.46), $p=0.001$).

Conclusions: This study shows that medications known to be associated with an altered cancer risk are identified in UK patients. It suggests other medications for which there may potentially be a clinically plausible causal relationship. Signals require examination elsewhere to determine whether these associations are replicated.

[RP1.17]

A case study of initial data analysis for longitudinal studies

Lara Lusa[1], Andrej Srakar[2], Marianne Huebner[3]

[1] *Department of Mathematics, University of Primorska and University of Ljubljana, Slovenia*
[2] *Institute for Economic Research (IER) and School of Economics and Business University of Ljubljana, Slovenia*
[3] *Department of Statistics and Probability, Michigan State University, United States*

Background: Systematic initial data analysis (IDA) and clear reporting of the findings is an important step towards reproducible research. A general framework of IDA for observational studies was proposed to include data cleaning, data screening, and possible refinements of the preplanned analyses. Longitudinal studies, where participants are observed repeatedly over time, have special features that should be taken into account in the IDA steps before addressing the research question. Our aim was to use a case study of data in decline of grip strength in aging populations to explore the specificities of IDA for longitudinal data.

Methods: Based on the IDA framework we describe a detailed IDA plan for the analysis of average rate of decline of grip strength with age in men and women 50 years or older, using data from 120,000 individuals included in five waves of the cohort longitudinal study Survey of Health Ageing and Retirement in Europe.

Results: We included in the IDA plan several steps that are specific to longitudinal data, or bear greater importance when data are longitudinal. For example, IDA was used to check the consistency of time-varying covariates, to summarize longitudinal average trends, to explore the variation between individuals, to characterize the correlation and the covariance, for the exploration of missing values and for the description of drop-out.

The IDA steps allowed us to refine the choices about the initial model specification or to confirm the appropriateness of the choices included in our statistical analysis plan. All the steps of IDA and statistical analyses were reported in a reproducible document and IDA findings were summarized.

Conclusion: Appropriate numerical and graphical tools for longitudinal data allow the researchers to conduct IDA in a reproducible manner to avoid non-transparent impact on the interpretation of model results. We provide an example on how to conduct IDA in the context of a longitudinal population cohort study.

[RP1.18]

Determining the influence of individual metabolic profiles on drug safety in Germany- the EMPAR study

Christoph Roethlein[1], Steffen Hess[2], Michael Steffens[3], Jochen Fracowiak[3], Tatjana Hübner[3], Daria Langer[4], Marco Garling[4], Felix Falkenberg[5], Willy Gomm[1], Roland Lindner[4], Julia Stingl[5], Britta Haenisch[2]

[1] *Department of Pharmacoepidemiology, German Center for Neurodegenerative Diseases, Germany*
[2] *Department of Pharmacoepidemiology, Federal Institute for Drugs and Medical Devices, Germany*
[3] *Department of Pharmacogenetics, Federal Institute for Drugs and Medical Devices, Germany*
[4] *Techniker Krankenkasse, Germany*
[5] *Department of Pharmacology and Toxicology, Institute of Clinical Pharmacology University Hospital of RWTH Aachen, Germany*

Background: Genetic variability in metabolic profiles can affect the individual response to drug treatment and the occurrence of adverse drug reactions (ADRs). Research on the magnitude of this effect in routine care is important to assess usefulness of preemptive pharmacogenetic testing in clinical practice.

Objectives: The main objective is to determine potentially modified risks for ADRs for pharmacogenetically relevant drug classes.in patient groups with different metabolic profiles. Therefore, the EMPAR project contributes to drug safety in routine care.

Methods: The cohort study EMPAR includes study participants who are insured by the large German health insurance Techniker Krankenkasse, users of a pharmacogenetically relevant drug (including statins and anticoagulants) and provide DNA samples for the analyses of metabolic profiles. Also, patients who received an ICD-10 Y57.9 diagnosis of ADRs are involved in the study. Participants are analyzed regarding their metabolic profile, examining pharmacogenetically relevant enzymes such as the cytochrome p450 enzymes CYP2C9, CYP2C19, and CYP3A4. The occurrence of ADRs is evaluated amongst different metabolic profile groups using routine care data. Additional pharmacoepidemiological research includes analyses with further parameters such as economic outcomes and health care utilization.

Results: Currently, metabolic profiles of 7668 participants have been determined and combined with relevant information from routine care data. Analyzing this information, we are able to determine differences in ADR patterns for distinct metabolic profiles.

Conclusions: We successfully implemented the first large scale study focusing on the effect of metabolic profiles on drug safety with German routine care data. Our results can provide insights in the healthcare potential of pharmacogenetic testing in clinical practice.

[RP1.19]

Joint modeling of growth, puberty and type 1 diabetes in a multi-cohort data

Essi Syrjälä[1], Riitta Veijola[2], Harri Niinikoski[3], Jorma Toppari[3], Mikael Knip[4], Helena E. Virtanen[5], Jorma Ilonen[3], Suvi M. Virtanen[6], Jaakko Peltonen[1], Jaakko Nevalainen[1]

[1] *Tampere University, Finland*
[2] *University of Oulu and Oulu University Hospital, Finland*
[3] *University of Turku and Turku University Hospital, Finland*
[4] *University of Helsinki and Helsinki University Central Hospital, Tampere University Hospital, Folkhälsan Research Center, Finland*
[5] *University of Turku, Finland*
[6] *Tampere University and Tampere University Hospital, National Institute for Health and Welfare, Finland*

Background: The incidence of type 1 diabetes (T1D) has increased worldwide with Finland having the highest rates but reasons for the increase remains still unclear. To our knowledge, studies concerning the influence of pubertal onset on T1D do not exist but since there occur various changes in body during puberty it could potentially affect to the disease development. We have a large dataset from three Finnish cohort studies that enables the study of this issue.

Objectives: We set out to 1) estimate the puberty-related individual growth parameters: age at peak height velocity (APHV) and PHV, 2) study if we can determine the pubertal onset ages based on APHVs and PHVs, and 3) study how can we use the predicted pubertal onset ages to study the influence of pubertal onset on the development of T1D.

Methods: Data were from Special Turku Coronary Risk Factor Intervention Project (STRIP), Type 1 Diabetes Prediction and Prevention (DIPP) Study and Boy cohort. Height was measured in all the cohorts and pubertal markers in STRIP and Boy cohort. DIPP included follow-up for T1D and related autoantibodies. APHVs and PHVs obtained via a Super-Imposition by Translation And Rotation model were used as covariates in a time-to-pubertal onset model to predict the pubertal onset ages for STRIP and Boy cohort. To study the success of the model, predictions were obtained using a cross-validation and compared to pubertal data. Prediction model was further used to obtain the pubertal onset ages for DIPP children to enable to study the influence on T1D.

Results: APHVs and PHVs were straightforwardly estimated for the majority of the children. Observed and model-predicted pubertal onset ages agreed with moderate accuracy. Individual predictions and prediction intervals could be obtained via prediction model and used to study the influence on T1D.

Conclusions: Longitudinal growth data can provide accurate estimates of individual PHV and APHV. These in turn can be used to estimate timing of puberty, but such predictions should be interpreted as intervals rather than point estimates. Pubertal onset prediction provides an opportunity to evaluate the influence of pubertal onset on T1D in DIPP Study with lacking pubertal follow-up.

[RP1.20]

Longitudinal Analysis of Medical Faculty Students' Empathy Scores With Latent Growth Curve Models

Funda Seher Ozalp Ates[1], Ipek Gonullu[2], Derya Gokmen[1], Sengul Erden[2]

[1] *Department of Biostatistics, Ankara University, Turkey*
[2] *Department of Medical Education and Informatics, Ankara University, Turkey*

Background: In the analysis of longitudinal data, both changes in individuals within time and inter-individual differences should be evaluated. Latent Growth Curve Models (LGCM) is one of the commonly used approaches to evaluate both differences within the scope of Structural Equation Models. While the model fit is evaluated with goodness of fit indexes, the intercept and slope values are also interpreted.

Objective: The aim of this study is to evaluate the change in empathy scores of Ankara University Faculty of Medicine students with time.

Methods: In this study; the Student Version of the Jefferson Scale of Physician Empathy (JSPE) was used. JSPE was applied to the students who started their medical education in 2012-2013 academic year (n=317), at the 7 different time points until the last year. JSPE consists of three subscales (perspective taking, compassionate care and standing in the patient's shoes) and LGCM was applied for the scores of these three dimensions separately. Model fit was evaluated by the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) and Root Mean Square Error of Approximation (RMSEA). The analyses were performed with R programming language v.4.0.2.

Results: For "Perspective Taking", the scores at initial (the intercept) was 3.620, this score decreased on average (the slope) by 0.043 over the 7 time points. The RMSEA, CFI and TLI was 0.079, 0.876 and 0.887, respectively. For "Compassionate Care", the scores at initial was 3.491, this score decreased, on average, by 0.019 over the 7 time points. The RMSEA, CFI and TLI was 0.078, 0.871 and 0.882, respectively. For "Standing in the Patient's Shoes", the scores at initial was 3.253, this score increased on average by 0.042 over the 7 time points. The RMSEA, CFI and TLI was 0.032, 0.968 and 0.971, respectively.

Conclusions: When the longitudinal data of student's empathy scores were evaluated with the LGCM, it was concluded that while scores for "Perspective Taking" and "Compassionate Care" subscales decreased, those for "Standing in the Patient's Shoes" subscale increased. As unsuitable learning environments, distress, negative role-models and patient factors affect the students, these results were compatible with the literature.

[RP1.21]

Estimating modifiable stroke risk in multiple regions of the world: the INTERSTROKE Modifiable Risk Score

Alberto Alvarez-Iglesias, John Ferguson, Conor Judge, Martin O'Donnell

School of Medicine, HRB Clinical Research Facility, NUI Galway, Ireland

Background: Stroke is a leading cause of death and disability, making the prevention of stroke a global health priority. The creation of objective and quantifiable measures of risk for patients who might potentially develop the disease is of vital importance to avoid subjective risk assessments based on physicians' examinations, in line with the principles of evidence-based medicine.

Objective: A derivation of a score for the risk of Ischemic stroke using an incidence prediction model that, unlike other existing stroke risk assessment tools, is based on both modifiable and non-modifiable risk factors as measured in the INTERSTROKE case-control study (n=26945).

Methods: Multivariate logistic models were used to generate a risk score for stroke (ischaemic stroke) combining the INTERSTROKE dataset and incidence estimates obtained from the Global Burden of Disease Epi Visualization tool (IHME 2017). All models were validated internally using a 3-fold split sample (2/3 training set and 1/3 test set) and final model's external validation was performed using an international prospective cohort study. Potential modifiable risk variables included drinking habits, eating, smoking, stress factors, lipids and physical activity.

Results: Final evaluation of model discrimination was assessed using the area under the curve (AUC) of the receiver-operating characteristic (ROC) function with an estimated AUC of 0.78 in the test set for the risk of Ischemic stroke. External model discrimination assessment resulted in an estimated AUC of 0.72 (95% CI 0.66 to 0.79). Variations of these findings were minimal across regions and ethnic groups.

Conclusions: A simple stroke risk prediction score has been developed using INTERSTROKE, an international case-control study, which has the potential to be used globally in the management of individuals at risk of stroke. Unlike other risk prediction tools that are mostly based on non-modifiable risk factors, the proposed model makes it practical for clinicians to intervene and modify the patients risk of developing the disease.

[RP1.22]

MEDIATION ANALYSIS IN RANDOMISED CONTROLLED TRIALS WITH ORDINAL OUTCOMES

Heather Riley

Bio-statistics, Manchester University, United Kingdom

In clinical trials the primary aim is to assess the efficacy of an intervention. Once this has been established, the underlying mechanism by which the intervention affects the outcome is of interest. Mediation analysis can be used to help us understand these mechanisms, to evaluate the possible pathways between treatment and outcome.

Aneurysmal subarachnoid hemorrhage (aSAH) is a cerebrovascular event with a poor long term prognosis. Current research highlights the limited treatments available and a failure of neuroprotection, demonstrating a pressing need to understand the underlying mechanisms.

Classical approaches to mediation are well developed for a continuous outcome, however in many settings the outcome of interest is measured on an ordinal scale, stroke studies in particular use scales such as the modifed Rankin Scale (mRS) or the Glasgow Outcome Scale - extended (GOSe).

The objective is to apply mediation analysis in the setting of a randomised controlled trial with binary exposure, continuous mediator and ordinal outcome. This is achieved by implementing a modern re-sampling approach on data from a clinical trial into aSAH.

[RP1.23]

Development and validation of a heart failure phenotype stratified prognostic model: an IPD meta-analysis

Yuntao Chen, Douwe Postmus

Department of Epidemiology, University Medical Center Groningen, Netherlands

Background: Many clinical prediction models have been developed with the goal of stratifying hospitalized patients with heart failure (HF). Some of these models have been developed in patients with a particular HF phenotypes, while others have been developed in more general populations with a mixture of HF phenotypes. While such heterogeneity population models are more widely applicable, they are also likely to yield larger prediction errors in settings where HF subtypes have substantially different baseline survival rates. In such situations, the use of a HF phenotype stratified statistical model may result in a higher predictive accuracy without limiting the model's applicability to a particular phenotype. In terms of model development, the vast majority of the existing prediction models were derived using data from a single HF cohort and then either internally validated or externally validated using data from a second HF cohort. An alternative approach that makes better use of the available data is to perform an individual participant data (IPD) meta-analysis.

Objective: The purpose of this study was to use an IPD meta-analysis to develop an HF phenotype stratified model for predicting all-cause mortality in patients admitted with acute HF.

Methods: We pooled data from three European cohorts for the development of a prediction model, with a fourth Scottish cohort for external validation. Cox model was used to derive the prognostic index. Weibull model was used to obtain the baseline hazard function with the prognostic index obtained from the Cox model included as an offset. The model performances were assessed by discrimination and calibration. Internal validation (bootstrap) was implemented to estimate optimism and correct measures of predictive performance. The internal-external cross validation (IECV) approach was used to evaluate the generalizability of the developed model across the three included cohorts. Finally an independent Scottish cohort was used for an external validation.

Results and conclusions: Our HF phenotype-stratified model validated well both within the derivation population and also within an independent Scottish dataset. It can be seen as a useful tool in quantifying and classifying the prognosis of patients with acute HF, allowing more targeted treatment and management of those patients.

[RP1.24]

Using elastic net regularised logistic regression to find key protective antibodies against placental malaria

Saber Dini[1], Elizabeth Aitken[2], Stephen Rogerson[2], Julie Simpson[1]

[1] Centre for Epidemiology & Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Australia
[2] Department of Medicine (RMH), Peter Doherty Institute, University of Melbourne, Australia

Background: Pregnant women are at great risk of suffering from complications of malaria, majorly due to development of placental malaria caused by Plasmodium falciparum infection. Studies have shown that pregnant women who are continually exposed to malaria infection can acquire immunity to this disease, protecting them from the development of placental malaria. However, identifying the key antibody characteristics that confer the protection is difficult because of the complex activation network of numerous antibodies, characterised by high-dimensional, highly correlated and noisy data.

Objective: To identify influential antibody characteristics associated with conferring immunity against placental malaria in pregnant women using machine learning methods that allow variable selection.

Methods: We used logistic regression with elastic net regularisation [1] to select the most influential antibodies associated with protection against placental malaria. Missing values were multiply imputed using chained equations. Cross-validation, with adjustment for class imbalance, was performed to find the parameter estimates that provided best classification performance. Robustness of the selected set of variables was evaluated by their selection frequency in the cross-validation. Regularisation paths of the parameters were examined to identify the antibody characteristics that impact protection collectively in groups. Finally, the results were compared with sparse Partial Least Squares method (sPLSDA) regularised with LASSO.

Results: We analysed a broad array of 169 antibody characteristics towards P. falciparum in a case-control study of 77 (27 placental and 50 non-placental malaria) infected pregnant women from Papua New Guinea. A total of nine antibody characteristics were identified as important protective antibodies against placental malaria. A complementary correlation network analysis confirmed that the selected antibody characteristics provide distinctive information for discriminating pregnant women groups. Comparing the results of elastic net and sPLSDA showed a large overlap between the antibody characteristics that were selected by the methods, confirming the robustness of the selected antibodies.

Conclusions: Elucidation of the antibody characteristics which confer protection to placental malaria will inform vaccine development and other prevention and control strategies for pregnant women living in malaria-endemic regions.

References:

1. *Zou, H. and Hastie, T. (2005), Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: SeriesB (Statistical Methodology), 67:301-320.*

[RP1.25]

Evaluation of Sample Size Requirements for Developing a Risk Prediction Model for Clustered Data

Chen Qu, Gareth Ambler, Menelaos Pavlou, Rumana Omar

Department of Statistical Science, University College London, United Kingdom

Background: Risk prediction models are increasingly being used in clinical practice. It is important that the dataset used to develop a risk model is of an appropriate size. Typically in multicentre studies, patients within a centre tend to be more correlated than patients from different centres (clustered data). The sample size calculation for these studies is either based on the rule of 10 (10 events per predictor, EPV) or the EPV based sample size calculation is inflated using a variance inflation factor to account for clustering. Regression models that accounts for clustering (e.g. marginal models with robust standard errors or random effects models) are typically used to develop risk models. Previous studies investigating sample size requirements for clustered data suggest that an EPV of at least 10 is sufficient provided that the random parameter is included in the EPV calculation. The effects of intracluster correlation (ICC) values on sample size was investigated with focus on 'average predictions' where the random intercept is ignored for the purposes of prediction.

Objective: We extend previous work to investigate the effects of prognostic strength of the model and the clinical objective (for example whether predictions will be made for the same clusters or for new clusters) on sample size for binary outcomes. We use random intercept logistic regression models and consider cluster specific, marginal and average predictions. We use a wider range of performance measures than examined in previous studies, to investigate sample size requirements to achieve a 90% probability that the risk model will have acceptable performance, defined on the basis of a range of values specified for the C-statistic, calibration in the large, calibration slope and the accuracy of individual predictions (AIP) in the validation data. The minimum cluster size required for each cluster in the development data are investigated.

Findings: Prognostic strength and clinical objective (marginal or cluster specific predictions) affect sample size requirements. Models with high prognostic strength require smaller sample sizes. For ICC values >0.1, models presenting marginal predictions require a larger sample size compared to models presenting cluster specific predictions.

[RP1.26]

Growth Mixture Modelling of Anxiety and Depression Symptoms from Electronic Records of Psychological Therapy

Megan Skelton[1], Ewan Carr[1], Alicia J. Peel[1], Janet Wingrove[2], Kimberley A. Goldsmith[1], Gerome Breen[1], Thalia C. Eley[1]

[1] Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom
[2] South London and Maudsley NHS Foundation Trust, United Kingdom

Background: The substantial heterogeneity in patient symptoms following psychological therapy for anxiety and depression is well-established. However, there may be additional inter-individual differences that occur throughout therapy itself. Identification of latent classes of symptom trajectories and associated patient characteristics could help to guide therapist and patient expectations and their decisions about the suitability and continuation of therapy.

Objectives:

1. To use latent growth mixture modelling within electronic health records of psychological therapy to determine whether there is heterogeneity in intra-individual anxiety and depression symptom change.
2. To compare the one-step and three-step estimation methods of identifying latent classes and predictors of class membership.
3. To estimate the growth curve function for individuals with missing data due to prematurely ending treatment, and assess how these missing data influence the overall model.

Methods: The sample consisted of de-identified health records from 17,411 individuals who had received at least three sessions of talking therapy for anxiety or depression symptoms. Linear, quadratic, log-linear and piecewise growth curve models were fitted to measures of depression (Patient Health Questionnaire-9) and anxiety (Generalised Anxiety Disorder-7) recorded at each session. Following model comparisons to identify the optimum growth curve function, classes were incrementally added to determine the best-fitting number of subgroups. Patient and treatment characteristics, as well as treatment outcomes, were added as covariates using both the one-step and three-step estimation methods. The three-step approach identifies trajectories independent of choice of covariates. Additional growth curve models were fitted to the subset of patients who were recorded as 'drop-outs'.

Results: Model comparisons show that symptom change is best modelled using a quadratic growth curve. Statistically significant variance in the growth factors indicate that multiple classes are a better model of the data.

Conclusions: Growth mixture models can identify subgroups of trajectories of depression and anxiety symptoms. Understanding the predictors of these distinct classes could inform clinical decision-making but identified classes depend on the choice of one- or three-step estimation.

[RP1.27]

On the hazard rate of death due to HIV/AIDS with missing covariates and lost to follow up as competing risk

Vinay Gupta[1], Gurprit Grover[2]

[1] *Department of Biostatistics, JSS Medical Research India Pvt Ltd, India*
[2] *Department of Statistics, Delhi University, India*

Background: Estimation of hazard rate when the survival time is interval censored and data has missing covariates along with loss to follow ups by using Multiple Imputation is not available in literature.

Objective: The objective is to estimate hazard rate of death due to HIV/ AIDS considering lost to follow ups as competing risk by using Multiple Imputation by Chained Equations (MICE) technique to impute exact survival time along with missing covariate values in the presence of competing risk.

Methodology: Records of 2052 HIV/AIDS patients on Antiretroviral Therapy at the ART centre of RML Hospital, New Delhi, India during April 2004 to December 2010 were collected retrospectively through their ART routine registers. Around 10% patients had missing values in weight and alcohol consumption. Multiple imputation was done at two levels. First, the exact survival times were imputed within the time interval and then the missing covariate values were imputed through MICE technique using sub-distribution hazard model considering lost to follow up as competing risk. Ten imputed data sets were created and each was analyzed with sub-distribution hazard model. Finally, the results from the analysis of all imputed data were combined. For comparison, two more imputation methods were applied; 1. Mid-point imputation for interval censored time and 2. Exclusion of missing covariates with imputing interval censored time through sub-distribution hazard model.

Results: Out of 2052 patients, the survival time due to HIV/AIDS was observed for 273 (13.3%) patients. At the end of study 81.8% of the patients were alive and 4.9% were lost to follow up. Other information recorded at baseline were age, sex, CD4 cell count, date of visit, mode of transmission, weight, hemoglobin, area of residence, marital status, smoking, alcohol consumption and opportunistic infection. Sex, CD4 cell count, weight and hemoglobin were found to be significantly associated with the death due to HIV/AIDS.

Conclusions: The proposed MICE technique for interval censored data using sub-distribution hazard model in the presence of lost to follow up as competing risk provided better results in terms of lower standard errors of estimates. It is also confirmed through graphical methods and simulation analysis.

[RP1.28]

Application of Cronbach's alpha in scientific studies on diagnosing and treating of neurological problems

Michal M. Skoczylas

Department of Diagnostic Imaging and Interventional Radiology, Pomeranian Medical University in Szczecin, Poland

Background: Cronbach's alpha classically used to assess the reliability of a psychometric test is also useful in many medical problems, especially for measurement of various substantive areas within a single construct.

Objectives and method: Basing on the review of articles in Medline/ PubMed from the year 1975 to 2017 general characteristics of the use of this coefficient in scientific studies in neurology was prepared.

Results:

Cronbach's alpha was used to:

- construct a short scale with high internal reliability to measure morning or evening disposition in human circadian activity (Torsvall and Akerstedt, 1980),
- characterize therapy expectations in chronically ill patients suffering from neurological and mental disorders (Schneider et al., 1990),
- study the carer satisfaction with healthcare services for stroke patients (Pound et al., 1993),
- describe the accuracy of the Pediatric Motor Activity Log-Revised in neurorehabilitation of impaired upper extremity in children (Uswatte et al., 2012),
- validate Stroke and aphasia quality of life-39 (SAQOL-39) into Kanarese language, shortly Kannada SAQOL-39 (Kiran and Krishnan, 2013),
- assess the validity of the International Cooperative Ataxia Rating Scale in children and adolescents with genetically confirmed phosphomannomutase deficiency (Serrano et al., 2015),
- validate the Cluster Headache Quality of life scale (CHQ) (Bakar et al., 2016),
- prove the validity of the Unified Multiple System Atrophy Rating Scale (UMSARS) in describing symptoms of MSA (Matsushima et al., 2017),
- and in two studies – to study neurological soft signs in psychiatric patients (Vitiello et al., 1988; Schröder et al., 1992).

The majority of ten articles in the set above is formed with papers with high value of Cronbach's alpha ($\alpha \geq 0.7$), which may result from the difficulties in publishing statistically insignificant results of scientific research.

Conclusion: Cronbach's alpha is widely used in scientific studies on neurological problems – in a range including clinical, imaging and molecular diagnostics together with physiotherapeutic and psychological issues – giving suitable information about common and rare diseases and general health state of patients and opinions of their carers.

[RP1.29]

Merging and analysing data from national registries – use of antibiotics (defined daily doses)

Milada Cvancarova Smastuen[1], Gunn Signe Jakobsen[2], Pavla Pecherkova[3], Asgeir Johannessen[1], Rune Sandbu[1], Jens K. Hertel[1], Jøran Hjelmesæth[1]

[1] *Public Health, Oslo Metropolitan University, Norway*
[2] *Morbid Obesity Centre, Vestfold Hospital Trust, Norway*
[3] *Faculty of Transportation Sciences, Czech Technical University, Norway*

Background: Large registry based data are rich sources of information and are extensively used to provide information in medical research. However, analyzing such data requires a high level of expertise both in biostatistics and data science. Data is often stored in various formats and importing such data can be a challenge.

Objectives: We used data from the national Norwegian Prescription Database (NPD) and a hospital patient registry to compare the use of antibiotics, measured as defined daily doses (DDD) in patients with severe obesity undergoing surgical (ST) or medical treatment (MT), and with the general population. DDDs provide a fixed unit of measurement enabling the assessment of trends in drug consumption and comparisons between population groups.

Methods: Data from patients referred to a tertiary care center from 2005-10 were linked to follow-up data retrieved from the NPD from 2005-15. All prescriptions collected by Norwegian citizens are registered and the coverage is almost complete. Each patient had up to 45 entries each year in the NPD. Average DDDs were computed for each patient and each year of follow-up and two years prior to inclusion. Use of antibiotics in the ST and MT groups was compared using Generalized mixed models for repeated measures using the Gaussian identity link. The results were depicted graphically using box-plots for each year and treatment group. The raw data were filtered using the Tukey outlier filter to improve the readability of the figures. As data on use of antibiotics in the general Norwegian population were available only on group level, we constructed age-, gender- and time-period weighted rates using open data from NPD to compare the use of antibiotics in patients with the general population.

Results: A total of 1995 consecutively included patients were analyzed. Our data did not reveal any differences in average yearly DDDs between the groups, however the consumption of antibiotics in patients was significantly higher compared with the general population.

Conclusion: Our analyses revealed large differences in use of antibiotics between the patients and the general population. However, to correctly interpret these results, a close collaboration between medical professionals and statisticians is warranted.

[RP1.30]

Parent-child discrepancies in the assessment of adolescent's emotional and behavioral problems

Agnieszka Pac, Renata Majewska, Natalia Nidecka, Agata Sowa

Chair of Epidemiology and Preventive Medicine, Jagiellonian University Medical College, Poland

Background: Adolescence is a stage of life characterized by physical, psychological and social vulnerability, distancing from the family of origin and greater approximation with peers, and can be considered a period of risk for the development of emotional and behavioral problems. Both, externalizing and internalizing problems, affect the psychic development of adolescents, besides being predictors of problems in adult life.

Objective: The aim of this study was to assess the discrepancies between adolescent and parent in the assessment of emotional and behavioral problems of people aged 15-17 years.

Methods: The study sample consisted of 105 mother-child pairs from Krakow birth cohort study. The data for 56 girls and 49 boys were analyzed. Two instruments from The Achenbach System of Empirically Based Assessment (ASEBA) were used in this study: Child Behavior Checklist for Ages 6-18 (CBCL/6-18) for parents and Youth Self Report (YSR). Truncated t-scores were (in respect to age and sex) were used for the comparability of self- and maternal reports.

Results: We have found than adolescents reported significantly more symptoms related to externalizing problems than their mothers (mean t-scores 52.3 (SD: 9.5) and 49.3 (9.6), respectively). Young people reported higher scores on four symptom scales: anxious/depressed (56.9(7.5) vs. 55.0(6.9)), attention problems (56.7 vs 53.8(5.2)), rule-breaking behaviors (55.4(5.8) vs 53.8(5.2)) and aggressive behaviors (55.1(6.5) vs 53.7(5.8)) while lower scores in somatic complaints (54.9 (5.2) vs 56.9 (7.1)). The highest agreement between maternal and adolescent's report, as measured by kappa coefficient (for clinical/borderline symptom presence) ranged from 0.00 for Attention Problems up to 0.497 for Aggressive behaviors.

Conclusions: The use of maternal report for assessment of adolescents emotional and behavioral problems may lead to underreporting of problems in this age group.

[RP1.31]

The use of logistic regression to identify risk factors in a clinical study

Calin Avram[1], Laura Avram[2], Marius Marusteri[1]

[1] *Department of Medical informatics and biostatistics, George Emil Palade University of Medicine, Pharmacy, Science and Technology of Targu Mures, Romania*
[2] *Universitatea „Dimitrie Cantemir" Targu Mures, Romania*

Background: Logistic regression calculation makes it possible to describe the relationship between the values of two or more random variables. In clinical research for decision-making the odds ratio (OR) it gives clear and direct information to clinicians about which treatment approach has the best odds of benefiting the patient. In logistic regression, the dependent variable is binary (dichotomous).

Objective: The aim of this paper is to demonstrate methods of correctly identifying the dependent variables which can be in relation to the independent variables, using the logistic regression method. This statistical approach also enables the calculation of the specific indicator odds ratio (OR) for clinical studies, allowing for a quick and correct interpretation of the results of the study.

Methods: The OR evaluates whether the odds of a certain event or outcome is the same for two groups of patients. Choosing the dependent variable correctly is in close connection to the method of defining the variable. If the variable is binary (dichotomous) from the start (e.g. positive/negative in a clinical test or yes/no for a questionnaire), logistic regression can be applied. For continuous numerical variables, a cut-off value is defined (e.g. in the case of blood glucose, a value under 110 mg/dL represents 'normal glycemia' and a value above 110 mg/dL represents 'hyperglycemia') in order to transform it into a binary variable.

Results: Odds ratio is that it is simple to calculate, very easy to interpret, and provides results upon which clinical decisions can be made. Sometimes it is useful, in clinical situations, to be able to provide information to patients on the results obtained using the odd ratio. The paper also presents practical examples based on previous work (articles published in journals).

Conclusions: Applying logistic regression as a method of calculation in statistical analysis allows for an accurate identification of the association between dependent and independent variables of the clinical study, making it possible to interpret the results in a quick and easy way.

[RP1.32]

Continuous Glucose Monitoring record length and minimum daily observations for clinical interpretation

Jędrzej Chrzanowski[1], Przemysław Kucharski[2], Arkadiusz Michalak[3], Konrad Pagacz[1], Beata Mianowska[3], Agnieszka Szadkowska[3], Wojciech Fendler[1]

[1] *Department of Biostatistics and Translational Medicine, Medical University of Lodz, Poland*
[2] *Institute of Applied Computer Science, Technical University of Lodz, Poland*
[3] *Department of Pediatrics, Diabetology, Endocrinology and Nephrology, Medical University of Lodz, Poland*

Background: Continuous Glucose Monitoring (CGM) - a technique used in the daily management of diabetes - can measure interstitial glucose concentration in 5 or 15 minutes intervals. The accumulated data can be used to investigate short- and long-term glycemic variability (GV) in patients with diabetes. Unfortunately, CGM records are still prone to data loss due to sensor- and patient-related issues. The effect of quantity and patterns of data losses in CGM records on calculated GV indices is currently unknown and should be investigated.

Objective: To determine the minimal length and quality of CGM records necessary for short and long-term GV assessment.

Methods: CGM records were collected from pediatric type 1 diabetes patients from 2015 to 2019. Data were processed using Python with pandas, scipy, fbprophet. Calculated GV indices included: mean, median, standard deviation (SD), coefficient of variation (CV), time below, in and above range (TBR, TIR, TAR).

Lengths for short and long-term GV assessments were found using 180-days CGM records with at least >80% daily measurements. Lengths were defined as the minimum for which GV was similar for local (short-term) and global (long-term) GVs.

The minimum required number of daily observations was investigated using records with at least 90% daily observations. CGM measurements were removed using the algorithm reconstructing patterns of data loss from real CGM traces. Minimum daily observations were found for GV values relative difference hold within 5%.

Results: We collected data from 451 patients, the median length of CGM records was 202 (98-368) days, altogether 331.96 years of data.

Using 21 records, length for short and long-term GV assessment was determined as 7 and 35 days respectively. The most robust GV for data loss were mean and CV, and the least robust were TBR <54mg/dL and TAR >250mg/dL. Minimum number of daily observations for which GVs were within 5% from the source record were 70% for 7-days and 30% for 35-days intervals.

Conclusions: Short and long-term GV could be investigated using shortened CGM records. It is possible to assess GV using records with >70% and >30% daily observations for 7 or 35-day recordings respectively, without a significant loss of information.

[RP1.33]

Population-Based Cancer Survival in the Czech Republic

Ivana Svobodova[1], Stanislav Katina[2]

[1] *Institute of Health Information and Statistics of the Czech Republic, Czech Republic*
[2] *Masaryk University / Institute of Mathematics and Statistics, Czech Republic*

Background: The Czech National Cancer Registry is a nationwide population-based cancer registry founded in 1976 and maintained by the Institute of Health Information and Statistics of the Czech Republic. In the Czech Republic it is mandatory to report all cases of oncological diseases to this registry for the purpose of surveillance and research, so the registry completely covers Czech population. Population-based cancer survival is an important part of assessing the effectiveness of cancer care.

Objectives: The main objective is to calculate 5-year population-based cancer survival (net survival using Pohar-Perme estimator [1]) in the Czech Republic and assess trends in time.

Methods: Data of patients diagnosed during 2013–2017 (period approach) were compared with data of patients diagnosed during 2008–2012 and 2003–2007 (cohort approach). Estimates were age-standardized with the International Cancer Survival Standard weights. Only adult patients with solid tumours were considered (age 15 to 99 years) in this analysis. Patients registered based on death certificates only or patients diagnosed at autopsy were excluded.

Results: The highest values of 5-year age standardized net survival during 2013–2017 were observed among patients diagnosed with the malignant neoplasm (MN) of thyroid gland (94.2 %), MN of testis (88.7 %) and MN of prostate (88.0 %). On the contrary, the lowest values were observed among patients diagnosed with MN of pancreas (8.0 %), MN of liver (8.3 %) and MN of oesophagus (10.2 %). The highest increase in the survival since the time period 2003–2007 was observed among patients diagnosed with MN of colon and rectum (+8.3 %), MN of prostate (+7.9 %) and MN of kidney (+7.0 %).

Conclusions: We identified the diagnoses with good and bad prognosis. These results can be used as key policy tool that can be used to evaluate the impact of cancer prevention strategies, the effectiveness of cancer care and health systems for all patients diagnosed with cancer.

References:

1. *Perme, M. P., Stare, J., & Esteve, J. (2012). On estimation in relative survival. Biometrics, 68(1), 113–120.*

[RP1.34]

swdpwr: A SAS Macro and An R Package for Power Calculation in Stepped Wedge Cluster Randomized Trials

Jiachen Chen, Xin Zhou, Fan Li, Donna Spiegelman

Department of Biostatistics, Yale School of Public Health, United States

Background: In stepped wedge cluster randomized trials (CRTs), all clusters start from the control condition and cross over to receive intervention in a unidirectional and randomly assigned order until all clusters are exposed to the intervention. Stepped wedge designs (SWDs) may be preferred for estimating intervention effects when it is logistically more convenient to roll-out intervention in a staggered fashion and when stakeholders or participating clusters perceive the intervention to be beneficial to the target population.

Objective: Binary outcomes are frequently seen in CRTs as endpoints. However, existing methods for sample size calculation of SWDs have been almost exclusively focused on continuous outcomes. Systematic reviews indicated that the majority of SWDs with binary outcomes used approximation methods based on continuous approaches, which may either overestimate or underestimate the power in different scenarios. To improve this approximation, Zhou et al.(2020) and Li et al.(2018) recently proposed two new methods for power calculations of SWDs with binary outcomes.

Methods: Zhou et al.(2020) developed a maximum likelihood method in SWDs with binary outcomes based on conditional models (GLMM). Li et al.(2018) proposed a marginal method for continuous and binary outcomes within the framework of GEE that employed a block exchangeable within-cluster correlation structure. These new methods have not been implemented in publicly available software such as SAS and R. Additionally, existing software for SWDs focused on limited settings and did not have accurate power calculations for binary outcomes. Hence, to make the new methods more accessible and incorporate as various settings as possible, we have developed user-friendly and computationally efficient software based on the methods proposed by Zhou et al.(2020) and Li et al.(2018) to implement power calculations for SWDs with binary as well as continuous outcomes.

Results: The core of the software is developed in Fortran and is built into a SAS macro swdpwr and an R package swdpwr, accommodating scenarios of cross-sectional and cohort settings, binary and continuous outcomes, marginal and conditional methods, different link functions, with and without time effects, etc.

Conclusions: The development of this software addresses the implementation gap between newly proposed methodology and applying them to obtain more accurate power calculations in SWDs.

[RP1.35]

Longitudinal complex survey methods: weighting for missing data

Janet L. MacNeil Vroomen[1], Heather G. Allore[2]

[1] *Amsterdam Unviversity Medical Center, Netherlands*
[2] *Yale School of Medicine, United States*

Background: Survey weights to be nationally-representative were developed for the National Health and Aging Trends Study, which is comprised of adults 65 years and older. Due to losses to follow-up in this older cohort, modification of the survey weights to account for death and missing data was required. Options included adjusting weights based on response to outcome across two time points or to adjust weights based on response to outcome through cumulative wave/yearly adjustments.

Objective: To determine the most appropriate weighting approach.

Methods: Weights calculated based on missing data on the 5th year of outcome data and for the cumulative missing data. Histograms of the predicted probabilities were evaluated for best fit.

Results and Conclusions: The cumulative weights were bimodal and skewed, whereas the weight constructed from two time points appeared normally distributed. Researchers should evaluate if the original weights when applying longitudinal modifications, maintain their representative nature of the target population.

[RP1.36]

Apply Joint model to life course epidemiological data -birth cohort in Japan-

Ayano Takeuchi[1], Yukihiro Ohya[2], Masami Narita[2]

[1] *Department of Preventive Medicine and Public Health, Keio University, Japan*
[2] *Department of Allergy and Clinical Immunology, National Center for Child Health and Development, Japan*

Birth cohort studies have been launched around the world in the last 20 years. Their target is to detect the effect of risk factors (from lifestyle factors to environmental chemicals) at fetal life to onset of diseases after grown up (ex. Neuropsychiatric developmental disorders, allergies, cancers, Metabolic disease etc.). In traditional cohort study, we consider exposures and confounders status only at baseline (starting time of follow-up for cohort study) but now we have to consider their life course exposure (changing exposure status in long-term life stage), because we are required to assess the environmental factors that influence on the specific time window of growth and development or multiple effects, rather than the strong single risk factors such as smoking. So it is not sufficient to regard baseline risk factors and covariates. Our study aim is to draw DAG (Directed Acyclic Graph) and apply joint models to evaluate direct and indirect effects of these risk factors.

Our birth cohort study has been conducted in Japan since 2003. Study subject is 1703 children. We focus on the relationship between start time of baby food and onset of allergy at 5 yrs. Covariates and intermediate factors are history of parent's allergy, frequency of cleaning and washing, holding pets…etc. in 1-4yrs. Comparing with crude analysis (include all factors into regression model simultaneously), estimated direct effect and indirect effects is small, and total effect is larger than crude results.

[RP1.37]

Use of Aromatase Inhibitors and Risk of Cardiovascular Disease: A Population-Based Study

Lauren McVicker, Christopher Cardwell, Úna McMenamin

Centre for Public Health, Queens University Belfast, United Kingdom

Background: Adjuvant endocrine therapy, including aromatase inhibitors (AI) or tamoxifen is recommended for women diagnosed with oestrogen receptor and/or progesterone receptor positive breast cancer as it significantly reduces the risk of recurrence and breast cancer-specific mortality. AIs are the preferred treatment for postmenopausal women, however concerns have been raised with respect to their cardiovascular safety. Some meta-analyses of randomised controlled trials report an increase in cardiovascular disease (CVD) risk with use of AIs compared to tamoxifen however 'real-world' observational studies have yielded inconsistent findings.

Objectives: We aimed to conduct a population-based study in Scotland to investigate whether breast cancer patients treated with AIs are at an increased risk of cardiovascular disease compared to patients treated with tamoxifen.

Methods: Women newly diagnosed with breast cancer, from 2009 to 2017, were identified from the Scottish Cancer Registry. AI and tamoxifen use were identified from the nationwide Scottish Prescribing Information System. The primary outcome was CVD, based on incidence or death from CVD obtained from hospital admission and death records, respectively. Cox regression models calculated hazard ratios (HRs) and 95% intervals (CIs) comparing AI use to tamoxifen use and risk of CVD, using an as treated exposure definition. Analyses were adjusted for a range of potential confounders and sub-group analysis investigated risk by CVD subtypes including myocardial infarction, heart failure, ischaemic stroke, venous thromboembolism and angina.

Results: The cohort contained 28,712 breast cancer patients, followed for 110,999 person years, during which there were 776 cardiovascular events. Compared to tamoxifen use, AI use was associated with an increased risk of any CVD event (adjusted HR 1.14, 95% CI: 1.02, 1.27) which was particularly marked for ischaemic stroke (adjusted HR 1.27, 95% CI: 0.99, 1.63). There were no clear associations for other CVD subtypes.

Conclusions: Our preliminary results suggest an increased risk of CVD, and specifically ischaemic stroke, in AI users compared to tamoxifen. Additional analyses will be conducted to further investigate these associations, including dose-response.

[RP1.38]

Multi-stage evaluation of mediation and effect modification in a Gambian birth cohort

Raymond Nhapi[1], Maia Lesosky[1], Brenda Kwambana-Adams[2], Martin Antonio[3]

[1] Division of Epidemiology & Biostatistics, University of Cape Town, South Africa
[2] Division of Infection and Immunity, University College London, United Kingdom
[3] Medical Research Council Unit The Gambia, London School of Hygiene and Tropical Medicine, Gambia

Background: Recent literature points out the role of air pollution as a major contributor to the risk of pneumonia acquisition especially in infants. Amongst these pollutants, biomass smoke and tobacco smoke exposures tend to be highly prevalent especially in low- to middle- income countries (LMICs). This study aims to investigate the associations between smoke exposure and infant pneumococcal carriage specifically focussing on the role of inflammation in this 'potentially causal' pathway.

Cohort and Data description: A cohort of 120 mother-infant pairs was recruited at birth in Gambia between March 2013 and September 2015. They were followed up monthly in the first 12 months then quarterly in the following year. Medical data (vaccinations, antibiotics etc.) were collected from infant health cards and prescriptions, nasopharyngeal swabs, blood and serum samples were collected at a number of visits and anthropometric and demographic data was collected by the field team at each visit. Pneumococcal carriage was measured by qPCR focussing on the detection of the lyta gene whilst inflammation was measured by using the Alpha-1 glycoprotein (agp) and C-reactive protein (crp).

Analysis Methods: Variable selection is done by making use of a combination of variable importance statistics from random forests, penalized generalized estimation equations (PGEE) and theory from a literature search. Directed acyclic graphs (DAGs) are then used to obtain the minimal confounding sets. Sensitivity analyses were implemented over a number of generalized linear mixed models (GLMMs) in order to assess the strengths of the hypothesized associations.

Results: Exposure to biomass smoke was significantly associated with a nearly 3-fold increase in the odds of pneumococcal carriage (OR 2.9, 95% CI: 1.13 - 7.5) and, in independent models, a 1/3-log10 increase in pneumococcal load (Coefficient 0.35, 95% CI: 0.11 - 0.59), compared to non-exposure. Inflammation (AGP) was significantly associated with an increased pneumococcal load (Coefficient 0.22, 95% CI: 0.03 - 0.41) in a model unadjusted for smoke exposure. Mediation analysis suggests there are age, inflammation and smoke exposure interactions that may modify the effects of smoke exposure on pneumococcal carriage.

Conclusions: Biomass smoke exposure may be an important environmental factor driving pneumococcal carriage and loads among PCV-vaccinated Gambian children.

[RP1.39]

Prediagnostic Circulating Sex Hormones and Risk of Gastrointestinal Cancers in the UK Biobank

Una McMenamin[1], Peipei Liu[1], Andrew Kunzmann[1], Michael Cook[2], Helen Coleman[1], Brian Johnston[3], Marie Cantwell[4], Chris Cardwell[1]

[1] Centre for Public Health, Queen's University Belfast, United Kingdom
[2] Division of Cancer Epidemiology and Genetics, National Cancer Institute, United States
[3] Department of Gastroenterology, Belfast Health & Social Care Trust, United Kingdom
[4] Institute for Global Food Security, Queen's University Belfast, United Kingdom

Background & Aims: The incidence of gastrointestinal cancers show an unexplained male predominance, suggesting sex hormone involvement but few prospective studies have investigated sex hormones and gastrointestinal cancer risk. We aimed to determine the impact of circulating sex hormones on risk of esophageal, gastric and colorectal cancer in men and women.

Methods: We included 219,425 men and 169,112 postmenopausal women from the UK Biobank. Sex hormone concentrations were quantified using chemiluminescent immunoassay. Gastrointestinal cancers were identified through linkage to cancer registries. Sex hormone concentrations and risk of gastrointestinal cancers were investigated using Cox proportional hazards regression.

Results: During 10 years of follow-up, 388 esophageal adenocarcinoma, 120 esophageal squamous cell carcinoma, 342 gastric and 3,023 colorectal cancer cases were identified. Increased hazard ratios (HRs) were found for sex hormone-binding globulin (SHBG) and risk of gastric cancer in men (Q4 v. Q1 HR 1.43, 95% CI 0.95, 2.17, P_{trend}=0.01). Free testosterone was inversely associated with esophageal squamous cell carcinoma in women (Q4 v. Q1 HR 0.33, 95% CI 0.12, 0.88, P_{trend}=0.05). For colorectal cancer, SHBG was associated with a reduced risk among men (Q4 v. Q1 HR 0.89, 95% CI 0.77, 1.03, P_{trend}=0.04) and free testosterone concentrations a reduction in risk among women (Q4 v. Q1 HR 0.83, 95% CI 0.69, 1.00, P_{trend}=0.03). No associations were found for esophageal adenocarcinoma.

Conclusion: In this large prospective investigation of prediagnostic sex hormones and risk of gastrointestinal cancers, men with higher SHBG concentrations had higher gastric, and yet lower colorectal, cancer risks while women with higher free testosterone levels had lower risk of esophageal squamous cell carcinoma and colorectal cancer.

[RP1.40]

Estimation of effect regardless of treatment discontinuation in trials with non-compliance and missing data

Lorenzo Guizzaro[1], Frank Petavy[1], Robin Ristl[2], Ciro Gallo[3]

[1] *European Medicines Agency, Netherlands*
[2] *Departmet of Medical Statistics, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Austria*
[3] *Statistica Medica, Universita' della Campania Luigi Vanvitelli, Italy*

Background: In Clinical Trials not all patients comply with the course of treatment they are assigned to. The potential impact of such deviations has been one of the reasons for a redefinition of the target of estimation (Estimand) in the ICH E9 Addendum. The effect of treatment assignment, regardless of compliance, appears to be an Estimand of practical interest. This is in line with the intention-to-treat principle and it reflects the occurrence of treatment discontinuation in clinical practice, including for the treatment of Major Depressive Disorder. The estimation of such effect would be straightforward in presence of complete post-discontinuation follow-up. However, a more realistic scenario needs to account for missing data.

Objective: This study aims at evaluating the performance of different estimation techniques in trials with incomplete post-discontinuation follow-up when a treatment-policy strategy is selected.

Methods: We have (i) modelled and visualised using directed acyclic diagrams reasonable data-generating models for placebo-controlled randomised trials in Major Depressive Disorders, where patients can discontinue the assigned treatment and have the option to continue follow-up after that; (ii) investigated which set of variables allows for identification and estimation of an Estimand that uses a treatment-policy strategy; (iii) simulated 10,000 trials in Major Depressive Disorder, with varying real treatment effects, proportions of patients discontinuing the treatment, and proportion of post-discontinuation follow-up data; (iv) tested the performance of different methods (imputation-, weighting-, and likelihood-based) in different configurations.

Results: Our results suggest that all estimation methods studied have their performance increased when a variable representing compliance is used to account for post-discontinuation missingness.

Conclusions: The Estimand framework can inform the handling of missing data, including the choice of variables for the Missing at Random (MAR) assumption. Further research on sensitivity analyses that do not rely on MAR given any set of variables is needed.

[RP1.41]

Application of an SEIRD model to track the spread of COVID-19 in Nigeria

Maya Mueller[1], Ashok Krishnamurthy[2], Bedřich Sousedík[3], Agatha Ojimelukwe[4], Jocelyn Boegelsack[2]

[1] *Department of Mathematics, University of Maryland, Baltimore County, United States*
[2] *Department of Mathematics & Computing, Mount Royal University, Canada*
[3] *Department of Mathematics & Statistics, University of Maryland, Baltimore County, United States*
[4] *Department of Biological Systems and Organisms, University of Port-Harcourt, Nigeria*

Background: The global coronavirus pandemic of 2019 (COVID-19) reached Lagos, Nigeria on February 27, 2020. Since then, the Nigeria Center for Disease Control (NCDC) is reporting a steady increase in the number of confirmed cases. Reliable information regarding the nature of the spread is imperative for Nigerian policy makers to make educated decisions on safety measures moving forward.

Objectives: We aim to use mathematical modeling of infectious diseases in order to support the NCDC and Nigerian government towards informed, data-driven decision making.

Methods: We present a spatial SEIRD (Susceptible, Exposed, Infectious, Recovered and Dead compartments) epidemic model to capture the transmission dynamics of the spread COVID-19 in Nigeria. Using the data from the Center for International Earth Science Information Network as a Gridded Population of the World (GPW) map, we assess the geographical spread of Nigeria's population. We represent the population as a gridded map. Each grid cell has a population count, which is divided into disease compartments. Each cell can transmit disease to its neighbors, with probabilities that decline exponentially with Euclidean distance.

Results: We use the spatial epidemic model to estimate and project the number of newly infected and death cases up to October 1, 2020. We present spatio-temporal disease maps for the infectious variable for the progress of COVID-19 in Nigeria.

Conclusions: Predicting the transmission dynamics of COVID-19 in Nigeria comes with a lot of uncertainty. First, we run the stochastic spatial simulations under the worst-case scenario, in which there are absolutely no governmental or public health interventions. Next, we account for mitigation efforts including lockdown and social distancing. Predictions for disease prevalence with and without mitigation efforts are presented via time-series graphs for the epidemic states.

[RP1.42]

Variable selection in mediation analysis

Awa Diop[1], Geneviève Lefebvre[2], Caroline S. Duchaine[3], Danielle Laurin[4], Denis Talbot[3]

[1] *Department of Mathematics and Statistics, Laval University, Canada*
[2] *Department of Mathematics, Universite du Quebec a Montreal, Canada*
[3] *Department of Social and Preventive medicine, Laval University, Canada*
[4] *Department of Pharmacy, Laval University, Canada*

Background: It is now well established that adjusting for pure predictors of the outcome, in addition to confounders, allows unbiased estimation of the total exposure effect on an outcome with generally reduced standard errors. However, no analogous results have been derived for mediation analysis.

Objectives: Determine the impact of adjusting for pure predictors of the exposure, mediator and outcome on the variance of popular natural direct and indirect effect estimators used by practitioners.

Methods: We considered both a conventional regression and a propensity-score perspectives. Theoretical results were first derived by considering the simplest linear regression framework, then simulation studies were used to extend the results.

Results: We obtained theoretical results showing that, in regression models, adjusting for pure predictors of the outcome, in addition to confounders, allows unbiased estimation of the natural indirect effect (NIE) and the natural direct effect (NDE) with reduced standard errors. Adjusting for pure predictors of the mediator increases the standard error of the NDE, but may either increase or decrease the variance of the NIE. Adjusting for pure predictors of the exposure increases the variance of both the NIE and the NDE estimators. Simulation studies were used to confirm and extend these results to the case where the mediator or the outcome is binary. Both a regression approach and an inverse probability weighting approach were considered in the simulation study. A real-data analysis employing data from the Canadian Study of Health and Aging is provided for illustration. This analysis is concerned with the mediating effect of vitamin D in the effect of physical activity on dementia and its results are overall consistent with theoretical and empirical findings.

Conclusions: It is desirable to adjust for pure predictors of the outcome, in addition to confounders, and to avoid adjusting for pure predictors of the exposure to reduce the standard errors of mediation analysis estimators. Our results are an important step to better understand variable selection in mediation analyses and will prove helpful to guide practitioners performing such analyses.

[RP1.43]

Clinical trials impacted by the COVID-19 pandemic: Adaptive designs to the rescue?

Cornelia Kunz[1], Silke Jörgens[2], Frank Bretz[3], Nigel Stallard[4], Kelly Van Lancker[4], Dong Xi[6], Sarah Zohar[7], Christoph Gerlinger[8], Tim Friede[9]

[1] *Boehringer Ingelheim Pharma GmbH & Co. KG, Germany*
[2] *Janssen-Cilag GmbH, Germany*
[3] *Novartis Pharma AG, Switzerland*
[4] *University of Warwick, United Kingdom*
[5] *Ghent University, Belgium*
[6] *Novartis Pharmaceuticals, United States*
[7] *Inserm, France*
[8] *Bayer AG, Germany*
[9] *University Medical Center Göttingen, Germany*

Background: Very recently the new pathogen severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was identified and the coronavirus disease 2019 (COVID-19) declared a pandemic by the World Health Organization. The pandemic has a number of consequences for the ongoing clinical trials in non-COVID-19 conditions.

Objectives: Motivated by four currently ongoing clinical trials in a variety of disease areas we illustrate the challenges faced by the pandemic and sketch out possible solutions including adaptive designs.

Methods: Guidance is provided on (i) where blinded adaptations can help; (ii) how to achieve type I error rate control, if required; (iii) how to deal with potential treatment effect heterogeneity; (iv) how to utilize early readouts; and (v) how to utilize Bayesian techniques. In more detail approaches to resizing a trial affected by the pandemic are developed including considerations to stop a trial early, the use of group-sequential designs or sample size adjustment.

Results and Conclusions: All methods considered are implemented in a freely available R shiny app (https://power-implications.shinyapps.io/prod/). Furthermore, regulatory and operational issues including the role of data monitoring committees are discussed (https://arxiv.org/abs/2005.13979).

[RP1.44]

A practical guide to causal discovery for cohort data

Ryan M. Andrews[1], Janine Witte[1], Ronja Foraita[1], Robin Denz[2], Vanessa Didelez[1]

[1] Department of Biometry and Data Management, Leibniz Institute for Prevention Research and Epidemiology - BIPS, Germany
[2] University of Duisburg/Essen, Germany

Background: Causal discovery can be an important tool for exploring cause-effect relationships, particularly in observational cohort data. There are many available software tools for causal discovery; however, there is currently little information on how to apply and tailor them to observational data, which frequently have missing data, repeated/time-ordered longitudinal measurements, or a mixture of continuous, categorical, and binary variables in the data ("mixed scales"). Therefore, there is a need to understand the relative merits and implementation of available causal discovery software in the face of these challenges.

Objective: We aimed to give a practical guide to the use and comparison of the features of the R packages pcalg and bnlearn and the Java application TETRAD in the presence of longitudinal/time-ordered measurements, and mixed or missing data.

Methods: We illustrated the use of causal discovery on an artificial data set that mimics the three waves of the pan-European I.Family/IDEFICS children's cohort study. We generated these data from an assumed "true" causal graph between 33 study variables roughly matching the observed summary statistics within each I.Family/IDEFICS study wave. Importantly, our artificial data also mimicked the cohort structure and the missing data patterns of the study, and we deliberately chose variables of interest on mixed scales. These data were then analyzed with pcalg, bnlearn, and TETRAD to illustrate how each software package can address these challenges and in what sense they recover important aspects of the "true" causal graph. We also assessed other aspects of the software, like the flexibility and user-friendliness of the software.

Results: We found notable similarities and differences between the three software packages considered. For example, only bnlearn had prebuilt options to handle missing data aside from complete case analysis, and TETRAD was the only software that allowed for an interactive GUI. All implementations struggled with mixed data and repeated measures, with pcalg being the least flexible.

Conclusions: Causal discovery software implementations are widely available. With minor to moderate modifications, they can be applied to typical observational cohort data. We found they have the potential to suggest/uncover new causal relations that may be interesting to explore in further studies.

[RP1.45]

Using multiple indicator cluster surveys to determine pneumonia burden in Lao People's Democratic Republic

Cattram Nguyen, Fiona Russell

MCRI, Australia

Background: For many low-resource settings, the burden of pneumonia is unknown, or modelled creating a national average. However, there can be substantial variability in risk factors for pneumonia within a country, leading to variable disease burden at sub-national level and within subpopulations. Given that pneumonia surveillance and prospective studies are not always feasible in many settings, we explore the use of Multiple Indicator Cluster Surveys (MICS) data to determine the incidence of pneumonia at national and sub-national levels in Lao People's Democratic Republic.

Objective: The aim of this project is to estimate pneumonia burden at national and sub-national levels using publicly-available survey data.

Methods: National representative data from MICS (2000, 2006, 2011, 2017), of individually collected data on health indicators, demographics and vaccination status, were used. We quantified and mapped pneumonia risk factors by province. We will estimate pneumonia incidence where both MICS data and prospective pneumonia data are available to validate the proposed methods. Hierarchical models incorporating survey weights will be used to estimate pneumonia incidence as a function of risk factors, obtaining stable province-level estimates by borrowing strength from larger provinces with similar characteristics.

Results: Interim results show variability in risk factors across 17 provinces of Laos. For example, in 2011 the percentage of children with malnutrition (weight-for-age Z score <2 SD) ranged between 16% and 46%, and the proportion of households with 7 or more members ranged between 31% and 61%, demonstrating substantial variability in known risk factors for pneumonia.

Conclusions: This project will generate local data that are relevant to decision-making, and will establish methods that can be used across other low-resource settings.

[RP1.46]

Comparison of national COVID-19 death rates after age adjustment with a weighting method

Lan Kelly[1], Greer Humphrey[1], Caroline Miller[2], Jacqueline Bowden[3], Joanne Dono[2], Paddy Phillips[4]

[1] Health Translation SA, South Australian Health and Medical Research Institute, Australia
[2] South Australian Health and Medical Research Institute, Australia
[3] School of Public Health, University of Adelaide, Australia
[4] SA Commission on Excellence and Innovation in Health, Australia

Background: Deaths due to COVID-19 are highly age dependent, with risk increasing with age. Comparisons of death rates between nations require age adjustment to be meaningful.

Objective: To age adjust COVID-19 death rates to the population profile of Australia and group countries into clusters with similar age profiles.

Methods: A weighting method was used for age adjustment which assumed that age-stratified death rates per population relative to the 80+ age group were the same as China for all nations. For every nation, relative rates were multiplied by the fraction of the population in each age group and summed to produce an overall relative rate. To age adjust other regions to Australia, the crude rate for each region was weighted by the ratio of the overall age adjusted relative rate for Australia compared with the region. Weights were assumed stable over time. Countries were clustered into groups using Gaussian finite mixture modelling on the weights (R package 'mclust').

Results: Countries with older or younger population profiles than Australia were weighted downwards and upwards, respectively. Weights ranged from 0.59 for Japan to 6.39 for Uganda. The best model fit produced three clusters, with mean weights 0.89, 2.05 and 4.40. The group with the lowest weights (0.59-1.13) included developed nations such as Australia, Japan, USA and European countries. The group with the highest weights (3.11-6.39) included many African, Middle Eastern and Pacific Islander nations.

Conclusion: Nations with younger age profiles than Australia are at lower risk of death from COVID-19 due to age alone. However, death rates depend on multiple factors and many of these nations have less developed health systems, limited access to ventilators and may have high population densities which aids the spread of the disease.

[RP1.47]

Implementing ECG Features into a Clinical Prediction Model to Assess the Individualized Risk in Arrhythmogenic cardiomyopathy

Laura Hofstee[1], Rob Roudijk[2], Folkert Asselbergs[2], Thomas Debray[1], Hae-Won Uh[3]

[1] Department of Epidemiology, University Medical Center Utrecht, Netherlands
[2] Department of Cardiology, University Medical Center Utrecht, Netherlands
[3] Department of Biostatistics and Research Support, University Medical Center Utrecht, Netherlands

Background: Arrhythmogenic cardiomyopathy is characterized by a predisposition for life-threatening ventricular arrhythmias (i.e. abnormal heartbeats) and heart failure. Sudden cardiac death can be the first symptom of disease in patients with arrhythmogenic cardiomyopathy, which can be prevented by implantable cardioverter-defibrillators. Selecting patients who will benefit from such a device during the early stages of the disease is of utmost importance.

Objective: This study aims to improve the currently used clinical prediction model to predict the patients' individualized risk of developing an arrhythmogenic cardiomyopathy related event by implementing electrocardiogram features into the model.

Methods: Two baseline models – clinical prediction models with and without an electrocardiogram variable – and a feature model combining clinical data with 1288 electrocardiograms features (i.e. 59 PQRST, 8 heart rate variability, 80 fast Fourier transformed and 14 continuous wavelet transformed features for all eight electrocardiogram signals) were built employing logistic regression using stepwise backward selection and ridge regression respectively. The outcome of interest was whether 'an event' occurred during follow-up. Events were defined as the composite of 'any arrhythmic event', 'death', or both. The models were internally validated through bootstrapping.

Results: The data consisted of 108 patients of which 60 patients had no event and 48 patients had an event. The feature model was shown to outperform the final baseline model, the optimism corrected R^2 [95% CI] = .553 [0.39, 0.71], C-statistic [95% CI] = .897 [0.84, 0.95], calibration slope [95% CI] = 0.76 [0.40, 1.13] and calibration intercept [95% CI] = 0.02 [-0.78, 0.71], and R^2 [95% CI] = .733 [0.50, 0.88], C-statistic [95% CI] = .949 [0.90, 0.99], calibration slope [95% CI] = 1.29 [0.89, 1.82] and calibration intercept [95% CI] = 0.25 [-1.20, 1.73] for the baseline and feature model respectively.

Conclusions: Implementation of electrocardiogram features in the clinical prediction model improves the accuracy of the prediction of ventricular arrhythmias and sudden cardiac death in patients with arrhythmogenic cardiomyopathy.

[RP1.48]

Prediction equations for blood concentration markers in the HCHS/SOL Nutrition and Physical Activity Study

Lillian Boe[1], Yasmin Mossavar-Rahmani[2], Daniela Sotres-Alvarez[3], Martha Daviglus[4], Ramon Durazo-Arvizu[4], Robert Kaplan[2], Pamela Shaw[1]

[1] *University of Pennsylvania, United States*
[2] *Albert Einstein College of Medicine, United States*
[3] *University of North Carolina at Chapel Hill, United States*
[4] *University of Illinois College of Medicine, United States*

Introduction: Measurement error is a major issue in self-reported diet and this error, observed to have subject-specific and random components, can distort diet-disease relationships. Serum biomarkers present the opportunity to measure diet objectively, but their potential requires further study. These biomarkers are typically only practical to obtain in a subset of participants in a large cohort study. We developed prediction equations for the serum biomarker level for several nutritional intakes assessed in the Study of Latinos: Nutrition and Physical Activity Assessment Study (SOLNAS) and examined whether the predictive accuracy would be high enough to reliably detect an underlying diet-disease association in the larger Hispanic Community Health Study/Study of Latinos (HCHS/SOL) cohort, for which the majority would have a predicted exposure.

Methods: In SOLNAS, blood concentration biomarkers and participant characteristics were collected at baseline for 447 participants from the multi-center HCHS/SOL cohort (n = 16,415), namely for alpha-carotene, beta-carotene, alpha-tocopherol, gamma-tocopherol, vitamin B-12, beta-cryptoxanthin, retinyl palmitate, folate, lycopene, retinol, and zeaxanthin. We build regression-based prediction equations for these 11 biomarkers based on participant characteristics, including self-reported 24-hour dietary recalls. We used simulations to study the power of detecting the association between a true average concentration marker and a hypothetical incident diabetes survival outcome using the predicted biomarker level.

Results: Based on preliminary regressions, we observe R^2 values ranging from 0.5028 to 0.1013. Correlations between the concentration biomarker and self-reported measures ranged from 0.4100 to 0.0481, the highest being for beta- cryptoxanthin and lowest for vitamin B-12. When using predicted intake, power reduced from 90% for the true underlying biomarker level to 67% when the calibration-model R^2=0.50 and to 29% with the calibration-model R^2=0.22.

Discussion: For several concentration biomarkers, there were important predictive associations, but the accuracy was not high enough for predicted intakes to provide reasonable power in the larger cohort study. For these biomarkers to fullfil their promise as measures for dietary intake for outcome-diet association studies they would need to be measured in more individuals or predictive accuracy improved.

[RP1.49]

Multi-level modeling of early COVID-19 epidemic dynamics in France to estimate lockdown impact on infection

Mélanie Prague[1], Linda Wittkop[2], Quentin Clairon[3], Dan Dutartre[4], Rodolphe Thiébaut[5], Boris Hejblum[3]

[1] *Inria SISTM / Université de Bordeaux / Inserm U1219 / VRI, France*
[2] *Université de Bordeaux / Inserm U1219 / Inria SISTM / VRI / CHU Bordeaux, France*
[3] *Université de Bordeaux / Inserm U1219 / Inria SISTM / VRI, France*
[4] *Inria Bordeaux Sud-Ouest, France*
[5] *Université de Bordeaux / Inserm U1219 / Inria SISTM / VRI, France*

We propose a multi-level approach to model the beginning of the French COVID-19 epidemic at the regional level. We rely on an extended Susceptible-Exposed-Infectious-Recovered (SEIR) mechanistic model, a simplified representation of the average epidemic process. Combining several French public datasets on the early dynamics of the epidemic, we estimate region-specific key parameters conditionally on this mechanistic model through Stochastic Approximation Expectation Maximization (SAEM) optimization using Monolix software. We thus estimate basic reproductive numbers by region before isolation (between 2.4 and 3.1), the percentage of infected people over time (between 2.0 and 5.9% as of May 11th, 2020) and the impact of nationwide lockdown on the infection rate (decreasing the transmission rate by 72% toward a Re ranging from 0.7 to 0.9). We conclude that a lifting of the lockdown should be accompanied by further interventions to avoid an epidemic rebound.

[RP1.50]

Antidepressants and the risk of death in northern Poland: registry-based cohort study

Paweł Zagożdżon

Department of Hygiene and Epidemiology, Medical University of Gdansk, Poland

Antidepressants (ADs) are commonly prescribed medications, but their long-term health effects are not fully understood. We conducted a cohort study assessing the effects of ADs on all-cause mortality in users of antidepressant medications.

Methods: Antidepressants prescriptions from National Health Fund database between 2008 and 2018 were analyzed. Mortality rates and Cox proportional hazard model were utilized to examine overall mortality and survival in different antidepressive drugs. There were 3 223 150 refunded prescriptions for 312 080 patients who initiated treatment with one of 23 antidepressants in analysed period.

Results: In total 43,447 individuals died during follow-up corresponding to a rate of 21.63 per 1,000 person years. After adjustment for age and sex the highest risk of death was found with the use of amitriptyline. Other ADs were characterised by similar mortality risk.

Conclusion: This study suggests that ADs used in our study population have a similar safety profile with regard to the risk of death, except amitriptyline. Further research is needed to investigate the risk of death for individual ADs in specific subgroups such as patients with diagnosed depression, anxiety or with concomitant cancer or cardiovascular disease.

[RP1.51]

Modeling of effectiveness of potential antibacterial and anti-inflammatory drugs

Renata Paprocka[1], Małgorzata Ćwiklińska-Jurkowska[2], Bożena Modzelewska-Banachiewicz[1]

[1] Faculty of Pharmacy, Department of Organic Chemistry, Nicolaus Copernicus University in Toruń, Poland
[2] Faculty of Pharmacy, Department of Biostatistic and Biomedical System Theory, Nicolaus Copernicus University in Toruń, Poland

Background: Searching for new therapeutic substances still consists actual ask and challenge for many scientists. Despite huge advance in medicine, chemistry and biochemistry still there are known diseases without effective therapy. Seeking for new drugs is long-standing and expensive process, while majority of promising substances is discarded for not fulfilling strict demands of several clinical phases studies. Useful tool in those processes might be statistical modeling.

Objective: Modeling biological properties of potential antibacterial and anti-inflammatory substances (amidrazone derivatives) by generalized linear models (GLM).

Methods: Among 79 compounds with determined biological activities we analyzed physicochemical properties like molecular weight, percent composition (oxygen, hydrogen, nitrogen and carbon), lipophilicity (LogP), number of acceptors and donors of hydrogen, medium melting point. Potential biological effects (eg. antibacterial, antituberculostatic, anti-inflammatory, antitumor, antieczematic) of tested compounds were calculated with PASS (Prediction for Activity Spectra for Substances) program. Above mentioned variables are potential explaining variables in the model.

The following variables are explained variables. Minimal inhibitory concentrations (MIC) were experimentally evaluated for *E. coli, Y. enterocolitica, P. aeruginosa* (all G- bacterial strains) and *S. aureus, M. smegmatis, Y. enterocolitica, E. faecalis, S. lutea* (all G+). Anti-inflammatory properties were tested by ability to suppress the production of TNF-a in mitogen-stimulated PBMC (peripheral blood mononuclear cells). Another important parameter was antiproliferative activity also tested in mitogen-stimulated PBMC culture.

Seventy nine examined compounds come from three groups depending on chemical structure (linear, 1,2,4-triazole derivatives and pyrrole-2,5-dione derivatives). Individual compounds differ in the substituents R1 and R2. These are qualitative variables in the models.

Results: Statistically significant models were obtained, and also high percentages of variance represented by the model, according to the R-squared value. For some continuous variables, squares or interactions with the type of R1 and R2 substituents are statistically significant.

Conclusions: Generalized linear models for predicting the activity of chemical compounds from 3 groups: (linear, 1,2,4-triazole derivatives and pyrrole-2,5-dione derivatives) on the basis of selected theoretical variables, lipophilicity and the type of R1 and R2 substituents have good predictive properties for the compounds as potential drugs.

[RP1.52]

A life-course model to estimate associations between measures of alcohol exposure and colorectal cancer risk

Ana-Lucia Mayen, Pietro Ferrari on behalf of EPIC PIs and scientists

Department of Nutritional Methodology and Biostatistics, International Agency for Research on Cancer, France

Background: Most prospective studies measure exposures at one time-point, usually at study recruitment to later assess associations with disease outcomes. Growing evidence on body fatness and cancer indicates measuring exposure changes during the life-course is relevant to capture key elements of cancer etiology.

Objective: To examine the relationship between alcohol consumption and colorectal cancer using data on alcohol intake during lifetime (ages 20, 30 and 40 years), at baseline and follow-up (7.3 years after baseline) within the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort.

Methods: A total of 184,956 participants (33% women), aged 35 to 65 years at baseline were included. Lifetime alcohol intake assessments were obtained by lifestyle questionnaires, baseline intake through Dietary questionnaires, and follow-up assessment by a mixture of both. Incident CRC cases were identified by linkage with national cancer registries, health insurance records or active follow-up. Sex-specific Cox proportional hazard models with five-year alcohol intake change between baseline and follow-up were modelled. Latent class growth analyses (LCGA) were used to estimate the association of lifetime, baseline and follow-up intake assessments with colorectal cancer risk.

Results: During 1,602,328 person years, 1,759 incident CRC cases (865 men and 894 women) were identified. Median intake in men during lifetime (1.7 drinks/day) was higher than at baseline (1.3 drinks/day) or follow-up (1.2 drinks/day) while median intakes remained stable in women (0.3 drinks/day). After modelling five-year change in alcohol intake with splines, a higher intake at follow-up tended to show an increased CRC risk in both sexes while a decreased intake at follow-up tended to show a decreased risk in women. In categorical analyses, men increasing intake from none to 1 drink at baseline to ≥2 drinks at follow-up showed an increased risk (HR 1.42, 95%CI, 1.06-1.90). Associations were not significantly associated with CRC for participants who lowered their intake between baseline and follow-up. LCGA analyses tended to show a reduced risk of CRC using a stable high intake trajectory as reference.

Conclusions: Alcohol change was positively associated with CRC risk in men that consumed ≥2 drinks/day at follow-up compared to baseline, with respect to men with stable intakes.

[RP1.53]

Seeking possible cross-allergens to C4MGH0 from common mugwort

Katarzyna Jagielnicka[1], Jan Jurkowski[2], Michał Jagielnicki[1], Małgorzata Ćwiklińska-Jurkowska[3]

[1] *Medical Department, Gedanensis Medical University, Poland*
[2] *Department of Mathematics, Informatics and Mechanics, University of Warsaw, Poland*
[3] *Department of Pharmacy, Nicolaus Copernicus University, Poland*

Background: Cross-reactivity is an important clinical issue. Lately, an opportunity to use bioinformatics tools and clustering to examinations in this subject has arisen.

Objective: The goal of this study was examining the similarities between breathable allergen from common mugwort (Artemisa Vulgaris Art v3) with different allergens, in particular food allergens.

Methods: Clustering of allergens based on Levenstien distance is presented. We specifically focused our efforts on analyzing allergens based on aminoacid similarities. Sequences of typical allergens, especially food allergens were compared. We have described with bioinformatic methodology, similarity between the pollen allergen Art v3.0201 and other potential allergens.

Our search was performed by combining BLAST and FASTA algorithms to achieve optimal results by first isolating allergens having the most similarity in aminoacid sequencing search and then by performing detailed query on these proteins. The dendrogram of similarities between potential across-allergens is construed on Levenstien distance. Table of percent identity matrix of non-specific lipid-transfer protein C4MGH0 of Artemisia vulgaris Art v 3 with food allergens and aero-allergens is studied. The dendrogram construed on Levenstien distance displays similarities between allergens.

Results: Allergenic proteins in the sequence obtained for clustering show possibility of cross-allergy. The sensitization to common mugwort (Artemisia Vulgaris) pollen could be considered as a marker for the related sensitization to vegetable and fruits allergens, in particular as celery, white mustard, lentil, peanut, bean, maize, common hazel, Armenian plum, European pear, orange, black mulberry and red raspberry and garden strawberry. Some results are concordant with cross-reactions on allergen discovered earlier clinically. Additionally, sensitization to Old World Sycamore and rubber tree can be examined. C4MGG9, similar in 94% to C4MGH0 indicate additional possible cross-reactivity with other edible fruits.

Conclusions: Similarity between aminoacid sequences between Artemisia Vuglaris pollen and other allergens is useful to consider possible cross-reactivity. This work may be a contribution to the analysis in observational studies for potential cross-reactivity allergens found.

The use of molecular-based diagnosis might deepen the knowledge of clinically relevant IgE sensitization to cross-reactive allergen ingredients from pollen allergen sources, fruits and vegetables. The results might be be examined or possibly confirmed by clinical and immunology methodology.

[RP1.54]

Evaluating the Establishment-based Risk Assessment Model for Salmonella spp. at Canadian Hatcheries (ERA – H)

Ornella Wafo Noubissie[1], Manon Racicot[2], Alexandre Leroux[2]

[1] *Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada*
[2] *Canadian Food Inspection Agency (CFIA), Canada*

Background: To modernize its regulatory and inspection structures, the Canadian Food Inspection Agency (CFIA) developed quantitative Establishment-based Risk Assessment algorithms (ERA). These algorithms model the burden of the 17 enteric and zoonotic pathogens accounting for 99.8% of the Canadian burden of foodborne diseases. Interdisciplinary and systematic, the algorithms integrate public health, epidemiology, supply chain and statistics. They revolutionize the CFIA's inspection and oversight structure as inspection frequency and rigour would be contingent on the determined risk category. The ERA-Hatchery (ERA-H) algorithm distributes the total Disability Adjusted Life Years (DALYs) of non-typhoidal Salmonella spp. across 83 of 95 federally-regulated Canadian hatcheries. The DALYs are adjusted using 96 risk criteria selected through literature review and expert elicitation using a specially-formed Scientific Advisory Committee (SAC) They are categorized into inherent, mitigation, and compliance. Four types of sensitivity analyses were conducted to formally assess risk criterion pertinence and impact on risk.

Methods: Four sensitivity analyses were applied, evaluating the effects of individual and grouped criterion exclusion, risk criterion adjustment exclusion and risk slope on the risk metrics. The assessed risk criteria data was collected in 2018. The ERA-H risk predictions and all sensitivity analyses were conducted in Lumina Analytica.

Results: 57.83% and 86.7% of evaluated hatcheries were low-risk (Category 4) and low or moderate risk (Category 3 and 4), respectively. Compliance and mitigation criteria respectively demonstrated the highest and lowest impacts on risk metrics. High-risk hatcheries (Category 1) were most responsive to risk criterion exclusion, while low-risk hatcheries (Category 4) experienced minimal changes. The opposite trend was observed in criterion adjustment analyses. Analyses of risk slope produced no changes in risk category.

Conclusions: The risk of non-typhoidal salmonella spp. varies across hatcheries. Therefore, oversight should be tailored to estimated risks. Adherence to compliance criteria and the effective implementation of mitigation factors require special attention.

RP2: Development of New Methods

[RP2.01]

Probabilistic data aggregation in healthcare data in order to solve content differences

Alexia Sampri, Nophar Geifman, Philip Couch, Niels Peek

Division of Informatics, Imaging and Data Sciences, University of Manchester, United Kingdom

Background: Putting data together from different sources into a homogeneous data resource would enable unprecedented opportunities to study human health. However, these disparate collections of data are inevitably heterogeneous and have made aggregation a difficult challenge. In this paper, we focus on the issue of content heterogeneity in data integration. Traditional approaches for resolving content heterogeneity map all source datasets to a common data model that includes only shared data items, and thus omit all items that vary between datasets.

Objectives: Our focus is on integration of structured data. We assume that each one of these datasets that needed to be integrated consists of a single table; and that each of these datasets describes a disjoint set of entities. Therefore, record linkage is not needed.

Methods: We propose the development of improved, probabilistic approaches for data integration, capable of advancing the timely utilisation of large-scale biomedical data resources. Our approaches aim to forego the need for perfect data standardisation by employing a probabilistic post-alignment of data items that is integrated with statistical inference. Using these approaches, missing or semantically ambiguous information is estimated from datasets potentially relevant for answering the research question. The main challenge is to quantify the additional uncertainty in statistical estimates that is due to incomplete harmonisation of the underlying datasets.

Results: The MAximizing Sle ThERapeutic PotentiaL by Application of Novel and Stratified approaches programme (MASTERPLANS) aims to improve care for Systemic Lupus Erythematosus patients by taking a precision medicine approach to identifying groups of patients that respond to particular biologic therapies. Based on dataset examples provided by MASTERPLANS we describe and evaluate the proposed probabilistic data integration approaches.

Conclusions: Our approaches insist on the future existence of health data heterogeneity. They strive for post alignment of Big datasets. As a post-alignment of heterogeneous data sources will be always imperfect and it is not a problem if we estimate the probability that they are. Our approaches are also pragmatic because they always provide an answer. However, the results of the probabilistic data integration would be the same as those that would result from analysing an integrated dataset.

[RP2.02]

Adjustments to network meta-analysis in the small number of studies by Bartlett-type corrections

Masahiro Kojima

Department of Statistical Science, Graduate University for Advanced Studies, Japan

Recently, when the number of studies in network meta-analysis is not enough, there is a problem that confidence intervals cannot hold nominal (1-alpha). As Noma et al point out, there is a case where conducting the subgroup analysis reduces the number of studies within every subgroup. So as to more increase the accuracy in such cases, Noma et alproposed Bartlett-type correction via bootstrap method. Bartlett-type correction can adjust the second-order bias in type I error. However, there seem to be three problematic factors: (1) the confidence intervals cannot express in closed form, one also has to be derived by numerical approach. (2) since they did not identify the mechanism of the bias with confidence intervals, they had to use the computational method such as bootstrap method. Thus, it remains unclear why the bias occurs and how the bias is adjusted. (3) the bootstrap calculation takes too much time to numerically solve ML or REML estimator since m sets of bootstrap samples need to be generated and m ML or REML estimators need to be calculated. Kojima and Kubokawa analytically derive Bartlett-type correction for hypothesis testing in general linear models via analysis approach. However, this research uses the modified Likelihood Ratio (LR) for using the general consistent estimators and the confidence interval of modified LR cannot express explicitly.

We introduce the three novel achievements which formulate the confidence intervals, identify the bias mechanism, and reduce the calculation time to derive Bartlett-type adjustment terms. Specifically, we take the following innovative efforts. We formulate the three explicit confidence intervals of Wald, LR, Score test statistic using ML and REML estimator to nuisance parameters. This formulation helps to clarify the bias mechanism or derive explicit Bartlett correction terms. We derive the explicit Bartlett-type correction terms via the asymptotic expansion. Here, we make improvements of the results of Kojima and Kubokawa. In other words, we also derive the Bartlett-type correction of LR using REML estimator. We do not have to numerically calculate the Bartlett correction terms and can execute by 1/m speed compared with the method with m bootstrap samples.

[RP2.03]

Rank-ordered logit model for confounder adjustment in the study of continuous outcomes

Afiq Daniel

Department of Statistics and Applied Probability, National University of Singapore, Singapore

Confounder adjustment is an important aspect of epidemiological studies, as the confounding effects by external factors can distort the true association of an exposure variable to the outcome. Experimental designs such as matching or stratification are commonly used to control such confounding effects. However, such designs may be impractical at times, and we have to resort to regression models to adjust for these confounding effects by including the confounders into the model. The rank-ordered logit (ROlogit) model, a logit-based model, was recently proposed to handle confounder adjustment via stratification in the study of continuous outcomes. There are, however, limitations to stratification, making the rank-ordered logit model unsuitable when stratification becomes impractical. All independent variables are important when identifying determinants of an outcome. This method extends the ROlogit model to be used without the need for stratification. Several different simulation studies were conducted for the model in comparison with the normal linear regression model and EVT1 regression model. The results are comparable and shows the feasibility of this new approach.

[RP2.04]

Outliers that don't out-lie: Identifying and eliminating implausible birthweight for gestational age

Konstantinos Pateras[1], Domniki Ladopoulou[2], Andriani Tsagkari[3]

[1] *Department of Biostatistics and Research Support, University Medical Center Utrecht, Netherlands*
[2] *Department of Statistics, Athens University of Economics and Business, Greece*
[3] *Health Center of Vyronas, Athens, Greece*

Background: Gestational age and birthweight are important variables for measuring fetal growth. Deviation from the appropriate birthweight for gestational age has been correlated with increased mortality and morbidity in the perinatal and neonatal period but also with increased incidence of health problems in adult life. The reported gestational age values are largely affected by reporting errors due to the lack of a gold standard for accurate assessment [1]. Previous clinical studies have utilized birthweight as a more valid variable than gestational age. However, erroneously reported implausible birthweight for gestational age values often lead to misclassification of neonates. These outliers greatly impact the birthweight by gestational age distributions. The resulting wrong assignment of neonates as at risk for adverse health outcomes (or not), heavily impacts clinical practice [1-2].

Objectives: Existing statistical approaches are solely unable to properly identify implausible birthweight for gestational age [2], therefore, a more flexible approach needs to be identified.

Methods: To this end, we suggest a modified expectation-maximization algorithm to identify and correct implausible values and/or misclassified neonates. A cohort of all singleton female liveborn neonates in Greece between 2011 and 2017 was used as a case study to compare the birthweight distributions at each gestational age based on the proposed and previously published approaches [2].

Results and Conclusions: In the simulation study that was performed, the suggested algorithm outperformed previously published approaches by producing results closer to the true mean birthweights and standard deviations for most gestational ages.

*DL and AT contributed equally.

References:

1. *Norris, T., Seaton, S. E., Manktelow, et al. (2018). Updated birth weight centiles for England and Wales. ADC Fetal & Neonatal, 103(6), 577-582.*
2. *Platt, R. W., Abrahamowicz, M., Kramer, M. S., and Joseph, K. S. (2001). Detecting and eliminating erroneous gestational ages : a normal mixture model. Statistics in Medicine,20(23), 3491-3503.*

[RP2.05]

Allocation-Adaptive Randomization Methods – Biased Coin Probability and Unequal Allocation Rates

Petra Ofner-Kopeinig, Maximilian Errath, Andrea Berghold

Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria

Unequal allocation rates are receiving more attention in clinical trials. Recent work [1] gives a good overview about the use of unequal randomization in clinical trials. They are motivated by various reasons: participant preference, ethical reasons, trial feasibility considerations or trial efficacy, and also because of costs or logistical reasons.

Several randomization designs already have been expanded to allow for unequal allocation rates, e.g. the mass weighted urn design by Zhao [2].

We focus on the biased coin design [3]. Here the underrepresented group gets a higher allocation probability, e.g. $p = 2/3$. We show how this randomization method can be applied to study designs with unequal allocation ratios.

In a simulation study using the simulation tool of the web-based randomization service "Randomizer for clinical trials" (www.randomizer.at) we explore different study designs and describe the allocation probabilities and achieved allocation rates. Furthermore, we compare simulation results to theoretical results.

References:

1. *Peckham, E., Brabyn, S., Cook, L., Devlin, T., Dumville, J. and Torgerson, D.J., 2015. The use of unequal randomisation in clinical trials - An update. Contemporary Clinical Trials; 10th Anniversary Special Issue, 45, Part A, pp. 113-122.*
2. *Zhao, W., 2015. Mass weighted urn design--A new randomization algorithm for unequal allocations. Contemporary clinical trials, 43, pp. 209-216.*
3. *Efron, B., 1971. Forcing a Sequential Experiment to be Balanced. Biometrika, 58(3), pp. 403-417.*

[RP2.06]

Using multidimensional scaling techniques to rank interventions taking multiple outcomes into account

Dimitrios Mavridis[1], Adriani Nikolakopoulou[2], Anna Chaimani[3], Philippe Ravaud[3]

[1] *Department of Education, University of Ioannina, Greece*
[2] *Institute of Social and Preventive Medicine, University of Bern, Switzerland*
[3] *INSERM, Paris Descartes University, France*

Background: Ranking competing interventions is one of the most appealing features in network meta-analysis. There is a plethora of ranking metrics and methods but these are typically prone to misinterpretation. Additionally, most ranking metrics focus on one outcome and ranking across the various outcomes is synthesized narratively.

Objective: To present graphical and quantitative ways to rank and group competing interventions taking multiple outcomes into account. To present methods taken from market research to weigh the various outcomes according to stakeholders' preferences.

Methods: We used individual differences multidimensional scaling (IDMDS) to visualize the extent of similarity of interventions and subsequently cluster them into meaningful groups. We used regression methods to weight the various outcomes borrowing methods from market research that are used to weight attributes of products (or in our case outcomes of interventions). Such a method is a conjoint analysis that helps determine how stakeholders view the various outcomes by asking them to score various combinations of performance on all outcomes. We provide R code that make the illustration straightforward.

Results: To illustrate the methods we used a network of 212 randomized controlled trials and 43049 participants comparing 15 antipsychotic drugs and placebo. We considered three outcomes (efficacy, acceptability and weight gain). We present various multidimensional scaling configurations of the antipsychotics using one to three dimensions regarding the number of outcomes considered. Dimensions seem to reflect ranking in outcomes. We present results with all outomes equally weighted and with weights computed from conjoint analysis. When weight gain is not considered or given small weight we see that clozapine, amisulpride, olanzapine, risperidone and paliperidone form a distinct class of drugs. When weight gain is considered it is only amisulpride, risperidone and paliperidone that look promising.

Conclusions: We can use IDMDS to visualise and cluster interventions taking multiple outcomes into account and with different weights attributed to each outcome.

[RP2.07]

One small clinical trial design to provide additional information than single arm trials

Akimitsu Miyake[1], Tomoharu Sato[2], Tomomi Yamada[1], Eisuke Hida[2]

[1] *Department of Medical Innovation, Osaka University Hospital, Japan*
[2] *Graduate School of Medicine, Osaka University, Japan*

Background: The traditional randomized controlled trial designs are difficult to conduct in small populations such as rare diseases and pediatric diseases area. In such small clinical trials, various methodological and statistical considerations have been reported (IOM. Small clinical trials. 2001, CHMP/EWP/83561/05. 2006). However, many single arm trials to assess within-patient comparisons are still performed due to the feasibility. Therefore, the efficacy of a test drug is evaluated based on a pre-specified threshold value. Even if it is a strictly controlled in single arm trial, biases will occur in estimating the effect. It is necessary to prespecify the threshold value as a more appropriate value from the evidence of the natural history or external information. Under these circumstances, it is desirable to be able to estimate the true effect size of the test drug without the influence of bias.

Objective: We propose one trial design that makes level of evidence strengthen than single arm trials without increasing the required sample size.

Methods: The proposed design can assign treatment at a different time points, which is especially the same structure as the delayed start design (D'Agostino RB. N Engl J Med., 2009). Individuals are randomly assigned to either placebo or test drug, and group assigned to placebo at first time period is switched to test drug at subsequent time period (e.g., Group1: placebo - test drug, Group2: test - test drug). This design allows the conventional assessment of treatment effect in the same way as single-arm trials with pre-specified threshold. Moreover, we can estimate the absolute effect size by evaluating the difference between a placebo and the test drug. To evaluate the validity of the proposed design, we performed simulations in various scenarios based on the actual clinical trials data.

Result: As a result, the proposed design has enabled direct comparison of test drug with placebo, while allowing for assessing a pre-specified threshold value without increasing the sample size.

Conclusion: This design seemed to be more useful than single arm trials to practically apply clinical trials of rare diseases and pediatric diseases area.

[RP2.08]

Partitioning the family history relative risk into genetic and environmental effects

Philippe Wagner

Department of Cancer Epidemiology, Lund University, Sweden

In epidemiological studies of putative inherited disease, focus is frequently on risk associated with a positive family history, measured in terms of the relative risk (RR) of having an affected first degree relative; father, mother, sibling and/or child. While valuable information has been extracted using this approach, it is unsatisfactory when interest is primarily in inherited genetic, as opposed to environmental, factors, because it fundamentally confounds the effect of the two. To separate them, the common approach has been to use a so-called ACDE model to partition the outcome variance into components of shared genetic and environmental factors [1]. However, this method may appear complicated to some as partitioning of outcome pertains to latent variables, where outcome is assumed to be the manifestation of a latent trait. The method also makes untestable distributional assumptions with regards to these latent traits.

We show that in the scenario where environmental factors are shared equally between parents and children, a simpler method, using conventional measures of risk, is possible under relaxed assumptions. When effects of genetic and environmental factors work in a multiplicative fashion on risk, the family history RR becomes a product of a family history RR due to environmental, RRc, and one due to genetic factors, RRg.

$$RR = RRg\ RRc$$

This means that we can adjust the family history RR for shared environmental factors by simply dividing the parent-parent outcome association by the parent-child outcome association. By simulation we further show that the resulting measure has a straight forward interpretation. It corresponds to the parent-child association as it would have been had it only been due to genetic factors.

When outcome prevalence is low, the same relationship holds for the odds ratio (OR). Furthermore, an analogous approach is possible when genetic and environmental effects work linearly on risk, by simply subtracting the parent-parent risk difference (RD) from the parent-child RD. We also present regression models for inferences with respect to these adjusted measures of association.

Reference:

1. *Neale & Cardon (1992). Methodology for Genetic Studies of Twins and Families (1st ed. Vol. 67). Netherlands: Springer.*

[RP2.09]

Multilevel Zero-inflated Censored Beta Regression Modeling for Proportions and Rate Data with Extra-zeros

Leili Tapak

Department of Biostatistics, Hamadan University of Medical Science, Islamic Republic of Iran

Background: Zero-inflated proportion or rate data nested in clusters due to the sampling structure can be found in many disciplines. Sometimes, the rate response may not be observed for some study units because of some limitations (false negative) like failure in recording data and the zeros are observed instead of the actual value of the rate/proportions (low incidence).

Objectives: In this study, we proposed a multilevel zero-inflated censored Beta regression model that can address zero-inflation rate data with low incidence.

Methods: We assumed that the random effects are independent and normally distributed. The performance of the proposed approach was evaluated by application on a three level real data set and a simulation study. We applied the proposed model to analyze brucellosis diagnosis rate data and investigate the effects of climatic and geographical position. For comparison, we also applied the standard zero-inflated censored Beta regression model that does not account for correlation.

Results: Results showed the proposed model performed better than zero-inflated censored Beta based on AIC criterion. Height (p-value <0.0001), temperature (p-value <0.0001) and precipitation (p-value $= 0.0006$) significantly affected brucellosis rates. While, precipitation in multilevel zero-inflated censored Beta regression model was not statistically significant (p-value $= 0.385$). Simulation study also showed that the estimations obtained by maximum likelihood approach had reasonable in terms of mean square error.

Conclusions: The results showed that the proposed method can capture the correlations in the real data set and yields accurate parameter estimates.

[RP2.10]

Causal Inference for Extreme Quantiles

Shuo Sun[1], Johanna Neslehova[2], Erica Moodie[1]

[1] *Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Canada*
[2] *Department of Mathematics and Statistics, McGill University, Canada*

Environmental hazards increase the risk of cancer, cardiovascular disease, asthma, and many other illnesses. Analyses of environmental health problems often are concerned with understanding extreme events such as excessive chemical exposures or high concentrations of air pollutants. Yet, the majority of the causal inference literature has focused on modelling means. We define a general estimator of the population quantile treatment (or exposure) effects (QTE) --- weighted QTE (WQTE) --- of which the population QTE is a special case, along with a general class of balancing weights incorporating the propensity score, allowing for the exposure to be binary, discrete or continuous. Asymptotic properties of the WQTE estimators are derived. We further propose and compare three methods of modelling the population QTE based on these balancing weights to understand the causal effect of an exposure on extreme quantiles. Finite sample behavior of the three estimators is studied in simulation.

[RP2.11]

Decomposition of an interaction effect between two independent exposures through a mediator

Helene Colineaux[1], Sebastien Lamy[2], Cyrille Delpierre[2], Michelle Kelly-Irving[2], Benoit Lepage[1]

[1] *EQUITY team & Epidemiology department, INSERM UMR 1027 - Toulouse III University & Toulouse Teaching Hospital, France*
[2] *EQUITY Team, INSERM UMR 1027 - Toulouse III University, France*

Background: When gender differences in health are observed, the question of the mechanisms (biological or social) of these differences may be raised. From this purpose, we have defined the gender effect (i.e. the effect of sex relying on social mechanisms) as an interaction between the sex assigned at birth S and the social environment E.

Objective: The objective was to decompose the interaction effect between these two independent exposures (E*S) on an outcome Y, into a direct effect, an indirect effect (i.e. mediated by an intermediate factor M), an interaction and a mediated interaction effects between E*S and M.

Methods: We propose a generalization of the VanderWeele's 4-way decomposition method to decompose an additive interaction effect between E and S which might affect Y through M, taking into account all possible interactions. The counterfactual writing of all corresponding effects has been specified. Under the identification assumptions, estimation of these effects by G-computation was tested on simulated datasets with binary E, S, M and Y, assuming different data generating mechanisms to simulate various decompositions of the total interaction effect.

Results: The total effect of the Sex*Environment interaction is decomposed into a controlled direct effect, a reference interactional effect, a pure indirect effect and a mediated interactional effect, according to the 4-way decomposition. In our approach to decomposing an interaction effect, the "mediated interactional" component can be interpreted by referring to additional indirect effects with secondary S*E interactions and a "cross indirect" effects. If the identification conditions are met, estimation of the effects is possible by G-computation, and the sum of all the components was equal to the total effect of the S*E interaction on the basis of simulated data.

Conclusion: The decomposition of an interaction effect between two independent factors may allow a finer analysis of the mechanisms of the sex and gender effect in health, but also of the mechanisms of gene-environment interactions. We consider how the approach can be generalized to situations with exposure induced mediator-outcome confounders, as well as the application of this method on real life data.

[RP2.12]

The Kilim plot: a tool for visualizing network meta-analysis results for multiple outcomes

Michael Seo[1], Toshi Furukawa[2], Areti Angeliki Veroniki[3], Toby Pillinger[4], Anneka Tomlinson[5], Georgia Salanti[1], Andrea Cipriani[5], Orestis Efthimiou[1]

[1] *Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland*
[2] *Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, University of Kyoto, Japan*
[3] *Department of Primary Education, University of Ioannina, Greece*
[4] *Institute of Psychiatry, Psychology and Neuroscience, King's College London, United Kingdom*
[5] *Department of Psychiatry, University of Oxford, United Kingdom*

Background: Network meta-analysis (NMA) can be used to compare multiple competing treatments for the same disease. In practice, usually a range of outcomes are of interest. As the number of outcomes increases, summarizing results from multiple NMAs becomes a non-trivial task, especially for larger networks. Moreover, NMAs provide results in terms of relative effect measures that can be difficult to interpret and apply in every-day clinical practice, such as the odds ratios.

Objectives: We aim to facilitate the clinical decision-making process by proposing a new graphical tool, the ‚Kilim plot', for presenting results from NMA on multiple outcomes.

Methods: We introduce a novel method for compactly visualizing results from NMAs on multiple outcomes. Our plot provides information regarding the strength of the evidence regarding treatment effects, while it illustrates absolute, rather than relative, effects of all interventions and all outcomes. Moreover, it can be easily modified to include considerations regarding clinically important effects (e.g. using the minimal clinically importance differences). To showcase our method, we use data from a network of studies in antidepressants. All analyses are performed in R and we provide the source code needed to produce the Kilim plot, as well as an interactive web application.

Results and Conclusions: The Kilim plot can be a valuable aid in summarizing and communicating results from NMAs on multiple outcomes. It can be especially useful for larger networks, for the case of many outcomes, and when aiming to communicate NMA results with patients and/or clinicians, so as to facilitate every-day clinical practice.

[RP2.13]

Comparison of AUC in clinical trials with follow-up censoring: Application to HIV therapeutic vaccines

Marie Alexandre[1], Mélanie Prague[1], Yves Lévy[2], Rodolphe Thiébaut[1]

[1] *Inria SISTM Team, INSERM U1219, University of Bordeaux, ISPED, France*
[2] *Vaccine Research Institute, France*

Background: We are interested in the analysis of longitudinal data from trial in which the area under the curve (AUC) is the primary endpoint. In such study, missing data affects the validity of AUC analysis. Bell et al. (2014) highlighted the interest of using maximum likelihood-based methods to infer AUC by reducing bias. However, their linear mixed effect model is not flexible enough to fit data other than patient reported outcome.

Objective: We aim at developing a broader parametric statistical test to compare the AUC normalized by the follow-up time (nAUC) estimated on right-censored data.

Methods: We developed a two-step method based on a non-linear mixed effects model (NLMEM). Firstly, individual data are fitted with NLMEM where marginal dynamics are described by B-spline curves at group level while polynomial time-dependent random effects are used to model inter-individual variability. Then, the difference of nAUC (DnAUC) between two groups is estimated using the marginal dynamics through trapezoid method. Finally, we propose to statistically test the null hypothesis of no difference of nAUC between groups. This is possible thanks to a modified wald-test computed on a linear combination of regression coefficients obtained by maximum likelihood in the NLMEM.

Results: We applied this method in the context of HIV therapeutic vaccine efficacy assessment through analytic treatment interruption (ATI). In such trial, patients do not take their treatment for a controlled period of time. Their HIV RNA load is monitored and right-censored when it gets too high because of patient treatment resumption. We evaluated the method on simulated data where DnAUC was varied from 0 to 0.5 log10 cp/ml and the level of right-censoring being driven by the cut-off value for HIV RNA load (4 or 5 log10 cp/ml). Moreover, we applied the method on real data from two HIV therapeutic vaccine trials: ANRS 149 LIGHT and 093 Vac-IL2 including respectively 98 and 71 patients randomized to receive either combined vaccine therapies or placebo, before 12 weeks of ATI.

Conclusion: We showed good statistical properties for the proposed test. Moreover, we demonstrated a significant difference in nAUC between treatment and placebo groups in available studies.

[RP2.14]

Adverse effect of diabetes on risk of myocardial infarction by gender and age: A methodological challenge

Grethe Albrektsen[1], Ivar Heuch[2], Tom Wilsgaard[3], Maja-Lisa Løchen[3], Dag Steinar Thelle[4], Inger Njølstad[5], Kaare Harald Bønaa[6]

[1] *Faculty of Medicine and Health Science, NTNU - Norwegian University of Science and Technology, Norway*
[2] *Department of Mathematics, University of Bergen, Norway*
[3] *Department of Community Medicine, UiT - The Artic University of Norway, Norway*
[4] *Department of Biostatistics, University of Oslo, Norway*
[5] *Department of Community Medicine, UiT - Artic University of Health Science, Norway*
[6] *Department of Circulation and Medical Imaging, NTNU - Norwegian University of Science and Technology, Norway*

Background: Diabetes is associated with increased risk of myocardial infarction (MI). The relative risk has been found to be higher in women than men, and the more pronounced adverse effect has even been claimed to eliminate the general female advantage in risk. Few previous studies have explored heterogeneity in risk estimates across age groups. A direct comparison of relative risks across subgroups with different absolute risk is difficult. Analysis of combined categories of interacting factors, using a common reference group, makes this easier. However, a potential problem with residual confounding arise when comparing across broad age groups. Traditional analytic approach cannot be applied.

Objective: Evaluate interaction between gender, age and diabetes in relation to risk of incident MI.

Methods: Incidence rate ratios (IRR) were calculated as estimates of relative risk in Poisson-regression analysis of person-years at risk. Gender- and age-specific results (3 broad age groups; 35-54, 55-74 and 75-94 years) for the association between diabetes and risk of MI, based on data from the Tromsř Study (33, 859 individuals at risk, 1063 with diabetes) are presented. To avoid residual confounding by age when comparing across broad age groups, additional indicator variables (1-year categories, original age variable recoded 1-20 within each age group) were included.

Results: Adjusted for age, gender and established risk factors, diabetes was associated with a doubling in risk of MI (IRR=2.18, 95% CI=1.86-2.55). The adverse effect was slightly more pronounced for women than men (IRR of 2.55 vs. 1.96, p=0.11, test for interaction). Considering combined categories of interacting factors, women with diabetes had a risk level close to men without diabetes, but men with diabetes had a risk about four times as high as women without diabetes. Some heterogeneity across age groups was seen, but risk estimates were imprecise. Results are preliminary. An extended follow-up is planned.

Conclusion: The adverse effect of diabetes on risk of MI was slightly more pronounced for women than men. Among persons with diabetes, women remained at lowest risk. Analysis of combined categories of interacting factors provide useful complementary information, but non-standard analytic strategies for dealing with residual confounding may be needed.

[RP2.15]

Bayesian models for early dose finding in phase I trials with multiple treatment courses

Moreno Ursino[1], Lucie Biard[2], Sylvie Chevret[2]

[1] *Centre de Recherche des Cordeliers, INSERM, Sorbonne Université, USPC, Université de Paris, F-75006 Paris, France*
[2] *ECSTRRA, INSERM U1153 CRESS, France*

Background: Dose-finding clinical trials in oncology of cytotoxic drugs, as well as molecularly targeted agents, aim at determining the maximum tolerated dose (MTD) of the new drug, generally defined by the number of patients with short-term dose-limiting toxicities (DLT). To better identify a safe dose with an acceptable toxicity profile, model-based approaches for such phase I oncology trials have been widely designed, mostly restricted to the DLTs occurring during the first cycle, although patients continue to receive treatment for multiple cycles. Such designs ignore the time to occurrence of DLT. However, the occurrence of DLT could be related to the time elapsed since the treatment onset and/or to the cumulative doses of the drug from the previous cycles of treatment.

Objective: This work aims at (i) estimating the probability of DLTs over the different cycles, extending the Bayesian conditional modelling approach, and (ii) proposing a design for dose-escalation and de-escalation according to these previously observed toxicities.

Methods: The conditional probability of toxicity, given the previous administered doses and in absence of any toxicity during the previous cycle, is modelled at each cycle via a 3-parameter logistic function. The first parameter is related to the toxicity probability of the dose at the first cycle. The second deals with assumed increase in individual tolerance to the treatment, after a previous cycle without toxicity. The last one captures the effect of cumulative dosages over the cycles. A two-stage design is then built, where dose-allocation rules and stopping rules are based on posterior and predictive probabilities. It is applied retrospectively to patients with myelodysplastic syndrome who received 6 cycles of chemotherapy.

Results: We performed an extensive simulation study comparing this approach to the time-to-event continual reassessment method (TITE-CRM). In general, our approach gave better or comparable operating characteristics (notably percentages of correct MTD selection), both when used retrospectively or prospectively.

Conclusions: The main interest of this approach is the possibility of predicting DLTs at further cycles. Therefore, it could be used to adjust and select dose sequences to increase the expected number of cycles linked to a target probability of toxicity.

[RP2.16]

The Power Prior with Multiple Historical Studies for the Linear Regression Model

Akalu Banbeta Tereda

Department of Statistics, Jimma University, Ethiopia

Combining historical control data with current control data may reduce the necessary study size of a clinical trial. However, this only applies when the historical control data are similar enough to the current control data. Several Bayesian approaches for incorporating historical data in a dynamic way have been proposed, such as the meta-analytic-predictive (MAP) prior and the modified power prior (MPP) both for single control as well as for multiple historical data. Here we discuss the generalization of the MPP approach for multiple historical controls with a Gaussian response in the presence of covariates, i.e. for a linear regression model. This approach is useful when the controls differ more than in a random way, but become again (approximately) exchangeable conditional on covariates. The proposed approach builds on the approach developed for binary outcomes (Banbeta et al, 2019). As for a binary outcome, two versions of the MPP approach for a linear regression model with multiple controls have been developed. The first assumes independent powers, while in the second approach the powers have a hierarchical structure. Analytical results will be presented. We conducted a simulation study to investigate the frequentist characteristics of these two versions of the MPP, and compare them with e.g. the MAP approach. Several scenarios were considered with a binary treatment indicator correcting for confounders with fixed or varying regression coefficients across the multiple controls. The effect of neglecting correction of such a covariate is also examined. Additionally, we illustrate our approaches on an Alzheimer data set. In scenarios with between-study variation in the regression coefficients, the MPP approach achieves approximately nominal type I error rates and greater power than the MAP prior, provided that the covariates are included in the model. While the intercepts vary, the MPP yields a slightly inflated type I error rate, whereas the MAP didn't. We conclude that our approach is a worthy competitor to the MAP approach for the linear regression case.

[RP2.17]

Prediction of chromatographic retention using Bayesian hierarchical modeling

Agnieszka Kamedulska, Paweł Wiczling, Łukasz Kubik

Department of Biopharmacy and Pharmacokinetics, Medical University of Gdańsk, Poland

Background: Reversed-phase high-performance liquid chromatography (RP HPLC) is a widely used analytical technique for separating samples in pharmaceutical quality control or life sciences research. Due to its selectivity, sensitivity, throughput and analysis time, this technique has found many applications for routine use in clinical laboratories. Optimization of chromatographic conditions leading to a desired separation is an important step in the development of any analytical method in RP HPLC. A mathematical model that is able to predict retention times and peak widths of analytes can greatly facilitate the search for the desired separation.

Objective: The purpose of this work was to develop a Bayesian hierarchical model based on retention times data collected for 1026 analytes under isocratic conditions at various acetonitrile contents using Agilent Eclipse Plus C18 stationary phase with 3.5 μm particles. The dissociation constant and lipophilicity was considered as a covariates explaining between analyte variability of model parameters.

Methods: The Stan program coupled with R was used for data analysis. These tools allow full Bayesian inference with Monte Carlo sampling by Markov chains.

Results: The developed model accurately described the available chromatographic data. The results were obtained in the form of posterior probability distribution, that quantifies uncertainty about model parameters and predictions. The posterior probability is also directly relevant for decision making. As a result of the analysis, we found that analytes form two clusters with different retention properties. The grouping of analytes was related to their dissociation constant values (degree of dissociation). To describe this phenomenon statistically, a mixed model was used, which assumes two data generation processes, each with its own set of parameters.

Conclusions: The proposed model gives insight into behavior of analytes in the chromatographic column and can be used to make predictions for structurally diverse set of analytes.

The study was supported by the project POWR.03.02.00-00-I035/16-00 co-financed by the European Union through the European Social Fund under the Operational Programme Knowledge Education Development 2014–2020

[RP2.18]

Double robust estimation of partially adaptive treatment strategies

Denis Talbot[1], Erica Moodie[2]

[1] *Le Département de Médecine Sociale et Préventive, Université Laval, Canada*
[2] *Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Canada*

Background: Personalized medicine tailor treatment decisions according to patients' characteristics. G-estimation and dynamic ordinary weighted least squares (dWOLS) are doubly robust statistical methods for personalized medicine. Both require modeling a treatment model as well as an outcome model. The latter of these is further divided in a treatment-free model, relating covariates to the outcome, and a blip model, which models treatment-effect modification. While it is recognized that the blip does not need to include the same covariates as the treatment-free component, it is underappreciated that this can lead to biased results. This would be the case when some confounders are true effect modifiers, but are not included in the blip.

Objective: Propose doubly robust estimators of partially adaptive treatment strategies (PATS), that tailor treatment decision according to a subset of the covariates.

Methods: Building on g-estimation and dWOLS, six estimators of PATS are introduced, including an approach that combines inverse probability weighting with g-estimation first introduced by van der Laan and Robins. Simulation studies are conducted to compare and evaluate their performance.

Results: Biased estimates were obtained when using standard g-estimation and dWOLS for PATS. All estimators we introduced yield unbiased estimates if the treatment model is correctly specified, regardless of whether or not the treatment-free model is correctly specified. When the treatment model is incorrectly specified, the estimators that combine inverse probability weighting with g-estimation or dWOLS yield biased results, while the other estimators remain unbiased. The standard deviations of all estimators were comparable.

Conclusions: Four of the estimators we introduced were effectively doubly-robust for estimating PATS. We have therefore provided several new estimators that allow tailoring to only a subset of covariates. Such partially tailored treatment strategies may be highly practical, especially in settings where some variables are not routinely available in clinical practice or to simplify the clinical application of decision rules.

[RP2.19]

Semi-parametric Bayesian Inference for Optimal Dynamic Treatment Regimes via Marginal Structural Models

Daniel Rodriguez Duque[1], Erica E. M. Moodie[1], David A. Stephens[2]

[1] *Department of Epidemiology, Biostatistics and Occupational Health, McGill, Canada*
[2] *Department of Mathematics and Statistics, McGill, Canada*

Background: Statisticians play a crucial role in using data to identify optimal dynamic treatment regimes (DTRs). Much of the statistical work done on DTRs is in the frequentist paradigm, but Bayesian methods may have a lot to offer in this setting as they allow for the appropriate representation and propagation of uncertainty.

Objectives: We develop robust semi-parametric Bayesian methods for identifying optimal DTRs, both in the singly and doubly robust setting. We apply these methods to an example in HIV care which seeks to tailor antiretroviral therapy to time-varying FIB-4 score, a measure for liver scarring, in order to reduce long-term liver damage.

Methods: We extend recently developed Bayesian methods for Marginal Structural Models (MSMs) to the DTR setting in order to propose Bayesian inference for dynamic regime MSMs. We do this by 1) linking the observational world with a world in which all patients are randomized to a DTR, thereby allowing for causal inference 2) maximizing a posterior predictive utility, where the posterior distribution has been obtained from non-parametric prior assumptions on the observational world data-generating process. Our approach relies on Bayesian semi-parametric inference where inference about a finite-dimensional parameter is made all while working within an infinite-dimensional space of distributions. Double robust inference is obtained by using posterior predictive inference and the non-parametric Bayesian bootstrap.

Results: We provide a precise formulation for these newly developed methods and demonstrate their performance via simulation. As expected, the double robust estimator is most efficient under correct model specification. All Bayesian methods perform as well as their frequentist counterparts, as measured by bias and variance. In the clinical example, we determine that there is no improvement in liver outcomes due to tailoring therapy on FIB-4 score.

Conclusions: We develop a new methodology that allows us to identify optimal DTRs via Bayesian semi-parametric inference and demonstrate its utility in HIV research.

[RP2.20]

Statistical integration of heterogeneous omics datasets using Group Sparse O2PLS (GO2PLS)

Zhujie Gu[1], Said el Bouhaddani[1], Jeanine Houwing-Duistermaat[2], Hae-Won Uh[1]

[1] *Department of Biostatistics and Research Support, Julius Centre, UMC Utrecht, Netherlands*
[2] *Department of Statistics, University of Leeds, United Kingdom*

Background: Integration of multiple correlated omics datasets is an open research topic. Various methods have been proposed for this purpose, such as PLS-related approaches, which decompose datasets into joint and residual parts. Omics data are heterogeneous (e.g. differences in scale, dimensionality, etc.) and the joint parts estimated in PLS contain data-specific variations. O2PLS was proposed to capture the heterogeneity using data specific parts and better estimate the joint parts. However, the latent components spanning the joint subspace in O2PLS are linear combinations of all the variables, hampering interpretation. For better interpretation, feature selection is needed, which can be enhanced by incorporating known group structures. Our motivating datasets are methylation (482,563 CpG sites) and IgG glycomics (22 glycan peaks) data from 392 samples in the TwinsUK study. IgG is an antibody whose functional diversity is mainly achieved by glycosylation. Methylation has an important role in the glycosylation pathways. Group structure is present in methylation data as CpG sites on the same CpG island tend to function together.

Objective: We extend O2PLS to GO2PLS which performs variable selection while incorporating group structures of variables. We apply GO2PLS on the TwinsUK datasets to identify groups of CpG sites that impact IgG glycosylation, and hence have influence on immune response.

Methods: To perform variable selection, L1 penalty is introduced on the loadings. Sparse solutions are obtained by retaining only variables with a large contribution to the covariance. GO2PLS imposes penalties on the sum of the group-wise L2 norms of loadings, which result in group-wise sparsity where variables of the same group are selected or dropped altogether based on the contribution to the covariance as a whole. If all the groups have size 1, the sum of group-wise L2 norms corresponds to L1 norm.

Results: A simulation study shows that GO2PLS performs better than O2PLS in terms of variable selection, joint score and loading estimation. Results of the data analysis show that the groups selected appear to be related to immune functions.

Conclusions: GO2PLS provides a framework to integrate two heterogeneous omics datasets and select relevant groups of variables, thereby facilitating interpretation.

[RP2.21]

Statistical integration of multi-omics data on Multiple System Atrophy

Said el Bouhaddani[1], Hae-Won Uh[1], Jeanine Duistermaat[2]

[1] *Department of Biostatistics and Research Support, Julius Centre, UMC Utrecht, Netherlands*
[2] *Department of Statistics, University of Leeds, United Kingdom*

Background: Multiple System Atrophy (MSA) is a rare neurodegenerative disorder. Almost 80% of patients are disabled within 5 years of disease onset. The key pathogenic event when developing MSA is an abnormal accumulation of harmful proteins. Molecular causes and consequences of this aggregation need to be elucidated, e.g. using multiple omics datasets. We have access to DNA-methylome, miRNome, transcriptome, and proteome data, measured in cell lines that show harmful protein aggregation and in negative controls. Standard sequential analysis of these data shows no overlap of the significant genes. A combined analysis can detect relevant features shared by all datasets, improving the understanding of MSA.

Objective: Our aim is to develop a data integration method to identify consistent molecular biomarkers that can classify cells with protein aggregation across all datasets. Apart from the high dimensionality ($p>N$), also platform-specific heterogeneity between the omics data need to be considered. Several algorithmic approaches to integrate multiple datasets have been proposed, for example, multi-group PLS (mg-PLS) and MINT. They decompose the datasets into joint and residual parts. The joint components capture the consistent effects of the molecular measurements on the outcome across all datasets. The optimal components are obtained by iteratively maximizing the covariance between the molecular measurements and a dummy matrix based on the binary outcome.

Methods: The drawbacks of mg-PLS and MINT are a lack of platform-specific parts in the decomposition, absence of a proper model for the binary outcome, and risk of overfitting when data are high dimensional. Therefore, we propose a novel Probabilistic multi-group OPLS (mg-POPLS) model for multiple datasets in terms of joint, platform-specific, and residual parts. Systematic differences between the omics data are incorporated in the model by including specific parts. The outcome is explicitly modeled via these components. The components and coefficients are estimated with maximum likelihood using an EM algorithm.

Results and conclusions: An extensive simulation study will be conducted to investigate the performance of mg-POPLS compared to mg-PLS. We apply the mg-POPLS method to the omics data measured in the cell lines. This will facilitate detecting the most relevant omics features for separating MSA cases from controls.

[RP2.22]

Spatiotemporal transmission dynamics of Ebola in Congo

Ashok Krishnamurthy[1], Loren Cobb[2], Gursimran Dhaliwal[1], Naseem Hijazi[1]

[1] Department of Mathematics & Computing, Mount Royal University, Canada
[2] Department of Mathematical and Statistical Sciences, University of Colorado Denver, United States

Background: Mathematical modelling of infectious diseases is an interdisciplinary area of increasing interest. We present a spatial variant of an S-I-R (Susceptible-Infectious-Recovered) compartmental model of epidemiology to capture the transmission dynamics and the spatial spread of the ongoing Ebola outbreak in the Democratic Republic of Congo.

Objectives: It is essential to understand what future epidemic trends will be, as well as the effectiveness and potential impact of government disease intervention measures. We aim to provide insight that supports the WHO and the Congo Ministry of Health towards data-driven decision making.

Methods: We present a spatial Susceptible-Vaccinated-Exposed-Infectious-Recovered-Dead (S-V-E-I-R-D) compartmental model of epidemiology to capture the transmission dynamics and the spatial spread of the 2018-2020 Ebola outbreak in two northeastern provinces (North-Kivu and Ituri) in Congo. We use registered data (province-wide weekly counts of total Ebola cases and confirmed dead) up to June 2, 2020 from the World Health Organization (WHO) situation reports.

Results: Our simulations show good correspondences between the stochastic model and the available sparse empirical data. A comparison between weekly incidence data set and our SVEIRD model coupled with Bayesian data assimilation highlights the role of a realization conditioned on all prior data and newly arrived data. In general, the SVEIRD model with data assimilation gives a better fit than the model without data assimilation for the same time period. We present spatio-temporal disease maps for the infectious variable for the progress of Ebola in the North-Kivu and Ituri provinces.

Conclusions: Our analyses may shed light more broadly on how the disease spreads in a large geographical area with places where no empirical data is recorded or observed.

The analysis presented herein can be applied to a large class of compartmental epidemic models. It is important to remember that the model type is not particularly crucial for data assimilation, the Bayesian framework is the key. Data assimilation neither requires nor presupposes that the model of the infectious disease be in the family of S-I-R compartmental models. The projected number of newly infected and death cases up to August 1, 2020 are estimated and presented.

[RP2.23]

A semiparametric joint binomial–Poisson model for the stimulation phase of an IVF cycle

Mark W. Donoghoe[1], Angelique H. Riepsamen[2], Robert B. Gilchrist[2]

[1] Stats Central, Mark Wainwright Analytical Centre, UNSW Sydney, Australia
[2] Fertility and Research Centre, School of Women's and Children's Health, UNSW Sydney, Australia

Background: During the process of in vitro fertilisation (IVF), a prediction of the number of oocytes that will be retrieved following a cycle of ovarian stimulation can be used to guide the intensity of treatment that will be delivered. Statistical modelling of this outcome is complicated by the fact that cycles may be cancelled – stopped without proceeding to oocyte retrieval – if monitoring reveals that too few follicles have developed. The majority of prior work in this area has either categorised the outcome variable, excluded cancelled cycles altogether, or treated them in the same way as a cycle that yielded zero oocytes. Each of these approaches discards information.

Objective: To develop a statistical model for the outcome of an IVF cycle that can take into account information from cancelled cycles, reflecting the underlying process.

Methods: We present a novel approach in which the outcome of the ovarian stimulation is modelled by a joint binomial–Poisson distribution. Specifically, the expected oocyte yield for a completed cycle is modelled as a function of covariates, and the probability of cycle completion is itself a monotonic function of the expected yield. This model can be fitted by defining a new extended family for the mgcv package in R, such that complex semiparametric relationships between the covariates and the outcomes can be estimated via maximisation of a penalised likelihood. We applied this model to data from 76 women undergoing IVF treatment.

Results: Six women in the sample had their IVF cycle cancelled. A model that includes two variables that are known to have strong relationships with ovarian response – age and serum anti-Müllerian hormone concentration – suggests that there is evidence of a complex non-linear relationship between these two factors in predicting both the probability of cycle completion and the oocyte yield of completed cycles.

Conclusions: The proposed model can help identify variables that are associated with the outcome of an IVF cycle, so can be used to validate new potential prognostic factors. It could be extended to allow for the possibility of cycle cancellation due to hyperstimulation, which did not occur in our sample.

[RP2.24]

Multiple imputation for constraint-based causal discovery with cohort data

Janine Witte, Ronja Foraita, Ryan M. Andrews, Vanessa Didele

Department of Biostatistics and Epidemiology, Rutgers School of Public Health, United States

Background: Observational data obtained e.g. from cohort studies, typically contains missing values and it is popular to address these by multiple imputation. When the research objective is to explore the relationships among many variables simultaneously, recent work has suggested combining constraint-based causal discovery with multiple imputation (Foraita et al. 2020).

Objective: On this poster, we extend and generalise the approach by Foraita et al. 2020 so as to deal with issues common for cohort data, namely variables of different scale (e.g. continuous, categorical, ordered categorical) and monotone missingness patterns.

Methods: Constraint-based causal discovery algorithms take as input the conditional independencies found in the data, where the type of independence measure is usually based on a test statistic chosen according to the scale of the considered variables. Popular tests include Fisher's z-test for continuous variables, the G-test for categorical variables, and the Conditional Gaussian likelihood ratio test for mixed variables. We give an overview about how multiple imputation analysis can be carried out where the key is to choose an appropriate pooling rule for such test statistics. We further investigate how the choice of the test type should inform the choice of the imputation model. Finally, we illustrate and compare with simulated cohort data the performance of different multiple imputation procedures with that of alternative approaches such as available-case analysis.

Results: Our comparisons suggest that the imputation model should be chosen such that it is compatible with, or congenial to, the model implied by the chosen conditional independence test, even if the imputed values appear not to match the observed values. Then, multiple imputation outperformed alternative approaches in many scenarios, including those with a monotone missingness pattern.

Conclusions: Our work provides further evidence of the usefulness and flexibility of multiple imputation especially as a promising alternative to available-case analysis for constraint-based causal discovery. R code for combining multiple imputation with different conditional independence tests will be made available.

[RP2.25]

Use of genetic variants as Instrumental Variables to be applied in mediation analysis: a new methodological proposal

Claudia Coscia[1], Núria Malats[2], Teresa Perez[2]

[1] *Spanish National Cancer Research Centre (CNIO), Spain*
[2] *Department of Statistics and Data Science, Complutense University (UCM), Spain*

Background: The use of Causal Mediation Analysis (CMA) is increasing in clinical and epidemiological settings. CMA estimates direct, indirect, and total effect of an exposure on an outcome, considering a third variable called mediator. Assumptions regarding CMA are strong and not always easy to prove. Mendelian Randomization Analysis (MRA) uses genetic variants as Instrumental Variables (IVs) to estimate the causal effect of exposure on the outcome. Extension of both methods can lead to unbiased estimates. We study this potential impact on a real and a simulation study on pancreatic cancer (PC), which is the 4th cause of death by cancer worldwide and presents a complex aetiology. Obesity and type 2 diabetes mellitus (T2DM) are stablished risk factors for PC.

Objective: To obtain unbiased results of the direct, indirect and total effect using IVs in several CMA scenarios, and to compare two statistical methodologies: Multiple-Stage-Least-Square (MSLS) and Structural Equation Models (SEMs).

Methods: We based on the PanGenEU case-control study to obtain simulation datasets and to estimate the causal effect of obesity on PC with the mediation of T2DM. We considered both exposure and mediator as continuous and binary variables, different sample sizes, and correlations. We extended the MSLS approach considering that each exposure and mediator present an IV and using predicted values of both variables to obtain the causal estimators. We also explored the use of SEMs technique, which is not described for IVs integration. We consider bias, mean squared error (MSE) and coverage as performance measures.

Results: For continuous variables, direct effect estimation is similar between methods ($bias_{MSLS}$ = 0.0011; $bias_{SEM}$ = 0.0011) and there is a difference in terms of coverage ($coverage_{MSLS}$ = 0.99 vs $coverage_{SEM}$ = 0.93) for the indirect effect. We obtained different results between the methods for binary variables, being SEMs the method with lowest bias and MSE results (MSE_{MSLS}=0.21 vs MSE_{SEM}=-0.12) in the direct effect estimation. Increasing sample size improve results obtained.

Conclusion: The variable type and the different scenario condition the performance of the methodology proposed. Our results show that, by extending CMA using IVs, we can address causal mediation analysis hypotheses.

[RP2.26]

Comparison regions to visualize uncertainty in two-parameter estimation problems in medical research

Maren Eckert[1], Werner Vach[2]

[1] *Institute of Medical Biometry and Statistics, Medical Faculty and Medical Center, University of Freiburg, Germany*
[2] *Basel Academy, Switzerland*

Background: Some research questions in the health sciences require considering the joint stochastic uncertainty of two parameter estimates. Balancing increased costs or an increased frequency of side effects vs a gain in a patient relevant outcome when comparing two treatments or balancing an increased rate of FP decisions against an increased rate of TP decisions when comparing two diagnostic tests are typical examples. It is very common to visualize this uncertainty by plotting two-dimensional confidence regions. This allows us to test post hoc null hypotheses about a single point in a simple manner. However, when balancing between two parameters the interest is not in rejecting hypotheses on single points. Typically we are interested in that some weighted average is above a certain threshold. The threshold and the weights may differ from stakeholder to stakeholder increasing the need for simple approaches to post hoc testing.

Methods: We introduce the idea of two dimensional comparison regions as a simple visual tool allowing directly such post hoc tests in a valid manner.

Results: A general construction principle could be derived. It allows us to demonstrate that comparison regions have the same shape as confidence regions. However, to conduct a post-hoc test at the 5% level, the corresponding comparison region is distinctly smaller than a 95% confidence region.

Conclusions: Comparison regions add a valueable tool improving the presentation of results in two parameter estimation problems. They may provide a more adequate description of stochastic uncertainty than confidence regions.

[RP2.27]

Machine learning approaches for the diagnosis of coeliac disease

Anna Fowler[1], Michael E. B. Fitzpatrick[2], Andrew D. Foers[3], M. Saad Shoukat[3], Shelley C. Evans[3], Paul Klenerman[2], Elizabeth J. Soilleux[3]

[1] *Department of Biostatistics, University of Liverpool, United Kingdom*
[2] *Department of Medicine, University of Oxford, United Kingdom*
[3] *Department of Pathology, University of Cambridge, United Kingdom*

Background: Coeliac disease is an autoimmune disease which is triggered by gluten. Diagnosis of coeliac disease is difficult due to the wide range of symptoms, and currently the gold standard involves examination of a biopsy by a pathologist and requires the patient to consume gluten. This process is subjective and often painful for the patient.

T-cell receptors (TCRs) are located on the surface of T cells and recognise pathogens; a wide range of receptors are required for a healthy immune response. However in coeliac disease patients, there are some TCRs which recognise and mount an immune response towards gluten. Each TCR has a unique genetic sequence which determines the pathogens it recognises, by capturing those genetic sequences, we can profile the TCR repertoire.

Objective: Our objective is to develop an algorithm for the diagnosis of coeliac disease from sequencing of the TCR repertoire.

Methods: We introduce a novel algorithm that takes a set of TCR repertoire sequences from healthy controls and coeliac patients as an input, and aims to classify these patients according to their coeliac disease status. The algorithm involves calculating the frequency of very short genetic motifs (kmers), performing principal component analysis then selecting the principal components which enable the best separation between the groups.

A machine learning approach is adopted to determine the optimal parameters of the algorithm and leave-one-out cross validation is used to evaluate the performance.

Results: On a dataset consisting of 31 coeliac patients and 21 healthy controls, our algorithm has a training accuracy of 95.0% and a testing accuracy of 87%. Importantly, we demonstrated that TCR repertoires from an additional 4 coeliac patients on a gluten free diet were correctly classified whereas they would be misclassified as unaffected by coeliac disease by all current tests.

Conclusion: Our algorithm has the potential to provide a novel diagnostic test for coeliac disease which is objective and does not require the patient to consume gluten. It may also be broadly applicable with application to a wide range of autoimmune diseases.

[RP2.28]

A Progressive Three-State Model to Estimate Time to Cancer: A Likelihood-Based Approach

Eddymurphy Akwiwu[1], Thomas Klausch[1], Henriette Jodal[2], Beatriz Carvalho[3], Magnus Loberg[2], Mette Kalager[2], Johannes Berkhof[1], Veerle Coupe[1]

[1] *Department of Epidemiology and Biostatistics, Amsterdam UMC, Netherlands*
[2] *Department of Health Management and Health Economics, University of Oslo, Norway*
[3] *Department of Pathology, Netherlands Cancer Institute, Netherlands*

Background: To optimize colorectal cancer (CRC) screening, information regarding the time-dependent risk of advanced adenomas (AA) to develop into CRC is crucial. However, since AA are treated after diagnosis, the time from AA to CRC cannot be observed in an ethically acceptable manner.

Objective: We propose a method to indirectly infer the time from AA to CRC using a three-state progressive disease model describing the natural history of CRC. We derive and maximize the likelihood function using longitudinal surveillance data.

Methods: We assumed all individuals to be successfully treated for adenoma at the start of surveillance and health outcomes to be observed at repeated surveillance visits. Time X from the adenoma-free state (AF) to AA and time Y from AA to CRC were estimated simultaneously, allowing inclusion of covariates on X. We assumed an exponential distribution for X and either exponential or Weibull for Y. Parameter estimates were obtained by maximizing the likelihood function. Model performance was assessed via simulation and the methodology was applied in a cohort of 834 individuals from the Norwegian Cancer Registry, participating in the nationwide longitudinal surveillance program after adenoma removal between 1993 and 2007.

Results: The model provided consistent parameter estimates, with bias dependent on sample size n and the proportion of CRCs. When we assumed a 5% CRC and 35% AA prevalence with a Weibull model for Y, bias of the scale decreased from 10% to 0.3% as n increased from 500 to 10,000. With a higher proportion of CRCs, bias was negligible from n= 500 onwards. Also, the variance of the scale was sensitive to the proportion of CRCs. In the Norwegian data (AF: 76%, AA: 22%, CRC: 2%), the Weibull scale and shape parameters for Y were 29.69 (95% CI: 6.94-126.95) and 0.72 (95% CI: 0.41-1.27), respectively, resulting in a median duration from AA to CRC of 17.8 years (mean 36.6 years).

Conclusion: When X is assumed to be exponential, any parametric distribution for the unobservable time-dependent risk from AA to CRC can be estimated. Using a Bayesian approach to this estimation problem allows further flexibility (see abstract 0545; [OC39.3]).

[RP2.29]

A Case Study of a Successful Design Adaptation based on an improved Group Sequential Holm procedure

Fredrik Öhrn

Early Biometrics and Statistical Innovation, AstraZeneca, Sweden

Our case study was a large cardiovascular outcomes trial (CVOT) where the pre-specified primary efficacy objective was to assess the effect on major adverse cardiovascular events (MACE) vs standard of care. After the first patient was randomized, but before the first interim analysis, the CVOT for another drug in the same class reported important findings with respect to MACE and other endpoints. Although the population in the other CVOT was different from that in the case study, the emerging information presented an opportunity to re-design the confirmatory testing strategy of our case study, with dual primary efficacy endpoints for which the familywise error rate (FWER) would have to be strongly controlled.

An updated protocol was submitted to regulatory authorities before the first interim analysis. At study read-out MACE did not meet statistical significance, while the result for the new primary endpoint was found to be both clinically meaningful and statistically significant. The re-design helped ensure fuller assessment and confirmation of potential clinical benefits while maintaining strong control of FWER.

The adaptation was optimised by exploring a broad range of scenarios and testing strategies. Power gains with the group sequential Holm [1] procedure that was ultimately applied will be illustrated. A generalization of the recently published method of Li et al. [2] to the group sequential Holm setting will also be shown, with sharper rejection regions and corresponding power gains if retrospectively applied to our case study.

References:

1. *Y. Ye, A. Li, and B. Yao. A group sequential Holm procedure with multiple primary endpoints. Statistics in Medicine, 32:1112-1124, 2013.*
2. *H. Li, J. Wang, X. Luo, J. Grechko, and C. Jennison. Improved two-stage group sequential procedures for testing a secondary endpoint after the primary endpoint achieves significance. Biometric Journal, 60:893-902, 2018.*

RP3: Method(s) Validation and/or Comparison

[RP3.01]

Actual predictive performances of Bayesian prediction intervals for meta-analysis

Yuta Hamaguchi[1], Hisashi Noma[2], Kengo Nagashima[3]

[1] *Department of Statistical Science, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies, SOKENDAI, Japan*
[2] *Department of Data Science, The Institute of Statistical Mathematics, Japan*
[3] *Research Center for Medical and Health Data Science, The Institute of Statistical Mathematics, Japan*

Context: The prediction interval has been increasingly used in meta-analyses as a useful measure for assessing the magnitude of treatment effect and between-studies heterogeneity. In calculations of the prediction interval, although the Higgins-Thompson-Spiegelhalter method is used most often in practice, it might not have adequate coverage probability for the true treatment effect of a future study under realistic situations. An effective alternative candidate is the Bayesian prediction interval, which has also been widely used in general prediction problems. However, these prediction intervals are constructed based on the Bayesian philosophy and their frequentist validities are only justified by large-sample approximations even if non-informative priors are adopted. There has been no certain evidence that evaluated their frequentist performances under realistic situations of meta-analyses.

Objective: To assess the frequentist performances of Bayesian prediction intervals in meta-analyses.

Results: We conducted extensive simulation studies to assess the frequentist coverage performances of Bayesian prediction intervals under general settings of meta-analyses in medical studies. We adopted 11 non-informative prior distributions. We found the frequentist coverage performances strongly depended on what the prior distributions were adopted, especially when the number of studies was smaller than 10. In addition, there were no prior distributions that retained fine coverage properties consistently under various settings of meta-analysis. The coverage performances became better when the number of studies gets larger; although a typical number of studies of meta-analyses in medical researches is smaller than 20.

Conclusions: The Bayesian prediction intervals have favorable frequentist performances when the number of studies is large, but they cannot be used as approximate frequentist prediction intervals under general settings of meta-analyses in medical researches. The development of more rich methods for accurate predictions is needed, and represents an important priority for future work in research synthesis methodology.

[RP3.02]

Evaluation by simulation of clinical trial designs for treatment efficacy during infectious diseases outbreak

Pauline Manchon[1], Drifa Belhadi[2], France Mentre[2], Cedric Laouenan[2]

[1] *Department of Epidemiology Biostatistics and Clinical Research, AP-HP, France*
[2] *IAME UMR 1137, Inserm, France*

Background: Emerging infectious diseases are characterized by irregular outbreaks with high mortality rate. Difficulties arise when implementing therapeutic trials in this context. The outbreak duration is hard to predict and can be short compared to delays of trial launch and number of subject needed (NSN) recruitment.

Objectives: To compare different trial designs for evaluation of an experimental treatment for various outbreak scenarios by clinical trial simulation.

Methods: Four designs were compared: fixed or group-sequential, single- or two-arm. The primary outcome was 14-day survival rate. For single-arm designs, results were compared to a pre-trial historical survival rate p_H. Treatments efficacy was evaluated by one-sided tests of proportion (fixed designs) and Whitehead triangular tests (group-sequential designs). We varied survival rates in the control arm p_C and expected survival rate differences Δ (including 0). Three scenarios were considered: standard (fixed p_C, reaching NSN for fixed designs and maximum sample size N_{max} for group-sequential designs); changing with time (increasing p_C over time); early stop of recruitment (epidemic ends). We calculated the proportion of simulated trials showing experimental treatment efficacy, with K=93,639 simulated trials to get a type- I-error $PI_{95\%}$ of [0.024;0.026].

Results: Under H_0 (p_C=0.50; Δ=0), for standard scenarios, the type-I-error was maintained regardless of trial designs. When $p_C \neq p_H$ type-I-error was inflated most of the time. For changing with time scenarios, wrong conclusions were more often observed for single-arm designs due to an increase of Δ over time and the assumption around p_H. Under H_1 (p_C=0.5; Δ=+0.2), for standard scenarios, power were similar between single- and two-arm designs when $p_C=p_H$. For stopping of recruitment scenarios, single-arm performed better than two-arm designs, and fixed designs reported higher power than group-sequential designs. A web R-Shiny application was developed.

Conclusions: At an outbreak beginning, group-sequential two-arm trials should be preferred, as the infected cases number will increase allowing to conduct a strong RCT. Group-sequential designs will allow early termination of trials in cases of harmful experimental treatment. After the epidemic peak, fixed single-arm design should be performed, as the cases number decreases but this assumes a high level of confidence on the pre-trial historical survival rate.

[RP3.03]

Machine Learning Algorithm to Identify a Metabolic Profile Able to Predict Biomarker Levels

Fabiola Del Greco M.[1], Chiara Volani[1], Johannes Rainer[1], Giuseppe Paglia[2], Peter P. Pramstaller[1]

[1] *Institute for Biomedicine, Eurac Research, Italy*
[2] *School of Medicine and Surgery, University of Milano-Bicocca, Italy*

Background: Metabolomics aims to measure small molecules (called metabolites) in cells, tissues, biofluids, that represent intermediates and/or end products of biochemical/cellular processes. As such, metabolomics has shown to be useful for making prediction about disease risks. Given the large and complex dataset, the Machine learning (ML) approach represents an appropriate statistical and computational tool for building predictive models.

Objective: By a data-driven investigation, the aim of this work is to provide a ML workflow to predict biomarker levels by using targeted metabolomics data from a general population. Specifically, we identify and discuss specific ML issues that could influence the results.

Method: We explore a supervised ML algorithm based on the feature selection using hypothesis testing, with particular attention on the following problems: the presence of confounding; non-independent observations (typical of family- or pedigree-based studies); high dimensionality of the dataset (feature space). The impact of those issues on the results is investigated through a dataset of 154 serum targeted metabolites determined by liquid chromatography (LC)–electrospray ionization–tandem MS and flow injection electrospray ionization–tandem MS profiling (AbsoluteIDQ p180 kit, Biocrates Life Sciences AG), and serum ferritin levels, as an example of biomarker, from a subsample of the Cooperative Health Research In South Tyrol (CHRIS) study with around 5,000 participants.

Results: In the context of a family-based study, we provide a workflow for predicting biomarker levels starting from a large and complex feature space, based on a supervised ML approach.

Conclusions: This workflow should provide a framework for greater integration of ML within metabolomics studies, taking into account the problem of confounding, the high complexity and dimensionality of the dataset and the presence of family related data.

[RP3.04]

Autoregressive linear mixed effects models in structural equation modeling

Ikuko Funatogawa[1], Takashi Funatogawa[2]

[1] *Department of Statistical Data Science, The Institute of Statistical Mathematics, Japan*
[2] *Clinical Science and Strategy Department, Chugai Pharmaceutical Co. Ltd., Japan*

Background: In the fields of statistics in medicine, we previously proposed autoregressive linear mixed effects models and showed the interpretation (Funatogawa et al., 2007; Funatogawa and Funatogawa, 2018). In an autoregressive linear mixed effects model, the current response is regressed on the previous response, fixed effects, and random effects.

Objective: Recently, development of models including autoregressive parameters in longitudinal data analysis has become active in fields other than medicine. We clarify the usage of these models.

Methods: We compare the autoregressive linear mixed effects models with the dual change models in the latent change score models in psychology and developmental, social, and behavioral research.

Results: The dual change models in the latent change score models are closely related with the autoregressive linear mixed effects models. Although the latent change score models focus on the latent changes, the models are easily re-parameterized to represent the latent responses itself as usual. The dual change models have a slope factor and a proportional change, and these correspond to the random intercept and an autoregressive term in the autoregressive linear mixed effects models. The latent change score models are based on structural equation modeling (SEM), whereas the autoregressive linear mixed effects models are based on the linear mixed effects models. Maximum likelihood (ML) methods are usually used in both models, and it is called full information ML in SEM. Vector representation differs between two models. For the latent change score models, SEM software, such as Mplus, OpenMx in R, lavaan in R, and LISREL are used, and the models are often expressed by path diagrams instead of vector representation. Both models are extended to multivariate longitudinal models. The latent change score models usually include no time-dependent covariates, whereas the autoregressive linear mixed effects models often include a time dependent covariate, such as treatment doses. When there are no time dependent covariates, the models are expressed by nonlinear mixed effects models with linear random effects. We analyze actual data to compare both models.

Conclusion: Closely related models are developed and used independently in different fields.

[RP3.05]

Transformations for estimation of restricted mean survival times in small sample size

Hiroya Hashimoto, Akiko Kada

Clinical Research Center, National Hospital Organization Nagoya Medical Center, Japan

Background: Recently, a growing number of clinical trials have been using a restricted mean survival time (RMST) for summary of survival analysis data. Although RMST is easy to understand, when the sample size is small or the number of events is small, the distribution of estimated RMST greatly deviates from a normal distribution.

Objective: We assessed the usefulness of using transformation methods for estimating the confidence interval (CI) of RMST in a small sample size by numerical simulation.

Method: Assumed that the survival time follows Weibull distribution, we evaluated the proportion of errors above and below the true value for transformed CIs of the RMST by numerical simulation. Several conditions with different parameters of Weibull distribution and sample size were examined. The number of simulations was 10,000 for each condition. The CIs of the RMST were estimated using no transformation and four transformation methods (logarithmic, arcsine-square root, logit, and complementary log-log).

Results: Under conditions with a small number of events, the CI without transformation was overestimated, and the logarithmic transformed CI was overestimated even more strongly. On the other hand, the logit and the complementary log-log transformed CIs were underestimated. Among those we considered, the arcsine-square root transformed CI was the most balanced.

Conclusions: The results of this simulation show that the CI of the RMST using the arcsine-square root transformation is better estimation than that without transformation. The results are important in single-arm clinical trials using the RMST as an endpoint to compare with historical controls.

[RP3.06]

A Bayesian Multivariate Joint Model for Multiple longitudinal and Multistate Processes

Elham Madreseh[1], Mahmood Mahmoudi[1], Mohssen Nassiri-Toosi[2], Taban Baghfalaki[3], Hojjat Zeraati[1]

[1] *Department of Epidemiology & Biostatistics, Tehran University of Medical Sciences, Islamic Republic of Iran*
[2] *Liver Transplantation Research Center, Tehran University of Medical Sciences, Islamic Republic of Iran*
[3] *Department of statistics, Tarbiat Modares University, Islamic Republic of Iran*

Motivated by a cirrhosis data example on liver transplant failure, we propose a new joint model which considers a multivariate linear mixed sub-model for longitudinal processes and a proportional intensity sub-model for multistate process. In this context, a semi-Markov model is defined for transition between states and also the time spent in each transition is assumed to have parametric distribution. The both sub-models linked by a function of shared random effects. From a practitioner's point of view, suitable parameterization are discussed, to capture dependency structure between the longitudinal and multistate outcomes. Parameters of this Multivariate Joint Multistate Model (MJMM), are estimated within the Bayesian framework using Markov Chain Mont Carlo (MCMC). The Deviance Information Criteria (DIC) is derived for model selection and comparison.

[RP3.07]

An overview on how measurements affected by medication use are handled and reported in observational research

Jungyeon Choi[1], Olaf Dekkers[2], Saskia le Cessie[3]

[1] *Department of Clinical Epidemiology, Leiden University Medical Center, Netherlands*
[2] *Department of Endocrinology and Metabolism, Leiden University Medical Center, Netherlands*
[3] *Department of Biomedical Data Sciences, Leiden University Medical Center, Netherlands*

In epidemiological research it is common to encounter measurements affected by medication use, such as blood pressure lowered by antihypertensive drugs. This may cause bias when one is interested in the relationship between the variables not affected by medication. We have previous shown [1] that effect estimates may largely defer depending on which method was use to handle the problem.

In this systematic review, we aim to provide an overview on how measurements affected by medication use are handled in observational studies. We further aim to evaluate whether the information about medication use is sufficiently provided, whether the justification for the chosen method is properly discussed and how often (in)valid methods are used.

PubMed database were searched for studies that performed logistic or linear regressions as their main analysis between January 1st 2015 to December 31st 2019 and where blood pressure, glucose or lipid measurements were used in the analysis. We searched two journal fields; hypertension journals and diabetes journals, and selected five journal for each field. In each journal, 10 articles were randomly selected to be reviewed.

The most commonly used method for handling medication use, regardless whether the exposure, outcome or a confounder was affected, was adjusting for medication use (42%), which is an invalid approach when the exposure or outcome is affected by medication. The second most used methods was excluding treated individuals (25%), which introduces bias when the outcome is affected. The problem of medication use was ignored in 22% of the studies. Justification for the chosen method was only given in 17% of the studies. Information about medication use was insufficiently reported in 31%. We evaluated that approximately 47% of the studies used invalid methods for handling medication use.

Our review shows that measurements affected by medication use were often handled with invalid methods, and justification is rarely given. We suggest researchers to critically consider how to handle medication use and provide clear information related to medication use.

References:

1. *Choi J, Dekkers OM, le Cessie S. A comparison of different methods for handling measurements affected by medication use (submitted)*

[RP3.08]

A comparison of methods for meta-analysis of continuous outcomes from randomised trials

Miriam Hattle, Richard Riley

Keele University, United Kingdom

Background: When conducting a meta-analysis for randomised trials with a continuous outcome, a challenge is that trials use different methods to produce their treatment effect estimates. Though an analysis of covariance (ANCOVA) model provides more desirable statistical properties, some trials only use a final score or change score approach. In such situations, it is standard in Cochrane for meta-analysts to combine a mixture of final score, change score, and ANCOVA results across studies.

Objectives: To propose a novel bivariate meta-analysis approach for combining a mixture of ANCOVA and final score results across trials, and to compare the approach with other methods.

Methods: We propose a bivariate meta-analysis of ANCOVA and final score results, so that a pooled ANCOVA result can be obtained whilst borrowing strength from correlated final score results. The approach can accommodate trials that provide either (i) only ANCOVA results, (ii) both ANCOVA and final score results, or (iii) just final score results, and thereby utilises all available information. A simulation study is used to compare performance of the bivariate approach versus the standard Cochrane approach and an approach that discards trials entirely if they do not report ANCOVA results.

Results: In the simulation study, for simulations where trials were well-balanced at baseline, the bivariate approach and the standard Cochrane method performed similarly with regards to pooled treatment effect in terms of bias, mean squared error, mean empirical standard error and coverage. For scenarios with baseline imbalance, a separate meta-analysis of just ANCOVA treatment effect estimates generally had better performance than the bivariate or standard Cochrane approaches. Real examples are used for illustration.

Conclusions: We recommend that the method recommended by Cochrane should be used when there are no concerns about baseline imbalance. If baseline imbalance is a concern, then we recommend a including a separate meta-analysis of only ANCOVA treatment effect estimates.

[RP3.09]

A Comparison of Survival Methods for the Analysis of Recurrent Childhood Wheeze

Patrick Hannan[1], Maia Lesosky[2], Heather Zar[2]

[1] Division of Epidemiology and Biostatistics, University of Cape Town, South Africa
[2] Department of Paediatrics and Child Health, University of Cape Town, South Africa

Background: Wheezing is common in young children. Data from high-income countries show that by the age of six, approximately 50% of children will have experienced at least one episode of wheezing, while some children may wheeze recurrently. Furthermore, childhood wheezing may be associated with reduced lung function and increased risk of asthma in later life. Multiple statistical approaches are available are available to evaluate recurrent events.

Objectives: This study aimed to use multi-state models to estimate the rate of transition between states of wheeze in children from birth to the age of three years. Furthermore, a comparison of traditional survival methods to multi-state models for the analysis of childhood wheezing was carried out.

Methods: This study was a secondary analysis of data from 1143 children from birth to three years, born to mothers enrolled in the Drakenstein Child Health Study. A time-inhomogeneous multi-state Markov model with three states was constructed. Baseline transition intensities were assumed to be piecewise constant. Direct transitions were not permitted between non-adjacent states. A proportional intensity model was constructed to model recurrent childhood wheeze using traditional survival methods. Model precision and inference were compared to assess performance.

Results and Conclusions: Of the 1143 children included in the study, 471 (41%) experienced at least one episode of wheezing, and 225 (20%) experienced more than one episode of wheezing in the first three years of life. A total of 944 episodes of wheezing were recorded in the 36 months of follow-up time. Lower respiratory tract infection (LRTI) associated wheeze (transition intensity: 0.0003) and wheeze not associated with LRTI (transition intensity: 0.00042) were equally likely to be the first wheezing event. Recurrent wheezing events were more likely to be associated with a LRTI. Male gender was associated with an increased likelihood of transitioning to the LRTI-associated wheeze state from never wheeze (HR: 1.736; 95% CI: 1.24 - 2.432) and had significantly higher likelihood of subsequent recurrent wheezing. Multi-state models and recurrent survival methods produced similar estimates for risk factors for wheeze, however multi-state models allowed for inclusion of a greater number of wheeze events as well as more flexibility in study design.

[RP3.10]

Modelling egg counts of Helminths to evaluate the influence of multiple infections and compare treatments

Michel Vaillant[1], Piero Olliaro[2]

[1] *Competence Center for Methodology and Statistics, Luxembourg Institute of Health, Luxembourg*
[2] *Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, United Kingdom*

Background: Treatment efficacy for schistosomiasis and soil-transmitted nematodes (STN) is assessed as egg reduction rate (ERR) based on the difference in geometric means between pre- and post-treatment egg counts. This however does not allow for comparison of treatment effects in controlled trials.

Objective: To assess whether Poisson-type modelling could be used to compare efficacy between treatments and whether multiple species infections could have an impact.

Methods: We used two databases made of 16 studies of 6693 individuals infected with *S. mansoni* and/or *S. haematonium*, and 13 studies enrolling overall 6,829 individuals infected with A. lumbricoides, T. trichiura, and/or Hookworms, conducted at African, Asian, and Latin American sites, and treated with different drugs. Gender, age, treatment, species, follow-up duration and baseline egg counts were entered in all models as factors. Since different individuals were measured using different numbers of slides, this number was used as a proportionality constant (offset) in the models to analyse the sum of counts. In case of over-dispersion alternative models such as the quasi-Poisson, negative binomial & zero-inflated Poisson, were fitted and compared. Random variation in risk between individuals and random variation between slides were explicitly allowed for in multi-level model with the same distributions in a further approach.

Results: 1050 patients with 3577 measurements of Schistosomiasis and 5832 patients treated for STH were analysed. 92% of the observed post-treatment data were zero. The Poisson model of the sum of egg counts and the quasi-Poisson model proved unsuited due to over-dispersion. The negative binomial model showed a better fit and predicted 92.07% of zeros. Zero-inflated model predicted 91.98% of zeros.

With the multilevel modelling strategy, the better model was the Poisson model (95.05% of zero counts predicted) which variable coefficients were higher than the sum egg counts' model. A 40 mg/kg, 60 mg/kg or 80 mg/kg praziquantel dose reduced the egg counts between baseline and post-treatment but with no significant difference between treatments. Baseline counts were significant predictors of post-treatments counts.

Conclusions: This study shows that adequate modelling of the sum of post-treatment egg counts or raw egg counts could be useful for assessing treatment effects of anthelmintic treatment.

[RP3.11]

Interim adaptation in individually randomised cross-over trials

Michael Grayling[1], James Wason[1], Adrian Mander[2]

[1] *Population Health Sciences Institute, Newcastle University, United Kingdom*
[2] *Centre for Trials Research, Cardiff University, United Kingdom*

Background: Cross-over designs are important to the evaluation of treatments in many disease areas, most notably chronic conditions. Typically, they are favoured because the fact that patients can act as their own control is highly efficient. However, cross-over designs are not without limitation. In particular, total cross-over trial length can be long, and specification of the required within-patient variance can also be challenging. Furthermore, in the case where included treatments have substantially different efficacy, commonly employed allocation sequences can lead to patients spending long periods of time on inferior treatments.

Objective(s): To resolve several of the issues that can be faced in cross-over trials, the use of adaptive methods offers promise. Therefore, we describe several recently developed approaches for interim adaptation in cross-over trials, including procedures to drop treatments for futility/efficacy, to re-specify the required sample size, and to continuously update the utilised treatment sequences.

Method(s): In each instance, the discussed procedures assume the most commonly employed method of analysis for a cross-over trial with a continuous outcome. Thus, the proposals function as an extension of more well-known methods for parallel arm designs to the requirements of repeated measure designs. The application of established methods of group-sequential design to cross-over trials is first categorised. Blinded methods of sample size re-estimation are then described, based on one-sample and block randomised based approaches. To adaptively specify patient sequence allocation, an objective function balancing power and ethical considerations is also proposed.

Results: Using design parameters motivated by a sleep-apnoea cross-over trial, we demonstrate that a group-sequential approach could reduce the average required number of observations under the null hypothesis by as much as 33%. In addition, for a wide range of true parameter combinations, the proposed sample size re-estimation procedures ensure approximately the desired type-II error-rate is attained. Response-adaptive sequence allocation can also substantially enhance the average patient outcome, often at only a small cost to the trial's error-rates.

Conclusions: Cross-over trials can be challenging to conduct in practice. The use of interim adaptation could help mitigate several issues that are routinely faced. We recommend that such methods be considered more often.

[RP3.12]

Effects of stratification on performance of methods for analyzing continuous data from stratified CRTs

Sayem Borhan[1], Alexandra Papaioannou[1], Jinhui Ma[1], Jonathan Adachi[2], Lehana Thabane[1]

[1] *Department of Health Research Methods, Evidence, and Impact, McMaster University, Canada*
[2] *Department of Medicine, McMaster University, Canada*

Background: Adoption of cluster randomized trials (CRTs) with stratified design is increasing. However, we have limited knowledge about the performance of methods under different scenarios of stratification, in terms of number of stratification variables and strata.

Objectives: In this study, we conducted a simulation study to examine the effects of stratification, in terms of number of stratification variables and strata, on the performance of methods for assessing the intervention effect from stratified CRTs. We evaluated several methods, in terms of type I error rate, empirical power, root mean square error rate, and width of 95% confidence intervals, for analyzing continuous data from stratified CRTs.

Methods: We compared 4 methods: mixed-effect, generalized estimating equation (GEE), cluster-level (CL) linear regression and meta-regression to analyze the continuous data from stratified CRTs using a simulation study with varying number of stratification variables, number of strata, number of clusters, cluster sizes, treatment effect, and intra-cluster correlation coefficients (ICCs). The performance of the methods was evaluated in terms of type I error rate, empirical power, root mean squared error (RMSE), and average width of the 95% confidence intervals.

Results: GEE and meta-regression methods yielded overly liberal type I error rates, approximately 10%, for study with small number of clusters and strata, while the type I error rates for CL linear regression and mixed-effect methods were approximately 5%. For all methods, the study with large number of strata required fewer number of clusters to yield satisfactory 80% power compared to study with small number of strata. All methods, except GEE, yielded similar power, when comparing among themselves, for studies with same number of strata. Meta-regression yielded the lowest power compared to other methods. Similarly, meta-regression had the highest average RMSEs and widest widths of 95% CIs compared to other methods.

Conclusions: The performance of all methods improved as the number of strata increased along with number of clusters and cluster size. There were very little or no difference among the methods themselves for studies with same number strata. Meta-regression was the least powerful and efficient method.

[RP3.13]

Backward elimination to select covariates for confounding adjustment: harmful practice or sometimes useful?

Kim Luijken[1], Rolf Groenwold[1], Maarten van Smeden[1], Susanne Strohmaier[2], Georg Heinze[2]

[1] *Department of Clinical Epidemiology, Leiden University Medical Center, Netherlands*
[2] *Section for Clinical Biometrics; Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria*

Background: Identification of causal effects from observational data relies on proper control of confounding. To select covariates, the common view among practicing statisticians is that automated covariate selection methods, such as backward elimination (BE), should be avoided as they can lead to seriously biased estimated effect sizes and underestimation of statistical uncertainty. Nevertheless, BE has occasionally been reported to possibly improve efficiency of estimation.

Objectives: To evaluate under which conditions efficiency of causal effect estimation in observational studies can be improved when BE is used to select covariates for confounding adjustment from a pre-specified set of covariates.

Methods: In a simulation study we assumed that a set of covariates can be pre-specified that represents a mixture of true confounders, true instruments, true predictors and noise variables. We simulated small samples with a single binary exposure and a binary outcome and applied BE of covariates with various elimination thresholds in comparison to no elimination. The causal effect of the exposure on the outcome was evaluated in terms of the mean squared error (MSE) of the logarithm of the marginal risk ratio (MRR) estimated using logistic regression models.

Results: In most scenarios, the MSE was lowest when the log(MRR) was estimated by a model that included all covariates in the initial set. Models selected using BE resulted in a lower MSE of the log(MRR) only in situations where the true confounding effect was weak, when the outcome was not well predictable by exposure and covariates or when instrumental variables were present.

Conclusions: While under correct model pre-specification (no confounder missed) estimates are unbiased without further selection and often more efficient than with BE, we could identify some scenarios under which BE may be able to improve efficiency. We conclude that the conflicting recommendations in the literature may have resulted from differences in the assumptions on the true data generating process, the compositions of the available covariate sets and the typically considered sample sizes.

[RP3.14]

Informative Presence and Observation in Clinical Risk Prediction: A review of methods

Rose Sisk, Glen Martin, Matthew Sperrin, Lijing Lin, Niels Peek

Health eResearch Centre, Division of Informatics, Imaging and Data Science, University of Manchester, United Kingdom

Background: Clinical prediction models (CPMs) are increasingly developed and validated using electronic health records (EHRs) since they provide rich, longitudinal information on a patient's interactions with healthcare services. The analysis of such data is not, however, without challenges. Specifically, the observation process of EHRs is dependent on the underlying health status of the individual, which not only leads to irregularly collected information, but importantly means that the type, timing, and frequency of data collection could be informative with respect to a patient's health status. This is referred to as "informative presence" and "informative observation". Informative presence/observation may be an opportunity, as the additional information contained within the observation process could improve accuracy of prediction models.

Aims: This project aims to synthesise the existing analytical methodology that could be used to allow CPMs to learn from "informative presence" and "informative observation". In doing so, we aim to identify remaining methodological challenges in this area.

Methods: A systematic literature search was conducted by two independent reviewers. Keywords were identified and used to search Embase, MEDLINE and Web of Science. Articles were screened based on title and abstract at stage one, and full texts at stage two for any remaining papers.

Results: All methods (within 36 papers) discovered during this review broadly fall under three categories; methods which use derived information about the observation process as model predictors (e.g. missing indicators and counts of observations or visits), methods which make indirect use of the observation process via a latent structure (e.g. through random effects in joint models), or methods that model under informed missingness (e.g. likelihood-based methods and pattern-specific models).

Conclusions: There is clearly a growing recognition of both informative presence and informative observation in CPMs, and there is compelling evidence to suggest that incorporating these phenomema into prediction models has the potential to improve the performance of prediction models. However this is still an underdeveloped area, and further methodological work should be conducted to explore under what conditions each method improves predictive accuracy.

[RP3.15]

Robustness of latent class models for diagnostic testing with no gold standard

Matthew Schofield[1], Michael Maze[2], John Crump[2], Matthew Rubach[3], Renee Galloway[4], Katrina Sharples[1]

[1] *Department of Mathematics and Statistics, University of Otago, New Zealand*
[2] *Centre for International Health, University of Otago, New Zealand*
[3] *Division of Infectious Diseases, Duke University, United States*
[4] *Special Pathogens Branch, Centers for Disease Control and Prevention, United States*

It is difficult to estimate sensitivity and specificity of diagnostic tests when there is no gold standard. Latent class models have been proposed as they provide estimates without the need for a gold standard. We argue that there is a lack of robustness to model assumptions that has not been well documented. Using a motivating example of the evaluation of point of care tests for leptospirosis in Tanzania, we show how a realistic violation of assumptions underpinning the latent class model can lead directly to substantial bias in the estimates of the parameters of interest. The violation is minor in the sense that it cannot be routinely detected with goodness-of-fit procedures, but is major with regard to the resulting bias. We show that use of latent class models can be as bad, or worse, than treating an imperfect test as a gold standard.

[RP3.16]

Incorporating follow-up lengths in sample size re-estimation to compare over-dispersed count data

Masataka Igeta[1], Shigeyuki Matsui[2]

[1] *Department of Biostatistics, Hyogo College of Medicine, Japan*
[2] *Department of Biostatistics, Nagoya University Graduate School of Medicine, Japan*

Background: In comparative clinical trials with recurrent events, such as exacerbations of chronic obstructive pulmonary disease (COPD), over-dispersion needs to be appropriately incorporated. Recently, blinded sample size re-estimation methods have been proposed to achieve stable power control by re-estimating the mean and variance parameters based on interim count data. However, these methods can substantially fail to control the power at the final analysis, especially when the variance function is misspecified.

Objectives: For the existing methods for sample size re-estimation for over-dispersed count data, under misspecification of the variance function, we show that the failure in power control can be brought by the difference in the length of patient follow-up between the interim analysis (with sample size re-estimation) and the final analysis.

Methods: We provide some formula and numerical results to show how the difference in follow-up lengths can impact on the power at the final analysis when the variance function is misspecified. Based on these observations, we also consider countermeasures to address this issue.

Results and Conclusions: The final analysis of clinical trial can be substantially under-powered (over-powered) for a re-estimated sample size based on a misspecified linear (quadratic) variance function when the form of the true variance is quadratic (linear). One effective approach to mitigate this issue would be to predict the distribution of the follow-up length at the final analysis and to perform sample size re-estimation based on the predicted distribution.

[RP3.17]

Evaluation of treatment effect using propensity score in single-arm clinical trial and external control group

Akiko Kada, Hiroya Hashimoto

Clinical Research Center, National Hospital Organization Nagoya Medical Center, Japan

Background: In rare disease areas, sometimes it is impossible to conduct randomized controlled trial, and small single-arm clinical trial would be conducted to evaluate the treatment effect explanatory. In such a situation, it would be useful to estimate the treatment effect with using external control groups. The issue is the similarity between single-arm clinical trial participants and the external control population. Propensity score methods works to create comparable groups, with assumption of strongly ignorable treatment assignment. In a limited target population, it is important whether the groups are comparable.

Objective: To evaluate treatment effect using propensity score method in various situation of single-arm clinical trial and external control groups.

Methods: We assumed a continuous outcome variable evaluated at two time points. We considered various settings of covariates of multiple normal distributions and unmeasured confounders. The sample size was determined from 20 to 100 for each group. We compared the inverse probability treatment weighted method using propensity score with the linear regression model. To evaluate performance of the methods, bias in estimating the change of outcome variable from baseline and the actual confidence interval coverage of nominal 95% confidence intervals were determined in simulation study. Generalizability index (Tipton 2014) was also confirmed to evaluate the similarity of the two groups.

Results: Misidentification of covariates in the outcome regression model or propensity score model, large differences in propensity scores between groups, and large gaps in covariate distributions between groups increased bias and reduced coverage. The smaller the sample size, the greater the bias of the method of propensity scores. The generalizability index was more stable according to sample size increased.

Conclusion: It is important to be able to identify the covariates strongly corelated to the difference between groups. For smaller sample size, propensity score may not be calculated or with large difference, then regression models would be a solution if covariates could be identified.

[RP3.18]

Prediction of Site Opening Times in Clinical Trials using Cox-type Survival Models and Random Forests

Aris Perperoglou, Kevin Quach, Youyi Zhang, Richard Dearden

Data Science and AI, AstraZeneca, United Kingdom

Background: Phase III clinical trials involve multiple countries and large number of clinical sites for screening and randomising patients into a trial. At the design phase of a study, each site's initiation time is important to estimate time to first patient arrival. This information determines trial costing, supply chain and logistics operations for the trial. Therefore, building predictive models to understand site initiation times at baseline and during the trial is essential to improve real-time site monitoring capabilities.

Objectives: The objectives were to train models on historical data from 110 past trials to predict a given site's time to first enrolled patient. Models are fitted at baseline and interim timepoints (25th, 50th, and 75th of all sites required) in the trial. Predictions of time to first patient also serve as an estimate on how many sites will be active in the trial at a given time in the future. This information is crucial for improving the prediction of recruitment rate of the trial.

Methods: Historical data were obtained from phase II/III trials sponsored by AstraZeneca between 2006 and 2018 within oncology, cardiovascular and respiratory diseases. Random forest model was trained on features engineered to incorporate site-, and trial-level details (planned no. of patients, sites, dates, therapeutic area) Survival models included a simple Cox proportional hazards model and landmarking for dynamic prediction, with information on similar covariates.

Results and Conclusions: We will present our findings on a test set of 30 trials. Preliminary results suggest Cox-type models have better predictive behaviour than random forests with boxplots of residuals confirm the Cox-type models are tightly bounded around zero. We will show a detailed comparison of predictions prior to study initiation and at given time-points once a study has started and showcase how these predictions influence the overall behaviour of recruitment prediction models. Models are implemented into clinical trials operations to improve site monitor and trial planning across organization.

[RP3.19]

A review and evaluation of methods for comparing clinical performance and detecting outliers

Menelaos Pavlou, Gareth Ambler, Rumana Omar

Department of Statistical Science, UCL, United Kingdom

Background: Routinely collected data are often used to compare hospital performance and to identify hospitals with unusual (poor) results. For example, it might be of interest to identify hospitals with unusually high in-hospital mortality following Percutaneous Coronary Intervention (PCI).

Objectives: Several statistical methods have been suggested for this purpose, including approaches that are based on simple funnel plots and others that are based on random effects modelling. The first make use of hospital-level data while the second can be either applied on hospital-level or patient-level data. Some methods are also able to incorporate risk-adjustment (ideally using a validated risk prediction model) to take into account differences in patient case-mix. We aim to review and evaluate the methods used for outlier detection using both real surgical data and simulated data.

Methods: We compare the performance of the methods by quantifying the numbers of true/false outliers detected in different scenarios. In simulations we vary the outcome prevalence, the degree of clustering, the number of hospitals and the number of patients per hospital. We discuss the use of two possible types of standard errors for the estimated random effects in a random effects model: the diagnostic and the comparative. The diagnostic standard SE has been suggested as the appropriate one for outlier detection. We also discuss the graphical presentation of the results from each method.

Results and Conclusions: There was no method that dominated the others in terms of performance across all scenarios. Random effects models fitted at the patient level are appealing because they inherently account for the clustering present in the data, they can easily accommodate risk-adjustment and they make better use of information by using patient-level data. Funnel plot methods are easy to apply and have a graphical representation that is appealing to a clinical audience. Analysts should be cautious in choosing the type of standard error for outlier detection. Incorrectly using the comparative standard error which is the default standard error provided in most software packages can result in substantial overestimation of the number of outliers.

[RP3.20]

Dealing with missing outcome and covariates in randomized trials: When and when not to use simple methods

Mutamba Tonton Kayembe, Shahab Jolani, Frans E. S. Tan, Gerard J. P . van Breukelen

Department of Methodology and Statistics, Maastricht University, Netherlands

Context: For linear regression analysis, there is an extensive literature on dealing with missing data in randomized controlled trials (RCTs), when there is missingness in the covariates only, or in the outcome only. But the more general case of missingness in both the covariates and the outcome has only received scant attention in the literature, by means of a limited scope of case-studies and/or of formal proofs, with no systematic simulation study. Further, for this case, the literature has only investigated the performance of a limited number of missing data methods under a small range of missingness mechanisms (ignoring several others that may be of interest in practice), with little interest on simple methods such as complete-case analysis (CCA) and mean imputation (ME). This leaves some important gaps in our knowledge on this topic.

Objective: In this presentation, we will compare the performance of multiple imputation (MI) and maximum likelihood (ML)-based methods with simple alternative methods, such as CCA and ME, for handling missing data on both the covariates and the outcome in RCTs under a wider range of missingness mechanisms.

Methods: First we formally described the various methods for handling missingness and intuitively discussed what should be their performance in terms of bias and precision of the treatment effect estimate. We then performed a simulation study and compared all the proposed methods across various missingness mechanisms, based on various performance criteria. Finally, a real case study of RCT data on chronic low back pain was utilized to illustrate the implementation of the methods.

Results: The results of the simulation study suggest that no missing data method is universally superior, but ML followed by MI is better under most missingness scenarios. CCA performs similarly to ML and MI when the missingness rate is negligible. ME is not suitable, as it always results in serious under-coverage. This was confirmed by the results from the application to the real data example too.

Conclusions: In practice, methods like MI and ML are recommended, because they generally perform better than all other methods. CCA can be used only if the missingness rate is negligible.

[RP3.21]

Learning Bayesian networks and regression for studying the relationship between scabies and rheumatic fever

Simon Thornley

Section of Epidemiology and Biostatistics, The University of Auckland, New Zealand

Background: Use of antibiotics to treat throat infections has proved disappointing as a preventive measure for acute rheumatic fever in New Zealand children. An alternative hypothesis is that scabies is the primary cause of acute rheumatic fever. This has been raised, since scabies is prevalent in Pacific island nations, who are also the principal ethnic group affected by rheumatic fever in New Zealand.

Objective: To examine statistical evidence the statistical evidence of association between rates of treatment for scabies and incidence of acute rheumatic fever at a neighbourhood level, using healthcare data.

Methods: A cross-sectional study of prescribing for scabies and incidence of acute rheumatic fever were obtained from dispensing and disease notification records in Auckland, by neighbourhood. Potential confounding factors included the ethnic make-up and the socioeconomic status of each neighbourhood.

Analysis consisted of traditional quasi-Poisson regression, where the outcome variable was the incidence rate of acute rheumatic fever and the independent variables were scabies treatment rates, ethnic composition and socioeconomic status of each neighbourhood. Learning Bayesian network analysis was conducted as an adjunct to the regression modelling.

Results: A total of 413 neighbourhoods were included. The incidence of ARF varied between 0 and 102 per 100 000 people per year. High correlation was present between all independent variables. The annual incidence of dispensing of permethrin varied between 0 and 3201 per 100 000 people per year. In an adjusted quasi-Poisson model, permethrin dispensing rates were strongly associated with acute rheumatic fever incidence, with a change from the 16th to the 84th centile associated with a 16.5-fold increase in incidence (95% confidence interval: 3.82–71.6).

The learning Bayesian network analysis showed strong conditional dependence between variables, with acute rheumatic fever dependent on both scabies treatment and Pacific ethnicity. Maori and socioeconomic status were 'explained' as influences on scabies treatment, but not directly on acute rheumatic fever.

Conclusion: The learning Bayesian network analysis complemented the regression findings, indicating that scabies is a likely biological mediator of the strong association between ethnicity, socioeconomic status and acute rheumatic fever. The Bayesian network analysis provided a plausible causal explanation for the findings of the regression model.

[RP3.22]

Patient-centric outcomes in clinical trials: The Desirability of Outcome Ranking

Diane Uschner, Toshimitsu Hamasaki, Qing Pan, Weixiao Dai, Michelle Earley, Scott Evans

Biostatistics Center, The George Washington University, United States

In clinical trials, investigators are often interested in studying the effects of interventions on more than one clinical outcome. For example in antibiotic trials, treatment strategies should have a favorable benefit:risk profile i.e., not only have high efficacy in curing the infection, but also an acceptable adverse event profile. While multivariate methods and multiple testing methods have been used to study multiple endpoints in clinical trials, these methods usually operate from a population-based perspective in contrast to a patient-centric perspective. Furthermore, standard approaches do not incorporate associations between outcomes of interest, cumulative outcomes, or competing risks for individual patients. Since efficacy and safety analyses are conducted on different analysis populations, the population to which these benefit:risk analyses apply, is unclear.

The desirability of outcome ranking methodology proposed by Evans et al. (2015) takes a patient-level approach. The method consists of ranking patients based on the desirability of their overall experience encompassing individual component outcomes. The probability of observing a lower rank in the experimental compared to the control group provides an intuitive measure of the superiority of an experimental treatment i.e., the probability of a better overall result in one intervention compared to another. The DOOR probability is a Mann-Whitney type statistic, for which several confidence intervals have been proposed in the literature. We conducted a simulation study to compare the coverage probabilities of different methodologies for confidence interval construction for the DOOR probability. We present a sample size calculation for a trial of evaluating diagnostic stewardship interventions using a DOOR outcome.

The results of the simulation study indicate that the bootstrap confidence interval has good coverage probability, also for small and medium sample sizes. With increasing sample size, the coverage probabilities of confidence intervals for different methodologies for confidence interval construction are similar. The DOOR outcome can reduce the sample size needed in a clinical trial compared to standard methods. The methods are implemented in the R package DOOR.

The DOOR methodology has good statistical properties, attractive interpretability, illuminates correlations between and the cumulative nature of different outcomes, and can lead to more efficient clinical trials.

[RP3.23]

Analysis of complex survey data: A comparison between SURVEY LOGISTIC and GLIMMIX with random effects

Tanvir Ahammed, Jamal Uddin

Department of Statistics, Shahjalal University of Science and Technology, Bangladesh

Background: Complex survey or Hierarchical data are common in health research. Most of the times information is likely to be observed on nested units at multiple levels. In this situation applying the classical logistic regression model may not valid, as it doesn't allow the complex design of the survey. Some researchers use SURVEY LOGISTIC in SAS. However, it cannot add a random effect in the model. The GLIMMIX procedure in SAS with random effects may answer this question.

Objectives: We aimed to compare between Survey Logistic and GLIMMIX procedure with the application of the Global Youth Tobacco Survey (GYTS) survey datasets.

Method: We used two nationally representative complex surveys, i.e. Global Youth Tobacco Survey (GYTS) conducted in Bangladesh (the year 2007 and 2013). GYTS was a school-based survey of students targeting adolescents age 13–15 years developed by the World Health Organization and the Centres for Disease Control and Prevention. Data were analyzed using logistic regression models using SURVEY LOGISTIC and GLIMMIX procedure in SAS.

Results: Our study showed that, AIC is smaller in the GLIMMIX procedure for both the survey datasets. Moreover, for 2007 GYTS survey data, amount of pocket money received, exposure to passive smoking, receiving free tobacco products were significant factors in SURVEY LOGISTIC procedure while in GLIMMIX procedure three additional factors (age, gender, in favor of banning smoking at public places) became significant too. For the 2013 GYTS survey gender, in favor of banning smoking at public places, exposure to passive smoking, the average costs of pack cigarettes were significant factors in both SURVEY LOGISTIC and GLIMMIX procedure. But age was significant in SURVEY LOGISTIC and receiving free tobacco products was significant factors in GLIMMIX procedure.

Conclusions: As GLIMMIX procedure can add a random effect, it is recommended to use in analysis of complex survey data.

[RP3.24]

A simple heuristics for detection of age-related increase of risk of rare congenital anomalies

Jan Klaschka[1], Marek Maly[1], Antonin Sipek[2]

[1] *Department of Statistical Modelling, Institute of Computer Science of the Czech Academy of Sciences, Czech Republic*
[2] *Department of Medical Genetics, Thomayer Hospital, Prague, Czech Republic*

Background: Both high and very low maternal ages are known to increase the risk of congenital anomalies. The increase size and time of onset depend on anomaly kind. Modelling of the relationship between maternal age and anomaly risk in the Czech population is among the focuses of our research. Methods like joinpoint regression are often used for such tasks, but they may, by our experience, fail when the anomaly is rare, and/or the risk increase takes place close to the end of age scale.

Objective: The present work aims at designing a simple heuristic method of analysis of relationship between anomaly risk and maternal age that could stand in for commonly used methods when these fail.

Method: Our method (inspired by tree-based methods) compares the anomaly risk in pairs of opposite upper and lower age scale tails, i. e. in intervals from a cutpoint to the upper and lower scale ends. For each cutpoint (usually a year), anomaly frequencies are compared using relative risk (RR) and Fisher test. Risk increase in upper tail T is suspect when RR exceeds 2 and unadjusted Fisher test is significant. Increased risk in T is confirmed when suspect, and T is included in a tail with significant Bonferroni-adjusted Fisher test. Evaluation of lower tails is analogous.

Results: As an example, our method was applied to the Czech epidemiological data on anencephaly in 1964 – 1991 and 1992 - 2016 periods. In the first period, a verified risk increase from the age of 41 was found. In the second period, there was a suspect risk increase from 42 years and a verified increase for maternal age of 18 or lower. Joinpoint regression (R package segmented) suffered from convergence problems and dependence on initial values on the same data. Exploratory CART class probability trees (Breiman et al., Classification and Regression Trees, Chapman & Hall, Boca Raton 1984, ch. 4.6) yielded (via regression trees in R package rpart) splits close to those found by our method.

Conclusions: Results obtained so far appear medically plausible, but more experience is needed.

Supported by Czech Health Research Council grant No. 17-29622A.

[RP3.25]

Re-evaluation of bootstrap-based optimism correction methods in multivariable clinical prediction models

Katsuhiro Iba[1], Tomohiro Shinozaki[2], Kazushi Maruo[3], Hisashi Noma[4]

[1] *The Graduate University for Advanced Studies, Japan*
[2] *Tokyo University of Science, Japan*
[3] *University of Tsukuba, Japan*
[4] *The Institute of Statistical Mathematics, Japan*

Background: Multivariable prediction models are important statistical tools for providing synthetic diagnosis and prognostic algorithms based on patients' multiple characteristics. Their apparent measures for predictive accuracy usually have overestimation biases (known as 'optimism') relative to the actual performances for external populations. Existing statistical evidence and guidelines suggest that three bootstrap-based bias correction methods are preferable in practice, namely Harrell's bias correction and the .632 and .632+ estimators. Although Harrell's method has been widely adopted in clinical studies, simulation-based evidence indicates that the .632+ estimator may perform better than the other two methods. However, these methods' actual comparative effectiveness is still unclear due to limited numerical evidence.

Objectives: To evaluate the comparative effectiveness of bootstrap-based optimism correction methods in the development of multivariable clinical prediction models.

Methods: We conducted extensive simulation studies to compare the effectiveness of these three bootstrapping methods, particularly using various model building strategies: conventional logistic regression, stepwise variable selections, Firth's penalized likelihood method, ridge, lasso, and elastic-net regression. We generated the simulation data based on the GUSTO-I trial dataset and considered how event per variable, event fraction, number of candidate predictors, and the regression coefficients of the predictors impacted the performances. Biases and root mean squared errors (RMSE) for assessing the internal validity of C-statistics were evaluated.

Results: Under relatively large sample settings (roughly, events per variable $\geq$ 10), the three bootstrap-based methods were comparable and performed well. However, all three methods had biases under small sample settings, and the directions and sizes of biases were inconsistent. In general, Harrell's and .632 methods had overestimation biases when event fraction become lager. Besides, .632+ method had a slight underestimation bias when event fraction was very low. Although the bias of the .632+ estimator was relatively small, its RMSE was comparable or sometimes larger than those of the other two methods, especially for the regularized estimation methods.

Conclusions: In general, the three bootstrap estimators were comparable, but the .632+ estimator performed relatively well under small sample settings. However, the .632+ estimator can perform worse under small sample settings when the regularized estimation methods are adopted.

[RP3.26]

Properties of Adjusted Per-Protocol Effect Estimators to Address Treatment Non-Adherence in Pragmatic Trials

Lucy Mosquera[1], M. Ehsan Karim[2]

[1] Department of Statistics, University of British Columbia, Canada
[2] School of Population and Public Health, University of British Columbia, Canada

Background: Pragmatic trials aim to assess the practical effect of treatment strategies. Participants are randomly assigned to treatment strategies at baseline, but patients may not adhere to their assigned treatments during follow-up. Non-adherence may confound the estimation of the per-protocol (PP) treatment effect, biasing conventional estimations methods towards a null treatment effect.

Aim: In this work, we focus on sustained treatment regimes, where non-adherence is quantified at regular intervals via follow-up visits. In the presence of non-adherence, we aim to compare the conventionally used naïve analyses (i.e., intention-to-treat (ITT) and naive per-protocol (PP)) with adjusted-PP effect estimates. We estimated two types of adjusted-PP effect estimates: unstabilized inverse probability weighted (IPW) PP effects based on cumulative survival estimates, and stabilized IPW based on pooled logistic regression model survival estimates. We aim to contrast these methods for a variety of treatment effects, measurement schedules, and rates of non-adherence.

Methods: We have extended Young and colleagues' simulation framework in a comprehensive way to generate two-armed pragmatic trial data with a baseline covariate and two post-randomization time-varying covariates. A variety of realistic scenarios were generated, including varying treatment effects (null and non-null effects), measurement schedules, and rates of non-adherence. For each scenario, the statistical properties of each estimator were calculated, including bias, coverage, mean squared error and power.

Results: In the presence of non-adherence, our simulations show that both ITT and IPW-PP estimates are unbiased for null treatment effects. For non-null treatment effects, only the IPW-PP estimates were reasonably unbiased. Stabilized IPW-PP methods, in general, were associated with better statistical properties (coverage, mean squared error). When receipt of treatment and post-randomization covariate measurements were infrequent, or the rate of non-adherence is quite high, IPW-PP estimates may be slightly biased, but are consistently less biased than naïve methods. Across all scenarios, naïve-PP estimates always fail to recover the true effect of treatment, highlighting the need for adjusting using pre- and post-randomization covariates.

Conclusion: This study demonstrates the necessity of measuring and utilizing both pre-and-post-randomization covariates in the design and analysis stages to reduce bias in the estimation of treatment effects in the presence of non-adherence.

[RP3.27]

How to effectively recruit patients for adaptive enrichment clinical trials

Ryuji Uozumi[1], Shinjo Yada[2], Atsushi Kawaguchi[3]

[1] Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine, Japan
[2] Biostatistics Department I, Data Science Division, A2 Healthcare Corporation, Japan
[3] Faculty of Medicine, Saga University, Japan

Background: Adaptive enrichment designs for clinical trials have great potential for the development of targeted therapies. In this context, an interim analysis is conducted to identify whether subsequent recruitment includes the overall population or is restricted to the subpopulation. To the best of our knowledge, however, patient recruitment within the adaptive enrichment design setting typically needs to be halted pending an interim decision, and is resumed after the interim decision has been made. Consequently, adaptive enrichment designs increase the total trial period owing to the stoppage in patient recruitment to make interim decisions. This is a major drawback; it results in delays in the submission of clinical trial reports and the appearance of drugs on the market.

Objective: We explore three types of patient recruitment strategy for the development of targeted therapies based on the adaptive enrichment design.

Methods: We consider two alternative methods which provide an option to continue recruiting patients from the overall population or only from the subpopulation even during the interim decision period. A simulation study was performed to investigate the operating characteristics by comparing an adaptive enrichment design using the recruitment methods with a non-enriched design.

Results: In settings in which the recruitment period is expected to be longer than the follow-up period, the adaptive enrichment design with continued recruitment from the overall population or only from the subpopulation even during the interim decision period conferred a major advantage, since the total trial period did not differ substantially from that of trials employing the non-enriched design. By contrast, the non-enriched design should be used in settings in which the follow-up period is expected to be longer than the recruitment period, since the total trial period was notably shorter than that of the adaptive enrichment design.

Conclusions: Adaptive enrichment designs that entail continued recruitment methods are beneficial owing to the shorter total trial period than expected in settings in which the recruitment period is expected to be longer than the follow-up period and the biomarker-positive population is promising.

[RP3.28]

Review of statistical methods to address treatment nonadherence in the pragmatic trial context

Belal Hossain, Mohammad Ehsanul Karim

School of Population and Public Health, University of British Columbia, Canada

Background: Pragmatic trials offer practical means of obtaining real-world evidence (RWE) to help improve decision-making in comparative effectiveness settings. Analyzing these trial data are particularly challenging in the presence of nonadherence (e.g., treatment discontinuation/switching). A wide variety of statistical methods are proposed to address nonadherence issue for randomized clinical trials, but most of these conventional methods are not suitable for pragmatic trial settings, producing biased estimates. Also, not all methods can deal with all types of nonadherence (initiation, implementation, and persistence), and all nonadherence patterns (random, explainable non-random, and non-random).

Objective: To review, compare/contrast new and conventional statistical methods to deal with nonadherence in pragmatic trial settings.

Methods: We conduct a methodological review in the literature to list the appropriate methods suitable for pragmatic trial scenarios, compare and contrast the assumptions, advantages and limitations.

Results: Naïve methods, such as intention-to-treat, naïve per-protocol, and as-treated are highly prone to give biased estimates in the presence of nonadherence, although these are the most popular methods to date. Advanced RWE methods, e.g., inverse probability censoring weighting (IPCW), inverse probability of adherence weighting (IPAW), g-estimation of structural nested failure time model (SNFTM),rank-preserving structural failure time model (RPSFTM), and instrumental variable (IV)-based models (two-stage least square, two-stage residual inclusion, and causal-bound) can be used to address various types of nonadherence in pragmatic trial settings. The inverse probability weighting-based methods (IPCW, IPAW) and SNFTM attempt to eliminate bias due to pre- and post-randomization (e.g., time-varying) prognostic factors, and can be used to address all types of nonadherence, but when adherence only depends on the measured factors (i.e., explainable non-random adherence). The IV-based methods can be used to address all types of nonadherence when adherence depends on both measured and unmeasured factors (i.e., non-random adherence).

Conclusions: Specially designed RWE methods have several advantages but often rely on strong assumptions. The IV-based methods produce biased estimates when the instrument is weak. IPCW, IPAW, and RPSFTM require longitudinal (post-randomization measurements of) data that predict adherence should be available, and careful data collection planning is necessary to apply these methods.

[RP3.29]

Challenges in integrating the different arms of the immune response against Mycobacterium tuberculosis

Tessa Lloyd[1], Pia Steigler[2], Francesca Little[1], Elisa Nemes[2]

[1] *Department of Biostatistics, University of Cape Town, South Africa*
[2] *Department of Pathology, University of Cape Town, South Africa*

Background: Tuberculosis (TB) is an airborne bacterial disease that is the leading cause of mortality due to a single infectious agent worldwide. To advance our understanding of the immune response against *Mycobacterium tuberculosis*, the causative agent of TB, we integrated three datasets each containing information on different arms of the immune response, namely adaptive and innate immunity, and "innaptive" immunity, which shares features of both adaptive and innate immunity and bridges between both arms.

Objectives: To integrate biological data such that interpretable, unbiased and valid results are yielded. The first challenge involved overcoming the inherent heterogeneity in the data that arose from using technological platforms, which yielded data on different scales. The second challenge was the high dimensionality of the data, where the number of predictors was significantly larger than the number of samples. Third, the complete dataset was not available for all individuals, and therefore imputation of missing data was required to account for the potentially different contributions of individual data types in final prediction models.

Methods: Several normalization techniques (variance stabilizing, pareto scaling, classical standardization and min-max normalization) and imputation methods (missForest, KNN, column median and Multiple Factor Analysis (MFA) imputation) were considered and their performances compared. In addition we used COMPASS (Lin et al., 2015) to identify biologically relevant cell subsets and pre-select variables to reduce dimensionality. For variable selection in predictive models, we used an Elastic Net model with multiple tuning parameters for each different data set (Liu et al., 2018).

Results: We found that variance stabilizing (VAST) standardization with a min-max normalization performed the best to transform all variables in the same range. For imputation we found that Multiple Factor Analysis Imputation (MFA) outperformed the other imputation methods. Lastly, COMPASS identified 270 out of 1217 biologically irrelevant cell subsets that were removed.

Conclusions: In order to integrate the data, several data pre-processing steps were required, increasing the risk of potentially affecting the results of our data integrated model. Therefore we emphasize the importance of testing different approaches to identify the best suited for a given dataset.

[RP3.30]

Copula-based modelling and analysis of semi-competing risk data, with application to renal transplant

Lexy Sorrell[1], Yinghui Wei[1], Malgorzata Wojtys[1], Peter Rowe[2]

[1] *Centre for Mathematical Sciences, University of Plymouth, United Kingdom*
[2] *Department of Nephrology, University Hospitals Plymouth NHS Trust, United Kingdom*

Semi-competing risk, time-to-event, bivariate endpoints occur when one terminal endpoint can censor another, non-terminal endpoint, but not vice-versa. These endpoints may be associated with one another, as both are occurring on the same individual. These individual events and the association between them may be influenced by covariates. Traditional methods of correlation analysis cannot be used directly as censoring can occur at one or both endpoints. Our aim is to estimate the correlation between the survival endpoints, examine the effects of covariates on these potential correlations and hazard ratios for both the terminal and non-terminal events.

We use conditional copulas to estimate the hazard ratios of covariates of the non-terminal and terminal events, along with the covariate effects on the association between them. We use the Normal, Clayton, Frank and Gumbel copulas to model the association structures between the survival endpoints. We perform simulation studies to compare the conditional copula models, with and without covariates included in the association parameter and the Cox proportional hazards model. We apply the methods to estimate the effects of covariates on the semi-competing risk endpoints of graft failure and death following renal transplant.

We estimate Spearman's rank correlation coefficient of graft failure and death following renal transplant to be moderately strong and positive using the Frank copula. The older age group and the deceased donor group have a higher risk of graft failure and death, and the female group a lower risk of death. The correlation between events is stronger for individuals in the higher age group and the living donor group.

The copula-based methods provide good estimation of the correlation between semi-competing events and the hazard ratios of both events. In this case, our methods perform better than the Cox model when estimating the hazard ratio of the non-terminal event, with coverage probability closer to 95% and a lower bias. The estimates of the hazard ratios are improved by the inclusion of the covariates in the correlation estimate.

[RP3.31]

Generalized Pairwise Comparison statistics: comparison of non-parametric and UMVUE estimators

Johan Verbeeck[1], Vaiva Deltuvaite-Thomas[1], Ben Berckmoes[3], Tomasz Burzykowski[1], Marc Aerts[1], Olivier Thas[1], Marc Buyse[2], Geert Molenberghs[1]

[1] *Data Science Institute, UHasselt, Belgium*
[2] *IDDI, Belgium*
[3] *Department of Mathematics, University of Antwerp, Belgium*

Background: In reliability theory, diagnostic accuracy, and clinical trials, the quantity $P(X > Y) + 1/2P(X = Y)$, also known as the Probabilistic Index (PI), is a common treatment effect measure when comparing two groups of observations. The PI is a member of the Generalized Pairwise Comparison (GPC) statistics and related to the other GPC members, the net benefit, win ratio, and win odds or success odds. Parametric estimation of the PI has received a lot of attention in the past 40 years, with the formulation of the Uniformly Minimum-Variance Unbiased Estimator (UMVUE) for many distributions. However, the non-parametric Mann-Whitney estimator of the PI, is also known to be UMVUE in some situations.

Objective(s): To understand the statistical properties of the parametric UMVUE and the nonparametric Mann–Whitney estimator of the GPC statistics.

Method(s): Compare, theoretically where possible and by simulations, the efficiency and the basic statistical properties, such as unbiasedness, sufficiency, and completeness of the parametric UMVUE and the nonparametric Mann–Whitney estimator for the PI.

Results: We show that the Mann–Whitney estimator is always an unbiased estimator of the PI with univariate, completely observed outcome data, while the parametric UMVUE is not when the distribution is misspecified. Additionally, the Mann-Whitney estimator is the UMVUE when observations belong to an unrestricted family of distributions. When observations come from a more restrictive family of distributions, the loss in efficiency for the non-parametric estimator is limited in realistic clinical scenarios, but is substantial in reliability and diagnostic accuracy settings. These results extend to the PI's linear transformation, net benefit, but not to the win ratio and win odds.

Conclusions: The non-parametric Mann-Whitney estimator is a simple to use and a reliable estimator for the PI and net benefit in realistic clinical scenarios, but is less efficient in reliability and diagnostic accuracy settings.

[RP3.32]

Prediction of vascular ageing based on smartphone acquired PPG signals

Lorenzo Dall'Olio[1], Nico Curti[2], Daniel Remondini[1], Yosef Safi Harb[3], Folkert W. Asselbergs[4], Gastone Castellani[2], Hae-Won Uh[5]

[1] *Department of Physics and Astronomy, University of Bologna, Italy*
[2] *Department of Specialised, Diagnostic and Experimental Medicine, University of Bologna, Italy*
[3] *Happitech, Netherlands*
[4] *Department of Cardiology, University Medical Center Utrecht, Netherlands*
[5] *Department of Biostatistics and Research Support, University Medical Center Utrecht, Netherlands*

Background: Photoplethysmography (PPG) measured by smartphone has the potential for a large scale, non-invasive, and easy-to-use screening tool. Ageing is a major non-reversible risk factor for cardiovascular disease. Vascular ageing is linked to increased arterial stiffness, which can be measured by PPG. Our motivating crowd-sourced data contains PPG recordings in large samples (ca. 10,000) and free access to raw PPG signals via the Happitech app, coming from the Heart for Heart (H4H) initiative. On the downside, these PPG signals can be very noisy compared to those in clinical settings, due to the uninstructed and not monitored PPG captures. On the other hand, to monitor cardiovascular status such as arterial blood pressure or arterial stiffness, the non-invasive PPG technique can be very useful.

Objectives: We investigate the feasibility of using PPG to predict healthy vascular ageing (HVA) based on two approaches: machine learning (ML) and deep learning (DL).

Method: We performed data preprocessing including detrending, demodulating, and denoising on the raw PPG signals. For ML, ridge penalized regression has been applied to 38 features extracted from PPG, whereas for DL several convolutional neural networks (CNNs) have been applied to the whole PPG signals as input to predict HVA, skipping the feature extraction step. The prediction performance of the final models is validated on the dataset held back from the training and testing of the model, such that AUC presents an unbiased performance measure for comparing final models.

Results: The analysis has been conducted using the crowd-sourced Heart for Heart data. The prediction performance of ML using two features (AUC of 94.7%) -- the 'a' wave of the second derivative PPG and 'tpr', including four covariates, sex, height, weight, and smoking -- was similar to that of the best performing CNN, 12-layer ResNet (AUC of 95.3%).

Conclusions: Without having the heavy computational cost of DL, ML might be advantageous in finding potential biomarkers for HVA prediction. The whole workflow of the procedure is clearly described, and open software has been made available to facilitate replication of the results.

[RP3.33]

Comparison of Imputation Methods for Multivariate Longitudinal data with Mixed-type Incomplete Variables

Yi Cao[1], Heather Allore[2], Roee Gutman[1]

[1] *Department of Biostatistics, Brown University, United States*
[2] *School of Medicine, Yale University, United States*

Context: When estimating relationships between multiple patient-centered outcomes of mixed type, methods for handling multivariate longitudinal missingness are required. Multiple imputation is one approach to fill-in missing data. Joint modeling (JM) and fully conditional specification (FCS) are two possible strategies. JM fits a multivariate distribution for the entire set of variables, but it may be complex to define and implement. FCS imputes data variable by variable from a set of conditional distributions, which is easier to define and implement, but it suffers from theoretical deficiencies. For longitudinal data, some imputation methods require wide-format data, which enjoys gains computationally. However, they are not compatible with the model of interests. Recently developed methods that apply multilevel modeling, show advantages of accounting for the clustered structure, but sometimes are too costly in computation to implement in practice.

Objective: This study aims to present a comprehensive comparison of currently available imputation methods for mixed-type incomplete longitudinal data, and provide advice to researchers on suitable methods to handle different missing data scenarios.

Method: We simulate datasets from the National Health and Aging Trends Study and apply ten imputation methods for imputing a mixture of normally-distributed, binary, and count variables. We assess the performance by modeling the associations between multiple health outcomes and chronic conditions through both univariate and multivariate generalized linear mixed models.

Results and Conclusion: Simpler methods such as Standard chained equation method without multilevel modeling, provide comparable performance compared to the complicated techniques with multilevel modeling. Overall, the best performance is presented in the linear mixed model with latent normal variable under the chained equations framework. The methods derived from generalized linear mixed models, though seem attractive theoretically, they led to biased random-effects estimates of the substantive analysis.

[RP3.34]

Preserving data privacy when using multi-site data to estimate individualized treatment rules

Coraline Danieli[1], Erica Moodie[2]

[1] *Department of Epidemiology, Biostatistics and Occupational Health, Research Institute of McGill University Health Center, Canada*
[2] *Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Canada*

Precision medicine is a rapidly expanding area of health research wherein patient level information is used to inform treatment decisions. Dynamic treatment regimens (DTRs) provide a statistical framework to formalize the individualization of treatment decisions that characterize personalized management plans. Numerous methods have been proposed to estimate DTRs that optimize expected patient outcomes, many of which have desirable properties such as robustness to model misspecification. However, while individual data are essential in this context, there may be concerns about data confidentiality, particularly in multi-centre studies where data are typically shared externally. To address this issue, we considered two approaches to privacy preservation: (i) data pooling and (ii) weighted least squares with data aggregation. These approaches were combined with the doubly robust yet user-friendly method of dynamic weighted ordinary least squares to estimate individualized treatment rules. In simulations, we evaluated the performance of the methods in estimating the parameters of the decision rule under different assumptions concerning (i) the distribution of the subject-specific covariates in each centre, (ii) the treatment variable type, (iii) the relationship between the treatment and the subject-specific covariate, (iv) the pool size when aggregating the data, and (v) model misspecification. The results demonstrate that double robustness is not maintained in data pooling setting and so which can result in bias, whereas the weighted least squares with data aggregation approach provides good performance. We illustrated the methods on an example of warfarin dosing using data from the International Warfarin Consortium.

[RP3.35]

Machine learning for variable selection with high dimensional data and time-to-event outcomes

Federico Ambrogi

Department of Clinical Sciences and Community Health, University of Milan, Italy

Background: Variable selection is a long debated subject. Guidelines are hard to be set for standard statistical applications and even more problematic with high dimensional data (HDD).

In such circumstances the instability of the selected model is a matter of fact, while the distinction between to explain (estimate the true sparsity pattern according to Wasserman and Roed er, 2009) or to predict is always subtle. In high dimensional settings, the problem of variable selection is crucial.

When considering the use of machine learning techniques it is also very important to uncover the black box in order for the regression model to be trusted and avoid unexpected biases.

Objective: The goal of this work is to evaluate some machine learning methods for variable selection in HDD and censored data with special emphasis on methods of post-selection inference.

Methods: Simulations were used in order to evaluate the different methods using parallel computations. We adopted the multi sample splitting procedure (Dezeure et al., 2015) in order to evaluate coefficient estimates, p-values and coverages. Moreover the ability of each method to select the important variables in terms of sensitivity and specificity was evaluated. Basically the model selected using Lasso, Boosting and Random Rorest was refit using validation data with a Cox regression model. This strategy is very general and can be applied to any kind of variable selection method provided it returns a sufficiently low number of selected covariates. Some low dimensional settings were also explored where traditional stepwise selection methods were used for comparison.

Results and Conclusions: The specificity of the metods is good while sensitivity is more variable. In general machine learning methods do not seem to outperform traditional methods. Random forest are very flexible but difficult to be tuned in proper way.

[RP3.36]

The impact of competing risks on individualised prediction of hip replacement outcomes using registry data

Alana Cuthbert[1], Lynne Giles[2], Gary Glonek[3], Stephen Graves[4], Nicole Pratt[1]

[1] *Quality Use of Medicines and Pharmacy Research Centre, Clinical and Health Sciences, University of South Australia, Australia*
[2] *School of Public Health, The University of Adelaide, Australia*
[3] *School of Mathematical Sciences, The University of Adelaide, Australia*
[4] *Australian Orthopaedic Association National Joint Replacement Registry, Australia*

Background: There is increasing interest in developing models that predict individualised patient outcomes following total hip replacement (THR) surgery, but outcomes such as implant failure are often precluded by patient death. Competing risks models are recommended in this scenario, but a common approach is to instead censor competing events and using standard survival models (e.g. Cox regression). Ignoring competing risks is well known to overestimate the absolute event risk for non-parametric population-level estimates, but the impact when predicting individualised risk following THR is unclear.

Methods: The predictive performance of a standard Cox model with censored death was compared to two competing risks approaches: the cause-specific Cox (CSC) and Fine-Gray models. Models were trained to predict peri-prosthetic fracture following THR using data on 274,618 surgeries recorded in Australia's Joint Replacement Registry. The performance of each model at 10 years was compared using c-indexes and calibration plots stratified by age. Predicted 10-year risks from each model were also compared in three hypothetical patients with different risk profiles, to determine whether differences in population-level performance metrics would translate into a meaningful difference for individual patients.

Results: The standard Cox, CSC and Fine-Gray models had near identical discrimination (c-index 0.66 [95% CI (0.66, 0.66] for all three models). Calibration plots showed that all models performed similarly in younger and averaged-aged patients, but for older age groups the standard Cox model overestimated the revision risk more than the CSC and Fine-Gray models. For the two hypothetical patients with low and average risk of death, the 10-year predicted risks from the models were very similar. However, a larger difference was observed in the predicted 10-year risks for a hypothetical elderly patient with high mortality risk (3.4% from CSC model versus 5.2% from standard Cox).

Conclusions: Failing to account for competing risks when developing prognostic models for long-term outcomes following THR will result in an overestimate of predicted risk for patients with high risk of mortality. However, given the risk of an adverse event following THR is very low, the magnitude of this overestimate is unlikely to be of consequence for individual patient predictions.

[RP3.37]

Mind the gap: A simulation study on the performance of naive imputation approaches for interval-censored data

Micki Hill[1], Paul Lambert[2], Michael Crowther[2]

[1] *Department of Biostatistics, University of Leicester, United Kingdom*
[2] *Department of Biostatistics, University of Leicester (United Kingdom) & Karolinska Institutet (Sweden)*

Background: Interval-censored data frequently occur but are often not analysed appropriately. A common approach is to impute a single event time and treat it as if it were observed exactly. Simulation studies have predominantly focused on the impact naive imputation has on the hazard ratio, while the impact on other useful estimands, such as survival probabilities, has been overlooked.

Objective: The aim of this simulation study was to investigate the performance of naive imputation on survival probabilities when data are interval-censored.

Methods: Event times were generated from a motivating dataset in breast cancer. Data were simulated from a Weibull distribution with a single binary covariate. Interval censoring was applied by assuming patients followed a regular schedule with a small amount of subject-specific jittering. At the end of follow-up, patients were right-censored. Aspects that were varied in the different scenarios included: shape parameter, amount of right censoring, treatment effect size, interval width and sample size. The estimand of interest was the survival probability in the control group at four evenly spaced time points. The methods compared were: analysis of exactly observed data prior to imputation (reference method), following naive imputation (left, mid or right) and the likelihood based method, which accounts for interval censoring. The performance outcomes considered were: bias, standard errors (empirical and model-based) and coverage.

Results: Bias and coverage were worse for the naive approaches for earlier time points and when the shape parameter was decreased, the interval width was increased and the amount of right censoring was decreased. The naive approaches frequently produced smaller empirical standard errors than the reference method, which resulted in poor coverage. Midpoint imputation was the overall best performing naive approach, while the likelihood based method performed well across all scenarios. Results are presented in nested loop plots.

Conclusions: Failing to account for interval censoring can lead to artificially precise estimates. In addition, it can cause estimates to be biased, especially when hazard rates are rapidly increasing, there is minimal right censoring and interval widths are large.

[RP3.38]

Evaluation of a Piecewise Linear Mixed-effects Model in the Analysis of Randomized Cross-over Trial

Moses Mwangi[1], Geert Molenberghs[1], Geert Verbeke[2], Edmund Njeru Njagi[3], Samuel Mwalili[4]

[1] *I-BioStat, Universiteit Hasselt, Belgium*
[2] *L-BioStat, KU Leuven, Belgium*
[3] *Department of Non-communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine University of London, United Kingdom*
[4] *School of Mathematical Sciences, JomoKenyatta University of Science and Technology, Kenya*

Cross-over designs are commonly used in randomized clinical trials to estimate efficacy of a new treatment with respect to a reference treatment (placebo or standard). The main advantage of using cross-over design over conventional parallel design is its flexibility, where every subject become its own control, thereby reducing confounding effect. Jones & Kenward (2015), discuss in detail more recent developments in the analysis of cross-over trials. We revisit the simple piecewise linear mixed-effects model, proposed by Mwangi et. al, (in press) for its first application in the analysis of cross-over trials. We compared performance of the proposed piecewise linear mixed-effects model with two commonly cited statistical models namely, (1) Grizzle model; and (2) Jones & Kenward model, used in estimation of the treatment effect, in the analysis of randomized cross-over trial. We estimate two performance measurements (mean square error (MSE) and coverage probability) for the three methods, using data simulated from the proposed piecewise linear mixed-effects model. Piecewise linear mixed-effects model yielded lowest MSE estimates compared to Grizzle and Jones & Kenward models for both small (Nobs=20) and large (Nobs=600) sample sizes. It's coverage probability were highest compared to Grizzle and Jones & Kenward models for both small and large sample sizes. A piecewise linear mixed-effects model is a better estimator of treatment effect than its two competing estimators (Grizzle and Jones & Kenward models) in the analysis of cross-over trials. The data generating mechanism used in this paper captures two time periods for a simple 2-Treatments x 2-Periods cross-over design. Its application is extendible to more complex cross-over designs with multiple treatments and periods. In addition, it is important to note that, even for single response models, adding more random effects increases the complexity of the model and thus may be difficult or impossible to fit in some cases.

[RP3.39]

Functional principal component analysis for high dimensional longitudinal omics data

Attila Csala[1], Kim-Anh Lê Cao[2], M. H. Hof[1], A. H. Zwinderman[1]

[1] *Amsterdam University Medical Canters, the Netherlands*
[2] *Integrative Genomics, University of Melbourne, Australia*

Background: The measurement of high dimensional biomedical data is becoming common practice in biomedical sciences due to new technological developments. Data is often measured from different biomolecular domains in the same subjects, and such data is often called collectively as omics data; for example data from the genome, epigenome, transcriptome, proteome or microbiome. These measurements can also be repeated at different time points, which results in data sets with multiple, high dimensional data sources (per biomolecular domains) with possibly repeated measurements (over time points). In this study we investigate how to extend longitudinal latent component based multivariate methods, such as high dimensional functional principal component analysis (fPCA), from analysing one data source to analysing multiple data sources in an integrated approach. We compare the fPCA approach to other latent component based approaches, such as sparse Redundancy Analyses, by applying these methods on both simulated and real omics data.

Method: fPCA is a multivariate latent component based method, that extracts subject specific random intercept and slopes from longitudinal data sources.

Results: Our simulation studies show that it is possible to identify groups of variables of different data sources that have highly correlated time trajectories („important" variables), via variable selection, and disregard variables that are not associated between data sources („noise" variables). It is also possible to extract „important" variable groups that are independent from each other, which results in models of different change patterns in different groups of variables. Additionally we analysed real omics data, including 459,341 epigenetic and 26 air pollution markers, measured repeatedly on 252 subjects.

Conclusions: Through simulation studies and real data analysis, we show that latent variable based models can simultaneously analyse multiple, high dimensional omics sets, with repeated measurements, and provide interpretable results via variable selection.

[RP3.40]

Review of interobserver variability studies in diagnostic imaging

Laura Quinn[1], Alice Sitch[1], Jon Deeks[1], Susan Mallett[1], Riekie De Vet[2], Wieneke Mokkink[2]

[1] *Institute of Applied Health Research, University of Birmingham, United Kingdom*
[2] *Department of Epidemiology and Biostatistics, Amsterdam UMC, Vrije Universiteit, Netherlands*

Background: Previously only one systematic review has been conducted of interobserver variability studies in diagnostic imaging [1]. However, only articles published from 2011-2012 in four radiological journals were included. This review identified poor reporting and provided a snapshot of design and methodological issues in these studies. Reporting guidelines [2] have been introduced since then, although no recent systematic reviews have been completed.

A key aspect of designing high quality interobserver studies is understanding the potential sources of bias, and making studies applicable to clinical practice. A Risk of Bias tool for interobserver variability studies is in development by the COSMIN (Consensus-based Standards for the selection of health Measurement Instruments) initiative and will be used in this review.

Aim: To provide an overview of current literature based on study design, patient population, variability measurements (e.g. Kappa, ICC), risk of bias, presentation and incorporation of results. This review will inform future methodological research.

Methods: The databases PubMed and Embase were searched for eligible studies, restricted to primary studies measuring interobserver variability in imaging test results for diagnosing conditions in patients. A predefined, piloted data extraction sheet including standards from the COSMIN risk of bias tool is being used for the assessment of these studies.

Results: The initial search from January 2018 to January 2020 identified 298 records. 205 records were included from title and abstract screening, 38 of these records were conference abstracts. The majority of records had sample sizes of less than 100 patients and a small number of observers, the main analysis methods used were ICC and Kappa statistics. The review of the published literature is currently being completed and will give an overview of the methods and risk of bias in these studies.

Conclusion: This review will be used to inform future methodological research in interobserver variability studies in diagnostic imaging.

References:

1. *Farzin, B., et al., Agreement studies in radiology research. Diagnostic and interventional imaging, 2017. 98(3): p. 227-233.*
2. *Kottner, J., et al., Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. International journal of nursing studies, 2011. 48(6): p. 661-671.*

Authors List

Author List

C

Ć

D

E

F

G

H

I

J

K

L

M

N

O

Ó

P

Q

R

S

T

U

V

W

X

Y

Z

Kraków

science centres

academic centres

startups

innovations

new technologies

business

development

The city of endless possibilities