38th Annual Conference of the International Society for Clinical Biostatistics

Vigo, Spain 9-13 July 2017

BOOK OF ABSTRACTS

More Info: **www.iscb2017.info**

ISCB

38th Annual Conference
of the International Society
for Clinical Biostatistics

Vigo, Spain 9-13 July 2017

BOOK OF ABSTRACTS

ISBN: 978-84-697-4704-9

www.iscb2017.info

Table of Contents

Table of Contents

Preamble

The 38th Annual Conference of the ISCB, hosted by the city of Vigo, brings together two plenary talks, eight oral invited sessions, and a number of oral and poster contributed sessions (Monday-Wednesday). Following the tradition of previous ISCB conferences, the Scientific Programme is completed with several pre-conference courses (Sunday) and mini-symposia (Thursday). This Book of Abstracts collects the information related to all these scientific activities.

The abstracts corresponding to the plenary talks, the invited and contributed presentations, the courses and the mini-symposia are gathered up day-to-day. Abstracts of ISCB President's Invited Speaker Stephen Senn and Keynote Speaker Francesca Dominici can be found in Monday's and Wednesday's sections, respectively. We thank very much Stephen and Francesca for their plenary talks, and all the members of the Scientific Programme Committee and the invited speakers and course providers for helping to devise an excellent scientific programme.

The number of contributed presentations raises to 363. They are distributed into 190 oral and 173 poster contributed presentations. All of them were selected by a peer review procedure, and we are happy to recognize the relevance and high-quality of all of them. Presenting authors are underlined throughout this book. Thanks to all the contributed authors!

Last but not least, we would like to express our deepest gratitude to the members of the Local Organising Commitee for making the conference possible.

Jacobo de Uña Álvarez
Chair of the Local Organising Committee

Guadalupe Gómez Melis
Chair of the Scientific Programme Committee

Committees

The **International Society for Clinical Biostatistics (ISCB)** is devoted 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.

ISCB Executive Committee

President
KyungMann Kim US

Vice-President
Vana Sypsa Greece

Treasurer
Zdenek Valenta Czech Republic

Secretary
Geir Egil Eide Norway

Past-President, Newsletter Editor
David W. Warne Switzerland

Chris Weir UK
Hein Putter The Netherlands
Katherine Lee Australia
Nadine Binder Germany
Stanislav Katina Czech Republic
Thomas Jaki UK
Tomasz Burzykowski Belgium
Toshiro Tango Japan

Scientific Programme Committee

Chair of Scientific Programme Committee
Guadalupe Gómez Melis Spain

Chair of Local Organising Committee
Jacobo de Uña Álvarez Spain

Carmen Armero Spain
Frank Bretz Switzerland
Carmen Cadarso Spain
M. Luz Calle Spain
Marie Davidian US
María Durbán Reguera Spain
Ronald Geskus The Netherlands
Els Goetghebeur Belgium
Jeanine Houwing Duistermaat UK
Rosa Lamarca Spain
Ian McGowan US
Ana Luisa Papoila Portugal
David Price UK
Geraldine Rauch Germany
María Xosé Rodríguez Álvarez Spain
Montse Rué Spain
Martin Schumacher Germany
Armando Teixeira Pinto Australia
Lee-Jen Wei US

Local Organising Committee

Chair
Jacobo de Uña Álvarez (SPC) University of Vigo, Spain

Secretary
Juan Carlos Pardo Fernández University of Vigo, Spain

Treasurer
Javier Roca Pardiñas University of Vigo, Spain

Public Relations
María Xosé Rodríguez Álvarez (SPC) Basque Center for Applied Mathematics BCAM IKERBASQUE, Basque Foundation for Science, Spain

Public Relations
Mar Rodríguez Girondo Leiden University Medical Center, The Netherlands

Sponsoring Support
María Jesús Longa Sayáns University of Vigo, Spain

Support Staff
M. Carmen Iglesias Pérez University of Vigo, Spain

Support Staff
Carla M. Moreira Institute of Public Health, University of Porto, Portugal

PhD Student
Marta Cousido Rocha University of Vigo, Spain

Programme Overview

From	To	Sunday 9	Monday 10	Tuesday 11	Wednesday 12	Thursday 13
08.00	08.30	Registration	Registration	Registration	Registration	Registration
08.30	09.00					
09.00	09.30		Welcome	IS OC PC	IS OC PC	
09.30	10.00		**ISCB President's Invited Speaker: Stephen Senn**			SY SD
10.00	10.30	CS				
10.30	11.00		Coffee Break	Coffee Break	Coffee Break	
11.00	11.30		IS OC PC	IS OC PC	**Keynote Speaker: Francesca Dominici**	Coffee Break
11.30	12.00	Coffee Break				SY SD
12.00	12.30	CS			ISCB Annual General Meeting	
12.30	13.00		OC	OC		
13.00	13.30					Lunch Break
13.30	14.00	Lunch Break	Lunch Break	Lunch Break	Lunch Break	
14.00	14.30					
14.30	15.00	CS		Excursions	IS OC PC	SY SD
15.00	15.30		IS OC PC			SD
15.30	16.00					
16.00	16.30	Coffee Break			Coffee Break	
16.30	17.00	CS	Coffee Break		IS OC PC	
17.00	17.30					SD
17.30	18.00		IS OC PC			
18.00	18.30				ISCB39 Presentation	
18.30	19.00					
19.00	19.30					
19.30	20.00	Students' Gathering				
20.00			Welcome Reception		Conference Dinner	

CS	IS	OC	PC	SY	SD
Pre-conference Courses	Invited Session	Oral Contributed Sessions	Poster Contributed Session	Mini-Symposia	Students' Day

From	To	Sunday 9				
10.00	11.30		CS2 Causal questions and principled answers: a guide through the landscape for practising statisticians	CS3 The analysis of recurrent event data in clinical trials		CS5 Advances in adaptive clinical trial design: Bayesian and frequentist approaches
11.30	12.00	Coffee Break				
12.00	13.30		CS2 Continued	CS3 Continued		CS5 Continued
13.30	14.30	Lunch Break				
14.30	16.00	CS1 Statistical methods for microbiome studies	CS2 Continued		CS4 Age-Period-Cohort modeling	CS5 Continued
16.00	16.30	Coffee Break				
16.30	18.00	CS1 Continued	CS2 Continued		CS4 Continued	CS5 Continued
19.30		Students' Gathering at Estrella Galicia brewery (Praza de Compostela)				

From	To	Monday 10					
09.00	09.30	Welcome					
09.30	10.30	ISCB President's Invited Speaker: Stephen Senn Competence Centre for Methodology and Statistics, Luxembourg Institute of Health The Revenge of RA Fisher: Thoughts on Randomgate and its Wider Implications					
10.30	11.00	Coffee Break					
11.00	12.30	IS1 Statistical methods for precision medicine	OC1 Bayesian methods in clinical research 1	OC2 Longitudinal data analysis and biomaker discovery	OC3 Machine learning for clinical data analysis	OC4 Multitaste survival analysis and dynamic prediction 1	PC1 Clinical trials 1
12.30	13.30	OC5 STRATOS 1	OC6 Bayesian methods in clinical research 2	OC7 Methods for handling missing data 1	OC8 Rare diseases and small populations	OC9 Survival analysis 1	
13.30	15.00	Lunch Break					
15.00	16.30	IS2 Methods in research on research	OC10 Biostatistics for high dimensional data 1	OC11 Joint modelling in practice 1	OC12 Statistical methods for systematic reviews	OC13 Causal inference and mediation analysis 1	PC2 Clinical trials 2
16.30	17.00	Coffee Break					
17.00	18.30	IS3 Beyond proportional hazards	OC14 Bayesian methods in clinical research 3	OC15 Complex survival data 1	OC16 Design and analysis of clinical trials 1	OC17 Statistical methods in epidemiology 1	PC3 Biomedical studies 1
20.00		Welcome reception at Pazo Quiñones de León (Castrelos Park)					

From	To	Tuesday 11					
09.00	10.30	IS4 Joint modelling in practice	OC18 Biostatistics for high dimensional data 2	OC19 Design and analysis of clinical trials 2	OC20 Statistical methods in epidemiology 2	OC21 Diagnostic tests	PC4 Survival analysis
10.30	11.00	Coffee Break					
11.00	12.30	IS5 Complex survival data	OC22 Bayesian methods in clinical research 4	OC23 Statistical methods for precision medicine 1	OC24 Design and analysis of clinical trials 3	OC25 Survival analysis 3	PC5 Statistical methods in epidemiology
12.30	13.30	OC26 STRATOS 2	OC27 Joint modelling in practice 2	OC28 Survival analysis 2	OC29 Topics in biostatistics 1	OC30 Methods for handling missing data 2	
13.30	14.30	Lunch Break					
14.30		Excursions					

From	To	Wednesday 12					
09.00	10.30	IS6 Biostatistics for high dimensional data	OC31 Design and analysis of clinical trials 4	OC32 Multitaste survival analysis and dynamic prediction 2	OC33 Heterogeneity in biomedical studies and meta-analysis	OC34 Statistical challenges of EHR (eHealth Records) analysis	PC6 Bayesian methods and joint modelling
10.30	11.00	Coffee Break					
11.00	12.00	Keynote Speaker: Francesca Dominici, Harvard University, USA Model Uncertainty and Covariate Selection in Casual Inference					
12.00	13.30	ISCB Annual General Meeting					
13.30	14.30	Lunch Break					
14.30	16.00	IS7 Bayesian methods in clinical research	OC35 Complex survival data 2	OC36 Statistical methods for precision medicine 2	OC37 Design and analysis of clinical trials 5	OC38 Statistical methods in epidemiology 3	PC7 Observational studies
16.00	16.30	Coffee Break					
16.30	18.00	IS8 Clinical trial simulations; the when, where and what	OC39 Bayesian methods in clinical research 5	OC40 Joint modelling in practice 3	OC41 Topics in biostatistics 2	OC42 Survival analysis 4	PC8 Biomedical studies 2
18.00	18.30	ISCB39 Presentation					
20.00		Conference dinner at Hotel Pazo Los Escudos (Beach of Alcabre)					

From	To	Thursday 13			
		Mini-Symposia			Students' Day
09:30	10.00	SY1 Choosing estimands in a clinical trial	SY2 Modelling personalised screening: a step forward on risk assessment methods	SY3 Statistical methods for medicall imaging and their use in clinical an epidemiological studies	Invited talk
10.00	11.00				Session 1
11.00	11.30	Coffee Break			
11.30	12.00	SY1 Continued	SY2 Continued	SY3 Continued	Invited Talk
12.00	13.00				Session 2
13.00	14.30	Lunch Break			
14.30	15.00	SY1 Continued	SY2 Continued	SY3 Continued	Open Questions
15.00	16.00				Session 3
16.00	16.30				
16.30	17.00				Invited Talk
17.00	18.00				Farewell

9th July
Sunday

Pre-conference Courses

Full-day Courses

CS2 Causal Questions and Principled Answers: a Guide Through the Landscape for Practising Statisticians

Sunday 9th July - 10.00-18.00 h. - Room: Sala Mar 2
Bianca DeStavola, *UCL Great Ormond Street Institute of Child Health, United Kingdom*
Els Goetghebeur, *Ghent University, Belgium*
Saskia Le Cessie, *Leiden University Medical Center, the Netherlands*
Erica Moodie, *Mc Gill University, Quebec, Canada*
Ingeborg Waernbaum, *Umea University, Sweden*

Outline: This course aims to support the practicing statistician in making it work: causal inference form observational data in a potential outcomes framework. How to chose among the many versions of exposure, the target population and indeed the estimand in a given setting? Different estimands serve different purposes while lending themselves (in)directly to natual constraints imposed on given datasets. We spend ample time exploring this in various contexts before explaining key features of estimation methods relying on either the no unmeasured confounders' assumption or the availability of `instrumental variables'. We review, apply and compare dedicated methods centered around outcome regression and stratification, inverse probability weighting and their double robust version, when relying on the no unmeasured confounders assumption, and two-stage-least-squares estimation when exploiting an instrumental variable. We illustrate key results and estimation properties with published case studies, using either the original data or simulations. Hands on sessions will guide participants in using R or Stata with the data. In the final session, the simulation engine is presented as a tool one can adapt to help support design and analysis, recognizing the consequences of various choices.

CS5 Advances in Adaptive Clinical Trial Design: Bayesian and Frequentist Approaches

Sunday 9th July - 10.00-18.00 h. - Room: Sala Terra 2
Rajat Mukherjee, *Cytel, USA*
Pantelis Vlachos, *Cytel, Switzerland*

Outline: Considerable interest has grown among pharmaceutical and other medical product developers in adaptive clinical trials, in which data collected during the course of a trial affects ongoing conduct or analysis of the trial. Following the release of the FDA draft Guidance document on adaptive design clinical trials in early 2010, expectations of an increase in regulatory submissions involving adaptive design features, particularly for confirmatory trials, were high. This course will introduce Bayesian and Frequentist approaches and hands-on techniques for the design, monitoring, and analysis of adaptive clinical trials. We will summarize the major statistical methods for adaptive trials by clinical development phase. You will learn different adaptive trial designs such as single and dual-agent model-based dose escalation, response-adaptive randomization, sample-size re-estimation and drop-the-loser designs through trial examples and the software tools provided (East and Compass). Challenges of adaptive trial implementations will be discussed and recommendations will be provided.

Half-day Courses

CS3 The Analysis of Recurrent Event Data in Clinical Trials

Sunday 9th July - 10.00-13.30 h. - Room: Sala Mar 4
Mouna Akacha, *Novartis Pharma AG, Switzerland*
Richard Cook, *University of Waterloo, Canada*

Outline: Recurrent event data arise when the same type of event can occur repeatedly over a period of observation. Such data are encountered increasingly often in clinical trials in areas such as chronic obstructive pulmonary disease where individuals may experience repeated infections, multiple sclerosis where repeated exacerbations of symptoms may occur, and cardiology where patients may undergo repeated hospitalizations for heart failure. In such studies, interest usually lies in understanding the underlying recurrent event process and how this is impacted by explanatory variables such as treatment. Characterizing the rate at which events occur and the extent of inter-individual variation are key objectives. In this tutorial, we will review common objectives and challenges in the analysis of recurrent events and critically discuss available methods with an emphasis on applications to clinical trials where causal inferences are of paramount importance. Nonparametric, parametric and semi-parametric methods will be discussed. Sample code will be provided through the analysis of data from recent trials.

Learning objectives:

- Upon completion of this course, the participants should be able to:
- Understand objectives, advantages and limitations of recurrent event endpoints and approaches
- Apply appropriate statistical methods for modelling recurrent events
- Interpret the results obtained from different approaches

Who should attend?:

- Statisticians with interest in the design and analysis of clinical trials with recurrent event endpoints

Half-day Courses

CS1 Statistical Methods for Microbiome Studies

Sunday 9th July - 14.30-18.00 h. - Room: Sala Mar 4
M. Luz Calle, *University of Vic - Central University of Catalonia, Spain*

Outline: Understanding human-microbiome relationship and how it can be modulated is a frontier for preventive medicine and for the medical management of chronic diseases and represents an opportunity for food and pharma industry.

Although the human microbiome has long been known to influence human health and disease, high-throughput DNA sequencing technologies have revolutionized this field by allowing the study of the genomes of all microorganisms of a given environment. Metagenomics provides an efficient and cost-effective tool for investigating the members of a microbial community and how they change. It represents a breakthrough in the study of the link between human microbiome and our health.

This course provides a self-contained introduction to microbiome data analysis. A brief introduction to concepts in microbial ecology and microbial DNA sequencing will be given but the main focus will be on illustrating the use of R packages, such as phyloseq, for managing and visualizing high-dimensional and structured data, normalization, filtering, and multivariate analysis of microbiome count data.

CS4 Age-Period-Cohort Modeling

Sunday 9th July - 14.30-18.00 h. - Room: Sala Terra 4
Jon Wakefield, *University of Washington, Seattle, USA*

Outline: In general, examination of disease rates may be carried out on three time scales: age (of the individual), period (time of diagnosis) and cohort (time of birth). Given any two of age, period, cohort, however, determines the third, and so one cannot uniquely identify the three different components. Despite numerous warnings (Clayton and Schifflers, 1987a,b, Carstensen, 2007, Smith and Wakefield, 2017) over interpretation continues to occur in the literature. In this course, the identifiability will be examined, and approaches to inference will be described. In particular, what can and what cannot be deduced from the data alone will be discussed.

Both frequentist and Bayesian methods will be presented. Throughout, ideas and modeling will be illustrated with examples. The examples will use publicly-available data, with methods implemented in the R programming language, with code provided, so that participants in the course will be able to carry out analyses with their own data.

Carstensen, B. (2007). *Age–period–cohort models for the Lexis diagram. Statistics in Medicine, 26:3018–3045.*
Clayton, D. and Schifflers, E. (1987a). *Models for temporal variation in cancer rates. I: age–period and age–cohort models. Statistics in Medicine, 6:449–467.*
Clayton, D. and Schifflers, E. (1987b). *Models for temporal variation in cancer rates. II: age–period–cohort models. Statistics in Medicine, 6:469–481.*
Smith, T.R. and Wakefield, J. (2017). *A review and comparison of age-period-cohort models for cancer incidence. Statistical Science. To appear.*

10th July
Monday

Plenary Speaker

ISCB President's Invited Speaker

Monday 10th July - 09.30-10.30 h. - Room: Auditorio
Chair: KyungMann Kim

The Revenge of RA Fisher: Thoughts on Randomgate and its Wider Implications

Stephen Senn, *Competence Centre for Methodology and Statistics, Luxembourg Institute of Health*

Stephen Senn has worked as a statistician but also as an academic in various positions in Switzerland, Scotland, England and Luxembourg. Since 2011 he has been head of the Competence Center for Methodology and Statistics at the Luxembourg Institute of Health in Luxembourg. He is the author of Cross-over Trials in Clinical Research (1993, 2002), Statistical Issues in Drug Development (1997, 2007), Dicing with Death (2003) and in 2009 was awarded the Bradford Hill Medal of the Royal Statistical Society. He is an honorary life member of PSI and ISCB.

Abstract: RA Fisher warned us years ago: as ye randomize so shall ye analyse. We ignored him and we haven't faithfully reflected the way we randomise in the way we analyse. In the pharmaceutical industry, we randomise in permuted blocks but we don't fit the block. Ex pharma researchers minimise but don't condition on the order of patients in the trial.

A recent interesting and heroic analysis by John Carlisle has identified dozens of clinical trials in leading journals as being either suspiciously imbalanced at baseline or having balance that's too good to be true. However, he analyses the baseline distributions as if the trials were completely randomized but nearly all trial are not. I consider Carlisle's analysis not only as regards whether or not the whole world of clinical trials is awash with fraud but also as regards the implication for analysis of outcomes, even if it is not.

I conclude that because we have arrogantly assumed that in theory we can do better than Fisher, in practice we often do worse.

Reference: Carlisle, J. B. "Data fabrication and other reasons for non random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals." Anaesthesia (2017).

Invited Session

IS1: Statistical Methods for Precision Medicine

Monday 10th July - 11.00-12.30 h. - Room: Auditorio
Chair: Annette Kopp-Schneider
Organised by Marie Davidian, *North Carolina State University, USA*

IS1-1 SMART Clinical Trial Design for Tailored Treatment Guidelines

Kelley Kidwell, *University of Michigan, USA*

A sequential multiple assignment randomized trial (SMART) design is an innovative clinical trial design that can construct and evaluate dynamic treatment regimens (DTRs). In turn these DTRs (i.e. tailored decision rules) can guide clinical practice and improve care. In this talk, we will highlight SMART designs that have been implemented to provide guidance on the types of questions they can address, sample size and power, and their accompanying analytic plans. We will discuss the advantages and disadvantages of a SMART design over standard designs in providing evidence for tailored treatment regimens in the era of precision medicine.

IS1-2 Sample Size Considerations for Precision Medicine

Eric Laber, *Carolina State University, USA*

A personalized treatment strategy formalizes evidence-based treatment selection by mapping patient information to a recommended treatment. Personalized treatment strategies can produce better patient outcomes while reducing cost and treatment burden. Thus, among clinical and intervention scientists, there is a growing interest in conducting randomized clinical trials when one of the primary aims is estimation of a personalized treatment strategy. However, at present, there are no appropriate sample size formulae to assist in the design of such a trial. Furthermore, because the sampling distribution of the estimated outcome under an estimated optimal treatment strategy can be highly sensitive to small perturbations in the underlying generative model, sample size calculations based on standard (uncorrected) asymptotic approximations or computer simulations may not be reliable. We offer a simple and robust method for powering a single and multiple stage randomized clinical trials when the primary aim is estimating an optimal personalized treatment strategy. The proposed method is based on inverting a plugin projection confidence interval and is thereby regular and robust to small perturbations of the underlying generative model. The proposed method requires elicitation of several clinically meaningful parameters from clinical scientists and uses data from a small pilot study to estimate nuisance parameters, which are not easily elicited.

Reference 1: Laber, E. B., Zhao, Y. Q., Regh, T., Davidian, M., Tsiatis, A., Stanford, J. B., ... & Kosorok, M. R. (2015). Using pilot data to size a two arm randomized trial to find a nearly optimal personalized treatment strategy. Statistics in medicine.

IS1-3 Personalizing Immunosuppressant Regimes Following Transplantation

Erica Moodie, *McGill University, Quebec, Canada*

The development of methods to determine personalized sequences of treatments is an area of active research in the statistical, computer sciences, and medical communities. Much of the methodological work done to date has focused on the continuous outcome settings; extensions to time-to-event settings have been relatively limited, and not all are easily applied in real-world settings. In this talk, I will consider the practical and analytic challenges involved in determining whether tailoring immunosuppressive therapy that depletes T cells in prophylaxis and treatment of graft-versus-host-disease to patient characteristics improves disease- free survival using a large cohort drawn from a transplantation registry.

Oral Contributed Sessionss

OC1: Bayesian methods in clinical research 1

Monday 10th July - 11.00-12.30 h. - Room: Sala Mar 2
Chair: Jordan Elm

OC1-1: Enhancing adaptive enrichment trials using short term endpoints

Authors: Thomas Burnett[1].
[1]*University of Bath, United Kingdom*.

Suppose a patient sub-population has been identified before conducting a confirmatory trial, where the sub-population is expected to respond well to the new treatment. Adaptive Enrichment trials initially recruit from the full patient population, then at an interim analysis choose how to recruit the remainder of the trial (recruiting patients exclusively from the sub-population or continuing to recruit from the full population).

The choice of population must be based on the available observations at the time of the interim analysis, however more patients will have already been recruited lowering the impact of any adaptation. If we collect short term observations in addition to the long term observations that will be used at the final analysis we may enhance the interim decision. This is done by assuming a joint model for the short term and long term observations, giving a better understanding of the expected future behaviour of the trial.

We use a Bayesian decision framework to optimise the decision at the interim analysis. Using this framework we may also assess the overall performance of different trial designs, allowing us to learn about the overall benefit offered by our enhanced interim decision. This method of optimisation may be applied to delayed responses, longitudinal observations and in some cases survival data.

OC1-2: Sample size recalculation in clinical trials incorporating historical data

Authors: Katharina Hees[1], Meinhard Kieser[1].
[1]*Institute of Medical Biometry and Informatics, Heidelberg.*

Recruiting sufficient patients within an acceptable time horizon is an issue for most clinical trials and is especially challenging in the field of rare diseases. It is therefore an attractive option to include historical data from previous (pilot) trials in the analysis of the current study thus reducing the recruitment burden. Various Bayesian methods for the incorporation of historical information in present trials have been proposed in the literature. In case that the current data match sufficiently well with the historical data, these approaches lead to increased power. However, if this assumption is not met, the gain in power may be much smaller than expected while at the same time a type I error inflation occurs. Therefore, so-called robust prior distributions are well-suited since in case

of a prior-data conflict they down-weight the extent to which the historical data is incorporated. When planning the sample size for trials incorporating historical data, not only the type I error rate, the power, and the treatment group difference but additionally the variance and the weight of the historical data have to be specified. However, there is usually some uncertainty in the planning phase about the value of these nuisance parameters. We present methods for blinded and unblinded sample size recalculation in the setting of two-arm superiority trials with historical control data where the variance – and in the unblinded setting additionally the extent to which the historical information is incorporated – is estimated mid-course and the sample size is recalculated accordingly. The operating characteristics of these methods are investigated in terms of actual type I error rate, power, and expected sample size.

OC1-3: Comparing drug development strategies with probabilities of success including benefit-risk assessment to inform decision-making

Authors: Gaelle Saint-Hilary[1], Giuseppe Luigi Lagrange[2], Veronique Robert[3], Mauro Gasparini[1], Giuseppe Luigi Lagrange[2].
[1]*Dipartimento di Scienze Matematiche (Disma),* [2]*Politecnico di Torino, Torino, Italy,* [3]*Department of Biostatistics, Institut de Recherches Internationales Servier (Iris), Suresnes, France.*

Evidence-based quantitative methodologies have been proposed to inform decision-making in drug development, such as metrics to make go/no-go decisions or predictions of success based on the statistical significance of future clinical trials. While these methodologies appropriately address some critical questions on the potential of a drug, they either consider the past evidence without predicting the outcome of the future trials or focus only on efficacy, failing to account for the multifaceted aspects of a successful drug development. As quantitative benefit-risk assessments could enhance decision-making, we propose a more comprehensive approach using a composite definition of success based not only on the statistical significance of the treatment effect on the primary endpoint, but also on its clinical relevance, and on a favorable benefit-risk balance in the next pivotal studies. For one drug, we can thus study several development strategies before starting the pivotal trials by comparing their predictive probability of success. The predictions are based on the available evidence from the previous trials, which could be combined with new hypotheses on the future development. The resulting predictive probability of composite success provides a useful summary to support the discussions of the decision-makers. We present a fictive, but realistic, example in Major Depressive Disorder inspired by a real decision-making case.

OC1-4: Partial pooling of patient subgroups in phase II trials: improving performance when treatments are only active in a minority of groups

Authors: Manuel Wiesenfarth[1], Annette Kopp-Schneider[1].
[1] *German Cancer Research Centre.*

Clinical trials in the era of precision medicine frequently require analyzing the efficacy of a treatment in several groups of patients with early stopping for futility at interim analyses.

If it is expected that some of the patient subpopulations respond similarly to the therapy, the use of Bayesian hierarchical models has been suggested which allow for outcome-adaptive borrowing of information across groups (Berry et al, 2013). Hierarchical modeling is known to be advantageous if all treatments are indeed similar. However, there is a risk of missing treatments that are only active in some groups of patients and of overestimating the efficacy of unpromising groups.

We consider early stopping based on posterior probabilities at interims and focus on binary outcomes, where typically the log-odds of response is assumed to be Gaussian and partial pooling is achieved by assigning a hyperprior on the variance term.

In simulations we study various model specifications with the aim to find models that are still capable of improving upon independent inference but are less prone to overlook treatments that are only active in a minority of subgroups. Heavy-tailed and mixture priors on the log-odds of responses and alternative priors on the associated variance terms are examined and power, type I errors and mean sample sizes are compared. Moreover, a strategy is proposed where at each interim analysis the independence and hierarchical model are compared based on the estimated out-of-sample prediction accuracy and the better performing model is used.

Reference 1: Berry et al (2013). Bayesian hierarchical modeling of patient subpopulations: efficient designs of phase II oncology clinical trials. Clin. Trials. 10, 720–734.

OC1-5: Quantitative approaches to design and analysis of experimental medicine clinical trials

Authors: Kirsty Hicks[1], Graeme Archer[1], Jacquie Christie[1], Sam Miller[1], Fabio Rigat[1].
[1]*GlaxoSmithKline, UK.*

Experimental Medicine (EM) clinical trials typically feature a relatively small number of patients and, because the objective of EM is to learn about the biological mechanisms of disease and their interaction with pharmacotherapy, a relatively large number of endpoints. EM trials then, even more than classical clinical trials, require careful thinking about their design and interpretation, to avoid the pit fall of post hoc reasoning. Power statements and Null hypothesis significance testing are of low utility; we need techniques to assess the fitness of many competing mechanistic models. The Bayesian paradigm is a natural framework for this multivariate set-up; we demonstrate how to use

multivariate Bayesian methods to design EM studies (and to estimate their probability of success), to assess the utility of the evidence from such studies, and to support Go/No-Go decision-making based on EM study read-outs.

OC2: Longitudinal data analysis and biomarker discovery

Monday 10th July - 11.00-12.30 h. - Room: Sala Mar 4
Chair: Paola Rebora

OC2-1: Modelling trajectories and causal links between anatomic, cognitive and functional domains in Alzheimer's disease: a latent process approach

Authors: Bachirou Taddé[1], Daniel Commenges[1], Hélène Jacqmin-Gadda[1], Cécile Proust-Lima[1].
[1]*Inserm U1219 - Bordeaux Population Health, Bordeaux, France.*

The Alzheimer's disease is progressive with a preclinical phase of more than 10 years and multidimensional with gradual impairments in the anatomical sphere (brain atrophies), the cognitive sphere (decline in various cognitive functions) and the functional sphere (activities of daily living affected).

We propose a new statistical approach which permits to model the dynamics of multiple domains involved in the disease and, for the first time, to simultaneously explore their dynamic causal relationships.

Our model, defined in discrete time and relying on latent processes, combines the features of multivariate linear mixed models and of Dynamic Bayesian Network models.

Each domain is defined as a latent process which is related to repeated measures of its markers in Gaussian equations of observation. The dynamics of the system of latent processes is modelled with separated linear mixed models and the dynamic interrelationships are assessed by regressing the network components at time t on the network components at time t-1. The program, based on Maximum Likelihood Estimation and written in R and C++, was validated in simulations. We applied our model to the data of the Alzheimer's disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) which contains longitudinal data on brain structure, cognitive functioning and functional dependency in three preclinical stages of Alzheimer's disease: control individuals, individuals with a mild cognitive impairment and diagnosed subjects. We specifically assessed the dynamic causal relationships between three dimensions (Cerebral atrophy, cognition and Functional dependency) and contrasted the causal structure between preclinical stages.

OC2-2: Methods for handling longitudinal outcomes truncated by dropout and death

Authors: Lan Wen[1], Shaun R. Seaman[1].
[1]*MRC Biostatistics Unit, United Kingdom.*

In cohort studies of ageing, missing data can arise from dropout or death. If there is a non-negligible amount of death, it is important to distinguish between inference for an immortal cohort and a mortal cohort. Immortal cohort inference makes no distinction between missing data due to death and missing data due to other reasons. An unconditional model (e.g. linear mixed model or LMM) provides immortal cohort inference, thereby creating a cohort that never dies. On the other hand, partly conditional models (e.g. generalized estimating equations with an independence working correlation or IEE) provide inference about a mortal cohort on the expected outcome at a given time among survivors at that time point (partly conditional inference). IEE makes strong assumptions about the longitudinal and dropout processes. To weaken this assumption, we look to popular missing data methods: inverse probability weighting (IPW) and multiple imputation (MI). When there are dropout and death, we should clarify the underlying assumption(s) of a missing data method to make sure that it is valid for making partly conditional inference. In this talk, we describe and compare the assumptions of MI and IPW methods. We also provide new augmented IPW estimating equations for making partly conditional inference, which offer double protection against model misspecification. Simulation studies show that a method will give negligibly biased estimates of the partly conditional mean when its assumptions are met, but potentially biased estimates if its assumptions are not met. We apply the missing data methods to the data from the OCTO study of ageing.

OC2-3: A criterion for selection of working correlation structure in generalized estimating equations

Authors: Gul Inan[1], Mahbub Latif[2], John Preisser[3].
[1]*Middle East Technical University, Turkey,* [2]*University of Dhaka, Bangladesh,* [3]*University Of North Carolina At Chapel Hill, U.S.A.*

Generalized estimating equations (GEE) are one of the most commonly used marginal models for the analysis of correlated responses, especially discrete outcomes. GEEs directly specify a generalized linear model for the correlated responses of a subject and model the association within the responses of a subject through a working correlation matrix. Irrespective of whether the working correlation structure type is correctly specified, GEEs produce consistent estimators for regression parameters, with a correctly specified marginal mean model. However, correct specification of the working correlation structure ensures efficient estimation of the regression parameters within the class of linear unbiased estimating equations. For this reason, developing methods for working correlation structure selection in GEEs, conditional on the correctly specified marginal mean model, has been an active area of research. This study proposes a working correlation selection criterion for GEEs based on a predicted residual sum of squares (PRESS) statistic. An extensive Monte Carlo simulation study is designed to assess the performance of the proposed criterion and to compare its performance with the existing criteria. Furthermore, the proposed PRESS statistic is illustrated via using the Coronary Artery Risk Development in Young Adults Study data.

OC2-4: CRP as a longitudinal biomarker for sepsis resolution

Authors: Armando Teixeira-Pinto[1], Rosa Oliveira[2].
[1] *University of Sydney, Australia,* [2] *Universidade do Porto, Portugal.*

The evaluation of the diagnostic or prognostic ability of clinical biomarkers usually focuses on a single measurement of the biomarker and its association with an outcome of interest. However, there are settings where the longitudinal profile of a biomarker may be more informative than the single value of a cross-sectional observation.

C-reactive protein (CRP) has been suggested as a potential marker of Sepsis resolution in patients admitted to intensive caRe. But clinical studies failed to show a good prognostic value of CRP at admission or maximum CRP during in-stay.

We discuss how the biomarker trend, over several days, can be used to model the probability of death for these patients. We propose a two-stage approach that first models CRP as a linear trajectory and then uses the estimates of the intercept and slope as predictor variables in the mortality model. However, these two quantities are estimates of the true intercept and slope, and using the ordinary least squares estimates directly in the mortality model produces biased estimates of their effect.

By recognizing this setting as a measurement error problem, we studied a pseudo-likelihood approach and adapted the regression calibration method to reduce the bias in the two-step approach. Simulations and results from a real data are discussed.

OC2-5: Beta-binomial mixed-effects model for analyzing longitudinal binomial data with overdispersion

Authors: Josu Najera-Zuloaga[1], Dae-Jin Lee[1], Inmaculada Arostegui[2].
[1]*Basque Center for Applied Mathematics - BCAM,* [2]*Universidad del País Vasco, UPV/EHU.*

It is usual in medical and biological framework the analysis of longitudinal binomial outcomes which present over-dispersion. Previous literature shows the adequacy of the beta-binomial distribution when fitting binomial variables with large over-dispersion in a regression context. In this work we present a mixed-effects model based on the beta-binomial distribution, which is able to accommodate random effects in the linear predictor of the variable, and hence, take into account the dependency over time in longitudinal studies.

The beta-binomial distribution does not belong to the exponential family, so most of the literature in mixed-effects models cannot be applicable to this methodology. One of the most widely used techniques when analyzing distributions that do not belong to the exponential family are the Generalized Additive Models for Location, Scale and Shape (GAMLSS). There exist relevant differences in the estimation of the model parameters between our model (based on hierarchical likelihood) and GAMLSS. Simulation-based comparisons were performed to illustrate the differences. For a real application, we considered data from the Mini-Mental State Examination questionnaire in patients with dementia.

Reference 1: Najera-Zuloaga J., Lee D.-J. and Arostegui I., Comparison of beta-binomial regression model approaches to analyze health-related quality of life data, Statistical Methods in Medical Research, (2017), DOI:10.1177/0962280217690413

Reference 2: Rigby, R. A. and Stasinopoulos D. M. (2005). Generalized Additive Models for Location, Scale and Shape, (with discussion). Appl. Statist., 54, pp 507-554

OC3: Machine learning for clinical data analysis

Monday 10th July - 11.00-12.30 h. - Room: Sala Terra 4
Chair: Hae Won Uh

OC3-1: A model-based subgroup learning strategy

Authors: Michael Leblanc[1].
[1]*Fred Hutchinson Cancer Research Center, United States.*

Developing simple descriptions of subjects with extreme risk, differing prognosis, or greater treatment effect has long been of interest in oncology. Using appropriate statistical learning tools can be helpful in the design of future targeted clinical trials. We will focus on a method that gives well defined and interpretable patient subgroups that are less variable than the results obtained from tree-based algorithms and that also controls the fraction of patients identified by the rule induction method.

Many statistical algorithms have been proposed that lead to simple logical or Boolean descriptions of subgroups of patients. The most widely utilized (and sometimes over-used) is the decision rule (or cut-point) developed on a single ordered biomarker strongly related to patient outcome. Extensions to multivariate learning methods include tree-based algorithms and rule induction methods sometimes called bump hunting or peeling (Dazard et. al, 2016).

We present a new algorithm based on combining generalized linear sub-model components to constructed Boolean decision rules. The method extends our prior work on Extreme Regression (LeBlanc et. al, 2006) and utilizes extrema functions (minimum and maximum) of sub-models involving clinical and biological patient attributes to construct rules/subgroups based on survival data. Additional variance control and variable selection for subgroups is obtained through Elastic-Net type regularization.

We present examples from clinical trials conducted by SWOG (formerly the Southwest Oncology Group), a United States National Cancer Institute (NCI) supported organization that conducts clinical trials in adult cancers.

OC3-2: Estimating inverse probability weights using super learner when weight-model specification is unknown in a marginal structural Cox model context

Authors: Mohammad Karim[1], Robert W. Platt[2].
[1]*St. Paul's Hospital, Providence Health Care, Canada,* [2]*Epidemiology, Biostatistics and Occupational Health, McGill University, Canada.*

Correct specification of the inverse probability weighting (IPW) model is necessary for consistent inference from a marginal structural Cox model (MSCM). In practical applications, researchers are typically unaware of the true specification of the weight model. Nonetheless, IPWs are commonly estimated using parametric models, such as the main-effects logistic regression model. In practice, assumptions underlying such models may not hold and data-adaptive statistical learning methods may provide an alternative. However, the optimal statistical learning approach for a given dataset is impossible to predict. Super Learner (SL) has been proposed as a tool for selecting an optimal learner from a set of candidates using cross-validation. In this study, we evaluate the usefulness of a SL in estimating IPW in four different MSCM simulation scenarios, in which we varied the specification of the true weight model specification. Our simulations show that, in the presence of weight model misspecification, with a rich and diverse set of candidate algorithms, SL can generally offer a better alternative to the commonly used statistical learning approaches in terms of MSE as well as the coverage probabilities of the estimated effect in an MSCM. The findings from the simulation studies guided the application of the MSCM in a multiple sclerosis cohort from British Columbia, Canada (1995-2008) to estimate the impact of beta-interferon treatment in delaying disability progression.

Reference 1: Karim M.E., et al. (2014) American Journal of Epidemiology; 180 (2): 160-171. DOI: 10.1093/aje/kwu125

Reference 2: Karim M.E., et al. (2016) Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2016.1248574, published online: 21 Oct 2016

OC3-3: Semivarying coefficient least squares support vector regression for analyzing high-dimensional gene-environmental data

Authors: Insuk Sohn[1], Jooyong Shim[2], Changha Hwang[3].
[1]*Samsung Medical Center, South Korea,* [2]*Inje University, South Korea,* [3]*Dankook University, South Korea.*

In the context of genetics and genomic medicine, gene-environment (G X E) interactions play a vital role in determining virtually the risk of human diseases. Some of the existing G X E interaction methods are limited by analyzing a small number of G factors at a time, by assuming linear effects of E factors and by adopting ineffective selection techniques. In this study, we propose a new approach for identifying important G X E interactions and main effects. This is based on a semivarying coefficient least squares support vector regression method, which is devised by utilizing flexible semiparametric least squares support vector regression technique for censored survival data. This semivarying coefficient model is adopted to accommodate possible nonlinear effects of E factors.

We also propose a generalized cross validation (GCV) method for choosing the hyperparameters which affect the performance of the proposed method. This GCV function is also used to identify important G X E interactions and main effects. The proposed method is evaluated through numerical studies.

Reference 1: J. Shim, C. Kim, C. Hwang, Semiparametric least squares support vector ma-chine for accelerated failure time model, Journal of the Korean StatisticalSociety 40 (2011) 75{83.

Reference 2: C. Wu, X. Shi, Y. Cui, S. Ma, A penalized robust semiparametric approachfor gene-environment interactions, Statistics in Medicine 34 (2015) 4016{4030.

OC3-4: Subgroup identification in dose-response studies via model-based recursive partitioning

Authors: Marius Thomas[1], Björn Bornkamp[1], Heidi Seibold[2], Torsten Hothorn[2].
[1]*Novartis Pharma AG, Switzerland,* [2]*University of Zurich, Switzerland.*

An important task in early phase drug development is to identify patients, which respond better or worse to an experimental treatment. While a variety of different subgroup identification methods have been developed for the situation of trials that study an experimental treatment and control, much less work has been done in the situation when patients were randomized to different dose groups.

The model-based recursive partitioning framework by Zeileis et. al (2008) has recently been applied to subgroup identification for two-arm studies (Seibold et. al, 2016). We extend this approach to trials with several dose groups and show that the method can be used to identify subgroups of patients with different dose-response curves and improves estimation of treatment effects and minimum effective doses, when heterogeneity among patients is present.

Reference 1: Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. Journal of Computational and Graphical Statistics, 17(2), 492-514.

Reference 2: Seibold, H., Zeileis, A., & Hothorn, T. (2016). Model-Based Recursive Partitioning for Subgroup Analyses. The international journal of biostatistics, 12(1), 45-63.

OC3-5: Segmentation methods comparison at events detection in physiological time-series: hidden Markov models versus spectrogram based techniques

Authors: Andrea Hita[1], Munshiimran Hossain[2], Rajat Mukherjee[1].
[1]*CYTEL INC, Spain,* [2]*CYTEL INC, India.*

Information encoded in physiological time series has been shown to be very useful in assessing all types of diseases. Specially, the extraction of information from biosignals is bringing the possibility to migrate to minimally invasive diagnostic tests when externally recorded physiological signals are available. However, the task of extracting the relevant information through the analysis of time-varying dynamical systems is not always trivial.

Here, we compare two different statistical processing approaches aiming the detection and segmentation of events from accelerometer time series recordings. Masked data from a real case-study is presented to illustrate both techniques behaviors and results. Firstly, a spectrogram based analysis was performed to find the event time bounds based on the power spectrum variability. Secondly, hidden Markov models (HMM) were used to recognize any dynamical switching on the temporal series density distribution and amplitude by processing the signal incrementally.

OC4: Multistate survival analysis and dynamic prediction 1

Monday 10th July - 11.00-12.30 h. - Room: Sala Terra 2
Chair: Hein Putter

OC4-1: Diagnostics in HIV/AIDS progression using a hidden Markov model

Authors: Tatek Getachew Habtemichael[1], Ayele T. Goshu[1], Henning Omre[2].
[1]*School of Mathematical And Statistical Sciences, Hawassa University, Hawassa, Ethiopia,* [2]*Department of Mathematical Sciences, Norwegian University of Sciences & Technology, Trondheim, Norway.*

In developing countries, like Ethiopia, medical capacity is limited and traveling is time consuming. The objective is to provide diagnostic indicators for individual patients to support decisions for reduced medication and monitoring frequency, such that limited medical resources can be reallocated. Data is HIV-infected patients in ART treatment from Hawassa University Referral Hospital in Ethiopia during 2006-2015. Longitudinal observations are collected every sixth month at t=(1, ..., T) for patient i as doctor's subjective classification of HIV-infection level and CD4-counts, for 544 patients. We use a hidden Markov model with the categorical hidden variable being the patients true WHO HIV-infection levels with time, being assigned a prior Markov chain model, parametrized by an initiation probability and a transition matrix. The observations, doctor's classified HIV-levels and CD4 counts are linked to the true HIV-levels by likelihood models which are parametrized by a mis-classification matrix

and log-Gaussian expectations and variances for each level. The model parameters are assessed by maximum marginal likelihood from the complete set of observations. Reliable and stable estimates are obtained and the estimation uncertainty is assessed by bootstrapping. The prior and likelihood models define the posterior model given the observations for patient i, which can be exactly assessed by the very efficient recursive Forward-Backward algorithm. We use diagnostic: Revise medication strategy for patient i if on HIV-level one for more than one year with probability 0.8, ie the posterior probability for being on HIV-level one at times (T-2, T-1, T) greater than 0.8, which increases with time due to medication.

OC4-2: Using point-prevalence studies for risk factor analyses of hospital-acquired infections: a multi-state point of view

Authors: Martin Wolkewitz[1], Micha Mandel[2], Mercedes Palomar-Martinez[3], Francisco Alvarez-Lerma[4], Pedro Olaechea-Astigarraga[4], Martin Schumacher[1].
[1] *Institute for Medical Biometry and Statistics, Medical Center University of Freiburg, Germany,* [2] *Department of Statistics, Israel,* [3] *Universitat Autonoma de Barcelona, Spain,* [4] *Service of Intensive Care Medicine, Spain.*

Point-prevalence studies are often used to investigate risk factors of hospital-acquired (nosocomial) infections [1]. In contrast to cohort studies, this design is easy to conduct as it requires less time and resources. However, it is well known that long-stay patients are more likely to be sampled (length-biased sampling); a patient who stays 10 days has the double chance to get sampled than a patient who stays 5 days [2]. As a result, very sick patients are overrepresented in prevalence studies due to the design. It is also well known that long stayers are at higher risk for nosocomial infections than short-stayers. And further, patients with nosocomial infections stay longer in hospital.

We study how these facts influence point-prevalence studies which are usedfor risk factor analyses. We used cohort data with full information (ie. exact date of the nosocomial infection)and an underlying multi-state model (which accounts for discharge/death as competing events) to discover mechanisms which are -due to the design- hidden in prevalence data. We mimic a point-prevalence study by picking a sample from this cohort study. The multi-state model plays the key role to understand relationships between the three association measures in risk factor analyses (prevalence-, risk- and hazard ratio of NI).

Reference 1: Vincent, JL, et al(2009). International study of the prevalence and outcomes of infection in intensive care units. JAMA, 302, 21:2323-9.

Reference 2: Mandel, M (2010). The competing risks illness-death model under cross-sectional sampling. Biostatistics, 11, 2:290-303.

OC4-3: Misspecification of cause of death in a progressive illness-death model

Authors: Michael Lauseker[1], Christine Zu Eulenburg[2].
[1] *Ludwig-Maximilians-Universität München, Institute for Medical Information Sciences, Biometry, and Epidemiology, Munich, Bavaria, Germany,* [2] *University of Groningen, University Medical Center Groningen, Dept Epidemiology, Med Stat & Decis Making, Groningen, Netherlands.*

Cause of death is increasingly becoming a topic in oncology. It is usually distinguished between disease-related and disease-unrelated death. In insufficiently documented data, a frequently used approach is to define death as disease-related when a progression to advanced phases has occurred before, otherwise as disease-unrelated. The data are analysed as competing risks, while the underlying model is in fact a progressive illness-death model. Therefore, the purpose of our work was to analyse if this misspecification leads to any systematic bias in the results.

Data were simulated according to different scenarios, following a Markov, semi-Markov and a non-Markov progressive illness-death model. This was done by varying the hazard for the transition from progression to death. This hazard was either simulated as independent of the time to progression, dependent on the time since progression or dependent on the time of progression. Censoring was added with the censoring distribution being independent of the events. We compared the cumulative incidences of the events "disease-related death" and "disease-unrelated death" in the competing risks analysis to the state occupation probabilities in the progressive illness-death model with regards to bias and variance.

Although both estimators are not equal when censoring is present, we could not prove any systematic bias. However, variance in the misspecified competing risk analysis was slightly larger than in the progressive illness-death model.

OC4-4: Multi state modelling for heart failure care path: a population-based study using administrative data

Authors: Francesca Leva[1], Francesca Gasperoni[1], Giulia Barbati[2].
[1] *Politecnico di Milano, Italy,* [2] *Università di Trieste, Italy.*

We investigate how different risk profiles of Heart Failure (HF) patients can influence multiple readmission rate and final outcomes (death). To date, several models for predicting adverse outcomes have been developed, but they are mainly focused on a single outcome We propose the application of two different multi state models in real world setting to jointly evaluate the impact of different risk factors on multiple hospital admissions, on Integrated Home Care (IHC) activations, on Intermediate Care Unit (ICU) admissions and on death. The first model (Model 1) concerns only hospitalizations as possible events in patients' clinical history. In the second one (Model 2), we consider both hospitalizations and ICU admission and IHC activation. Through Model 1, we want to detect the determinants of repeated hospitalizations, while, through Model 2, we want to evaluate which patients' profiles are associated with transitions in intermediate care with respect to repeated hospitalizations or death. Both models are characterized by transition specific covariates, adjusting

for patient's risk factors. The application of multi state models enables a better identification of implications of different patterns of care, once adjusted for patients' risk profiles. The results of the paper provide useful healthcare support to patients management in real word context. Our study suggests that the epidemiology of the considered clinical characteristics are more nuanced than traditionally presented with single event by demonstrating their important and versatile role on different patterns of care.

OC4-5: A bidirectional multi-state model for panel data on bone mineral density among HIV-infected patients

Authors: Klaus Langohr[1], Guadalupe Gómez Melis[1], Nuria Pérez[2], Eugenia Negredo[2], Anna Bonjoch[2].
[1]*Universitat Politècnica de Catalunya/ Barcelonatech, Spain,* [2]*Fundació Lluita Contra la Sida, Spain.*

Bone mineral density (BMD) measurements are used to determine bone health and can help to identify the risk of fracture. The most widely recognized BMD scan, which measures bone density at different parts of the body, is called dual-energy x-ray absorptiometry (DXA). The DXA measures are compared to the BMD of a healthy 30-year-old adult of the same gender and are converted into T-scores: T-scores above -1 are considered normal, values between -1 and -2.5 indicate low bone mass (osteopenia), and values below -2.5 indicate osteoporosis. The main goals of the present study are to study the evolution of BMD over time in a cohort of more than 700 HIV-infected persons with at least two DXA scans and to determine the risk factors for the progression of bone loss.For this purpose, a bidirectional multi-state model with states normal BMD, osteopenia, and osteoporosis is fitted to the data. The model considers four possible transitions -normal BMD to osteopenia, osteopenia to normal BMD, osteopenia to osteoporosis, and osteoporosis to osteopenia- which are studied as a function of age and antiretroviral treatment. Due to the nature of the panel data available, all transition times are interval-censored.This multi-state model allows us to estimate the transition probabilities and predict the percentages of patients in every health state as a function of age and treatment. The clinical relevance of building such a model is to guide the clinical practice and to rationalize DXA scans measurements.

Poster Contributed Session

PC1: Clinical trials 1

Monday 10th July - 11.00-12.30 h. - Room: Hall
Chair: Rosa Lamarca

PC1 - M1: A comparison of different ways of incorporating the baseline count in count response models for count data from falls prevention trials

Authors: Han Zheng[1], Alan Kimber[1], Victoria Goodwin[2], Ruth Pickering[1].
[1] *University of Southampton, UK,* [2] *University of Exeter Medical School, UK.*

The goal of a falls prevention trial is to test whether an intervention reduces the count of falls experienced by participants during follow-up. Typically the outcome count follows a right skewed distribution possibly with outlying large values, and variance greater than the mean. Negative binomial (NB) regression accommodates this overdispersion and is recommended for the analysis of falls preventions trials. If a count of falls during a baseline period is available, including it will lead to a more powerful analysis, and consideration of the log link function in the NB model leads to inclusion following log transformation. Cook and Wei [1] proposed an alternative, Conditional Negative Binomial (CNB) model assuming baseline and follow-up counts come from a joint distribution with shared random subject effect. We conducted a simulation study (2000 repeats) to compare: 1) NB regression excluding the baseline count (NB-null); NB regressions including the baseline count as a covariate 2) untransformed (NB-unlogged); 3) after log transformation (NB-logged); and 4) as an offset after log transformation (NB-offset); and finally 5) the CNB model. Simulations showed that in most scenarios NB-null had the lowest power. NB-unlogged only improved the power slightly when data were greatly overdispersed. NB-logged and NB-offset had considerably greater power. Though CNB performed best, our simulations matched the assumptions underlying the model. Regression incorporating the baseline count correctly is likely to increase power in the analysis of a falls prevention trial.

Reference 1: Cook RJ, Wei W. Conditional analysis of mixed Poisson processes with baseline counts: implications for trial design and analysis. Biostatistics 2003;4(3):479–94.

PC1 - M4: Do cluster randomised and individually randomised trials estimate the same intervention effect? A meta-epidemiological study

Authors: Clemence Leyrat[1], Agnès Caille[2], Sandra Eldridge[3], Sally Kerry[3], Agnès Dechartres[4], Bruno Giraudeau[2].
[1] *London School of Hygiene And Tropical Medicine, UK,* [2] *Université de Tours, Université de Nantes, Inserm, Sphere U1246, Tours, France,* [3] *Queen Mary University of London, UK,* [4] *Centre D'epidémiologie Clinique, Hôpital Hôtel-Dieu, Assistance Publique-Hôpitaux de Paris, France.*

In the medical literature, some interventions have been assessed using both individually (IRT) and cluster randomised trials (CRT). CRTs are thought to be more pragmatic but may be at higher risk of bias. It is still unknown if the intervention effects estimated in CRTs and IRTs are the same which raises the question of the relevance of meta-analysing them together. We aim to assess whether intervention effects differ between IRTs and CRTs.

We conducted a meta-epidemiological study of meta-analyses (MA) published between 2010 and 2014 in the Cochrane database. Intervention effect estimates were expressed as odds ratios (OR) or standardised mean differences (SMD) for binary or continuous outcomes. Ratio of OR (ROR) or difference of SMD (DMSD) were estimated for each MA, expressing difference in intervention effect between CRTs and IRTs and then combined using random-effects MAs.

76 MAs (917 trials, 183 CRTs/734 IRTs) with a binary outcome and 45 MAs (541 trials, 131/410) with a continuous outcome were selected. We observed no difference in intervention effect estimates between CRTs and IRTs for binary outcomes (ROR 1.00[0.93-1.08]). Conversely, the pooled DSMD was 0.13[0.06-0.19] showing larger estimates in IRTs as compared to CRTs. The DSMD was no longer significant among MAs with a pharmacological intervention, with an objective outcome or when sample size was accounted for.

We did not observe any difference in estimates between CRTs and IRTs for binary outcomes, whereas we observed slightly larger effects in IRTs as compared to CRTs for continuous outcomes. Overall, the impact of the trial type was very limited and observed differences between IRTs and CRTs might be confounded by some characteristics of the trials.

PC1 - M7: Interventions to improve reporting guidelines adherence: a scoping review

Authors: David Blanco de Tena-Dávila[1], Erik Cobo[1], Jamie Kirkham[2], Isabelle Boutron[3].
[1] *Universitat Politècnica De Catalunya, Spain,* [2] *University Of Liverpool, UK,* [3] *Université Paris-Descartes, France.*

Reporting Guidelines (RGs) have been developed since early 1990s to help improving the completeness and transparency of published articles, which helps decision makers to judge the applicability of the research, and enhances reproducibility. There is evidence that the use of some RGs, such as CONSORT, is associated with

improved standards of reporting. However, the current levels of adherence to RGs are poor, far from desired. For this reason, different actions aiming to improve RGs adherence have been taken over the last years.

We have performed a scoping review of interventions aiming to improve RGs adherence. After developing a suitable search strategy, we got 891 hits, from which we selected 18 articles reflecting different actions that have been taken to improve RGs adherence. Most of them (78%, 14 of 18) about journal policies, including weak or strong endorsement of RGs, compulsory trial registration or active implementation strategies of RGs in the editorial process. Others target authors (17%, 3 of 18) and peer reviewers (5%, 1 de 18). In this presentation, we will show in detail the findings of this review and comprehensively summarize the results in terms of adherence to RGs of the interventions found. Furthermore, we will analyze other possible interventions that have been reflected in the literature but never performed.

This review could send a message to funders, authors and editors about how the problem of adhering better to RGs has been tackled from different perspectives. Moreover, it could be a major first step towards developing future strategies to improve RGs adherence, which could have a decisive impact in informing the treatment and preventing the disease.

PC1 - M10: A non-parametric information-theoretic approach for phase I/II clinical trials

Authors: Pavel Mozgunov[1], Thomas Jaki[1].
[1] *Lancaster University, UK.*

Recently, a joint Phase I/II clinical trials have became more common for the novel drug investigation instead of two independent Phase I and Phase II clinical. In this paper we propose the general escalation criteria for Phase I/II clinical trials that can be used for both parametric and non-parametric methods. However, we advocate that due to special properties of the proposed criteria, one can use no parametric (or even monotonic) assumptions and achieve comparable or even better performance than existing model-based methods. The derivation of criteria is based on the recent developments in the weighted information measures. The proposed criteria explicitly reflects an trade-off between toxicity and efficacy that simplifies its practical implementation. To demonstrate the appealing properties of the proposed approach we consider Phase I/II dose-finding clinical trials for targeted therapies. It is shown how partial and missing information can be incorporated in the non-parametric approach. We compare the operational characteristics of the novel design with recently proposed model-based methods in various different scenarios. To address safety and futility issues corresponding time-varying constraints are introduced. Constraint parameters' influence is also extensively studied.

PC1 - M13: Comparison of superficial surgical site infection between delayed primary versus primary wound closure in complicated appendicitis: a randomized controlled trial

Authors: Ammarin Thakkinstian[1], Boonying Siribumrungwong[2], Chumpon Wilasrusmee[1], Pinit Noorit[3], Winai Ungpinitpong[4], Pratya Chotiya[5], Anuwat Chanthip[6].
[1] *Mahidol University,* [2] *Mahidol University and Thammasat University,* [3] *Chonburi Hospital,* [4] *Surin Hospital, Thailand,* [5] *Pathum Thani Hospital, Thailand,* [6] *Lampang Hospital, Thailand.*

Importance: Superficial surgical site infection (SSI) is common in appendectomy of complicated appendicitis. Delayed primary wound closure (DPC) is routinely used but efficacy is still conflict.

Objective: To compare superficial SSI between DPC and primary wound closure (PC).

Design, setting, participants: A multicenter, randomized clinical trial of complicated appendicitis (i.e., gangrenous and rupture) was conducted involving 607 adult patients from 6 hospitals in Thailand during November 2012 to February 2016.

Intervention: A stratified-block randomization sequence was generated to assign patients receiving PC or DPC.

Main outcome and measure:Superficial SSI by the Center for Disease Control criteria. Secondary outcomes included postoperative pain, length of stay, recovery time, quality of life and cost of treatment.

Results: Among 303 and 304 patients in PC and DPC, 5 and 4 patients were lost follow up, leaving 300 and 298 patients for intention-to-treat (ITT) analysis. The superficial SSI rate was lower in PC than DPC (i.e., 7.3% (4.4%, 10.3%) versus 10% (95%CI: 6.6%, 13.3%) with a risk difference (RD) of-2.7% (-7.1%, 1.9%). Applying instrumental-variable analysis to adjust protocol-violation (9 and 6 for PC and DPC) yielded similar results as ITT approach with RD of -2.8% (-7.6%, 1.9%). Postoperative pain, length of stay, recovery times, and quality of life were not significantly different with corresponding RDs of 0.3 (-2.5, 3.0), -0.1 (-0.5, 0.3), -0.2 (-0.8, 0.4), and 0.02 (-0.01, 0.04). However, costs for DPC were 2083 (1410, 2756) Baht higher than PC.

Conclusion and Relevance: PC should be applied for appendectomy in complicated appendicitis with right lower quadrant incision.

PC1 - M16: An efficient early phase trial design for dose-finding and activity evaluation in two patient populations simultaneously.

Authors: Daniel Slade[1], Josef Vormoor[2, 3, 4, 5, 6], Britta Vormoor[2, 3, 4, 5, 6], Tobias Menne[7], Deborah Bird[1], Laura Llewellyn[1], Lucinda Billingham[1].
[1]*Cancer Research Clinical Trials Unit, University of Birmingham, UK,* [2]*Wolfson Childhood Cancer Research Centre,* [3]*Northern Institute for Cancer Research,* [4]*Newcastle University, UK,* [5]*Great North Children's Hospital,* [6]*Newcastle Upon Tyne NHS Foundations Trust, UK,* [7]*Freeman Hospital, Newcastle Upon Tyne NHS Foundations Trust, UK*

The inclusion of both adult and paediatric patient groups in the same trial has many advantages, not limited to the potential borrowing of information across patient groups. There is a clear benefit to both the costs and time that would be required to assess both groups in independent trials. We explore the benefits of a design incorporating two patient populations in an early phase trial scenario.

SeluDex is an international phase I/II trial of selumetinib in combination with dexamethasone used for the treatment of relapsed/refractory RAS-pathway mutated Acute Lymphoblastic Leukaemia (ALL) in both adult and Paediatric patients. In Phase I SeluDex will utilise a modified two-stage Bayesian Continual Reassessment Method (CRM) to determine the maximum tolerated doses in both patient groups using independent models. The Phase II component will assess clinical response independently in each group of patients within a Bayesian framework.

The two-stage Bayesian CRM is modified such that it incorporates stopping for excess toxicity if the probability of toxicity at the lowest dose is greater than a defined critical value. The CRM has many merits over alternative rule based designs (e.g. 3+3), both in terms of practicality and more efficient use of accrued trial information in any inference. The use of a Bayesian design in Phase II was chosen as it can provide greater flexibility than early phase frequentist approaches. In SeluDex a Bayesian approach was implemented as it provided the opportunity to borrow information and gain efficiency in a rare disease setting.

This trial illustrates the potential to determine a maximum tolerated dose and assess early evidence of activity in two patient populations within the same trial.

PC1 - M19: Progressive dose-finding phase I trials in cancer: can results from one trial inform the next?

Authors: Lucinda Billingham[1], Kristian Brock[1], Christina Yap[1], Caroline Kelly[2], James Paul[2], Anthony Chalmers[3].
[1] *Cancer Research UK Clinical Trials Unit, University of Birmingham, UK,* [2] *Cancer Research UK Clinical Trials Unit, University of Glasgow, UK,* [3] *Beatson West of Scotland Cancer Centre, Glasgow, UK.*

Development of treatments for cancer is often based on investigating the addition of one active intervention to another. Typically an experimental drug is added to an established intervention such as other drugs or radiotherapy. In such situations, the dose-finding phase I trial of the experimental drug in combination is likely to have been preceded by a similar trial of the drug as a single agent. Using data on the dose-toxicity relationship from the first trial could provide a starting point for the subsequent trial and allow efficiency through cumulative learning.

The OPARATIC study was a phase I dose-finding trial that established the maximum tolerated dose (MTD) of olaparib in combination with low-dose oral temozolomide in patients with relapsed glioblastoma using a 3+3 rule-based design. It involved 32 patients in cohorts at 5 different dose/schedule combinations selected from two doses of olaparib (100mg and 150mg) in 3 different weekly schedules (days 1-3, 1-5 and 1-7). A new phase I dose-finding trial, the PARADIGM-2 study, again aims to determine the MTD of olaparib but this time taken concomitantly with low-dose oral temozolomide and partial brain radiotherapy in newly diagnosed glioblastoma. Olaparib treatment is specified as combinations of 3 different doses (50, 100 and 150mg) and 5 different weekly schedules (days 1, 1-2, 1-3, 1-4, 1-5). Although the trial has been planned as another 3+3 rule-based design, there are major overlaps between the trials and this paper explores how the results from OPARATIC might be used to provide a potentially more efficient model-based design for PARADIGM-2. Our motivating example has the added complexity of considering different doses and schedules for the drug.

PC1 - M22: Application of multilevel mixed model in randomized controlled trial of Mitomicin C for postoperative endoscopic sinus surgery

Authors: Pawin Numthavaj[1], Thongchai Bhongmakapat[2], Boonsarm Roongpuwabhat[2], Kangsadarn Tanjararak[2], Atiporn Ingsathit[1], Ammarin Thakkinstian[1].
[1] *Section For Clinical Epidemiology And Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand,* [2] *Department of Otolaryngology, Faculty of Medicine Ramathibodi Hospital Mahidol University, Bangkok, Thailand,*

Background: Mitomicin-C (MMC) is used for prevention of postoperative synechia in endoscopic sinus surgery (ESS). Meta-analysis was performed based on randomized controlled trials (RCTs) with poor quality. We therefore conduct a RCT to compare efficacy of MMC versus placebo.

Method: An ongoing RCT of 200 adult patients with chronic rhinosinusitis scheduled for ESS was conducted since April 2015 at Ramathibodi Hospital, Bangkok, Thailand. A block-randomization was

generated to assign patients to receive MMC (0.4mg/ml) or placebo, which was applied after ESS. The primary outcome was synechia rate assessing at 1 week, 1, 3, and 6 months postoperatively. Sinonasal outcome test-22 (SNOT-22) symptom scores were also assessed. A multilevel mixed-effect logit and linear regressions were applied to determine the treatment efficacy by intention-to-treat approach. Analysis was performed by Stata/SE v.14.2.

Results: A total of 161 patients completed 6-month follow-up until January 2017. In blinded complete-case preliminary analysis, 81 and 80 were respectively assigned to arm A and B. From multilevel mixed-effect logit regression model, risks of synechia for corresponding groups were 0.197 and 0.181, with risk ratio of 0.922 (95%CI: 0.483-1.561). Marginal means of SNOT-22 scores were 23.30 and 27.42, with mean difference of 4.13 (-0.52, 8.77).

Conclusion: Preliminary result showed that the two interventions were non-significant. Further follow-ups should be performed as planned and multiple imputation technique should be applied to replace visits with missing data.

PC1 - M25: Effect of denoising on longitudinal volumetric measurements of brain atrophy measurements based on MRI for Alzheimer's disease

Authors: Mojmir Vinkler[1], Stanislav Katina[1], Miroslav Smisek[2], Petr Novak[2], Norbert Zilka[3], Reinhold Schmidt[4], Eva Kontsekova[3], Peter Dal-Bianco[5], Martin Brunner[6], Wolfgang Staffen[7], Michael Rainer[8], Roman Sivak[2], Bengt Winblad[9], Michal Novak[10].
[1] *Institute of Mathematics and Statistics, Masaryk University, Brno, Czech Republic,* [2] *Axon Neuroscience CRM, Bratislava, Slovakia,* [3] *Axon Neuroscience R&D, Bratislava, Slovakia,* [4] *Clinical Division Of Neurogeriatrics, Department Of Neurology, Medical University Graz, Graz, Austria,* [5] *University Clinic Of Neurology, Medical University Of Vienna, Vienna, Austria,* [6] *University Clinic Of Clinical Pharmacology, Medical University Of Vienna, Vienna, Austria,* [7] *University Clinic Of Neurology, Christian-Doppler-Clinic, Salzburg, Austria,* [8] *Social And Medical Centre East, Danube Hospital, Karl Landsteiner Institute For Memory And Alzheimer Research, Vienna, Austria,* [9] *Karolinska Institute Alzheimer Disease Research Centre And Clinical Trial Unit, Geriatric Clinic, Karolinska University Hospital, Huddinge, Sweden,* [10] *Axon Neuroscience Se, Larnaca, Cyprus.*

Alzheimer's disease(AD) is a neurocognitive disorder with various level of atrophy of cerebral cortical regions and subcortical structures– mainly hippocampus but also other selected cerebral structures, denoted as regions of interest(ROIs). Volumetric measurements of hippocampus and its atrophy rate are promising tools in the determination of effect of disease-modifying treatment. Automatic segmentation of various ROIs visualized by MR and measurement of their volume is now freely available for clinical research of the natural course of the disease and effect of the treatment. However, due to inherent noise, variance of these measurement is high, making difficult reaching statistically valid conclusions. Using 140 T1 MRI scans(28 patents, 5 visits) from randomized, placebo-controlled, parallel group, double-blinded, multi-centre Phase I clinical trial(paper about this clinical trial was published recently in Lancet Neurology) we found that denoising with state of the art method prior to running FreeSurfer 5.3.0, automatic segmentation reduced measurement error from 6.79% to 3.54% without introducing processing bias. Furthermore, additional temporal information reduced error to 2.50%. Denoised volumetric data were statistically analyzed by mixed-effect linear regression

model, e.g. to find out differences in hippocampus atrophy across time between control(placebo) and treated group and also between left and right hippocampus. Finally, these results were used in power analysis and sample size calculation which could be applied during design of further Phase II and III clinical trials.

PC1 - M28: Covariates-adjusted comparison of two treatments with ordinal responses

Authors: Tong-Yu Lu[1], Ping Yang[2], Wai-Yin Poon[2], Siu Hung Cheung[2], Peifang Su[3], Huang-Tz Ou[3].
[1] *China Jiliang University,* [2] *The Chinese University Of Hong Kong,* [3] *National Cheng Kung University, Taiwan.*

In clinical studies, treatment responses are frequently measured on an ordinal scale. To compare treatment efficacies after adjusting for covariates, the proportional odds model is a popular choice when compared with other ordinal regression models. However, the proportional odds model relies on the proportional odds assumption, limiting its applications in practice. To provide a more general statistical method, we propose the latent covariate-adjusted regression model. The proposed procedure assumes that the ordinal response outcomes of a treatment are associated with a continuous latent variable. To compare two treatments by our approach, the underlying latent variables are not required to have homogeneous variance (equivalent to the proportional odds assumption in the proportional odds model). Moreover, the treatment-by-covariate interactions can be easily estimated and interpreted. Clinical examples are also provided to illustrate our method.

Reference 1: Lu, T.Y., Poon, W.Y., and Cheung, S.H. (2014). A unified framework for the comparison of treatments with ordinal responses. Psychometrika, 79, 605-520.

Reference 2: McCullagh, P. (1980). Regression models for ordinal data. Journal of the Royal Statistical Society, Series B, 42, 109-142.

PC1 - M31: Statistical methods for screening adverse drug interactions, with an application to spontaneous reporting systems database

Authors: Masahiko Gosho[1].
[1]*Department Of Clinical Trial And Clinical Epidemiology, Faculty Of Medicine, University Of Tsukuba-Japan.*

Early detection of adverse drug reactions for drug-drug interactions (DDIs) is an important issue for public health and individual patient safety using two or more drugs. A valid statistical method is needed to detect adverse drug reactions resulting from potential DDIs in spontaneous reporting systems because DDIs are often not identified until the drugs are widely used in clinical practice. Norén et al. (2008) have developed a novel criterion for detecting DDIs in spontaneous reporting systems. In addition, we have proposed a criterion based on chi-square statistics calculated from

observed number of events with drug combinations and the expected number of events. We applied the two criteria to the Food and Drug Administration Adverse Event Reporting System database, and some recognized DDIs were detected using the criteria.

Reference 1: Norén, GN, Sundberg, R, Bate, A, and Edwards, IR. (2008). A statistical methodology for drug-drug interaction surveillance. Statistics in Medicine 27: 3057—3070.

PC1 - M34: Evaluation of statistical methods for pharmacovigilance signal detection from social media

Authors: Xiaoyi Chen[1], Pierre Karapetiantz[1], Armelle Guenegou-Arnoux[1], Yannick Girardeau[1,2], Anita Burgun[1], Sandrine Katsahian[1].
[1] *Inserm, Umrs1138, Équipe 22, Centre De Recherche Des Cordeliers, Paris, France,* [2] *Université Paris Descartes, Sorbonne Paris Cité.*

Introduction: Pharmacovigilance aims to monitor marketed drugs' adverse effect (AE). Spontaneous Reporting System (SRS) is the principal database through which adverse drug reaction (ADR) signals can be detected. Recently, patients' posts on social media are being considered as a complementary source to SRS. However, these two sources differ significantly. Our objective is to evaluate the signal detection (SD) methods in the context of social media, which is essential for their generalization.

Methods: SRS tends to include more reports on newer drugs with serious or lesser-known AE. With these traits, several approaches have been proposed to detect disproportionate reporting signal: frequentist methods are sensitive but often with low specificity, while Bayesian methods show generally better results, especially for less frequent signals, but are still limited by concerns regarding coprescriptions and masking effect. Logistic regression type methods have been explored in order to overcome the drawbacks of disproportionality methods, but their ability to detect specific drug-event association is limited. On social media, data contains more information about common drugs with common and less serious AEs. We thus improved the signal generation modeling and adapted it to include the characteristics of social media. The three categories of approaches were compared via an intensive simulation study.

Results: Each simulated drug-event pair holds a Poisson probability considered as the ground truth. All methods provide a statistical score indicating the strength of each signal. The main result is shown in terms of ranked correlation. The advantages and limitations of each method are discussed within the context of social media.

PC1 - M37: The use of historical controls in randomized clinical trials in rare diseases

Authors: Thomas Zwingers[1].
[1]*Cros De. 86150 Augsburg. Germany.*

The use of prospectively randomized clinical trials is the gold standard to show efficacy in clinical research when evaluating new compounds. Especially in rare diseases, e.g. in cancer subtypes like brain tumours, it is sometimes very difficult to recruit the requested number of patients in a reasonable timeframe. On the other hand, there are often registers on such patients which have been treated with the standard of care or individual treatment regimens.

Stuart Pocock (1) suggested a compromise in the way to combine historical controls with a prospective control group in a randomized clinical trial. We implemented this design in a clinical trial in patients with a rare tumour in order to overcome the problems recruitment and limited resources. We will discuss the problems which are inherent to historical data not collected prospectively according to a study protocol and the assumptions of the sample size calculation. We will show that the inclusion of historical controls substantially reduced the number of patients to be recruited and the total duration of study.

Reference 1: (1) Stuart J. Pocock: The combination of randomized and historical controls in clinical trials. J Chron Dis 1976, Vol.29

PC1 - M40: The choice of a randomization procedure for randomized controlled clinical trials

Authors: Diane Uschner[1], Ralf-Dieter Hilgers[1], Nicole Heussen[1, 2].
[1] *Rwth Aachen University, Germany,* [2] *Sfu Vienna, Austria.*

Background: Randomization is used in clinical trials with the objective of balancing known and unknown covariates and to reduce the risk of selection bias. It has been remarked that randomization procedures differ with respect to avoiding selection bias. Even so, the choice of a randomization procedure at the design stage of a clinical trial is often arbitrary, and no software exists that guides the researcher to select a randomization procedure based on scientifically sound evaluation.

Methods: We propose the R package randomizeR that allows the comparison of a large number of randomization procedures with respect to scientifically sound criteria including imbalance, loss and control of type-I-error rate and power. In this talk, we focus on the susceptibility of small population group trials to selection bias, and conduct a simulation study comparing various randomization procedures using a sequence based approach. Particularly, we propose an exact formula for the probability of false rejection of the null hypothesis of no difference if selection bias is present. We illustrate the approach with a case study using parameter estimates from the SPR Study.

Results: Some randomization procedures are substantially more susceptible to false rejection of the null hypothesis due to selection bias than others. The randomizeR R package aids the user

in choosing a randomization procedure on the basis of established criteria from the literature. In addition, randomizeR facilitates the reporting of the randomization procedure according to the CONSORT guideline. All in all, the use of the randomizeR package can lead to unbiased outcomes and better reporting, and thus increase the credibility of the results of the clinical trial.

PC1 - M43: Impact of interim analyses in time-to-event randomized trials in rare diseases on a long-term horizon

Authors: Mohamed Amine Bayar[1, 2, 3], Gwénaël Le Teuff[1, 2, 3], Stefan Michiels[1, 2, 3], Marie-Cécile Le Deley[1, 2, 3, 4].
[1] *Service De Biostatistique Et D'epidémiologie, Gustave Roussy And Cesp, Faculté De Médecine,* [2] *Université Paris-Sud, Faculté De Médecine,* [3] *Uvsq, Inserm, Université Paris-Saclay, France,* [4] *Unité De Méthodologie Et Biostatistique Et D'epidémiologie, Centre Oscar Lam,*

In the rare diseases context, when traditional large trials may not be doable, we previously considered an approach based on a series of randomized controlled trials performed over a long research horizon. We concluded, under reasonable assumptions, that designs with a relaxed alpha-level and smaller sample sizes outperformed traditional ones. The aim of this work was to extend this approach to evaluate on the long term trials with a dynamic sample size calculation and interim analyses (IA).

We simulated a series of 2-arm superiority trials over a 15-year period with no IA or IA for futility and/or efficacy. We considered different disease severities, accrual rates, hypotheses of how treatments improve over time, and more or less 'aggressive' stopping rules and spending functions. To compare the designs performance with traditional ones, we estimated the total survival benefit as the relative difference in hazard rates, at year-15 vs. year-0, and the probability of selecting at year-15 a treatment inferior to the initial control (risk).

The simulation study suggests that designs with a relaxed alpha-level (1-sided 10%) and smaller sample sizes still outperform traditional ones in maximizing long-term survival benefit but are associated with higher risks (For the main scenario, benefit: 43% vs. 36%, risk: 1.0% vs. 0.3%). Including IAs yields an additional moderate increase of benefit and decrease of risk, notably when including both futility and efficacy rules (benefit: 46%, risk: 0.6%). We plan to further include multi-arm trials with a treatment selection at interim.

Designs based on a series of trials with IAs over a long research horizon could be an interesting alternative to traditional designs for rare diseases.

PC1 - M46: Evaluation of effect by using different control of internal clinical trial in rare disease registry

Authors: Akiko Kada[1], Akiko M. Saito[1].
[1] *National Hospital Organization Nagoya Medical Center, Japan*

It is important for rare disease to establish disease registries to understand a disease prevalence as a first step in the process of therapeutic development. It is desirable to develop a new treatment for rare disease as the next step. However, it is a challenge to conduct internal clinical trial in disease registries, because it is difficult for rare disease to have enough patients in the trial and to set a concurrent control. In such a setting, single-arm clinical trial will be one solution, and the candidates of control to compare are i) subjects without entering the clinical trial, ii) retrospective data of the subjects who enter the clinical trial, and iii) retrospective data of all subjects in particular period of the disease registry. In this situation, our purpose is to get better estimation of effects of a new treatment by using appropriate control in a disease registry. We supposed several different settings of control from disease registries of rare disease. And estimation of effects according to patterns of control were determined. We show examples using the disease registry of epilepsy syndrome of Japan.

PC1 - M49: A computationally simple central monitoring procedure for the detection of data fabrication in multi-center clinical trials

Authors: Rutger Van Den Bor[1, 2], Petrus W.j. Vaessen[2], Bas J. Oosterman[2], Nicolaas P.a. Zuithoff[1], Diederick E. Grobbee[1, 2], Kit C.B. Roes[1, 2].
[1] *Julius Center For Health Sciences And Primary Care, University Medical Center Utrecht, The Netherlands,* [2] *Julius Clinical LTD., Zeist, The Netherlands.*

In the conduct of clinical trials, central monitoring becomes an ever more feasible quality assurance tool, and may, by allowing center-by-center comparisons, in particular be useful for the detection of deviating data patterns in multi-center trials. We aimed to develop a central monitoring procedure that is specifically focused on detecting deviations characteristic of data fabrication, and that can be applied repeatedly to accumulating trial data in a semi-automated manner. Based on a range of anticipated characteristics of fabricated data (e.g., fabricated data may show limited variability and may contain few missing values), we propose a relatively simple algorithm that returns a ranking of centers, enabling more targeted inspection. The algorithm only utilizes the trial baseline data and can easily be applied to the accumulating trial data on frequent intervals during the progress of the trial. It was developed using data from the Second European Stroke Prevention Study, a large international multi-center trial in which fraud was discovered in one of the 60 centers. Applied to this trial, the algorithm consistently returns a high ranking for the fraudulent center, a finding observed for only few of the other centers. Thus, the procedure is a promising method to detect data fabrication, even early in the trial. Application to independent empirical datasets (preferably with known fraud) is required for further testing and refinement.

PC1 - M52: A novel decision-theoretic framework for adaptive multi-arm clinical trials

Authors: Andrea Bassi[1], Peter Van De Ven[1], Hans Berkhof[1].
[1] *Vu University Medical Center Amsterdam, Netherlands*

Current medicine sees a strong increase in the number of drugs hypothesized to be effective for treatment of specific diseases. Instead of conducting separate trials for each candidate drug, multiple drugs can be compared in a single trial , but standard approaches still require a large number of patients.We introduce a novel framework for adaptive multi-stage multi-arm clinical trials. Within a Bayesian setting, a decision is made regarding continuation of the trial after each stage. The decision to continue or stop the trial is made on the basis of a cost-utility evaluation, where the expected decrease in loss related to an incorrect decision is traded off against the cost of continuing the trial for one more stage. The framework is flexible in the sense that it allows for trial modifications such as early dropping of treatment arms.Through numerical simulation, we compare the performance of our adaptive design to non-adaptive trials. The proportion of simulated trials in which the best drug is selected is 0.8 to 14.3 percent higher for our adaptive design than for non-adaptive designs with the same expected number of patients enrolled.

PC1 - M55: Network meta-analysis of disconnected networks: how dangerous are random baseline treatment effects?

Authors: Audrey Béliveau[1], Sarah Goring[2], Robert W. Platt[3], Paul Gustafson[1].
[1] *University of British Columbia, Canada,* [2] *Icon Plc., Canada,* [3] *McGill University, Canada.*

In network meta-analysis, the use of fixed baseline treatment effects in a contrast-based approach is regularly preferred to the use of random baseline treatment effects because, often, there is not a need to model baseline treatment effects and the modeling of baseline treatment effects carries a risk for model misspecification. However, in disconnected networks, fixed baseline treatment effects do not work because there is not enough information in the data to update the prior distribution on the contrasts between disconnected treatments. In this talk, we investigate to what extent the use of random baseline treatment effects is dangerous in disconnected networks. We take two publicly available datasets of connected networks, and disconnected them in multiple ways. We then compare the results of treatment comparisons obtained from a Bayesian contrast-based analysis of each disconnected network using random normally-distributed baseline treatment effects to those obtained from a Bayesian contrast-based analysis of their initial connected network using fixed baseline treatment effects. For the two datasets studied, we found that the use of random baseline treatment effects in disconnected networks was appropriate. Because those datasets were not cherry-picked, we believe that there are other disconnected networks that would benefit from being analysed using random baseline treatment effects. However, there is also a risk for the normality assumption to be inappropriate in other datasets even though we have not observed this situation in our case study. We provide code so other datasets can be investigated.

PC1 - M58: Addressing treatment contamination in trials of complex interventions: results from a simulation study comparing two design options

Authors: Nicholas Magill[1], Khalida Ismail[1], Paul McCrone[1], Sabine Landau[1].
[1] *Institute Of Psychiatry, Psychology, And Neuroscience, King's College London, United Kingdom.*

In mental health trials there is concern that the control treatment might be contaminated, meaning that a participant in the comparator arm receives the active intervention. This can often be avoided by design through the use of cluster randomisation, with clusters defined by the level at which contamination is thought to take place. An alternative approach is to apply random treatment allocation at the level of the individual, measure treatment receipt for each individual and account for contamination in the analysis. However, very little is currently known about which design option is optimal under particular levels of clustering and contamination. We investigated whether it is more efficient to address contamination through the use of cluster randomisation with analysis accounting for clustered data or individual randomisation with adjustment for contamination in the analysis. We predicted that cluster randomisation will only be favoured under high levels of contamination, modest intraclass correlation coefficients and small cluster sizes. Contamination is commonly conceptualised as present or absent, but partial contamination, for example receipt of a few therapy sessions, is also possible. We report simulation findings under both scenarios. Our considerations are motivated by a trial of nurse-delivered psychotherapy for people with poorly controlled diabetes which used a continuous measure of treatment receipt for patients.

PC1 - M61: Why cluster size matters – lower boundaries for the expected intra cluster correlation coefficient in cluster randomized controlled trials

Authors: Hao Zhang[1], Tibor Schuster[1].
[1] *McGill University, Montréal, Canada.*

Cluster randomized controlled trials (cRCTs) are gaining increasing popularity in research studies in various domains such as global preventive health and primary care research. The intra cluster correlation coefficient (ICC) is an important parameter in the planning stage of a cRCT as it determines the necessary inflation of the required study sample size compared to a conventional RCT setting. The biggest practical challenge is the specification of a plausible range of expected ICC values to be used in the sample size calculations. It is widely accepted that, for a fixed total sample size and equal cluster sizes, an increase of the number of clusters is preferable over an increase of the cluster sizes. This notion may suggest that cluster size is a secondary parameter in the design of a study without specifically desirable constraints other than those imposed by feasibility aspects of the research.

We demonstrate that the minimum expected ICC in cluster randomized trials only depends on cluster size and not on the number of clusters. Both the theoretical derivations and the Monte-Carlo simulation studies demonstrate that the minimum expected ICC of a cRCT is equal to the inverse of the sum of cluster size and one. These findings suggest, for example, that the minimum expected ICCs for cluster sizes of 25, 50, 100 and 1000 take values of 0.038, 0.02, 0.01 and 0.001 respectively.

Designs of future cRCTs should consider cluster size as an important parameter to lower the minimum expected ICC. Failure to do so can lead to sensitive decrease of statistical power even if the ICC can be, in theory, assumed to be low. Empirical ICCs should be reported along with cluster sizes and interpreted accordingly.

Oral Contributed Sessions

OC5: STRengthening Analytical Thinking for Observational Studies (STRATOS) 1

Monday 10th July - 12.36-13.30 h. - Room: Auditorio
Chair: Bianca De Stavola

OC5-1: On 'state-of-the-art' for selection of variables and functional forms in multivariable analysis

Authors: Willi Sauerbrei[1], Patrick Royston[2], Michal Abrahamowicz[3], Aris Perperoglou[4].
[1] *Medical Center - University of Freiburg, Germany,* [2] *MRC Clinical Trials Unit London, UK,*
[3] *Department of Epidemiology and Biostatistics, McGill University, Montreal, Canada,*
[4] *Dept of Mathematical Sciences University of Essex, UK for TG2 (of the Stratos Initiative)*

Research questions have become more complex, stimulating continuous efforts to develop new and even more complex statistical methods. Tremendous progress in methodology for analyzing clinical and epidemiological studies has been made, but has it reached researchers who analyze observational studies? Part of the underlying problem may be that even experts (whoever they are) do often not agree on how to analyze a complex study in an acceptable way and on potential advantages and disadvantages of competing approaches. However, many analysts are required de facto to make important modelling decisions and would be delighted to receive help from 'state-of-the-art' documents.

In most observational studies, many variables are measured; for a multivariable analysis, selection of variables with influence on the outcome is necessary, and determination of the functional form for continuous variables is required. What would constitute a 'state-of-the-art' analysis?

More than twenty variable selection strategies have been proposed, and for dealing with continuous variables at least four strategies, with variations, are used in practice (http://stratos-initiative.org/). Regarding the latter, most analysts choose either (1) step functions (based on categorization) (2) assume that the functional form is linear (3) use fractional polynomials or (4) use one of the many approaches based on spline functions. All approaches for selection of variables and of the functional form are criticized, but for different reasons. We will discuss key issues, concluding that considerable research is required to gain better insight into advantages and disadvantages of competing strategies. Large and informative simulation studies are needed.

OC5-2: Accounting for complex measurement error in fractional polynomial models, with an application to alcohol and mortality

Authors: Christen Gray[1], Raymond J. Carroll[2], Ruth Keogh[1].
[1] *London School of Hygiene and Tropical Medicine, Department of Medical Statistics, UK,* [2] *Texas A&M, Department of Statistics, USA.*

The effect of alcohol intake on all-cause mortality is of great public health importance. Correct assessment of the relationship depends on: 1) appropriately accounting for inherent errors in the measurement of usual alcohol intake, which result in biased estimates; 2) appropriately modelling the shape of the relationship.

This work is motivated by the EPIC-Norfolk cohort where alcohol intake was measured using 7-day diet diaries on up to three occasions for 25,504 individuals. Short term assessments of alcohol intake are subject to excess zero measurements due to 'episodic consumers', in addition to other measurement error. The alcohol-mortality relationship has in some cases been found to be U- or J-shaped, though it is known that non-linear relationships typically appear more linear when there is measurement error.

Fractional polynomials are one popular method for flexibly modelling non-linear relationships. We present a method for mitigating the effects of measurement error in fractional polynomial models. The episodic consumption model (EC model) (Kipnis et al 2009) for alcohol intake measures is used to predict each individual's usual alcohol intake, which we have extended to the context of fractional polynomials. We present a likelihood-based approach and a fully Bayesian approach to estimation of the EC model and estimation of the alcohol-mortality association.

Reference 1: Kipnis V et al. Modeling Data with Excess Zeros and Measurement Error: Application to Evaluating Relationships between Episodically Consumed Foods and Health Outcomes. Biometrics 2009; 65: 1003-1010.

Reference 2: Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. Appl Stata 1994; 43: 429-67.

OC5-3: A comparison of spline methods in R for building explanatory models

Authors: Aris Perperoglou[1], Willi Sauerbrei[2], Matthias Schmid[3], Michal Abrahamowicz[4].
[1] *University of Essex, UK,* [2] *University of Freiburg, Germany,* [3] *University of Bonn, Germany,* [4] *McGill University, Montreal, Canada.*

Since the introduction of generalized additive models, splines are regularly used for building explanatory models in biomedical research. Splines offer high flexibility for modelling complex variable forms for continuous covariates, but this flexibility requires the user to have a good understanding of how to select an appropriate spline function and how to tune parameters to obtain an optimal fit. In a previous project we have identified all available R packages that can be used to fit

splines within a regression model. R packages often come with examples and vignettes, that offer users a cookbook approach in spline fitting. That may add to the confusion of non-experts. Topic Group 2 of the STRATOS initiative is working on a project to thoroughly compare approaches and provide guidance on best practice for using splines to correctly identify functional forms and variable selection in model building. In this work we will present a comparison of the most commonly used spline procedures and their corresponding packages. We will focus on thin plate regression splines, natural splines, b-splines and p-splines under mgcv, splines, rms and gamlss packages, respectively. We will use simulated and real datasets in order to investigate under which conditions specific each one of these approaches should be preferred, we will evaluate their ability to identify the correct functional form, consider prediction errors and investigate how these approaches can be used in the context of selection of variables into a multivariable model. Finally, we will compare ease of use and computational efficiency.

OC6: Bayesian methods in clinical research 2

Monday 10th July - 12.36-13.30 h. - Room: Sala Mar 2
Chair: Heinz Schmidli

OC6-1: Combining nested risk-prediction models

Authors: Philip Boonstra[1], Ryan Barbaro[1].
[1] *University of Michigan, USA.*

The notion of combining information from multiple risk prediction models is attractive but potentially challenging, because models may be based upon different sets of risk factors or covariates. In this work, we propose a Bayesian approach for constructing a prediction model for an outcome, Y, given two vectors of covariates, X and W, i.e. Y|X+W. We have access to the posterior distribution of risk factors for a nested sub-model predicting Y given X alone, i.e. Y|X, but we cannot directly incorporate this posterior from the Y|X model into the prior for the Y|X+W model. Because the former does not adjust for W, the risk factors will not in general be equal across models, either in interpretation or numeric value. We propose an approach that adjusts for these differences and thus properly incorporates the information from this sub-model into the prior for the new model. We conduct a simulation study to quantify the resulting gain in efficiency. We also demonstrate this method in our motivating example. We are seeking to identify pediatric patients who would have a high risk of dying while on ECMO, a highly invasive form of cardiac or respiratory support that can be life saving for some children. We would like to augment an existing model with a set of potentially prognostic risk factors that were not available in that model, but we only have these additional risk factors measured on a small cohort of additional patients. Our method is able to incorporate the existing model into the larger framework, thus achieving our goal of efficiently identifying risk factors and discriminating risk.

OC6-2: A Bayesian approach to modelling drug interactions

Authors: Andrea Cremaschi[1, 2], Arnoldo Frigessi[1], Kjetil Tasken[2], Manuela Zucknick[1].
[1] *OCBE, Norway,* [2] *NCMM, Norway*

The combination of two drugs administrated simultaneously may lead to different effects on the same kind of cell culture.For instance, two drugs at doses x1 and x2 providing monotherapy effects y1 = f1(x1) and y2 = f2(x2) can be combined to obtain the effect yc = fc(x1,x2),that can be representative of synergistic or antagonistic behaviour.A primary issue with this evaluation is to establish a reference value representing the condition where the compounds do not interact, also called zero-interaction level.To this aim, several heuristic approaches have been proposed that define such baseline level, and then compare it with the response from combination yc.However, these approaches rely on different modelling assumptions for the dose-response curve, and may provide conflicting outcomes in several situations.In order to overcome these issues, we propose a Bayesian regression framework for modelling the response surface when two drugs are combined, and apply it to a real dataset from cancer therapy.Posterior computations are obtained via MCMC algorithms, providing estimates of the zero-interaction level.Interestingly, the Bayesian framework allows for density estimation and prediction for those dose combinations that have not been tested.Additionally, a comparison with existing methodologies is provided.

OC6-3: Bayesian predictions for clinical trials with time-to-event endpoints

Authors: Iván Navarro[1], Rajat Mukherjee[1].
[1] *CYTEL INC. Spain.*

In clinical trials with time-to-event endpoints predicting land- mark event times and the probability of success can help in planning for interim and final analyses as well as for decision making. In this work we show under a Bayesian framework how to carry out blinded prediction of landmark event times and both blinded and unblinded predictions used for computing the probability of the trial success. The proposed Bayesian meth- ods can be used even if the trial design is a classical frequentist design.

We propose the use of piece-wise exponential model for the time-to-event endpoint while specifying Gamma process priors to the piece-wise hazards. A main advantage of the Gamma process priors are conjugacy while obtain- ing a smooth estimate for the hazard and survival functions. The general procedure involves drawing from the posterior predictive distribution in order to complete the data that may be incomplete at times where the predictions are made or at the time of an interim analysis.

Real but masked case studies will be used for illustration of the proposed computations.

OC7: Methods for handling missing data 1

Monday 10th July - 12.36-13.30 h. - Room: Sala Mar 4
Chair: Nadine Binder

OC7-1: Practical guidance for handling convergence issues in multiple imputation

Authors: Cattram Nguyen[1], John Carlin[1], Katherine Lee[1].
[1] *Murdoch Childrens Research Institute, Australia.*

Multiple imputation is a popular method for handling incomplete data problems. One of the challenges when using multiple imputation is the selection of variables for inclusion in the imputation model. The literature generally recommends including predictors of non-response and predictors of incomplete variables in the model, as well as employing an "inclusive" strategy to avoid omitting important variables. However, this inclusive approach can often lead to large imputation models and problems with convergence of estimation algorithms, especially with the popular approach of fully conditional specification or "chained equations".

We will provide an overview of strategies for handling problems with imputation model failure, after first describing methods for diagnosing issues with model convergence. Strategies for overcoming these issues include data reduction methods, augmented regression for perfect prediction, and checks for collinearity and sparse data. The main causes of convergence problems and strategies for addressing them will be reviewed and compared using simulations and case-study evaluations based on data from the Longitudinal Study of Australian Children.

Given that non-convergence of imputation algorithms is a common issue that hampers the implementation of multiple imputation, it is important that practical guidance is made available to users of this method.

OC7-2: Relative efficiency of joint-model and full-conditional-specification multiple imputation when conditional models are compatible

Authors: Shaun Seaman[1], Rachael Hughes[2].
[1] *Cambridge University, United Kingdom,* [2] *University of Bristol, United Kingdom.*

Estimating the parameters of a regression model of interest is complicated by missing data on the variables in that model. Multiple imputation (MI) is commonly used to handle these missing data. Joint model MI and full-conditional specification (FCS) MI are known to yield imputed data with the same asymptotic distribution when the conditional models of FCS are compatible with that joint model. We show that this asymptotic equivalence of imputation distributions does not imply that joint model MI and FCS MI will also yield asymptotically equally efficient inference about the parameters

of the model of interest, nor that they will be equally robust to misspecification of the joint model. When the conditional models used by FCS MI are linear, logistic and multinomial regressions, these are compatible with a restricted general location (RGL) joint model. We show that MI using the RGL joint model (RGL MI) can be substantially more asymptotically efficient than FCS MI, but this typically requires very strong associations between variables. When associations are weaker, the efficiency gain is small. Moreover, FCS MI is shown to be potentially much more robust than RGL MI to misspecification of the RGL model when there is substantial missingness in the outcome variable.

OC7-3: Using multiple imputation to account for missing data in marginal structural models: a simulation study

Authors: Emily Karahalios[1], Simon Turner[2], Dallas R English[1], Anne Kavanagh[1], Julie A Simpson[1].
[1] *The University of Melbourne, Australia,* [2] *Monash University, Australia,*

Time-dependent confounding and missing data are sources of problems in longitudinal cohort studies with repeated follow-up waves. Marginal structural models (MSMs) account for the changes in the nature of the associations between exposure and covariates over time (i.e. time-dependent confounding) to produce unbiased estimates of the effects. Multiple imputation is a method commonly used to handle missing data but this method has limited application in MSMs. Motivated by a longitudinal study of Australian adults with 13 waves of data, our interest is in estimating the net causal effect of the cumulative number of years living with disability on mental health, in the presence of missing data. We will present the results of a simulation study, based on the example, which compares the performance of using complete case analysis and multiple imputation to handle the missing data in the exposure and covariate data. Multiple imputation was implemented using fully conditional specification and the two-fold fully conditional specification algorithm, which imputes a time-dependent variable using information from only the specified and adjacent times. We generated datasets to resemble the distribution of the time-dependent exposure (i.e. presence/absence of a disability) and covariates (e.g. income, employment and relationship status), and the outcome (i.e. mental health), their inter-relationships and magnitudes of associations using the observed data. Varying proportions of missing data were imposed according to three different missingness mechanisms on the exposure of interest and the time-dependent covariates.

OC8: Rare diseases and small populations

Monday 10th July - 12.36-13.30 h. - Room: Sala Terra 4
Chair: Geert Molenberghs

OC8-1: Comparison of stratified randomization to unstratified randomization in small sample size trials with rare events on the basis of empirical type I error rates

Authors: Christina Fitzner[1], Ralf-Dieter Hilgers[1].
[1] *RWTH Aachen, Germany.*

Background: Stratified randomization is often recommended in small population trials to balance important covariates by the CPMP guidance. Consequently, a stratified randomization implies a stratified analysis. If the response is dichotomous a multivariate logistic regression model is used to analyze study data. Validity problems as complete or quasi complete separation may occur in case of small sample size trials with rare events.

Methods: In a Monte Carlo simulations study, we compare the empirical type I error rate from an analysis with exact logistic regression as well as logistic regression with Firth's bias reduction whether stratification is used in the randomization or not for small sample size trials (N<100) with rare events. Investigated randomization procedures are complete randomization, random allocation rule respectively with stratification or not, and minimization.

Results: In case of stratified data, a non-stratified analysis following a stratified randomization the empirical type I error rates are smaller compared non stratified randomization. In case of stratified analysis, empirical type I error rates do not deviate whether the randomization is stratified or not. Of course, in case of unstratified analysis and unstratified randomization, the empirical type I error rate can exceed the significance level.

Conclusion: We show for small sample size trials with rare events it is more important to use a stratified analysis than a stratified randomization considering the empirical type I error rate.

OC8-2: Assessing the design of our trial in an ultra-rare condition by the Parmar et al. framework for trials in smaller populations

Authors: Kristian Brock[1], Lucinda Billingham[1], Zsuzsa Nagy[1], Tamara Hershey[2], Darren Barton[1], Rebecca Storey[1], Timothy Barrett[1].
[1] *University of Birmingham, UK,* [2] *Washington University, USA.*

Wolfram syndrome (OMIM 222300) is an ultra-rare, monogenic, neurodegenerative disorder of young people with no effective treatment. We designed an international, double-blind RCT of sodium valproate, hypothesised to slow disease progression. Co-primary outcomes are visual acuity (VA), measured on the

logMAR scale; and ventral pons volume (VPV), measured by MRI. Hypothesis testing by conventional design was infeasible due to disease rarity. We will randomise 70 patients 2:1 to valproate vs placebo. Longitudinal analysis by mixed effects models will yield 80% power to detect 50% reduction in the natural rate of deterioration in VA, and 86% power to detect 70% reduction in deterioration of VPV, each at 5% significance. Calculations assume 15% of data is lost. After our submission, Parmar, Sydes & Morris published a framework on designing RCTs in smaller populations. We assess our design by their framework.

Firstly, PSM advocate increasing what is feasible. Using longitudinal analysis was vital and detecting clinically meaningful differences would necessitate three years' follow-up. Eligibility criteria could not be broadened. International recruitment was necessary.

PSM then advocate commonly considered approaches to attain feasibility. Only one experimental treatment was available. Analysis of Prof Hershey's patient data revealed two outcomes most amenable to analysis. One of these (VA) was also highly prioritised by patients. We relaxed power to $\geq$ 80%. We required none of PSM's less common approaches like relaxing alpha or one-sided tests.

Encouragingly, our design has high affinity with PSM's framework.

Reference 1: Parmar, Sydes & Morris (2016) How do you design randomised trials for smaller populations? A framework BMC Medicine

OC8-3: Statistical properties of hypothesis tests for a goal attainment scaling endpoint

Authors: Susanne Urach[1], Martin Posch[1], Robin Ristl[1], Gerd Rosemkranz[1], Kit Roes[2], Bernd Jilma[1], Hanneke Van Der Lee[3], Charlotte Gaasterland[3].
[1] *Medical University of Vienna, Austria,* [2] *University Medical Center Utrecht, The Netherlands,* [3] *University Of Amsterdam, The Netherlands.*

Goal Attainment Scaling (GAS) is a measurement instrument to evaluate the effect of anintervention on the basis of individual, patient-specific goals. The effect of a treatment on a particular goal is mapped in a pre-specified way to a common ordinal scale. The advantages of this measurement approach are the utilization of patient-centered outcomes and the possibility to combine the information from patients in heterogeneous populations. The latter is of particular interest in rare disease research, because it allows for as large as possible samples. Here we focus on the statistical aspects of using GAS data for the comparison of two treatment groups in a randomized clinical trial. A data generating model is set up based on the assumption of underlying latent multivariate normal responses, which are assumed to be connected with the unknown treatment effect on some common underlying physiological process. The actual ordinal observations are obtained by discretizing these continuous outcomes via thresholds. An extensive simulation study was carried out to find the optimal weighting strategy and parameter settings for GAS data fullfilling the proposed model assumptions. We discuss the scope of possible null hypotheses for the between-group comparison and review methods to aggregate the data on multiple goals within each patient. The results will be illustrated with a clinical trial example concerning children with cerebal palsy. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013-603160. ASTERIX Project - http://www.asterix-fp7.eu

OC9: Survival analysis 1

Monday 10th July - 12.36-13.30 h. - Room: Sala Terra 2
Chair: Dimitris Rizopoulos

OC9-1: Flexible hazard rate estimation with interval sampling

Authors: Carla Moreira[1, 2], Jacobo De Uña-Álvarez[2].
[1] *Institute of Public Health, University of Porto, Portugal,* [2] *University Of Vigo, Spain.*

In Survival Analysis, the observed lifetimes often reduce to those individuals with event (infection, death and so on) occurring within a specific calendar time interval, leading to the so-called interval sampling scenario (Zhu and Wang 2014). A typical target is the hazard rate function, which represents the instantaneous probability of the event of interest. In this work we introduce a flexible estimation approach for the hazard rate under interval sampling. Following Moreira and Van Keilegom (2013), a kernel smoother is considered, in both a fully nonparametric setting and a semi-parametric setting in which the incidence process fits a certain parametric model. An asymptotic expression of the mean integrated squared error is derived, leading to a data-driven bandwidth for the estimator. Properties of the proposed method are investigated both theoretically and through simulations. Applications to medical data are included.

Reference 1: Moreira, C. and Van Keilegom, I. (2013): Bandwidth Selection for Kernel Density Estimation with Doubly Truncated Data. Computational Statistics and Data Analysis, 61, 107-123.

Reference 2: Zhu H., Wang, M-C (2014): Nonparametric inference on bivariate survival data with interval sampling: association estimation and testing. Biometrika, 101, 1-15.

OC9-2: Solving the Fine-Gray riddle

Authors: Hans Van Houwelingen[1], Hein Putter[1].
[1] *LUMC, Netherlands.*

The Fine-Gray approach to the modelling of competing risks data is based on a proportional hazards model for the sub-distribution hazard of the cumulative incidence function. This approach has puzzled the bio-statistical community for three reasons:

1. The sub-distribution "hazard" is hard to interpret.
2. It cannot be used dynamically.
3. The proportional hazards model is fitted by means of the partial likelihood, where the risk set includes those patients who had no event yet and those who had already "died" from one of the other competing risks.

The alternative approach is multi-state models. The advantage is that the cause-specific hazards in a multi-state model are easier to understand. The disadvantages are that the computations and the prognostic models get complicated in case of more than two competing risks

In this presentation we introduce a different approach. It is based on decomposing the sub-distribution hazard in two multiplicative components:

Sub-distribution hazard (t)= r(t) · cause-specific hazard (t)

Here, r(t) is the reduction factor that describes which fraction of the population is still at risk for the competing risk of interest. Somehow it summarizes all information from the multi-state that is relevant for this specific competing risk.

Different models for the effects on covariates x on r(t) (r(t|x)) will be presented and it will be shown how simple models for the cumulative incidence function can be obtained without having "dead" individuals in the risk set.

Reference 1: Fine, J. P. & Gray, R. J. (1999), `A proportional hazards model for the Sub-distribution of a competing risk', Journal of the American Statistical Association 94, 496-509.

OC9-3: Two approaches to estimation of the subdistribution hazard with time-varying covariables

Authors: Ronald Geskus[1, 2, 3, 4].
[1] *Oxford University Clinical Research Unit, Centre For Tropical Medicine, Ho Chi Minh City, Vietnam,*
[2] *Nuffield Department Of Clinical Medicine, Oxford University, Oxford, United Kingdom.*
[3] *Academic Medical Center, University Of Amsterdam, Amsterdam, The Netherlands,*
[4] *Public Health Service Of Amsterdam, Amsterdam, The Netherlands.*

In the classical survival setting with a single event type, the relation between a time-varying covariable and the hazard is easy to conceptualize. For estimation, the changing value of a covariable can be represented by creating pseudo-individuals, such that each row represents a period during which the value remains constant. The start of this interval can be seen as a form of late entry. It has been called internal left truncation, because the individual was already under observation before. In the presence of competing risks, the same applies for the cause-specific hazard. Things are different for the subdistribution hazard, the hazard that uniquely determines the cause-specific cumulative incidence. Since an individual that experiences a competing event remains in the risk set, the definition of an internal time-varying covariable may become problematic: we don't have any value after he has died. With late entry and right censored data, the subdistribution hazard is estimated using inverse probability weights. This allows for two different approaches in the presence of time-varying covariables. In the pseudo-individual approach, we interpret the rows with constant value as coming from different persons. We use weights to correct for the late entry and right censoring as induced by the interval of constant value. Importantly, the covariable itself becomes time-fixed. In the internal approach, we consider the rows as continuing follow-up from the same individual. No weights are used while he is under event-free observation. Using a simple example, we contrast the interpretation of the estimates as obtained via both approaches. We explain why the pseudo-individual approach is more realistic and feasible in general.

Invited Session

IS2: Methods in Research on Research

Monday 10th July - 15.00-16.30 h. - Room: Auditorio
Chair: Els Goetghebeur
Organised by Els Goetghebeur, *Ghent University, Belgium*

IS2-1: Lessons from Meta-Epidemiological Assessments

John Ioannidis, *Stanford University, USA*

There is increasing interest in understanding research methods, their performance and their biases by empirical evaluations of large numbers of studies and data. These assessments are often called meta-epidemiological, since they are typically observational studies where the unit of analysis may be a study, dataset, protocol, grant, or other research document or piece of information. A large and expanding literature of meta-research is currently being published, with currently over 2,000 articles published every year. Most of these are conceptual and theoretical or simulation work, but also many empirical investigations are part of this corpus. The talk will discuss the typical strengths and limitations of meta-epidemiological assessments, the types of designs used, and cutting-edge issues in quantitative meta-meta-analyses and will provide some examples of potentially fruitful applications. It will also discuss how advances in data sharing, open science, transparency, and large-scale evidence may improve some of these efforts.

IS2-2: The Peer Review Process. Why Evidence Based Practices Are Needed

Isabelle Boutron, *Paris Descartes University, France*

The peer review process is a cornerstone of biomedical research publication. Nevertheless, the effectiveness of the system is questioned. Worldwide, peer review costs an estimated £1.9 billion annually and accounts for about onequarter of the overall costs of scholarly publishing and distribution. The human cost was estimated at around 15 million hours by 2010. Furthermore, the peer-review system may fail to identify important flaws in the quality of manuscripts and published articles.

In this presentation, we will highlight the limitations of the current system and the need to have an evidence-based approach to improve the system.

IS2-3: Twisting "Spin": How to Avoid or Attenuate Exaggerated Claims in Reports of Biomarker Studies

Patrick Bossuyt, *University of Amsterdam, Netherlands*

Researchers sometimes cannot avoid the tendency to report a completed study with a rosy presentation or a generous interpretation of the study findings. Several studies have now systematically documented manifestations of "spin", in randomized clinical trials, non-randomized studies, diagnostic accuracy studies, and others.

In this presentation, we present a grouping of "spin control" practices. One category consists of selective or incomplete reporting, with an emphasis on secondary outcomes or on results in subgroups, for example. Another manifestation is a selective emphasis in the title or abstract, with a mismatch between the body and the abstract results, with stronger wording in title or abstract.

A second set of practices relates to the interpretation of study findings. Frequently observed are mismatches between the study design and the main conclusion of the study, and between the study findings and the main conclusion. Other manifestations of "spin" are unwarranted generalizations, or recommendations based on an incomplete balancing of benefits and harms.

As spin can harm patients and constitutes a source of avoidable waste in research, it should be attenuated or avoided altogether. We present a series of strategies to reduce mismatches and counter spin control by authors. These include the strengthening of peer review processes, the construction of balanced research teams, and suggestions to move control over some elements of the study report, such as title and the abstract, from authors to other scientists, without conflicts of interest.

Oral Contributed Sessions

OC10: Biostatistics for high dimensional data 1

Monday 10th July - 15.00-16.30 h. - Room: Sala Terra 2
Chair: M. Luz Calle

OC10-1: Regularized cause-specific hazards models for competing risks

Authors: Maral Saadati[1], Jan Beyersmann[2], Annette Kopp-Schneider[1], Axel Benner[1].
[1] *German Cancer Research Center (DKFZ), Division Of Biostatistics, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany,* [2] *University Of Ulm, Institute Of Statistics, Helmholtzstr. 20, 89081 Ulm, Germany,*

We investigate regularized competing risks models given a high-dimensional covariate space. In particular, we compare penalized cause-specific hazards (CSH) approaches and subdistribution hazards (SDH) models with lasso penalties. Our focus is two-fold: (1) variable selection and (2) predictive accuracy.

Furthermore, a new method is proposed as a compromise between independently penalized CSH models and the SDH approach. The idea is to find optimal tuning parameters of the CSH model with respect to best prediction of the cumulative incidence of the event of interest. Simulation experiments and a real-life bladder carcinoma data set serve to contrast the given methods.

We conclude that penalized CSH models can generally be useful in high-dimensional competing risks settings. As most of the current literature for competing risks in high-dimensions is concerned with SDH models, one should keep in mind the immense advantages of penalized CSH models. Firstly, CSH methods provide a powerful tool to further our understanding of biological mechanisms that govern the transition hazards, which is particularly desirable when analyzing molecular data. Secondly, penalized CSH models can be extended to penalized multi-state models.

OC10-2: Detecting interactions in GWAS with the gene-gene eigen-epistasis approach

Authors: Virginie Stanislas[1], Cyril Dalmasso[1], Christophe Ambroise[1].
[1] *Université d'Evry-Val-d'Essonne, France*

In past years, numerous methods have been proposed for studying epistatic interactions in genome-wide association studies (GWASs). They vary in terms of data analysis and statistical methodology. Most of them focus on single locus interactions, but considering interactions at gene level may offer many advantages. In the past few years several gene-gene methods have been proposed, they rely on a summarizing step to obtain information at the gene level and a modeling phase to represent interactions. For the most recent methods, filters or penalized models are used to make the method applicable to a large number of genes.

Here we propose a new approach that takes into account the group structure of each gene in order to detect epistasis at the gene level. We introduce the Gene-Gene Eigen Epistasis (G-GEE) [1] as a new approach to compute the gene-gene interaction part of the model. The method first compute interaction variables for each gene pair by finding its Eigen-epistasis Component defined as the linear combination of Gene SNPs having the highest correlation with the phenotype. The selection of the significant effects results from a group LASSO penalized regression method combined to an adaptive ridge approach [2] allowing to control the False Discovery Rate.

We conduct a simulation study to compare G-GEE with recent alternative proposals and demonstrate the power of our approach by detecting new gene-gene interactions on genome-wide association studies.

Reference 1: Stanislas V, Dalmasso C, Ambroise C. Eigen-Epistasis for detecting gene-gene interactions. BMC Bioinformatics. 2017; 18:54

Reference 2: Bécu JM, Grandvalet Y, Ambroise C, Dalmasso C. Beyond Support in Two-Stage Variable Selection. Statistics and Computing. 2017;27:169–79

OC10-3: Efficient generalized canonical correlation method to perform multiomic data integration of big data matrices with missing individuals

Authors: Juan R. González[1, 2, 3], Harry Odell[1, 2, 3], Isaac Subirana[4].
[1] *Bioinformatic Research Group In Epidemiology (BRGE),* [2] *Barcelona Institute For Global Health (Isglobal),* [3] *Ciber Epidemiología y Salud Pública (CIBERESP), Spain,* [4] *Cardiovascular Epidemiology and Genetics Research Group, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain.*

Principal component analysis (PCA) is one of the standard methods to visualize big genomic, transcriptomic or epigenomic data. Current personalized medicine studies include more than one omic data. Therefore, extensions of PCA methods have been proposed to properly describe how omic features relate to each other. These methods include Generalized Canonical Correlation Analysis (GCCA). As in the case of PCA, the methods for parameter estimation of GCCA are straightforward since they rely on performing standard data matrix operations. However, a practical limitation appears when analyzing whole genome data. In that case, making computations with big matrices can be almost impossible using standard software. Another important problem arises when combining omic data from individuals that have been obtained from different projects. In that case it is normal to have missing data of some individuals for a particular omic since, for instance, the budget only allows getting data of a new omic for a subset of samples. As with most of the omic integration methods, GCCA requires complete cases of data analysis that normally reduces the statistical power. In order to overcome this limitation and provide an efficient tool to integrate multiple omic data, we have created an R/Bioconductor package that implements a method to deal with missing individuals by using parallel libraries to speed up the computing processes. Our method has been evaluated through simulation studies and we show that it increases the power with respect to methods based on analyzing complete cases. Our tool has also been used to analyze real data of TCGA project where different types of omic data are available having different levels of missingness.

OC10-4: Statistical methods for detecting gene-gene and gene-environment interactions in genetic association studies

Authors: Marianne Jonker[1], Armin Rauschenberger[2], Renee X. De Menezes[2], Jakub Pecanka[3].
[1] *Department for Health Evidence, Section Biostatistics, Radboudumc Nijmegen, The Netherlands,* [2] *Department of Epidemiology and Biostatistics, Vu University Medical Center Amsterdam, The Netherlands,* [3] *Department Of Medical Statistics And Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands.*

Undiscovered gene-gene and gene-environment interactions are possible explanations for the "missing heritability" of complex traits and diseases. A simple full screen for interaction effects on a genome-wide scale is still unfeasible: testing all SNP-SNP interactions for association with the disease of interest is computationally too expensive and multiple testing correction will wipe out all effects. In this presentation we explain some novel statistical methods to identify SNP-SNP and SNP- environment interactions that can be used on a genome wide scale. For example, several researchers proposed to use a two-step approach (e.g. Pecanka et al (2017) for an overview and new methods). In the first step, pairs of SNPs that are a priori more likely to be involved in SNP-SNP interactions are ascertained and only these are considered when testing for association with a complex phenotype. Rauschenberger (2017) developed a statistical method to detect SNPs that are involved in interactions by fitting a semi-supervised mixture model and testing for interaction within this framework.

Reference 1: Pecanka J, Jonker MA, Bochdanovits Z, van der Vaart AW (2017). A powerful and efficient two-stage method for detecting gene-to-gene interactions in GWAS. Accepted for publication in Biostatistics.

Reference 2: Rauschenberger A, Menezes RX, van de Wiel MA, van Schoor NM, Jonker MA (2017). Detecting SNPs with interaction effects on a quantitative trait. Preprint.

OC10-5: Partitioned learning of deep Boltzmann machines for linking complex SNP patterns to clinical endpoints

Authors: Harald Binder[1], Stefan Lenz[1], Tamara Blätte[2], Lars Bullinger[2], Moritz Hess[1].
[1] *University Medical Center Mainz, Germany,* [2] *University Hospital of Ulm, Germany.*

Deep learning approaches have become popular for learning complex structure in data. Yet, there are hardly any applications to high-dimensional molecular data. Furthermore, such approaches mostly are applied as black box tools for prediction, with few attempts to extract the learned structure, e.g. for investigating links to clinical endpoints. For modeling the connection between single nucleotide polymorphism (SNP) measurements and clinical phenotypes, we address both fitting to high-dimensional data as well as pattern extraction based on deep Boltzmann machines (DBMs). Specifically, we propose a sparse regression approach to coarsely screen the joint distribution of SNPs, followed by training several DBMs on SNP partitions that were identified by the screening. Aggregate features representing SNP patterns and the corresponding SNPs are extracted from the DBMs by a combination of statistical tests and sparse regression. In simulated case-control data, we

show how this can uncover complex SNP patterns and augment results from univariate approaches, while maintaining type 1 error control. Time-to-event endpoints are considered in an application with acute myeloid leukemia patients, where SNP patterns are modeled after a pre-screening based on gene expression data. The proposed approach identified three SNPs that seem to jointly influence survival in a validation data set. This indicates the added value of jointly investigating SNPs compared to standard univariate analyses and makes partitioned learning of DBMs an interesting complementary approach when analyzing SNP data.

OC11: Joint modelling in practice 1

Monday 10th July - 15.00-16.30 h. - Room: Sala Mar 2
Chair: Koos Zwinderman

OC11-1: Joint modelling of multiple longitudinal markers for clinical classification purposes

Authors: David Hughes[1], Marta García-Fiñana[1], Cheyne P. Chris[1], Arnošt Komárek[2], Simon P. Harding[1].
[1] *University of Liverpool, UK,* [2] *Charles University, Czech Republic.*

Frequent clinical interest is in being able to classify patients into various groups corresponding to severity of their disease status or disease progression, based on the evolution of biomarkers observed over time. We have recently developed a flexible and dynamic approach [1] in which available longitudinal data on multiple biomarkers of various types can be used to accurately classify patients into groups (such as diagnosis groups).

In this talk I will provide an overview of the multivariate longitudinal discriminant analysis (LoDA) method where longitudinal markers of different types (continuous, counts, binary) are jointly modelled. We will explore the benefits of taking into account the level of uncertainty of group membership probabilities to improve classification (promising results have already been shown when applied to predict patients with epilepsy who fail to achieve adequate seizure control).

This methodology will be illustrated using data from patients with diabetes with the aim of predicting patients who develop sight threatening diabetic retinopathy (STDR) within a year. Diabetic retinopathy is a common complication of diabetes and early detection of STDR is important to avoid significant vision loss. Levels of retinopathy are modelled using bivariate linear mixed-effects models that are subsequently used in the multivariate discriminant scheme. The allocation rules and classification accuracy will be presented and discussed.

Reference 1: Hughes D, Komárek A, Czanner G, García-Fiñana M. (2016) Dynamic longitudinal discriminant analysis using multiple longitudinal markers of different types. Statistical Methods in Medical Research 2016 (available on–line ahead of print).

OC11-2: Analysing unmeasured baseline covariates in studies with delayed entry using a joint model

Authors: Regina Stegherr[1], Jan Beyersmann[1], Peter Bramlage[2], Tobias Bluhmki[1].
[1] *Institute of Statistics Ulm University, Germany,* [2] *Institute of Pharmacology and Preventive Medicine, Germany.*

The natural choice for 'time zero' (baseline) in a randomized clinical trial is study entry, in particular, covariate information is available at that time point. A major challenge in observational cohorts is that study entry may happen some time after time-origin leading to left-truncated data. One specific example is the analysis of diabetes register-based data, where a relevant timescale is 'time-since-first-antidiabetic-medication'. Since such data is collected in calendar time, some patients may enter the study upon their first medication, but others may have a known date of therapy initiation before start of data collection. A relevant baseline covariate would be glycated haemoglobin (HbA1c) level in diabetes patients. The challenge is that such data is typically measured upon study entry and, hence, notat baseline for those with a delayed study entry, but, e.g., HbA1c will have changed in the random and patient-specific time interval between start of medication and study entry. The problem has been succinctly summarized in a letter by Keiding and Knuiman, Statistics in Medicine, Vol. 9, 1221-1222 (1990). In this work, we propose a joint model to investigate the impact of the baseline value of a longitudinal marker, possibly unmeasured due to delayed entry, on the time to the event of interest. This is in contrast to the standard use of a joint model where the typical aim is to analyse the effect of the current value of the marker on the hazard of an event. Our approach shows proper performance in an extensive simulation study and was applied to data from a German diabetes register. The aim was to evaluate the effect of the HbA1c level at therapy initiation on the risk of failure of standard antidiabetic therapy.

OC11-3: Individualized dynamic prediction of survival under time-varying treatment strategies

Authors: Grigorios Papageorgiou[1], Dimitris Rizopoulos[1], Mostafa M. Mokhles[2], Johanna J. M. Takkenberg[2].
[1] *Department of Biostatistics, Erasmus MC, The Netherlands,* [2] *Department of Cardio-Thoracic Surgery, Erasmus MC, The Netherlands.*

Our work is motivated by a study conducted at the department of Cardio-Thoracic Surgery of the Erasmus University Medical Center in the Netherlands. This study concerns patients who received an allograft for Right Ventricular Outflow Tract (RVOT) reconstruction after previous Tetralogy of Fallot (ToF) correction and were thereafter monitored echocardiographically. Cardio-thoracic surgeons are interested in studying the change in the longitudinal profile of the echocardiography measurements after RVOT reconstruction, and utilizing this change in obtaining more accurate risk probabilities of survival for these patients. To achieve this goal we propose here a flexible joint modeling framework for the longitudinal echocardiography measurements and the hazard of death that includes RVOT as a time-varying binary covariate in both the longitudinal and survival submodels. We consider a set of joint models that postulate different effects of RVOT in the longitudinal profile and the risk of death, and different formulations of the association structure. Based on these models we derive dynamic predictions of conditional survival probabilities, adaptive to time-varying RVOT reintervention strategies. The predictive accuracy of these

predictions is evaluated with a repeated cross-validation procedure using a time-dependent ROC analysis. The results suggest that it is important to account for the change in the longitudinal profiles of RVOT.

OC11-4: A novel approach to joint modelling inflammatory markers through flexible copula regression models

Authors: Jenifer Espasandín-Domínguez[1], Carmen Cadarso-Suárez[1], Arturo González-Quintela[1], Vanessa Allende[1], Óscar Lado-Baleato[1], Francisco Gude[1].
[1] *University of Santiago de Compostela, Spain.*

There is accumulating evidence that inflammation is an important risk factor in cardiovascular disease (CVD) and diabetes.

There are two widely used methods for detecting the inflammatory reaction in the clinical practice -erythrocyte sedimentation rate (ESR) and the c-reactive protein (CRP). However, CRP and ESR may not correlate highly in some conditions or diseases, and the extent to which CRP and ESR results agree with each other in the population is not clear. To better understand the possible causes of the CRP/ESR discordance, flexible copula regression models for bivariate responses are then proposed.

Specifically, the bivariate copula additive models for location, scale and shape (CGAMLSS, Marra and Radice, 2016) will be used. This modern approach extends the use of GAMLSS regression models to situations in which each parameter of a multivariate response is modeled simultaneously conditional on some covariates using different copula functions. This approach permits the multiple responses and copula parameter to be modelled using additive predictors that allow for several types of covariate effects: nonlinear effects of continuous covariates, random effects or interactions.

In our biomedical study, the use of CGAMLSS revealed hitherto unreported effects about the association between ESR and CRP: such association varies not only with age and body mass index, but also with important clinical variables as the hematocrits levels of the patient.

Reference 1: Marra G., Radice R., (2016). A Bivariate Copula Additive Model for Location, Scale and Shape. Cornell University Library. arXiv:1605.07521 [stat.ME].

OC11-5: Analysis of complex longitudinal data arising from multistage interventions

Authors: Jack Wilkinson[1], Andy Vail[1], Stephen A. Roberts[1].
[1] *University of Manchester, United Kingdom.*

Longitudinal data usually comprises repeated observations of a particular outcome measure on study participants. However, multistage treatments such as in vitro fertilisation (IVF) give rise to sequences of different outcome measures, which include different types of response variables (continuous, count, ordinal etc). In IVF, the patient's ovaries are stimulated for a period culminating

in the collection of eggs. These are then mixed with sperm and are cultured for several days to allow them to develop as embryos. One or more embryos are then selected to be transferred to the patient's uterus, where it is hoped that they will implant and develop into a healthy baby. This results in a longitudinal sequence of mixed outcome types (number of eggs collected, fertilisation rate, measures of embryo quality, type of transfer procedure, successful birth). Further complexity is introduced both by the fact that embryo quality is measured at the level of the individual embryo rather than at the level of the patient, and by the fact that attrition occurs as patients pass through each stage.

We present flexible joint modelling approaches for complex longitudinal data of this ilk, with applications to routinely collected IVF data. We include distinct submodels for each outcome measure and induce dependency between these through correlated random effects or by including outcomes as covariates in submodels for downstream responses. We simultaneously model the dropout process, allowing for different dropout mechanisms at each stage. The models can be fitted in standard Bayesian software (rstan).

OC12: Statistical methods for systematic reviews

Monday 10th July - 15.00-16.30 h. - Room: Sala Mar 4
Chair: Stephen Senn

OC12-1: Meta-analysis of external validation studies

Authors: Thomas Debray[1], Johanna Damen[1], Richard Riley[2], Gary Collins[3], Karel Moons[1].
[1] *Julius Center for Health Sciences and Primary Care, Netherlands,* [2] *Keele University, UK,* [3] *University Of Oxford, UK.*

It is widely recommended that developed prediction models are externally validated across different settings and populations. When multiple validations have been performed, a systematic review may help to understand whether and under what circumstances the model remains accurate or requires further improvements. Recently, we described statistical methods for extracting and summarizing performance estimates of prediction models with a binary outcome (BMJ, 2017). In this work, we discuss how to summarize the performance of prediction models with a time-to-event outcome. Hereto, we present statistical methods for dealing with incomplete reporting, and to obtain time-specific summary estimates of the c-statistic, the calibration-in-the-large and the calibration slope. In addition, we provide guidance on the implementation of a Bayesian estimation framework, and discuss several prior distributions. Finally, we illustrate all methods in two example reviews where we evaluate the predictive performance of EuroSCORE II and Framingham Wilson.

Reference 1: Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, Riley RD, Moons KG. A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017 Jan 5;356:i6460.

OC12-2: Random (Bayesian) main effect of treatment meta-analysis

Authors: Stephen Senn[1], Susanne Schmitz[1], Anna Schritz[1], Samir Salah[2].
[1] *Luxembourg Institute Of Health, Luxembourg,* [2] *L'Oreal, France.*

The term random effect is ambiguous when used to qualify meta-analysis[1]. It usually refers to a random treatment-by-trial interaction but it sometimes refers to a random main effect of trial. There is, however, a third type of random effect that is possible: the random main effect of treatment. Of course, a Bayesian analysis, has a random effect in the form of a prior distribution for every parameter, including the main effect of treatment, but even where a network meta-analysis is carried out, Bayesian practitioners have usually chosen independent uninformative priors for treatments and this means that the shrinkage of treatment effects does not occur and results may be too optimistic.Taking the example of a network meta-analysis of 44 treatments in 10 trials, we illustrate how a hierarchical approach to modelling a random main effect of treatment can be used to produce shrunk (towards the overall mean) estimates of effects for individual treatments. As a related problem we also consider the issue of using a random effects model for the within-trial variances from trial to trial.We provide a number of possible graphical representations of the results and discuss the advantages and disadvantages of such an approach.

Reference 1: Senn, S., A note regarding 'random effects'. Statistics in Medicine, 2014. 33(16): p. 2876-7.

OC12-3: A comparison of arm-based and contrast-based approaches to network meta-analysis

Authors: Ian White[1], Rebecca Turner[2].
[1] *UCL, MRC Clinical Trials Unit, UK,* [2] *University Of Cambridge, MRC Biostatistics Unit, UK*

Network meta-analysis is an effective and increasingly popular way to combine the evidence on all available treatments in order to guide clinical decision making. The choice between arm-based and contrast-based approaches to network meta-analysis has recently become controversial. Both approaches, in this debate, model arm-level data (for example, using a binomial likelihood for binary outcome data). However, the contrast-based approach has separate submodels for (i) the mean outcome under reference treatment and (ii) contrasts between treatments, whereas the arm-based approach jointly models all the arm-level mean outcomes.

This talk describes and assesses the claims made by each approach. First, we show that all arm-based models, and some contrast-based models, use between-study information (they "break randomisation"), but that the contribution of this information is typically very small. Second, we show that both approaches make a missing at random assumption, but the nature of this assumption differs between different models and may be more plausible in arm-based models. Finally, we discuss claims made about estimands and show that useful estimands can be reported from each approach.

To explore how big the differences between approaches can be, we use analyses of artificial data. We show that the approaches primarily differ when the underlying mean outcome is associated with both the treatments studied in a trial and the trial-specific treatment effects. This setting therefore requires special care in choosing an analysis model.

OC12-4: Bivariate meta-analysis for two binary outcomes with low incidence: a bivariate beta-binomial model using copulas

Authors: Yusuke Yamaguchi[1], Kazushi Maruo[2], Richard D. Riley[3].
[1]*Astellas Pharma INC., Japan,* [2]*National Center Of Neurology and Psychiatry, Japan,* [3]*Keele University, UK.*

A strength of meta-analysis is its capability of accumulating evidence from several studies. For example, if the interest is in safety assessments regarding serious drug-induced adverse events with very low incidence (e.g. 0.5%), the meta-analysis enables estimations and quantifications of the risk of such rare adverse events in situations where individual studies do not have a large enough sample size for reliable evaluations. However, a concern may arise due to the occurrence of studies with no event, where traditional approaches of adding a correction factor or omitting these studies are known to result in misleading conclusions. Moreover, studies involved in the analysis often report results for more than one outcome, though there has not been sufficient discussion on a multivariate extension in the context of meta-analysis for rare events. We consider a joint synthesis of two binary outcomes with low incidence, and propose a novel bivariate meta-analysis method using copulas. The method assumes marginal beta-binomial models for the two outcomes, and links these margins by a bivariate copula which identifies an overall dependence structure between outcomes. The use of copulas allows a flexible modeling of dependences, which enables us to obtain an unbiased estimation for the incidence of rare events, as well as potential benefits of bivariate meta-analysis such as a reduction of outcome reporting bias. We illustrated the method through an application to a real example, and performed a simulation study to examine a relative performance of the method in comparison to existing approaches. We discuss a possibility of expanding the method to indirect and mixed treatment comparisons.

OC12-5: Machine learning methods for systematic literature reviews: automated abstract selection based on study design

Authors: Bryony Langford[1], Sara Steeves[1], Stephanie Beaver[1].
[1]*Costello Medical Consulting LTD, United Kingdom.*

Objectives: Screening abstracts for eligibility in systematic reviews is resource-intensive. We compared various classifiers for automating abstract screening based on study design, and determined how many training examples would be required for classification accuracy to stabilise.

Methods: Multinomial naïve Bayes, logistic regression and linear support vector machine classifiers were applied to abstracts of three study designs: randomised controlled trial (RCT, n=3215), observational study (n=594) and economic evaluation (n=914). 75% of the abstracts were reserved for training, parameter tuning and selecting the best performing classifier (BPC). Due to imbalances in the numbers of abstracts between groups, resampling with replacement was used to increase the numbers of observational study and economic evaluation abstracts to match the number of RCT abstracts. The final accuracy of the BPC was calculated on the remaining 25% of abstracts (test set). Accuracy was measured using the F2 score, which favours false positives over false negatives. To evaluate variation in performance with training set size, the classifier was trained with truncated versions of the training set, and evaluated on the full test set.

Results: The BPC (logistic regression) achieved F2 scores of 0.98 and 0.93 for RCTs and economic evaluations, reached after around 1000 and 4000 abstracts, respectively. The score for observational studies fluctuated, reaching 0.83 on the full training set (n=7266).

Conclusions: It is possible to sort abstracts by study design with reasonable accuracy. RCTs and economic evaluations could be classified particularly well, likely as these had more training abstracts compared to observational studies.

OC13: Causal inference and mediation analysis 1

Monday 10th July - 15.00-16.30 h. - Room: Sala Terra 4
Chair: Helene Jacqmin-Gadda

OC13-1: Multiple questions for multiple mediators

Authors: Bianca L. De Stavola[1], Rhian M. Daniel[1], George Ploubidis[2].
[1] *London School of Hygiene and Tropical Medicine, UK,* [2] *UCL Institute of Education, UK.*

Investigating the mechanisms that may explain the causal links between an exposure and a temporally distal outcome often involves multiple interdependent mediators. Until recently, dealing with multiple mediators was restricted to settings where mediators relate to exposure and outcome only linearly. Extensions proposed in the causal inference literature to allow for interactions and non-linearities in the presence of multiple mediators initially focussed on natural direct and indirect effects. These however are not all identifiable, with the rest requiring stringent, and often unrealistic, assumptions. More recent developments have focussed interventional (or randomised) direct and indirect effects to deal with these issues (Vansteelandt and Daniel, 2017). They can be identified under less restrictive assumptions, with generalizations dealing with time-varying exposures, mediators and confounders also possible (VanderWeele and Tchetgen Tchetgen, 2017).

The mediation questions that can be addressed when estimating interventional effects differ from those asked by natural effects in subtle ways. In this talk we will review them, discuss their differences in emphasis, assumptions, and interpretation, and propose ways of exploiting thesedifferences to assess the robustness of conclusions. We will use an epidemiological investigation of the mechanisms linking maternal pre-pregnancy weight status and offspring eating disorders behaviour to illustrate these points.

Reference 1: VanderWeele TJ, Tchetgen Tchetgen EJ. Mediation analysis with time-varying exposures and mediators. J R Stat Soc B (in press).

Reference 2: Vansteelandt S; Daniel RM. Causal mediation analysis with multiple mediators Epidemiology, 2017; 28 (2): 258–265.

OC13-2: Mediation analysis for trials using Stein-like estimators with instrumental variables

Authors: Sabine Landau[1], Cedric Ginestet[1], Richard Emsley[2].
[1] *King\'S College London, UK,* [2] *University of Manchester, UK.*

Causal mediation analysis aims to estimate natural direct and indirect effects under clearly specified assumptions. Traditional mediation analysis in trials based on Ordinary Least Squares (OLS) relies on the absence of unmeasured common causes of the putative mediator and the clinical outcome. Instrumental variables estimators such as Two-Stage Least Squares (TSLS) have been proposed to produce asymptotically unbiased estimators when this assumption cannot be justified. However, bias removal comes at the cost of variance inflation. A Semi-Parametric Stein-Like (SPSL) estimator strikes a natural trade-off between the unbiasedness of the TSLS procedure and the relatively small variance of the OLS estimator. Moreover, the SPSL also has the advantage that its shrinkage parameter is automatically estimated from the data. We will demonstrate how this Stein-like estimator can be implemented in the context of causal mediation analysis. The performance of the competing methods is studied in a simulation study, in which both the strength of the statistical issue (hidden confounding) and the strength of the solution (predictive power of the instruments) are varied. These considerations are motivated by a trial in mental health evaluating the impact of a primary care-based intervention to reduce depression in the elderly. In this trial we are interested in determining the part of the causal effect of the intervention that comes about for reasons other than changes in the adherence with prescribed antidepressant treatment.

Reference 1: Ginestet C., Emsley R. and Landau S. (2017) Dose-response modeling in mental health using Stein-like estimators with instrumental variables. Statistics in Medicine, in press

OC13-3: Bias, bounds and sensitivity analysis for unobserved confounding when using regression imputation and double robust estimators

Authors: Minna Genbäck[1], Xavier de Luna[1].
[1] *Umeå University, Sweden.*

When estimating average causal effects of a treatment with observational data, scientists often rely on an assumption of no unobserved confounders. We propose a sensitivity analysis to unobserved confounders, for outcome regression imputation estimators and doubly robust estimators, based on identification bounds for the causal effect of interest. The bounds are derived from the bias of the estimators, expressed as a function of a sensitivity parameter. We describe how such bounds together with sampling variability yield an uncertainty interval with desired coverage. We are also able to contrast the size of the bias due to violation of the unconfoundedness assumption, with bias due to misspecification of the models used to explain potential outcomes. While the latter bias can in theory be made arbitrarily small with increasing sample size (assuming "sparsity", and by increasing the flexibility of the models used), the bias due to unobserved confounding will not disappear with increasing sample size. This is illustrated through numerical experiments where bias due to moderate unobserved confounding dominates misspecification bias for typical situations in

terms of sample size and modeling assumptions. We also study empirical coverage of the uncertainty interval introduced and apply the resulting sensitivity analysis method to real data.

Reference 1: Genbäck, M., E. Stanghellini, and X. de Luna (2015). Uncertainty intervals for regression parameters with non-ignorable missingness in the outcome. Statistical papers, 56, 829-847.

Reference 2: Vansteelandt, S., E. Goetghebeur, M. G. Kenward, and G. Molenberghs (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. Statistica Sinica,16(3), 953-979

OC13-4: Sensitivity analysis for unobserved confounding of direct and indirect effects using uncertainty intervals

Authors: Anita Lindmark[1], Xavier de Luna[1], Marie Eriksson[1].
[1] *Department of Statistics, Usbe, Umeå University, Sweden.*

To estimate direct and indirect effects of an exposure on an outcome from observed data strong assumptions about unconfoundedness are required. Since these assumptions cannot be tested using the observed data, a mediation analysis should always be accompanied by a sensitivity analysis of the resulting estimates. We propose a sensitivity analysis method for parametric estimation of direct and indirect effects when the exposure, mediator and outcome are all binary. The sensitivity parameters consist of the correlation between the error terms of the mediator and outcome models, the correlation between the error terms of the mediator model and the model for the exposure assignment mechanism, and the correlation between the error terms of the exposure assignment and outcome models. These correlations are incorporated into the estimation of the model parameters and identification sets are then obtained for the direct and indirect effects for a range of plausible correlation values. We take the sampling variability into account through the construction of uncertainty intervals. The proposed method is able to assess sensitivity to both mediator-outcome confounding and confounding involving the exposure. To illustrate the method we apply it to a mediation study based on data from the Swedish Stroke Register (Riksstroke).

OC13-5: Causal models to evaluate the role of anticoagulant therapy on mortality in haemodialysis patients

Authors: Paola Rebora[1, 2], Simonetta Genovesi[2], Maria Grazia Valsecchi[1, 2].
[1] *Center Of Biostatistics For Clinical Epidemiology,* [2] *School of Medicine and Surgery, University of Milano, Bicocca, Italy.*

The evaluation of the effect of oral anticoagulant therapy (OAT) on mortality in patients with atrial fibrillation and end-stage renal disease is complicated by the time dependent nature of this treatment and by the presence of time dependent confounders (such as bleeding events and the international normalized ratio), that can influence the interruption of the treatment. The sequential Cox model and marginal structural models are able to deal with time dependent confounding. By using data from a prospective study, where detailed information on treatment intake and time varying covariates

were collected beyond baseline data, we tackled these issues by a causal approach comparing the performance of these models. The sequential Cox model has the limit than does not deal with intermittent treatments. In our application, the main analysis only considered the first switch of therapy (stopping OAT) and the sequential Cox model was able to evaluate the effect of stopping treatment for different values of an important covariate, that is the international normalized ratio. The two models gave similar results showing an advantage of OAT on mortality.

Reference 1: GRAN, J.M., RØYSLAND K., WOLBERS M. et al (2010): A sequential Cox approach for estimating the causal effect of treatment in the presence of time-dependent confounding applied to data from the Swiss HIV Cohort Study. Stat Med. 2010 Nov 20;29(26):2757-68. doi: 10.1002/sim.4048.

Reference 2: HERNÁN M.A., BRUMBACK B., ROBINS J.M. (2000): Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology. 2000 Sep;11(5):561-70.

Poster Contributed Session

PC2: Clinical trials 2

Monday 10th July - 15.00-16.30 h. - Room: Hall
Chair: Armando Teixeira-Pinto

PC2 - M2: Exploration on non-inferiority test for continuous variable with smooth margin

Authors: Arsene Brunelle Sandie[1], Anthony Wanjoya[1], Jules Brice Tchatchueng-Mbougua[2].
[1] *Kenya,* [2] *Cameroon.*
[1] *Pan African University-Institute of Basics Science, Technology and Innovation (PAUSTI).*

The non-inferiority test procedure with the binary endpoint has been developed in the literature for fixed margin, linear and non-linear margin. However, when the endpoint is continuous, most test procedures that have been developed consider the cases when the margin is fixed or linear. In this work, we have proposed non-inferiority test procedures with smooth margin when the primary endpoint is continuous and based on mean difference. Two tests procedures have been proposed based on asymptotic test and confidence interval. Simulations have been used to carry out the procedures. The two procedures have good performance with large sample. For the procedure based on confidence interval, a linear relationship has been found between the level of confidence interval and the level of significance.

PC2 - M5: Challenges in using historical controls in drug development

Authors: Olivier Collignon[1], Anna Schritz[1], Stephen Senn[1].
[1] *Luxembourg Institute of Health, Luxembourg.*

There has been increasing interest in recent years in the possibility of increasing the efficiency of clinical trials by using historical controls. There has been a general recognition that in replacing concurrent by historical controls the potential for bias is serious and requires some down-weighting to the apparent amount of historical information available. However, such approaches have generally assumed that what is required is some modification to the standard inferential model offered by the parallel group trial. In our opinion, the correct starting point that requires modification is a cluster-randomised trial.

This immediately shows that the amount of information available is governed not just by the number of historical patients but also by the number of centres and of historical studies. Furthermore, once, one accepts that external patients may be used as controls, this raises the issue as to which patients should be used. Thus, abandoning concurrent control has implications for many aspects of design and analysis of trials, including:

- identification, pre-specification and agreement on a suitable historical data-set an agreed, enforceable and checkable plan for recruiting the experimental arm a finalised analysis plan prior to beginning the trial use of a hierarchical model with sufficient complexity

We discuss these issues and suggest approaches to design and analysis making extensive reference to the partially randomised TARGET study (Senn, S., 2008). We conclude that effective use of historical data will require considerable circumspection and discipline.

Reference 1: Senn, S. (2008). Lessons from TGN1412 and TARGET: implications for observational studies and meta-analysis. Pharmaceutical statistics, 7, 294-301

PC2 - M8: The registry-based trial-opportunities, challenges and solutions

Authors: Lehana Thabane[1, 2], Guowei Li[1, 2], Tolulope T. Sajobi[3, 4], Bijoy K. Menon[3, 4], Lawrence Korngut[5], Mark Lowerison[3], Matthew James[3, 6], Stephen B. Wilton[7], Tyler Williamson[3], Stephanie Gill[4, 8], Lauren L. Drogos[7], Eric E. Smith[5, 8], Sunita Vohra[9], Michael D. Hill[3, 4, 8].
[1] *Department of Clinical Epidemiology & Biostatistics, Mcmaster University, Hamilton, On, Canada,* [2] *St. Joseph's Healthcare Hamilton, Mcmaster University, Hamilton, On, Canada,* [3] *Department of Community Health Sciences & O'brien Institute For Public Health, University of Calgary, Calgary, Ab, Canada,* [4] *Department of Clinical Neurosciences, University of Calgary, Calgary, Ab, Canada. University of Calgary Hotchkiss Brain Institute, Calgary, Ab, Canada,* [5] *Department of Clinical Neurosciences, University of Calgary, Calgary, Ab, Canada. University of Calgary Hotchkiss Brain Institute, Calgary, Ab, Canada,* [6] *Departments of Medicine, University of Calgary, Calgary, Ab, Canada,* [7] *Libin Cardiovascular Institute of Alberta, University of Calgary, Calgary, Ab, Canada,* [8] *University of Calgary Hotchkiss Brain Institute, Calgary, Ab, Canada,* [9] *Department of Pediatrics, University of Alberta, Edmonton, Ab, Canada*

Registry-based randomized controlled trials (RRCTs) are defined as pragmatic trials that use a registry as a platform for case records, data collection, randomization or follow-up. Recently, the application of RRCTs has attracted increasing attention in health research to address comparative effectiveness research questions in real-world settings--mainly due to their lower costs, enhanced generalizability of trial findings, rapid enrolment and potential completeness of follow-up for the reference population, compared to conventional effectivess trials. However, RRCTs present several ethical, operational and methodological challenges which need to be taken into consideration. In this talk, I will review the challenges and potential solutions. I will use the CHAP (Cardiovascular Health Awareness Program)--a community-based RRCT designed to enhance management of cardiovascular health in mid-sized communities in Ontario, Canada--to illustrate the issues.

PC2 - M11: Relative risk in indirect treatment comparisons

Authors: Manjula Schou[1].
[1] *Macquarie University, Australia.*

In randomised clinical trials (RCT), comparative effectiveness of therapies measured on a dichotomous outcome are often reported as a risk difference (RD), risk ratio (RR) or odds ratio (OR). In cost-effectiveness (CE) evaluations in the context of health technology assessment (HTA), it is not uncommon for evaluations to consider consistency of results across an absolute difference (RD) and a relative measure (RR and/or OR). We explore the implications of using this approach via an indirect treatment comparison (ITC), a comparison of experimental treatments from two different RCTs via a common comparator present in both trials, conducted in the absence of direct head-to-head evidence. A drawback of an ITC is the increase in estimation error resulting from the contribution each trial makes to the overall estimation error. This increase in variance can lead to vague hierarchies of therapeutic relativity in disease indications where multiple treatments are available, but for which direct head-to-head evidence is not available. Establishing unequivocal therapeutic relativity between treatments can be advantageous to sponsor companies in price negotiations within the context of CE analyses in HTA. We explore the implications of the lack of symmetry of the RR measure on statistical efficiency in the context of an ITC. As the statistical efficiency of the RR is dependent on the response rates, we conclude that under certain circumstances, efficiency of an ITC is increased by considering a RR based on non-response, rather than response. An ITC of two treatments for severe chronic plaque psoriasis which is refractory to treatment with non-biological disease modifying anti-rheumatic drugs is used to demonstrate our findings.

PC2 - M14: Simulation study to estimate bias of treatment effect in pain scores in diabetic neuropathy

Authors: Marion Procter[1].
[1]*Frontier Science Scotland (LTD), UK.*

In clinical trials in pain management, 24-hour average pain score may be measured throughout the acute therapy period. However, in practice not all intended pain scores will be recorded, for example after discontinuation of therapy. This leads to a concern of potential bias in the treatment effect from repeated measure analysis. A simulation study was performed to estimate the bias of the treatment effect under various scenarios for discontinuation of therapy.

The context of the simulation study is pain management in diabetic neuropathy. A complete simulated dataset was created of weekly mean average 24-h pain scores from 400 patients during the 12 week acute therapy period, similarly to Goldstein et al. (2005). Patients were randomized 1:1 to placebo and active treatment.

It was assumed approximately 25% of patients discontinue therapy. The scenarios considered were:

- Patients with higher pain scores are more likely to discontinue therapy
- Patients who have not reached a 25% reduction in pain score from baseline are more likely to discontinue therapy
- Scenarios 1 and 2 both apply

A repeated measures analysis of change in pain score from baseline was considered. The bias in the treatment effect was 0.7%, 2.3% and 2.1% for scenarios 1, 2 and 3 respectively compared to the complete treatment effect of -0.957. While these biases are small, the simulation study indicates that the bias in the

treatment effect from scenarios 2 and 3 are numerically larger than the bias in scenario 1. This highlights the importance of considering potential reasons for missing data in analysing pain scores.

Reference 1: 1 Goldstein et al. (2005). Duloxetine vs. placebo in patients with painful diabetic neuropathy, Pain 116; 109-118.

PC2 - M17: Sequential estimation-adjusted adaptive design for three-arm non-inferiority trials

Authors: Siu Hung Cheung[1], Wenfu Xu[2], Feifang Hu[3].
[1] *The Chinese University Of Hong Kong, China,* [2] *Renmin University Of China, China,* [3] *George Washington University, USA.*

The major objective of a non-inferiority clinical study is to identify potential alternatives to a standard treatment for reasons such as detrimental side-effects and exorbitant treatment costs. The new treatment could serve as a replacement of the standard treatment only if the benefits of adopting it exceed a possible clinically insignificant reduction in treatment efficacy. Statistical procedures have recently been developed for treatment comparisons in non-inferiority clinical trials that have multiple experimental (new) treatments. An ethical concern for non-inferiority trials is that some patients undergo the less effective treatments and this problem is more serious when multiple experimental treatments are included in a balanced trial in which the sample sizes are the same for all experimental treatments. With the aim of giving fewer patients the inferior treatments, we propose a response-adaptive treatment allocation scheme. The proposed adaptive design is also shown to be superior to the balanced design in terms of testing power.

PC2 - M20: Estimating effects in two-stage randomised trials when some participants are indifferent to treatment choice

Authors: Stephen Walter[1], Robin M. Turner[2], Petra Macaskill[3], Kirsten J McCaffrey[3], Les Irwig[3].
[1] *McMaster University, Canada,* [2] *University of New South Wales, Australia,* [3] *University of Sydney, Australia.*

Clinical trial outcomes may be affected by preferences that participants might have between the treatments under comparison. Preference effects can be substantial, but they are unobservable in standard trial designs; however, they are estimable in two-stage randomised trials. In this design, a random subset of patients are permitted to choose their own treatment, with the remainder randomised in the usual way.

In earlier work on two-stage randomised trials, we showed how to optimise the proportion of participants to be assigned to the choice arm, but we ignored the possibility of some participants being indifferent to their choice of treatments. We have now extended this work to cover the possibility of some participants having no treatment preference, even if they are in the choice arm and allowed to choose.

We considered alternative assumptions about the undecided participants, and identified an approach to obtain unbiased estimates and tests of the impact of preferences on study outcomes, as well as the usual direct treatment effect. These methods will be illustrated with data from a clinical trial in which 69% of participants in the choice arm had no preferred treatment.

We conclude that use of the two-stage design can provide important insights into determinants of study outcomes that are not identifiable with other designs. It can remain attractive even if some participants have no stated treatment preference.

This work is in press with BMC Medical Research Methods.

Reference 1: Walter SD et al. Beyond the treatment effect. Stat Methods Med Res. 2014

Reference 2: Walter SD et al. Optimal allocation of participants... in the two-stage randomised trial design. Stat Med 2012;31:1307-22.

PC2 - M23: A bias correction method using prior information for interim analysis

Authors: Masashi Shimura[1], Kazushi Maruo[2], Masahiko Gosho[3].
[1] *Taiho Pharmaceutical. CO, LTD, and University of Tsukuba, Japan,* [2] *National Center Of Neurology And Psychiatry, Japan,* [3] *University Of Tsukuba, Japan.*

Group sequential designs are widely used in clinical trials to determine whether a trial should be stopped early or not. The maximum likelihood method is frequently used to estimate the treatment effect. However, it is well known that the maximum likelihood estimate (MLE) has the conditional bias in the group sequential designs. Koopmeiners et al. (2012) proposed the conditional mean-adjusted estimator (CMAE) to reduce the bias, but the bias for the CMAE was not negligible when the trial stopped at the interim analysis for efficacy. We proposed a new estimator to adjust the conditional bias, that is a weighing approach using a prior information of the treatment effect. We evaluated the performance of the proposed estimator via simulation studies in settings where the trial stopped at the interim analysis for efficacy. We found that the conditional bias of the proposed estimator was relatively smaller than that of the MLE and CMAE.

Reference 1: Koopmeiners JS, Feng Z, Pepe MS. Conditional estimation after a two-stage diagnostic biomarker study that allows early termination for futility. Statistics in Medicine 2012; 31(5):420–435.

PC2 - M26: Using a hierarchical score to account for composite outcomes in clinical trials: a re-analysis of the EMPIRICUS randomized clinical trial

Authors: Sébastien Bailly[1], Stéphane Ruckly[2], Wafa Ibn Essaied[1], Claire Dupuis[1], Michel Wolff[1], Pierre-Emmanuel Charles[3], Elie Azoulay[4], Jean-François Timsit[1].
[1] *Inserm, France,* [2] *Outcomerea, France,* [3] *Chu Dijon, France,* [4] *Hôpital Saint Louis, France.*

Composite outcomes are often used in clinical trials. A composite outcome focuses only on the first event which occurred. Although there is some interest for the use of a composite outcome, it can lead to statistical biases mainly because each event is considered with the same importance. The win ratio is a new tool, based on risk-matched pairs of patients, which optimizes statistical power by considering hierarchical outcomes. The objective is to re-analyse the EMPIRICUS randomized clinical trial by using the win ratio.

Materials and methods: The EMPIRICUS randomized clinical trial studied whether an empirical systemic antifungal treatment (micafungin), compared to placebo, reduces invasive fungal infection (IFI)-free survival at day 28. A risk score was computed using a Cox model adjusted on variables associated with the main outcome. Patients were matched according to the risk score and classified as winner or loser according to each outcome: 1) 28-day death 2) 28-day IFI. Finally the win ratio was computed as the ratio of winners in the micafungine arm and winners in the placebo arms.

Results: From the 251 patients of the RCT, 123 pairs were constituted. The win ratio computed on 68 pairs (55 pairs – 45% were tied pairs), was 1.27 [0.78-2.09]. It was close to the initial hazard ratio using a composite outcome (1.35 [0.87-2.08]) and higher than the HR based only on survival at day 28 (1.04 [0.364-1.67]).

Conclusion: the win ratio is a simple tool to avoid the use of composite outcome in RCT. An improvement could be made by introducing the duration of antifungal treatment in the case of tied pairs.

Reference 1: Timsit JF et al. JAMA 2016 Oct;18;316(15):1555-64.

Reference 2: Pocock SJ et al. Eur Heart J. 2012 Jan;33(2):176-82.

PC2 - M29: A reasonable adaptive Simon's two-stage design

Authors: Kosuke Kashiwabara[1], Yutaka Matsuyama[1].
[1] *The University Of Tokyo, Japan.*

For a single-arm phase II trial with a binary outcome, if little prior knowledge of the treatment effect exists, adaptive Simon's two-stage design is an attractive option, in which the final sample size, L, M or N (L < M < N), is determined based on the interim result of the trial at L. One assumes a single null hypothesis, whereas two alternative hypotheses, the one for a conservative and the other for an optimistic treatment effect, are assumed. Generally, a feasible search for the whole design space results in eligible designs, from which the "optimal" design is chosen. Lin and Shih originally proposed

to be eligible for all designs satisfying the Type I and Type II error probability conditions (Biometrics 2004; 60: 482-90). However, such eligible designs may include uncomfortable designs, for example, that a larger sample size is required for a better interim result. In this talk, we introduce a set of additional conditions on eligibility so that only "reasonable" designs are resulted. The conditions are (i) less final sample size M is required for better interim result, (ii) if M = N, the number of responses required to reject the null hypothesis is independent of the interim result and (iii) for each sample path which chooses M and reaches the rejection (acceptance) region of the null hypothesis at M, the conditional probability that the null hypothesis would have been accepted (rejected) had the trial continued until N is strictly less than 1. We show that all eligible designs (a) have the number of responses required to reject the null hypothesis at N equal to or larger than that at M, as the direct consequence of (iii), and (b) includes sufficiently many candidates for practical consideration.

PC2 - M32: Planning adaptive population selection design with survival endpoints

Authors: Ryuji Uozumi[1], Chikuma Hamada[2].
[1] *Kyoto University Graduate School Of Medicine, Japan,* [2] *Tokyo University Of Science, Japan.*

We consider that the full population comprises biomarker-positive and biomarker-negative populations categorized based on a promising biomarker in a setting in which the targeted therapy is beneficial only for the biomarker-positive population. In this setting, the use of adaptive population selection designs has spread in response to the emergence of numerous targeted therapies. Such a design provides an opportunity to stop recruitment for a population in midcourse when this population does not benefit from the treatment being tested; e.g., see Uozumi and Hamada (2017). However, there are no established approaches to setting the thresholds in an interim decision when applying a design to a clinical trial. We propose a novel utility-based approach to guide the construction of the interim decision rule of an adaptive population selection design for the setting of the survival endpoint. The utility functions are motivated by the work of Graf et al. (2015), in which several trial designs for the development of targeted therapies were compared. We use utility functions that consist of gain and loss functions solely to help in formulating the interim decision rule to determine whether the entire population is continued or only the promising population is selected. Simulation studies indicated that the proposed approach assists in setting optimal thresholds for selecting the appropriate population at the interim analysis.

Reference 1: Uozumi R, Hamada C. Interim decision making strategies in adaptive designs for population selection using time-to-event endpoints. J Biopharm Stat 2017;27:84-100.

Reference 2: Graf AC, Posch M, Koenig F. Adaptive designs for subpopulation analysis optimizing utility functions. Biom J 2015;57:76-89.

PC2 - M35: Hierarchical modeling for cluster randomized cross-over trial with a discrete outcome

Authors: Hiroaki Iijima[1].
[1]*Hokkaido University Hospital, Japan.*

Cluster randomized cross-over (CRXO) trial gains popularity in sociological and medical study where individual randomized trial is not practically possible. CRXO is a mixture of methods between Cluster randomized trial (CRT) and crossover trial. CRT is a powerful trial design if contamination may occur in individual randomization; as for cross-over trial, it is considered as efficient design in terms of reducing the impact of within subject variability and study contamination. In crossover design, the same subject receives both interventions: test and control in different periods; this enables gaining proficiencies reducing within subject variability. CRXO design can take advantages from both CRT and crossover design, however, within cluster correlation and period-period effects may deteriorate the efficient estimation; otherwise it may result in an incorrect typeerror rates. Morgan et al. [1] proposed hierarchical model in CRXO design with a binary outcome that accounts for the within cluster and between periods correlation. The proposed approach can successfully account for those correlations in the model and reduced the typeerror rate. In this study, we propose a hierarchical model that extends Morgan's hierarchical model to CRXO design with a multinomial outcome. Multinomial logistic model with a within cluster and between periods effect as a random effect is implemented in the model; it is compared with multiple regression in the simulation study. The simulation result will be shown at the presentation.

Reference 1: Morgan KE, Forbes AB, Keogh RH, Jairath V, Kahan BC. Choosing appropriate analysis methods for cluster randomised cross-over trials with a binary outcome. Stat Med 2017; 36: 318-333.

PC2 - M38: Analysis methods for partially nested randomised controlled trials

Authors: Jane Candlish[1], Laura Flight[1], Munyaradzi Dimairo[1], Laura Mandefield[1], Stephen J. Walters[1], Dawn Teare[1].
[1] *Scharr, University of Sheffield, UK.*

Individually randomised trials often have the added complication of a partially nested design: clustering of outcomes occurs in some trial arms and not others. Partially nested trials (pnRCTs) are commonly used in complex intervention research where the clustering of outcomes is defined by the nature of the intervention itself, for example, a comparison of group therapy intervention and drug therapy control. Small clusters, small intracluster correlations (ICCs), and differential variance between the control and intervention arms are often present in pnRCTs . If not accounted for in the design or analysis this may result in bias effect estimates with spurious precision.

We evaluate methods to analyse pnRCTs and provide practical advice on the analysis of pnRCTs using a simulation study.

The simulation study explores varying scenarios of cluster size, number of clusters, ICC, and differential variation between the two trial arms and their impact on bias, power, precision and ICC estimation. We compare several analysis methods, including: ignoring the clustering; imposing clusters in the control arm and fitting random effects model, or using an approach which only models the clustering in the treatment arm. In theory, the mixed effect models for partially nested trials do not model clustering in the control arm, however, when fitting these models in statistical software it is necessary to impose clustering in the unclustered control arm. We will provide advice on at what point are mixed effects models necessary, the most appropriate method for imposing clustering in the unclustered control arm, and the impact of differential variability between trial arms.

PC2 - M41: Comparing baseline as response and missing indicator methods for missing baseline data in a mixed design cluster randomised control trial

Authors: Lesley-Anne Carter[1], Chris Roberts[2], Karina Lovell[3].
[1] *Centre for Biostatistics, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, UK,* [2] *Centre for Biostatistics, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester UK,* [3] *Division of Nursing, Midwifery and Social Work, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester UK.*

When conducting a randomised control trial (RCT) with repeated measurements, a commitment is required of the cohort of participants to remain in the study for a potentially long follow-up period. The design is thus open to both recruitment issues and attrition. A cross-sectional design may be used in conjunction with the cohort design to protect against these problems, recruiting additional participants who only contribute once to the study, resulting in a 'mixed' design.

The EQUIP cluster RCT was designed to evaluate the efficacy of a training intervention for community mental health teams (CMHT), employing such a mixed design. The 'cluster cohort' sample provided data on service users at baseline prior to randomisation and at a six month follow-up interview. The 'cluster cross-sectional' sample involved all service users under the care of the CMHTs not in the cohort sample, who were sent a postal questionnaire six months after randomisation. The combined sample was intended to increase power should retention rates not meet expectations. Comparison of the results of the two designs would allow external validity of the intervention to be investigated.

As data were only collected in the cross-sectional design after randomisation, baseline data were missing in this sample posing a problem for the combined analysis. Two methods for overcoming this issue were considered: using baseline as response, where a joint model of baseline and response is fitted to all observed data, and a missing indicator method in which an indicator variable for the missing data is include in the model as a covariate. These two methods will be presented with a discussion of the challenges encountered in the application of each to the EQUIP study.

PC2 - M44: Inference on median difference for longitudinal skewed data in randomized clinical trials

Authors: Kazushi Maruo[1], Yusuke Yamaguchi[2], Hisashi Noma[3], Masahiko Gosho[4].
[1] *National Center of Neurology And Psychiatry, Japan,* [2] *Astellas Pharma Inc., Japan,* [3] *The Institute of Statistical Mathematics, Japan.* [4] *University Of Tsukuba, Japan.*

In randomized clinical trials with longitudinal continuous outcomes, right skewed data are often observed. Applying normal-distribution-based statistical methods such as the mixed-effects models for repeated measures (MMRM) method for skewed longitudinal data may result in inefficient inferences. Additionally, the estimated mean may be inadequate as a representative value for the skewed data. Maruo et al. (2015) proposed the model median inference on the original scale based on the linear model with the Box-Cox transformation. However, this approach does not consider longitudinal correlated data. In this study, we derived results for inference on parameters of the marginal model from the MMRM analysis with the Box-Cox transformation based on the asymptotic theory approach. Using these results, we developed an inference procedure for the difference of the model median between treatment groups at the specified occasion in the context of MMRM analysis for randomized clinical trials, which provided interpretable estimates of the treatment effect. From simulation studies, it was shown that our proposed method controlled the type I error for the model median difference in almost all the situations, and the power for our method was generally higher than that for existing methods.

Reference 1: Maruo, K., Isogawa, N., and Gosho, M. (2015). Inference of median difference based on the Box-Cox model in randomized clinical trials. Statistics in Medicine, 34, 1634-1644.

PC2 - M47: Admissible multi-arm stepped-wedge cluster randomized trial designs

Authors: Michael Grayling[1], Adrian Mander[1], James Wason[1].
[1] *MRC Biostatistics Unit, UK.*

Numerous publications have now addressed the principles of designing, analyzing, and reporting, stepped-wedge cluster randomized trials. In contrast, there is little research available pertaining to the design and analysis of multi-arm stepped-wedge cluster randomized trials, applied to evaluate the effectiveness of multiple experimental interventions with a natural ordering. Here, we address this by explaining how the required sample size in these trials can be ascertained when data are to be analyzed using a linear mixed model. We then go on to describe how the design of such trials can be optimized to balance between minimizing the cost of the trial, and minimizing the variance of the treatment effect estimators. Using a recently commenced trial evaluating the effectiveness of sensor monitoring in an occupational therapy rehabilitation program for older persons after hip fracture as an example, we demonstrate that our designs could reduce the number of observations required for a fixed power level by up to 75%.

PC2 - M50: Consideration of hybrid non-inferiority trials design using network-meta-analysis

Authors: Eisuke Hida[1], Kazue Yamaoka[2], Toshiro Tango[3].
[1] *Hiroshima University Hospital, Japan,* [2] *Teikyo University, Japan,* [3] *Center For Medical Statistics, Japan.*

In two-arm non-inferiority trials, choice of a non-inferiority margin and assurance of assay sensitivity are important keys to prove the efficacy of an experimental treatment. The assessment of assay sensitivity in a non-inferiority trial is based on (1) historical evidence of the efficacy of treatment effects, (2) constancy assumption and (3) quality of the non-inferiority trial. Thus, three-arm non-inferiority trials including a placebo are strongly recommended for assessing assay sensitivity and carrying out internal validation.

We proposed that the three-arm non-inferiority trial can be termed successful if and only if the magnitude relationship among three-arm is satisfied. In other words, we have to simultaneously show (a) the non-inferiority of the experimental treatment to the reference, and (b) the superiority of the reference treatment to the placebo by more than. However, there seems to be some confusion, misunderstanding on the appropriate design of three-arm non-inferiority trials. Therefore, it is necessary to prove non-inferiority (a) with assay sensitivity from another viewpoint.

To resolve these problems, we develop a method to assess assay sensitivity by using the Network-Meta-Analysis approach in two-arm non-inferiority trials.That is, to evaluate the substantial superiority of (b) as the historical evidence (1), we propose a procedure for simultaneous comparison of evidence from both current non-inferiority trial and multiple comparative trials in the same therapeutic area.

Reference 1: FDA Guidance 2016. Non-Inferiority Clinical Trials to Establish Effectiveness.

Reference 2: Hida E., Tango T. On the three-arm non-inferiority trial including a placebo with a prespecified margin. Statistics in Medicine 2011. 30: 224-231.

PC2 - M53: Measuring discrimination of models that predict treatment benefit: two examples guiding cardiovascular interventions

Authors: David Van Klaveren[1], Ewout W. Steyerberg[1], Patrick W. Serruys[2], David M. Kent[3].
[1] *Leiden University Medical Center, The Netherlands,* [2] *Imperial College London, United Kingdom,* [3] *Tufts Medical Center, Boston, USA.*

It is impossible to directly observe treatment benefit in individual patients since their (counterfactual) outcome under the alternative therapy is unknown. Thus, clinical prediction models that are used to support treatment decisions are usually evaluated for their ability to predict the risk of an outcome rather than treatment benefit. We aimed to define performance metrics for describing a model's ability to predict treatment benefit.

We analyzed data of Non-acute coronary artery disease patients in the SYNTAX trial (n=1800), and of acute ischemic stroke patients in 3 recombinant tissue plasminogen activator (rt-PA) trials (n=1205). We assessed alternative prediction models with a conventional c-statistic for outcome risk and a novel treatment benefit c-statistic. We defined observed treatment benefit by the outcomes in pairs of patients matched on predicted benefit but discordant for treatment assignment.

In the SYNTAX trial, compared to a model without treatment interactions, the SYNTAX Score II had improved ability to discriminate by treatment benefit (treatment benefit c-statistic 0.590 versus 0.552), despite having similar discrimination for outcome risk (conventional c-statistic 0.725 versus 0.719). However, for the simplified Stroke TPI versus the original Stroke TPI, the conventional c-statistic (0.790 versus 0.811) and the treatment benefit c-statistic (0.584 versus 0.578) were both similar, indicating no loss of discrimination for treatment benefit prediction with the simplified model.

The proposed methodology has the potential to measure and communicate information about a prediction model's ability to predict treatment benefit not captured with conventional performance metrics.

PC2 - M56: A non-parametric method to evaluate predictive biomarkers for patient treatment selection: the area between curves

Authors: Yoann Blangero[1, 2], Fabien Subtil[1, 2], Muriel Rabilloud[1, 2], Karine Le Malicot[3], Julien Taieb[4, 5], Pierre Laurent-Puig[4, 6], Come Lepage[7, 3].
[1] *Lbbe, Umr CNRS 5558, Université Claude Bernard Lyon 1,* [2] *Service de Biostatistique des Hospices Civils de Lyon, France,* [3] *Fédération Francophone de Cancérologie Digestive (FFCD), France,* [4] *Sorbonne Paris Cité, Université Paris Descartes,* [5] *Department of Gastroenterology and Gi Oncology, Hopital Européen Georges Pompidou, France,* [6] *Assistance Publique Hôpitaux de Paris, Department of Biology, Hôpital Européen Georges Pompidou, Inserm Umr-S1147, France,* [7] *Hepato-Gastroenterology Department, Dijon University Hospital, Inserm U 866.*

Markers commonly named "predictive" or "treatment selection" markers are assessed in randomized studies when evaluating a new treatment against a reference one. Such markers have the potential to identify subgroups of patients that will benefit the most from a treatment. Predictive markers became essential in the personalized medicine field as it relies on allowing a treatment based on the patients' characteristics. When the benefit is not constant among the marker values, the difference in prognostic effects between the two treatment groups is not null. Hence, a predictive marker can be characterized as its difference in prognostic effects between the two treatment groups. A non-parametric approach is proposed to assess the predictive effect of a quantitative marker in a clinical trial context. This approach relies mainly on ROC curves methodology by computing a ROC curve for each treatment group and provides a useful graphical tool already well-known by clinicians and an easy-to-use method to detect and evaluate predictive markers. An indicator: the Area Between Curves (ABC) was developed to quantify and test the predictive effect of the marker and the properties of the associated estimation method were investigated with a simulation study. The results show the low bias in finite samples and the good coverage probability of the confidence interval of the method. The method was applied to the data of the PETACC-8 trial. The ABC should be used as a first step in the identification and assessment of predictive markers.

PC2 - M59: Are enrichment designs for reducing the placebo response sound?

Authors: Dominik Grathwohl[1]
[1] *Nestlé Research Center*

It's better to solve the right problem approximately than to solve the wrong problem exactly." This is a quote of John Tukey and I will explain in this article why enrichment designs for reducing the placebo response fall into the category of solving the wrong problem exactly. I will also outline an alternative design which solves the right problem approximately. In several clinical domains, clinical trials face the problem of placebo response. This is the phenomena that in a randomized, double blind, placebo controlled clinical trial, the placebo group shows an improvement which approaches clinical relevance. There are enrichment designs where the placebo group is chosen bigger than the active treatment group and which are intended to identify the placebo-non-responders. Once identified, placebo-non-responders get re-randomized. E.g. sequential parallel comparison design, SPCD; Fava et al., 2003. We suggest alternatively a two by two factorial design, where one factor is heavy or light study procedure and the other factor is placebo or active treatment. For the SPCD an overall p-value is deduced which controls formally the experiment wise false positive rate. The corresponding treatment effect is modified by the choice of the allocation ratio and by the cut-off criteria of the response. The two by two factorial design enables to measure the effect modification by study procedure. Clinical trials are artificial procedures which may cause a placebo response more likely than the treatment of outpatients at the day to day work of a practitioner. Protagonists of the SPCD design argue that the treatment effect is corrected for the downward bias, due to the placebo response. In our opinion, the SPCD produces biased treatment effects as a function of the allocation ratio and the cut-off criteria. This treatment effects cannot be generalized to a patient population. The two by two factorial design enables to quantify the placebo effect by study procedure and to extrapolate to a patient population.

PC2 - M62: Nested case-control and full cohort approaches for evaluating biomarkers in RCTs – comparison of methods using data from the TACT2 trial

Authors: James Morden[1], John Bartlett[2], David Cameron[3], Judith M Bliss[1].
[1] *The Institute of Cancer Research, United Kingdom,* [2] *Ontario Institute for Cancer Research, Canada,* [3] *University of Edinburgh.*

Phase III oncology trials increasingly include parallel collection of biological material. These large cohorts comprise an invaluable tissue resource with which the prognostic and predictive value of biomarkers can be assessed. Ideally the whole cohort would be assessed for all biomarkers of interest; however this may not be feasible for financial and logistical reasons.

The nested case-control (NCC) design involves retrospectively identifying "cases" – patients with outcome event of interest – and matching to "controls" – patients without the event. The NCC design may be an attractive option to reduce the number of patients for whom a particular expensive test needs to be performed. The aim of this work is to compare prognostic estimates produced from a full cohort and NCC approach using data from a large phase III early breast cancer trial.

TACT2 (CRUK/05/019) is a multicentre phase III trial assessing different chemotherapy regimens in early breast cancer. 4391 patients were randomised (129 UK centres); tumour samples were collected prospectively from 3803 patients, of whom 3269 had data from central assessment of all IHC4 markers (ER, PgR, HER2 and Ki67). Time to distant recurrence (TTDR) was the endpoint considered here; 502/3269 (15%) had a TTDR event.

The cohort was randomly split into test and validation sets. Parameter estimates for each IHC4 variable and risk scores combining all four variables were obtained using data from the test set with three approaches: 1) Full cohort – models fitted using all test set data; 2) NCC (one control per case) and 3) NCC (two controls per case). Estimates/scores were then compared between approaches and applied to the validation set to investigate prognostic accuracy.

Invited Session

IS3: Beyond Proportional Hazards

Monday 10th July - 17.00-18.30 h. - Room: Auditorio
Chair: Tomasz Burzykowski
Organised by Lee-Jen Wei and Tianxi Cai, *Harvard T.H. Chan School of Public Health, USA*

IS3-1: On testing based on restricted mean survival time for time-to-event outcomes

Hajime Uno, *Dana-Farber Cancer Institute, USA*

Between-group summary measures based on restricted mean survival time (RMST) provide robust and clinically interpretable information about the treatment effect [1,2]. The conversion test of the difference or ratio of RMSTs may be useful as the primary analysis in comparative clinical trials because we can have a quantitative summary of the treatment effect (e.g., 0.95 confidence interval) that corresponds to the primary test result. Although a choice of the truncation time (tau) for RMST is a challenge in practice, this coherence still stands whichever tau may be used. Motivated by this, we developed a pre-specified versatile test based on RMST, where tau is selected data-dependently. We confirmed the validity of the new test by numerical studies. Our numerical studies also demonstrated that the new test is more powerful than logrank, Wilcoxon tests, and RMST-based tests with a pre-specified fixed tau, under a pattern of the difference in survival function seen in recent cancer clinical trials. The new test is inferior to the logrank test under the proportional hazards difference by theory. However, the new test would be preferable when investigators expect the similar pattern of difference seen in those cancer trials.

Reference 1: Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. Journal of Clinical Oncology. 2014 Aug 1;32(22):2380–5.

Reference 2: Uno H, Wittes J, Fu H, Solomon SD, Claggett B, Tian L, et al. Alternatives to Hazard Ratios for Comparing the Efficacy or Safety of Therapies in Noninferiority Studies. Ann Intern Med. 2015 Jul 21;163(2):127–112.

IS3-2: Non-Parametric Estimation of LongTerm Treatment Benefit from a Randomized Controlled Trial

Brian Claggett, *Harvard Medical School, USA*

While clinical trials can provide substantial evidence regarding the effectiveness of a new therapy relative to a control over a finite period of follow-up time, estimation of long-term treatment effects extending beyond the duration of the trial is often difficult to obtain and heavily reliant on unverifiable modeling assumptions.

The PARADIGM-HF trial demonstrated that the new therapy LCZ696 was superior to enalapril in reducing cardiovascular (CV) death, all-cause mortality and heart failure (HF) hospitalization in patients with heart failure and reduced ejection fraction. Taking advantage of the wide age range and large number of events in PARADIGM, we employed Kaplan-Meier methods to produce actuarial estimates of age-specific event rates and expected survival times in order to estimate the projected long-term effects of LCZ696 vs enalapril. We validated the methodology by comparing long- term survival estimates from the SOLVD-Treatment trial to actual long-term follow- up data from the SOLVD extension study. These methods may be useful in long-term estimation of benefit in clinical trials in which trial duration is substantially shorter than the potential projected lifespan of the patients.

IS3-3: Non-Inferiority Trials with Time-to-Event Outcomes: Design Based on the Restricted Mean Survival Times

Ludovic Trinquart, *Boston University School of Public Health, Departament of Biostatistics, USA*

The design of non-inferiority randomized trials is complex. Most trials with a time-to-event endpoint are designed by using the hazard ratio as a measure of treatment effect. Building upon previous works, I will describe the use of the restricted mean survival time to design non-inferiority trials with efficacy or safety time-to-event outcomes. Through the reanalysis of 38 non-inferiority trials published between 2013 and 2016 in leading medical journals, I will illustrate the current approaches for the choice of the non-inferiority margin and the sample size calculation, and the benefits of using the restricted mean survival times.

Reference 1: Jahn-Eimermacher A, Ingel K, Ozga AK, Preussler S, Binder H: Simulating recurrent event data with hazard functions defined on a total time scale. BMC Medical Research Methodology 2015; 15:16.

Oral Contributed Sessions

OC14: Bayesian methods in clinical research 3

Monday 10th July - 17.00-18.30 h. - Room: Sala Mar 2
Chair: David Dejardin

OC14-1: Subject randomization with optimal unequal allocation in Bayesian adaptive trials

Authors: Jordan Elm[1], Wenle Zhao[1].
[1] *Medical University of South Carolina, USA*

Unequal allocations are used in Bayesian adaptive trials aiming to optimize the trial efficiency. Commonly used permuted block randomization faces a trade-off between the allocation accuracy and the imbalance control. For example, in the Established Status Epilepticus Treatment Trial (ESETT NCT01960075), a randomized comparative effectiveness study of Fosphenytoin, Valproic Acid, and Levetiracetam in the treatment of Benzodiazepine-refractory Status Epilepticus, an optimal allocation ratio of 0.12:0.22:0.66 is demanded for the next 100 subjects if the response rate among the first 100 subjects is 0.51, 0.55, and 0.64 respectively for the three arms. To accurately target the desired allocation, the minimal block size is 50, which is too large for allocation imbalance control in the middle of the block. Using small block sizes, such as 10, will dilute the original study design. To solve this problem, the mass-weighted urn design is proposed. It uses an urn model for the conditional allocation probability. The urn starts with one ball for each treatment arm, with the mass proportional to the target allocation. The chance a ball being randomly selected is proportional to its mass. After each treatment assignment based on the color of the selected ball, a part of the mass of the selected ball is re-distributed to all balls based on the target allocation. This design allows any desired optimal unequal allocations to be accurately targeted while maintaining a consistent imbalance control. The statistical properties of the mass-weighted urn design are evaluated and compared with those of the permuted block randomization. This new randomization design has been implemented in the ESETT trial.

OC14-2: A combined proof of concept and dose finding study with multiple endpoints: a Bayesian adaptive design in chronic prostatitis/chronic pelvic pain syndrome

Authors: Reynaldo Martina[1], Jos Houbiers[2], Joost Melis[2], Olivier Van Till[2].
[1] *University of Liverpool, England,* [2] *Astellas Development, The Netherlands,*

Background: There is a need for identifying effective drugs early or terminating ineffective drugs sooner. We simulated and executed an adaptive allocation design to investigate the effects of a drug on male patients suffering from chronic prostatitis/chronic pelvic pain syndrome. The aim of the trial was to simultaneously establish Proof of Concept (PoC) and dose finding for a new drug with a novel mode of action in a new indication. A wide dose range was included in the study. This abstract describes the clinical trial simulations and primary analysis results.

Methods: The proposed combined PoC/dose finding trial was optimized through extensive clinical trial simulations.A Bayesian adaptive allocation procedure was used to allocate patients to treatment (placebo and 5 dose groups) using a normal dynamic linear model. The study was to stop early for efficacy if the probability that the clinically significant difference between experimental drug and placebo was at least 90%. The study was to stop for futility if the maximum effective dose was not better than placebo by at least 20%. An IDMC assessed the stopping rules regularly.

Results: The study was stopped early, i.e. 35% less than planned maximum sample size, due to futility. The final results confirmed that the predefined stopping rules were met and that the probability that the most likely effective dose was superior to placebo was 18%.

Conclusions: The simulations showed that this adaptive design could accomplish both the goals of PoC and dose finding. The study stopped early for futility in line with the simulation predictions for stopping. This resulted in the early stopping of a trial recruiting patients on ineffective treatment for their condition.

OC14-3: Bayesian test-based design for phase I clinical trial with late-onset toxicity

Authors: Yuh-Ing Chen[1], Chia-Ju Cheung[1].
[1] *National Central University, Taiwan.*

The major purpose of phase I trials is to estimate the maximum tolerated dose (MTD) of the drug under study which is the highest dose producing an acceptable toxicity, that is, the probability of experiencing the dose-limiting toxicity (DLT) is less than the pre-specified target toxicity probability (TTP). To protect patients from the risk of overdose with possibly late-onset toxicity, we suggest use Bayesian tests (BT) for the acceptable toxicity based on the available response and time-to-event data in the dose-escalation procedure. When the maximum number of patients is reached, the mode of the posterior distribution of the MTD is then recommended for the further phase II trial. Hence, the proposed design is denoted by TITE-BTMD. A simulation study is then conducted to investigate the performances of the TITE-BTMD and competitive designs on the toxicity probability and MTD estimation under different dose-toxicity curves and distributions of time-to-event. A sensitivity study is also implemented to explore the performances to miss-specified dose-toxicity models. The results indicate that, in general, the TITE-BTMD design produces reasonable toxicity probability and gives an accurate MTD estimation comparing with the competitive trial designs. Moreover, the performance of the TITE-BTMD design is relatively robust to the miss-specified dose-toxicity curves.

Reference 1: Cheung YK and Chappell R. Sequential designs for phase I clinical trials with late-onset toxicities. Biometrics 2000; 56:1177-1182.

Reference 2: Mauguen A, Le Delay MC and Zohar S. Dose-finding approach for dose escalation with overdose control considering incomplete observations. Stat Med 2011; 30:1584-1594

OC14-4: Bayesian design for clinical trials with composite endpoints consisting of survival and longitudinal endpoints

Authors: Rajat Mukherjee[1], Cyrus Mehta[2], Iván Navarro[1].
[1] *CYTEL INC, Spain,* [2] *CYTEL INC, USA.*

The Finkelstein-Schoenfeld (FS) statistic [1] is growing in popularity in clin- ical trials in which the primary assessment of therapeutic benefit is made not only using survival endpoints but also longitudinal measures of clinical endpoints. Here each patient in the treatment group is compared with each patient in the control group based on several endpoints in a hierarchy. For any two patients one from each group the treatment patient gets a score of +1 if his/her survival is larger than that of the control patient, otherwise -1. If due to censoring the winner cannot be determined then the comparison is made with respect to the second endpoint in the hierarchy and this process continues till all the endpoints in the hierarchy have been covered. Suppose that there are m patients in the treatment group and n in the control group. The FS-statistic is then the sum of the mn scores (+1/-1/0) and is an extension of the Gehan-Wilcoxon statistic and has been shown to be asympotically normal [1].

Here we present a Bayesian design based on testing for treatment effect using the FS-statistic. We extend this to a two-stage design where at the interim the sample size may be adapted based on the predictive power obtained using the interim data. We discuss this using simulations related to showing that for such designs the design parameters can be chosen in such a way that the frequentist propoerties of type-I error and power requirements are satisfied. We illustrate the procedure by designing a medical device clinical trial for the treatment of heart failure.

Reference 1: D. M. Finkelstein and D. A. Schoenfeld. "Combining mortality and lon- gitudinal measures in clinical trials". In: Stat Med 18.11 (June 1999), pp. 1341–1354.

OC14-5: Bayesian estimation of median survival times versus Kaplan-Meier estimates in single arm phase II clinical trials

Authors: Peter Fletcher[1], Daniel Slade[1], Kristian Brock[1], Laura Llewellyn[1], Dee Wherton[1], Gary Middleton[1], Lucinda Billingham[1].
[1] *University of Birmingham, UK.*

In phase II clinical trials of cancer treatments it is often desirable to use continuous time-to-event data in place of binary variables for evidence of activity. The need for interim analyses to assess safety and futility requires that follow-up times will be short and sample size numbers low. If a summary statistic such as median survival time is used then there will be a large degree of uncertainty around any estimates.

The National Lung Matrix Trial uses an umbrella design consisting of multiple phase II single treatment arm trials. It has a Bayesian adaptive design and one of the treatment arms uses median progression-free survival time > 3 months as a criterion of success for the treatment. Following Thall et al (2005), the design uses an exponential-inverse gamma conjugate analysis to generate a posterior

distribution for the true median PFS time. In reporting associated posterior probabilities alongside conventional Kaplan-Meier estimates of the median, some discrepancies emerged between the resultant inferences. The discrepancies were most pronounced in cohorts with a small number of long survival times plus several censored values with short survival times due to limited follow-up.

We use simulation to investigate differences between the two methods for estimating median survival time, with a variety of different parameters, including: true median survival, sample size, level of censoring, rate of patient recruitment, and minimum follow-up time. We report on the sensitivity of each method under the conditions that an early phase II trial might typically encounter.

Reference 1: Thall PF, Wooten LH, Tannir NM. Monitoring event times in early phase clinical trials: some practical issues. Clinical Trials 2005; 2: 467-478

OC15: Complex survival data 1

Monday 10th July - 17.00-18.30 h. - Room: Sala Terra 2
Chair: Philippe Lambert

OC15-1: Testing the Markov assumption in general multi-state models

Authors: Andrew Titman[1].
[1] *Department Of Mathematics & Statistics, Lancaster University, United Kingdom.*

Recently there has been interest in the development of estimators of the transition probabilities for right-censored data that are robust to departures from the Markov assumption. The landmark Aalen-Johansen (LMAJ) [1] estimator is robust to non-Markov processes, but this robustness comes at the cost of a loss of efficiency compared to the standard Aalen-Johansen (AJ) estimator, making it important to identify when it is necessary to use LMAJ.

A similar principle to the construction of the LMAJ can be used to build a test of the Markov property. For a given starting state and time, the set of patients who were in that state at that time can be identified and treated as a distinct group to those who were not. If the process is Markov, the transition intensities in the two groups will be equal. The log-rank test statistics from the transition intensities can be combined to produce a test at that time. Moreover, the statistics across time and starting state form a stochastic process allowing the construction of a global supremum test. A wild bootstrap procedure is proposed to approximate the null distribution in finite samples.

For the illness-death without recovery, the performance of the test is compared to the Kendall's tau based test [2]. The performance is also investigated for more complicated processes, including those allowing recovery.

Reference 1: Putter, H., Spitoni, C. (2016). Non-parametric estimation of transition probabilities in non-Markov multi-state models: the landmark Aalen-Johansen estimator. Statistical Methods in Medical Research. DOI:10.1177/0962280216674497

Reference 2: Rodriguez-Girondo, M., and Una-Alvarez, J. (2012). A nonparametric test for Markovianity in the illness death model. Statistics in Medicine 31, 4416-4427.

OC15-2: The role of frailty in the evaluation of treatment delay

Authors: Nan Van Geloven[1], Hein Putter[1], Saskia Le Cessie[1].
[1] *Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, The Netherlands.*

Doctors often have to choose between starting treatment immediately or first introducing a wait-and-see period during which a patient might recover without the treatment. The effect of a treatment delay period on the time to recovery depends on the heterogeneity between patients' recovery chances, i.e. the frailty. We analytically show that in a frailty model with a time-constant treatment effect that is common over patients, a treatment delay period hardly compromises cumulative recovery rates if the population is heterogeneous. In a homogeneous population however, cumulative recovery rates are directly compromised by treatment delay. Our modelling approach follows the lines of previous studies on the impact of frailty on discontinuing treatment (1).

Estimating the effect of treatment delay from data can be done in several ways. We show that the conventional Cox model overestimates the effect of treatment delay in case of heterogeneity. Including a frailty term in the model could improve the estimation, but frailties are generally hard to estimate in univariate survival data. We present an alternative approach accommodating the effect of heterogeneity on treatment delay using a treatment by time interaction term. Estimation results are presented both through simulations and in a motivating application evaluating the effect of delaying the start of treatment with intra-uterine insemination on time-to-pregnancy in unexplained subfertile couples.

Reference 1: (1) Aalen OO, Borgan Ø, Gjessing HK (2008). Survival and event history analysis: a process point of view. Springer, New York

OC15-3: Bias in rate-based analysis of recurrent exacerbations due to the risk-free period

Authors: Jooyoung Lee[1], Richard J. Cook[1].
[1] *University of Waterloo, Canada.*

In many chronic diseases interest lies in reducing the occurrence of recurrent exacerbations of symptoms each of which may last for an appreciable time. Examples include infections in individuals with chronic obstructive pulmonary disease, recurrent flares of symptoms in lupus, and repeated bouts of depression.

While marginal rate-based analyses have considerable appeal for recurrent event analyses in clinical trials, relatively little work has been carried out on how best to handle the duration of the symptomatic periods of exacerbations. We formulate a copula-based model to link a subject-specific multiplicative random effect acting on conditionally Markov intensity for the onset of exacerbations, with a random effect acting on a conditionally semi-Markov intensity for exacerbation durations. We then derive the asymptotic bias of regression coefficients from standard rate-based analyses under an Andersen-Gill model. Guidelines for the use of standard methods are offered which depend on the primary goal of analyses. An expectation-maximization algorithm is described for fitting the specified semiparametric model for an alternating two-state process. An application to a study of recurrent hospitalizations in patients with psychiatric illnesses is given for illustration.

OC15-4: Inference in the competing risks setting via constrained nonparametric maximum likelihood estimation

Authors: Paul Blanche[1].
[1]*University Of South Brittany, Vannes, France.*

Competing risks settings are common in medical research. Specific methods to deal with such settings are more and more used in clinical research, especially those which relate to the Cumulative Incidence Function (CIF).Methods to compute confidence intervals and to perform hypothesis tests for the CIF based on asymptotic results are well-established. They provide us with powerful tools. However, they can perform poorly in settings in which the number of events of interest which are observed is too small, either because the sample size is small or because the censoring rate is high or because the probability of observing a competing event is high. Unfortunately, these settings are common.We argue that inference procedures based on constrained Nonparametric Maximum Likelihood Estimation can perform better in these settings, while they are asymptotically equivalent to those which are currently well established. The reason is that they enjoy the nice features of profile likelihood methods.We present methods which extend those of Thomas and Grunkemeier [1] and of Barber and Jennison [2] to the competing risks setting. The theoretical developments are not straightforward but what is interesting from an applied statistician point of view is that (i) they are simple to use and reliable and that (ii) computation is fast and easy enough.We present some simulation results and real data applications. They illustrate how the new methods work and provide us with comparisons to benchmark methods implemented in popular software.

Reference 1: Thomas & Grunkemeier (1975). J Am Stat Assoc, 70(352), 865-871.

Reference 2: Barber & Jennison (1999). Biometrics, 55(2), 430-436.

OC15-5: Empirical simulation of complex time-to-event data incorporating time-dependent exposures: a multistate approach

Authors: Tobias Bluhmki[1], Hein Putter[2], Arthur Allignol[1], Jan Beyersmann[1].
[1] *Institute of Statistics, Ulm University, Germany,* [2] *Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, The Netherlands*

A renewed interest in simulating time-to-event data in the presence of time-dependent exposures hasrecently been observed. The common approach is to a priori generate the covariate trajectory andsubsequently apply the inversion method with respect to the Cox proportional hazards model. This issuitable for external covariates, but leads to a misleading population hazard if internal covariates areconsidered. We propose to simulate from multistate models where exposure categories are modeledas separate transient states. To obtain plausible real-world data, we present an 'empirical' modificationof a mathematically well-established, but rarely used hazard-based simulation algorithm. It has beendescribed in detail for competing risks [1] and briefly for general multistate models in the context ofprediction in reduced rank Cox models [2]. The idea is to work with the estimated cumulative hazards,which may be obtained from published data. Our formulation allows for complex survival outcomes be-yond the standard survival and competing risks setting, independent right-censoring and left-truncation,non-degenerated initial distributions, and non-Markov situations. Further, it can be used for individualprediction or model-based bootstrap. We present a simulation study using data of liver cirrhosis patients.

Reference 1: A. Allignol, M. Schumacher, C. Wanner, C. Drechsler, J. Beyersmann, Understanding competing risks: a simulation point of view, BMC Medical Research Methodology 11 (2011)

Reference 2: M. Fiocco, H. Putter, H. C. van Houwelingen, Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models, Statistics in Medicine 27 (2008) 4340–4358.

OC16: Design and analysis of clinical trials 1

Monday 10th July - 17.00-18.30 h. - Room: Sala Mar 4
Chair: Gillian Lancaster

OC16-1: Optimal sample size allocation and go/no-go decision rules for phase II/III programs where several phase III trials are performed

Authors: Stella Preussler[1], Meinhard Kieser[1], Marietta Kirchner[1], .
[1] *Institute of Medical Biometry and Informatics, University of Heidelberg, Germany.*

The conduct of phase II and III programs is costly, time consuming and, due to high failure rates in late development stages, risky. There is a strong connection between phase II and III trials as the go/

no-go decision and the sample size chosen for phase III are based on the results observed in phase II. An integrated planning of phase II and III is therefore reasonable. The performance of phase II/III programs crucially depends on the allocation of the resources to phase II and III in terms of sample size and the rule applied to decide whether to stop or to proceed to phase III.

Recently, a utility-based approach was proposed, where optimal planning of phase II/III programs is achieved by taking fixed and variable costs of the drug development program and potential gains after a successful launch into account [1]. However, this method is restricted to programs with a single phase III trial, while regulatory authorities generally require statistical significance in two or more phase III trials. We present a generalization of this procedure to programs where two or more phase III trials are performed. Optimal phase II sample sizes and go/no-go decision rules are provided for time-to-event outcomes and scenarios, where at least one or two phase III trials need to be successful. We investigate the consequences of the biased treatment effect estimate induced by the go/no-go decision and the effects of different strengths of correlation between the phase III trials. The correlation is due to the common information emerging from the phase II trial. Application to different practical examples illustrates the proposed method.

Reference 1: Kirchner M, et al. Utility-based optimization of phase II/III programs. Stat. Med. 2016; 35:305–316.

OC16-2: Split-plot designs: sample size considerations

Authors: Beatriz Goulao[1], Graeme MacLennan[1], Craig Ramsay[1].
[1] *Health Services Research Unit, University of Aberdeen, UK.*

The split-plot design is historically associated with agriculture studies, but more recently used in healthcare research. The split-plot is a complex design that has both cluster randomised and factorial elements, but is distinguished by two levels of randomisation: one at a cluster-level and one at a lower, often individual, level. In a previous review, we identified nine split-plot randomised controlled clinical trials with a sample size calculation based on the cluster randomisation element, ignoring the individual randomisation in the design.

To estimate the optimal sample size for a split-plot and how to report it, we used simulations to investigate the relationship between the number of clusters and statistical power for the cluster and patient-level interventions. A number of assumptions were varied, including the intra-cluster correlation and intervention target differences at the cluster and individual-level. A constant cluster size and no interaction between treatments were assumed. Simulated data sets were analysed using a mixed-effects model with a random-effect at the cluster level.

Power for the cluster and individual-level depended on the intervention target differences expected: researchers should indicate both explicitly; base sample size calculations on the cluster-level intervention if the target differences expected at both levels are similar; and use simulation, if a smaller target difference at the individual-level is expected, to estimate the number of participants that need to be recruited.

OC16-3: Type I error rates of Multi-Arm Multi-Stage (MAMS) platform clinical trials: impact of adding new research arms

Authors: Babak Choodari-Oskooei[1], Daniel J. Bratton[2], Melissa Spears[2], Matthew R. Sydes[1], Mahesh Kb Parmer[1].
[1] *MRC Clinical Trials Unit at UCL, UK,* [2] *Clinical Statistics, Glaxosmithkline, UK.*

Experimental treatments typically pass through varying stages of development. Once a treatment has passed the early phase developments, the investigators might wish to assess it in an ongoing late phase randomised trial. An efficient way of doing this is to add it as a new research arm in an ongoing trial. An example is the STAMPEDE trial. STAMPEDE is a multi-arm, multi-stage (MAMS) platform trial based on the design proposed by Royston et al. in 2003. This design allows new arms to be added part way through its life while allowing an arm to be stopped part-way through recruitment if the accumulating data suggests a lack-of-benefit of a research arm. Such interim decisions can be made using data on an available `intermediate' (I) outcome. The familywise type I error rate (FWER) is often a key parameter of interest in a MAMS trial. However, it is unclear how the FWER is estimated when new research arms are added to the trial some time after the trial has started. We show how this can be done in a MAMS trial with time-to-event outcomes. Our results indicate that the FWER depends on the shared primary outcome events from the common individuals (in the control arm) and the allocation ratio. Also, the FWER is mainly driven by the number of pairwise comparisons rather than the timing of the addition of a new research arm. Finally, the FWER can be estimated using Sidak correction if the correlation between the test statistics of pairwise comparisons is less than 0.30. Consequently, the correlation threshold of 0.30 or, more practically, the overlap between the pairwise comparisons can be used to define families of hypothesis in a MAMS platform trial.

Keywords: MAMS platform trials; familywise type I error rate, FWER.

OC16-4: Sample size derivation for binary composite endpoints

Authors: Marta Bofill Roig[1], Guadalupe Gómez Melis[1].
[1] *Universitat Politècnica De Catalunya (UPC) Spain.*

Composite binary endpoints (CBE), defined as the union of several binary endpoints, are frequently used as the primary endpoint in a clinical trial. The specification of the treatment effect on the composite endpoint requires the information of their components and the degree of association between them. We summarize the treatment effect on the composite endpoint by means of the odds ratio and show that is determined by six parameters including the degree of association between components, the event proportion and the corresponding odds ratio of the individual endpoints.

The purpose of this talk is to explore sample size formulations for CBE in terms of the marginal odds ratios, the event proportions and the degrees of association. We discuss the influence of each of the parameters of the composite in the required sample size. While anticipated values for the marginal parameters are often easier to guess and most of the sample size formulations depend on those, anticipating the degree of association between the components is a much harder task. Within the framework of multiple co-primary binary endpoints, several authors have addressed the influence of association on sample size [1]. However, approximations to the needed sample size of a CBE with

partial or null knowledge of the correlation are limited. We aim to develop alternative formulas for the computation of the sample size for CBE.

Reference 1: Sozu T., Sugimoto T. and Hamasaki T. (2010). Sample size determination in clinical trials with multiple co-primary binary endpoints. Statistics in Medicine. 29(21), 2169-79.

OC16-5: Power and sample size for the S:T repeated measures design combined with a linear mixed-effects model allowing for missing data

Authors: Toshiro Tango[1].
[1]*Center For Medical Statistics, Japan.*

Tango (Biostatistics, 2016) proposed a new repeated measures design, called the S:T repeated measures design combined with generalized linear mixed-effects models and sample size calculations for a test of the average treatment effect that depend not only on the number of subjects but on the number of repeated measures before and after randomization per subject used for analysis. The main advantages of the proposed design combined with the generalized linear mixed-effects models are 1) it can easily handle missing data by applying the likelihood-based ignorable analyses under the missing at random assumption and 2) it may lead to a reduction in sample size compared with the simple pre-post design.

In this presentation, we present formulas for calculating power and sample sizes for a test of the average treatment effect allowing for missing data within the framework of the S:T repeated measures design with a continuous response variable combined with a linear mixed-effects model. Examples are provided to illustrate use of these formulas.

OC17: Statistical methods in epidemiology 1

Monday 10th July - 17.00-18.30 h. - Room: Sala Terra 4
Chair: Saskia le Cessie

OC17-1: Missing disease information due to death may impact results of cohort studies reporting decline in dementia incidence

Authors: Nadine Binder[1], Martin Schumacher[1].
[1] *Institute For Medical Biometry And Statistics, Medical Center-University Of Freiburg, Germany.*

The possible decline of dementia incidence is currently a matter of intensive debate. For example, Satizabal et al. [1] found the risk of dementia to steadily decline in four, non-overlapping five-year epochs representing three decades within the Framingham Heart Study. We suspect that a statistical bias might to some extent contribute to these findings: Because of discrete-time follow-

up examinations, dementia status might not have been observed in individuals who died between visits; 'censoring' these death cases at their time of death or at the last visit observed disease-free can severely underestimate the disease incidence [2]. In a numerical investigation mimicking the dementia data, we illustrate the potential effect of conducting such ad-hoc model approaches and explain how the bias arises. We also highlight a problematic feature of the ad-hoc 'censoring' strategies, which is conditioning on the future, i.e., on being dead, thereby not fulfilling the independent censoring assumption. Furthermore, we describe a more sophisticated analysis approach based on the full likelihood from a multi-state illness-death model. Given the potential severe biases that become apparent in the dementia application at hand and the ready availability of the multi-state model approach (e.g., implemented in R package SmoothHazard), we emphasize that the latter should be considered whenever there is suspected missing disease information due death. Only such a comprehensive analysis would provide a sufficient basis for deciding whether the observed decrease in incidence is real or an artefact.

Reference 1: Satizabal CL. N Engl J Med. 2016;374(6):523-32.

Reference 2: Binder N (with reply Satizabal CL). N Engl J Med. 2016;375(1):92-94.

OC17-2: Estimation of time to statistical cure

Authors: Lasse Jakobsen[1], Tarec Christoffer El-Galaly[1], Therese Andersson[2], Martin Bøgsted[1].
[1] *Aalborg University, Denmark,* [2] *Karolinska Institutet, Sweden.*

Previously, the time to statistical cure (TTSC) has been defined as the point in time from which the patient mortality rate (hazard) is smaller than or equal to that of the matched general population. However, as hazard rates can be difficult to communicate to patients and hazard-based TTSC is sensitive to sample size we suggest to consider when the loss of lifetime (LOL) is less than a small but clinically insignificant value as a more clinical relevant measure for TTSC. The LOL is estimated as the difference in mean residual lifetime between the patients and the general population, which generally requires extrapolation of the patient and general population survival. Extrapolation of the general population survival is easily conducted whereas extrapolation of the patient survival is done by fitting a flexible parametric relative survival model and evaluating this model in all relevant time points. The TTSC is estimated as the time point where the LOL function reaches the aforementioned clinically insignificant threshold. This method was applied to a series of simulated data and a diffuse large B-cell lymphoma (DLBCL) data set retrieved from the Danish lymphoma registry. The simulations illustrated a fairly robust TTSC when varying the threshold in cases with an early cure point, while in cases where the true cure point is later, the estimated cure point becomes highly variable even for small changes in the threshold. With a threshold of 2 years, the estimated TTSC for the DLBCL patients was 9.8 years. In conclusion, determining the exact TTSC involves choosing a clinically relevant threshold. This should be done carefully since small changes in the threshold may cause major changes in the cure point.

OC17-3: Survival bias in studies with time-varying exposures: an application to dietary patterns and age-related macular degeneration

Authors: Myra McGuinness[1], Jessica Kasza[2], Amalia Karahalios[1], Robyn H. Guymer[3], Robert P. Finger[4], Julie A. Simpson[1].
[1] *University of Melbourne, Australia,* [2] *Monash University, Australia,* [3] *Centre For Eye Research Australia, Australia,* [4] *University of Bonn, Germany.*

Diet has been proposed as a modifiable risk factor for the development of age-related macular degeneration, a leading cause of vision impairment in developed countries. However, quantification of the association between diet and age-related macular degeneration is challenging due to changing dietary patterns over time. It is further complicated by attrition due to death, since age-related macular degeneration is a disease associated with aging. Both diet and age-related macular degeneration are known to be associated with mortality; survival bias is therefore likely to be influential when estimating the magnitude of association between the two, especially if a large number of participants die before the final study wave. Numerous statistical approaches have been proposed to investigate exposure-outcome relationships in the presence of survival bias; however the majority of these approaches only consider the effect of an exposure at a single time-point. We employ the survivor average causal effect as a sensitivity analysis to investigate the association between diet and age-related macular degeneration. Our approach involves the estimation of the marginal probability of late age-related macular degeneration (the binary outcome) in order to estimate the survival average causal effect. We use inverse probability of exposure weighting to account for the time-varying exposure (Mediterranean diet ascertained at two study waves) in the estimation of this marginal probability. This novel approach explores the influence of survival bias, and also addresses bias produced through loss to follow-up. We use data from the Melbourne Collaborative Cohort Study to demonstrate these methods.

OC17-4: Challenges in updating published survival prediction models with a large cohort

Authors: Christine Wallisch[1], Georg Heinze[1], Daniela Dunkler[1].
[1] *Medical University Of Vienna, Austria.*

Risk prediction models are important elements of health screening programs, facilitating the computation of individualized disease risk for a patient. Since many such models were published recently, some researchers advocated against the development of new models as an inefficient way of improving risk prediction. Updating of existing models with new cohorts allows to combine existing and new evidence. Various proposals were made for updating survival prediction models, including the re-estimation of the intercept (or updating the baseline survival function of Cox models), the re-estimation of regression coefficients, or their extension by further predictor variables. Here we report on our experiences in updating four cardiovascular disease (CVD) risk prediction models: the Framingham CVD models 2008 and 1991, the ACC/AHA ASCVD model and the SCORE model for fatal CVD. Our updating cohort comprised 1.5M individuals who participated in the Austrian health screening program where CVD risk is regularly assessed. In theory, model

updating should be a simple task but in practice several challenges arose. First, some prediction models were reported in insufficient detail, and additional assumptions had to be introduced for updating. Second, the complexity of computations may quickly border on capabilities of contemporary computers if the updating cohort is very large. Finally, we encountered difficulties in updating parametric survival models which were caused by intransparent reporting of their exact parametrization both in published models and software packages. We critically discuss our proposed solutions and possible alternatives. Finally, we compare the performance of the original and updated models in a test data set.

OC17-5: Modifiable risk factors for depression in adolescence – understanding the causal role of obesity

Authors: Nicholas Turner[1].
[1]*University of Bristol, UK.*

Introduction: The prevalence of adolescent depression is alarmingly high and there is evidence that it is increasing. A population-based preventive approach to improving adolescent depression is required, this should focus on potentially modifiable risk factors – one such factor may be obesity. The aim of this project was to identify the causal relationships between obesity and depression in adolescents, through an across-cohort analysis of longitudinal observational data collected by 3 cohorts.

Method: Linear regression was used to examine the effect of obesity on later depressive symptoms. Generalized Estimating Equations were used to model the average effect of obesity on future risk of depression over time. Mendelian Randomisation analysis was used to overcome the problem of residual confounding. Cross-lagged structural equation modelling was used to investigate whether there is a bi-directional relationship between obesity and depression. Analyses were carried out within and across-cohorts.

Results: In all of the analyses (except for the MR) there was evidence of an association between level of obesity and depression sporadically throughout adolescence in females but not in males.

Discussion: The results of the project were inconsistent, however, the results suggest that the relationship between levels of obesity and adolescent depression in females warrants further investigation. The finding that there may be a relationship in females and only observed sporadically throughout adolescence may be related to pubertal stage, a potential link puberty, obesity and depression in female adolescents should be investigated.

Poster Contributed Session

PC3: Biomedical studies 1

Monday 10th July - 17.00-18.30 h. - Room: Hall
Chair: María Xosé Rodríguez Álvarez

PC3 - M3: Explanatory factors of uric acid levels in patients with chronic heart failure based on multivariate adaptive regression splines and fuzzy c-regression models

Authors: Aleksander Owczarek[1], Piotr Choręza[1], Jerzy Chudek[1], Habibullah Arabzada[1], Romuald Wojnicz[1].
[1] *Medical University of Silesia, Katowice, Poland,*

Background: Accumulated clinical and experimental data suggest that hyperuricemia may be responsible for severity and progression of the chronic heart failure. The aim of this study was to assess the role of potential factors explaining variability of serum uric acid levels in patients with chronic heart failure not receiving xanthine oxidase inhibitors and acetylsalicylic acid.

Material and methods: We analysed prospectively collected data enclosed 294 patients who underwent right ventricular endomyocardial biopsy due to unexplained HF with New York Heart Association (NYHA) functional class II and III heart failure symptoms and decreased left ventricular ejection fraction < 40% on radionuclide study. The following factors were analysed: concentrations of serum creatinine, high-sensitivity C-reactive protein (hs-CRP), B-type natriuretic peptide (NT-proBNP), uric acid (SUA), glucose, total cholesterol, fraction HDL and LDL of cholesterol, triglycerides, fibrate and D-dimers. Multivariate Adaptive Regression Splines, local polynomial smoothing with Epanechnikov kernel function as well as the Time Domain Constrained Fuzzy c-Regression Models were used.

Results: SUA concentration variability is explained by BMI and eGFR values as well as by serum NT-pro BNP levels. The SUA declined with eGFR and increased with BMI values and serum NT-proBNP levels. A weak negative correlation between log10(SUA) levels and the LVEF was also observed ($r = -0.130$; $p < 0.05$).

Conclusions: Kidney function, nutritional status and the severity of left ventricle dysfunction reflected by serum NT-proBNP levels explain the serum uric acid levels variability in patients with heart failure.

PC3 - M6: Data-driven cluster analysis for identifying groups within users of anti-osteoporosis medication, using real-world primary care data

Authors: Sara Khalid[1], M. Sanni Ali[1], Alan Silman[1], Daniel Prieto Alhambra[1].
[1] *University of Oxford, UK.*

Background and Objectives: Data-driven methods can be used for pattern recognition within a clinical population, enriching the existing analytical tools for clinical data analysis. We clustered

anti-osteoporosis drug users with similar risk factors, to better determine the influence of therapy on their fracture risk. Methods: Using the SIDIAP Database (anonymized primary care records for >80% of Catalonian population), complete data from 37,996 incident users (2007-2014) of anti-osteoporosis drugs (AODs) were analysed. Hierarchical clustering was used to derive sub-groups based on risk factors including age, gender, body mass index, smoking, drinking, Charlson index, steroid /sedative use, and fracture history. For each sub-group, on-treatment incident fracture rate (/100 person-years) was estimated. Results: Patients could be stratified into one of five clusters: 1) elderly multi-morbid men with high prevalence of smoking and drinking; 2) elderly women with high co-morbidity; 3) systemic steroid users; 4) secondary prevention (previous fracture history); and 5) younger (early post-menopause) women with low-medium co-morbidity. Group 4 had the highest fracture incidence (1.05 (95%CI 0.88-1.22), and 4.63 (95%CI 4.29-4.97), for hip and non-hip fractures, respectively); whilst Group 5 had lowest fracture incidence (0.15 (95%CI 0.11-0.20), 1.72 (95%CI 1.58-1.87), for hip and non-hip fractures, respectively). Conclusion: Cluster analysis identified sub-groups within AOD users, including expected patient groups but also a surprising cluster of younger women with low fracture risk, where therapy is probably not recommended. Further work should explore the usefulness of such data-driven algorithms for clinical data analysis.

PC3 - M9: Mapping the modified Rankin Scale (mRS) measurement to the Assessment of Quality of Life (AQoL) utility values

Authors: Mohammadreza Mohebbi[1], Lauren Sheppard[2], Marj Moodie[2], Lan Gao[2], Janice Collier[3], Julie Bernhardt[3], Leonid Churilov[3].
[1] *Biostatistic Unit, Faculty of Health, Deakin University, Australia,* [2] *Deakin Health Economics, Centre for Population Health Research, Deakin University, Australia,* [3] *Florey Institute of Neuroscience & Mental Health, Melbourne Brain Centre, Melbourne, Australia.*

Background. Reliable and accurate mapping techniques that translate the modified Rankin Scale (mRS) data into the Assessment of Quality of Life (AQoL) utility values are in high demand in health economic evaluation. This research constructs and compares three algorithms to translate the mRS into the AQoL. Methods. mRS and AQoL information was derived from 'A Very Early Rehabilitation Trial' (AVERT), a parallel-group, single-blind, randomised controlled trial conducted in 56 acute stroke units in five countries. 2104 patients with ischaemic or haemorrhagic stroke were randomly assigned to receive usual stroke unit care alone or very early mobilisation in addition to usual care. Ordinary least squares (OLS) regression with Box-Cox transformation, generalized additive model (GAM) with spline smother and CART regression tree were used to predict AQoL from pre-intervention mRS, mRS at 3 and 12 month, by age and sex. The performance of the models was evaluated using mean absolute, mean squared errors (MAE and MSE) and R2.10-fold cross-validation was implemented for model validation. Results. The OLS, GAM and CART regression tree yielded similar MAE, MSE and R2 in the internal and external validation with slightly better results for the CART. The model explained more than 74% of the variance in the individual AQoL scores with a MAE of 0.13 and MSE of 0.18. Conclusions. Our results suggest that models mapping the mRS onto the AQoL have similar predictions. The mapped mRS yields usable AQoL utility values however, due to excessive variability in observed AQoL within each mRS category, the use of these mapping tools cannot be recommended as an adequate alternative for explicit collection of utility data in stroke populations.

PC3 - M12: Predicting treatment response in personalized cancer therapy – method comparison and a neural network approach using prior information

Authors: Dorothea Weber[1], Manuela Zucknick[2].
[1] *Heidelberg University Hospital, Institute of Medical Biometry and Informatics, Germany,*
[2] *University Of Oslo, Institute Of Basic Medical Sciences, Department Of Biostatistics, Norway.*

Melanoma is the most dangerous type of skin cancer. The prediction of drug responses according to mutation profiles is essential in identifying promising treatment approaches for melanoma. Therefore, there is increasing interest in strategies and methods treating this type of data while at the same time reaching good prediction performance.

We applied tree-based models, neural networks, logic regression, and generalized linear models with an elastic net penalty in their original form and combined them with different ensembling approaches (bagging and boosting). This research points out the results of prediction performance, furthermore, it deals with the variable importance of the individual genes and its connection to processes that lead to cancer. A main result is that no method consistently outperforms all competitors across all experiments. Instead, the different cancer compounds seem to favor different drug response prediction models. Moreover, we reach good prediction performance by using independent data sets for training and testing our models in the case of drugs with similar response distribution. Furthermore, we use information about genes that are linked to the drug target through pathway sharing as prior information in neural networks. We discuss whether this information can increase the prediction performance and the possibility to give a better biological interpretation of neural network models. A novel finding is that those neural networks may indicate which target genes make the greatest contribution to the drug response. Therefore, models to predict the drug responses according to the mutation profile are an achievement, which could be a step towards personalized cancer therapy.

PC3 - M15: Diagnostic process from a classification subsets approach

Authors: Pablo Martínez-Camblor[1], Juan Carlos Pardo-Fernández[2].
[1] *Dartmouth College, USA,* [2] *Universidad de Vigo, Spain.*

Receiver operating characteristic, ROC, curve is a valuable and popular tool used for studying and comparing the diagnostic capacity of a given marker. Besides, the area under the ROC curve, AUC, is frequently used as index of the global discrimination capacity. However, diagnostic process is mainly related with certain classification subsets. Final decision is based on whether the subject is within or without of one selected subset. In this work, we revise the ROC curve definition for setting the classification subsets in the spotlight of the analysis. Considering classification subsets, we study the behavior of the non-parametric estimate of the ROC curve when both lower and larger values are associated with more probability of having the studied characteristic.

PC3 - M18: Confidence interval for the risk ratio in case of joint misclassification of exposure and outcome

Authors: Jenő Reiczigel[1].
[1]*University of Veterinary Medicine Budapest.*

Reiczigel et al (2017) proposed a confidence interval for the risk ratio (RR) in case of misclassification of the outcome but assuming no misclassification error in the exposure. There are situations, however, in which even the exposure status may be classified wrongly. Correction of the point estimate of RR in case of misclassification of both exposure and disease is studied by some authors (e.g. Brenner et al, 1993) but no confidence interval procedure has been published yet.Here a profile likelihood CI is proposed for this problem and its properties are examined by simulation. The method assumes that misclassification probabilities are known, allowing different sensitivity and specificity for the exposure and outcome. Misclassification in exposure and outcome are assumed to be independent. The procedure can also be applied for testing the dependence of two diseases and for quantifying the increase or decrease of the risk of a certain disease given another disease is present.According to our simulation results, the profile likelihood CI performs quite well, it maintains well the confidence level in a wide range of scenarios.This research was supported by the Hungarian National Research Fund (grant number OTKA K108571).

Reference 1: Reiczigel J, Singer J, Lang Zs (2017) Exact inference for the risk ratio with animperfect diagnostic test, Epidemiology and Infection, 145, 187-193.

Reference 2: Brenner H, Savitz DA, Gefeller O (1993) The effects of joint misclassification of exposure and disease on epidemiologic measures of association. Journal of Clinical Epidemiology, 46, 1195-1202.

PC3 - M21: Testing the equivalence of two regressions

Authors: Yi Hao[1], Dr. Sonja Grill[1].
[1]*Roche Diagnostics Gmbh, Germany.*

In the context of an FDA request for a simultaneous equivalence test of slope and intercept for the clearance of coagulation assays, the measurements of three different devices are investigated: an investigational device, a predicate device and a reference method. The slope and intercept refer to two regressions: (1) measurements of the investigational device to the reference method and (2) measurements of the predicate device to the reference method (for FDA request see [1]).

The challenge in this context is mainly the repeated measurement structure for each sample and the choice of the regression method. The literature does not offer an analytical solution for an equivalence test and its power addressing this specific framework. Therefore, the results of a bootstrap approach with data from three different measurement devices (simulated, based on real data) will be presented and discussed, considering the covariance structure of the two regressions. Special attention is also paid to power considerations of the equivalence test.

The current analysis is approximated by an equivalence test applying a linear regression (see [2]), and the power of the equivalence test is obtained by a bootstrap approach.

Reference 1: Public Workshop - Point of Care Prothrombin Time/International Normalized Ratio Devices for Monitoring Warfarin Therapy, March 18, 2016, http://www.fda.gov/MedicalDevices/NewsEvents/WorkshopsConferences/ucm476561.htm

Reference 2: Arne Henningsen and Jeff D. Hamann (2007). systemfit: A Package for Estimating Systems of Simultaneous Equations in R. Journal of Statistical Software 23(4), 1-40. URL http://www.jstatsoft.org/v23/i04/.

PC3 - M24: Predictive models of malignant transudative pleural effusions

Authors: Carla Díaz-Louzao[1], Lucía Ferreiro[2], Carmen Cadarso-Suárez[3], Francisco Gude[4], Luís Valdés[2].
[1] *Biostatech, Advice, Training & Innovation in Biostatistics, S.L., Santiago de Compostela, Spain,* [2] *Department of Pulmonology, University Clinical Hospital of Santiago de Compostela, Santiago de Compostela, Spain,* [3] *Department of Statistics and Operations Research, University of Santiago de Compostela, Santiago de Compostela, Spain,* [4] *Department of Clinical Epidemiology, University Clinical Hospital of Santiago de Compostela, Santiago de Compostela, Spain.*

Background: There are no firm recommendations when cytology should be performed in pleural transudates, since some malignant pleural effusions (MPEs) behave biochemically as transudates. The objective was to assess when would be justified to perform cytology on pleural transudates.[1].

Methods: 281 consecutive patients with transudative pleural effusion (PE) were enrolled and divided in two groups: MPE and non-MPE. Two prognostic models (logistic regression) were considered: (I) clinical-radiological variables; (II) combination of clinical-radiological and analytical variables. Calibration and discrimination (ROC curves and AUC) were performed [2].

Results: ModelI (left PE, radiological images, absence of dyspnea, serosanguinous appearance), and ModelII (ModelI variables + CEA) AUC's were 0.973 and 0.995, respectively. By applying bootstrapping techniques to not find false negatives in 95% of other possible samples, the cut-off points for the probabilities were 3% (ModelI) and 4% (ModelII). The false positive results were 32 (ModelI) and 18 (ModelII), with no false negatives.

Conclusions: The applied models have a high discriminative ability to predict when a transudative PE may be of neoplastic origin, being superior to adding an analytical variable to the clinic-radiological variables.

Reference 1: FerreiroL, GudeF, ToubesME, LamaA, Suárez-AnteloJ, San-JoséE, González-BarcalaFJ, GolpeA, Álvarez-DobañoJM, RábadeC, Rodríguez-NúñezN, Díaz-LouzaoC, ValdésL (2017). Predictive models of malignant transudative pleural effusions. Journal of Thoracic Disease; 9: 106--116.

Reference 2: SteyerbergEW (2009). Clinical prediction models: a practical approach to development, validation and updating. Springer, New York.

PC3 - M27: Biases incurred from non-random repeat testing of haemoglobin levels in blood donors

Authors: Ryan Chung[1], Michael J. Sweeting[1].
[1] *University of Cambridge, UK.*

Before donation, it is a requisite for blood donors to undergo a haemoglobin (Hb) test to ensure levels are not too low. Recently, it has been suggested that if a donor's measured Hb level is below the level required for donation, another reading should be taken and if that is below the threshold, then a third reading should be used [1]. A simulation study was performed to investigate any possible biases that may incur from this approach and the effect it has on the estimated diagnostic characteristics (e.g. sensitivity and specificity).

To simulate the distribution of Hb levels, data from an existing study of 17,000 donors were used, where levels were measured using a gold-standard haematology analyser. Five different scenarios were investigated. The first used selective testing as mentioned above, the second was the same but based on a maximum of only two measurements, the third took a single measurement and the fourth and fifth took the average of two and three measurements, respectively, for all donors. We studied the performance of each scenario in donors close to the Hb threshold, together with the overall population diagnostic characteristics.

Preliminary results using the selective testing strategy shows consequences on the performance of the device, including a clear upward bias for donors whom have a 'true' Hb level close to the threshold. We also see that taking the average of multiple measurements improves the sensitivity and specificity, and that taking a single measurement provides marginally better performance to the selective testing scenario.

Reference 1: Baart, A. Mireille, et al. "Hemoglobin assessment: precision and practicability evaluated in the Netherlands—the HAPPEN study." Transfusion 56.8 (2016): 1984-1993.

PC3 - M30: Data generating models of dichotomous outcomes: heterogeneity in simulation studies for a random-effects meta-analysis

Authors: Konstantinos Pateras[1], Stavros Nikolakopoulos[1], Kit Roes[1].
[1] *University Medical Center Utrecht, The Netherlands.*

Increasingly, simulation studies are used to assess properties of statistical methods in more complex settings. Except for the statistical methods, the Data Generating Model (DGM) may impact the inference. This model is essential to interpret the results and to arrive at proper conclusions. A case in point is the random-effects meta-analysis of dichotomous outcomes.

We reviewed a number of simulation studies that evaluated the normal-normal model for dichotomous outcomes and we assessed the DGMs used to generate events for a series of (heterogeneous) trials. So far, at least three alternative DGMs are utilized in the literature for generating individual

dichotomous trial data. The first makes the assumption of homogeneity in the control arm and places all the between-study variance in the event rate of the treatment arm. The second assumes a fixed average trial risk, with which we can calculate the event probability in each arm, based on a simulated overall treatment effect. The third incorporates the between-study variance in both treatment arms via the use of logits.

In this work we use three common meta-analysis methods as motivating examples to compare the implications of the three DGMs. We demonstrate how different choices for a DGM substantially affect the results of simulation studies and stress the importance of thorough evaluation and reporting in such studies.

PC3 - M33: Meta-analysis of Vancomycin MIC Creep in Staphylococcus aureus infections: sensitivity analysis of the method

Authors: Raquel Diaz[1], Carmen Rodrigues[1], Vera Afreixo[1], Bruno Gago[1].
[1] *University of Aveiro, Portugal.*

Vancomycin is currently the primary option treatment for methicillin-resistant Staphylococcus aureus (MRSA). However, an increasing number of MRSA isolates with high minimum inhibitory concentrations (MICs), within the susceptible range (vancomycin MIC creep), are being reported worldwide.

Resorting to a systematic review and meta-analysis, it was assessed the evidence of vancomycin MIC creep. To accomplish this, studies were retrieved from Pubmed database. The inclusion criteria for study eligibility include the possibility of retrieving, from the reported data, values of vancomycin MIC and information concerning the applied MIC methodology.

The vancomycin MICs values of all S. aureus isolates were reported by some methods, in this work we include the two most used and acceptable methods: Etest and BMD.

We consider the mean values of vancomycin MIC and the proportion of S. aureus isolates with vancomycin MIC $\geq$ 2 mg/L over time, as meta-analysis effect sizes. The MIC values are separated by year classes and the pooled effects were obtained by each class. Then the Spearman's correlation coefficient was used to evaluate the possibility of MIC creep.

The results showed no evidence of MIC creep. In order to assess the sensitivity of the analysis method applied, it was pooled the mean of vancomycin MIC and the proportion of vancomycin MIC $\geq$2 mg/L, using ten random effects meta-analysis methods. For all meta-analysis methods applied there was no statistically evidence of MIC creep phenomenon.

PC3 - M36: Treatment efficacy of cyclooxygenase inhibitors and acetaminophen in preterm neonates with symptomatic PDA: a network meta-analysis

Authors: Sudarat Eursiriwan[1], Sakda Arj-Ong Vallibhakara[1], Chusak Okascharoen[2], Ammarin Thakkinstian[1].
[1] *Section for Clinical Epidemiology and Biostatistics, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand,* [2] *Department of Pediatrics, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand.*

Background: Patent ductus arteriosus (PDA) is a common problem in preterm neonates. Cyclooxygenase (COX) inhibitors are effectiveness for closing PDA with prevalent adverse drug events. We conducted a systematic review and network meta-analysis (NMA) to compare treatment efficacy between COX and acetaminophen.

Methods: We searched randomized controlled trials (RCTs) from PubMed and Scopus up to December 2016. RCTs were eligible if they compared PDA closure rates between any pair of COX, acetaminophen, and placebo in preterm neonates with symptomatic PDA. Data were independently extracted by two reviewers. A NMA with consistency model was applied to assess treatment efficacy. Probability of being best treatment was estimated using surface under the cumulative ranking curves (SUCRA).

Results: A total of 40 RCTs (n=2506 patients) were included. Interventions were placebo (N=11), oral (N=5)/intravenous indomethacin (N=21), oral (N=16)/intravenous brufen (N=14), oral (N=5)/intravenous acetaminophen (N=1). A NMA showed that for oral administration, brufen, indomethacin, and acetaminophen were significantly higher rates than placebo with pooled RR of 2.30 (95% CI: 1.77, 2.98), 2.28 (1.55, 3.37), and 2.28 (1.60, 3.25), respectively. Among intravenous route, only indomethacin and brufen were significant difference compared to placebo with RRs of 2.16 (1.68, 2.78) and 1.93 (1.48, 2.52). Probability being first-three rank treatments for oral routes were brufen, acetaminophen and indomethacin whereas indomethacin and brufen were the best rank for intravenous route.

Conclusions: Indomethacin, brufen, and acetaminophen are drugs of choice for PDA closure. Intravenous route of COX may be an option for abnormal absorption.

PC3 - M39: Analysing missing data in longitudinal binary outcomes using generalized linear mixed-effects model with Markov correlation structure

Authors: María Helena Gonçalves[1], M. Salomé Cabral[2].
[1] *Ceaul and Fct, Universidade do Algarve, Portugal,* [2] *Ceaul And Deio, Faculdade De Ciências Da Universidade De Lisboa, Portugal.*

Longitudinal binary data are routinely collected in medical studies in which repeated observations of response variable are taken over time on each individual in one or more groups of treatments. A common problem in these studies is the presence of missing data since it is difficult to have complete records of all subjects for a variety of reasons.The use of the generalized linear mixed-

effects model (GLMM) approach is widely used to analyse longitudinal binary data when the goal is a subject-specific interpretation because it allows missing values on the response, provided they are missing at random (MAR), and accounts the correlation among the repeated observations of the same subject by the inclusion of random effects in the linear predictor. However in GLMM it is assumed that the observations of the same subject are independent conditional to the random effects and covariates which may be not true. In the R package bild the methodology implemented overcomes this problem using a generalized linear mixed-effects model with binary Markov chain (GLM3C) as the basic stochastic mechanism to accommodate serial dependence and odds-ratio to measure dependence between successive observations.In GLM3C approach missing values on the response are allowed provided they are MAR but some restrictions exist for the presence of missing data when they occur in the middle of the profile due to the imposed correlation structure. Taking into account these restrictions a simulation study was performed to give a statistical assessment of the impact of intermittent missing in terms of properties such as efficiency and coverage probability when compared with the GLMM approach. The R packages bild and lme4 were used.

PC3 - M42: Using the past to predict the future: how useful are longitudinal measurements?

Authors: David Hughes[1], Marta García-Fiñana[1].
[1] *University of Liverpool, UK.*

Over recent years methods that combine multivariate longitudinal data in a classification model have been proposed, including joint modelling of longitudinal outcomes and survival data, multistate models and longitudinal discriminant analysis. However, until recently, most prediction models only consider either a single longitudinal marker or only the most recent value of multiple markers.

In this talk we assess the benefit of models that incorporate longitudinal marker information. We compare models which use simple summaries of longitudinal data (including survival models, linear discriminant analysis and landmarking) with more complex methods to investigate the effect on classification accuracy.

We present results from an analysis of patients with epilepsy where the aim is to identify patients with drug resistant epilepsy. First, we use simple methods containing only summaries of longitudinal data to estimate the probability that a patient has drug resistant epilepsy. Then, novel multivariate approaches which model changes over time in the patients seizure and treatment history (including longitudinal discriminant analysis and joint models) are applied.

We compare the discrimination ability of each of the methods to assess the value of models including multiple longitudinal measurements. We also contrast and discuss the different information that can be obtained from each of the model types.

PC3 - M45: Clustering of individual disability progression trajectories of multiple sclerosis patients using clinical and imaging data

Authors: Ceren Tozlu[1], Françoise Durand-Dubief[2], François Cotton[3], Sandra Vukusic[4], Dominique Sappey-Marinier[5], Delphine Maucort-Boulch[1].
[1] *Laboratoire Biométrie et Biologie Evolutive (Lbbe, Umr Cnrs 5558), Université Claude Bernard-Lyon 1, Université De Lyon; Service De Biostatistique, Hospices Civils De Lyon, France,* [2] *Centre de Recherche en Acquisition et Traitement de L'image Pour La Sante, (Creatis, Umr Cnrs 5220 & U1044 Inserm), Université Claude Bernard-Lyon 1, Université de Lyon, France; Service de Neurologie A, Hôpital Neurologique, Hospices Civils De Lyon, Fra,* [3] *Centre de Recherche en Acquisition et Traitement de L'image Pour la Sante, (Creatis, Umr Cnrs 5220 & U1044 Inserm), Université Claude Bernard-Lyon 1, Université de Lyon; Service de Radiologie, Centre Hospitalier Lyon-Sud, Hospices Civils de Lyon France,* [4] *Service de Neurologie A, Hôpital Neurologique, Hospices Civils de Lyon, France,* [5] *Centre de Recherche en Acquisition et Traitement de L'image Pour la Sante, (Creatis, Umr Cnrs 5220 & U1044 Inserm), Université Claude Bernard-Lyon 1, Université de Lyon; Cermep-Imagerie du Vivant, Université de Lyon, France.*

Multiple sclerosis (MS) is the most frequent disabling neurological disease in young adults. Clinical exams allowing clinical scores and imaging exams using tools as Magnetic Resonance Imaging (MRI) help to the diagnosis and to monitor the disease evolution. The MS patients are classified actually into 4 major subtypes such as clinically isolated syndrome (CIS), remitting-relapsing (RR), secondary progressive (SP) and primary-progressive (PP). The major challenge of today's neurologist is to classify MS patients in these subtypes regarding their clinical course. An extension of the k-means method was proposed to classify patients using longitudinal imaging data beside of clinical data.

80 MS patients divided in 4 subtypes were followed up with clinical and MRI examination every six months during the first 3 years and every year during the last 2 years. The clustering method is an unsupervised method proposed in kml and kml3d packages and able to cluster the trajectories calculating the trajectory distance the at each time point. The method was performed on clinical and jointly clinical and imaging data.

The classification results were quite similar on clinical and jointly clinical and imaging data. The best cluster number was 3; one cluster was composed of CIS and RR patients as well as another one was mainly composed of PP and SP patients.

Although 4 subtypes are predefined, 3 clusters may be explained by the evolution of CIS patients afterwards as RR. This clustering method was able to give the clusters of the patients with a progressive evolution (PP and SP) and a moderate evolution (CIS and RR). An additional use of imaging data didn't have a significant effect to improve the results obtained on clinical data.

PC3 - M48: Unequal intra-cluster variances in trajectory classification

Authors: Amna Klich[1], René Ecochard[1], Fabien Subtil[1].
[1] *Lbbe, Umr Cnrs 5558, Université Claude Bernard Lyon 1 - Service de Biostatistique des Hospices Civils de Lyon.*

Classifying patients according to longitudinal measures (i.e., trajectory classification) has become frequent in clinical research. Two main approaches have been proposed: the mixture model and the classification model. In the mixture model, the trajectory of each subject is a mixture of typical trajectories with different weights. In the classification model, the trajectory of each subject is the typical trajectory of the cluster to which the subject belongs. The current version of the latter model supposes often that intra-cluster variance is the same in all clusters. This assumption is sometimes inappropriate; e.g., the measurements in diseased subjects are more heterogeneous than in healthy ones.

We developed a new version of classification model that allows for different intra-cluster variances. The classification EM algorithm was used; it alternates between EC-step, which assigns each subject to the cluster which provides the maximum posterior probability, and M-step, which estimates the parameters of the typical trajectory of each cluster by the generalized least-square method for unequal variances. Simulation studies showed that, in case of unequal intra-cluster variances, the misclassification rate is higher in the model with common variances than in the one with different variances when the simulated typical trajectories are not well separated.

The two models were applied to a randomized trial that compared the effects of low vs. standard dose of cyclosporin A on creatinine levels in the post-cardiac-transplant period. The creatinine typical trajectories were different between the two models; with different dose effect estimates. Classification models should thus consider unequal intra-cluster variances.

PC3 - M51: Modeling individual disability course of multiple sclerosis patients using clinical and imaging data

Authors: Ceren Tozlu[1], Françoise Durand-Dubief[2,4], François Cotton[3], Sandra Vukusic[4], Dominique Sappey-Marinier[2,5], Delphine Maucort-Boulch[1].
[1] *Laboratoire Biométrie et Biologie Evolutive (Lbbe, Umr Cnrs 5558), Université Claude Bernard-Lyon 1, Université de Lyon; Service de Biostatistique, Hospices Civils de Lyon, France,* [2] *Centre de Recherche en Acquisition et Traitement de L'image Pour La Sante, (Creatis, Umr Cnrs 5220 & U1044 Inserm), Université Claude Bernard-Lyon 1, Université de Lyon,* [3] *Centre de Recherche en Acquisition et Traitement de L'image Pour la Sante, (Creatis, Umr Cnrs 5220 & U1044 Inserm), Université Claude Bernard-Lyon 1, Université de Lyon; Service de Radiologie, Centre Hospitalier Lyon-Sud, Hospices Civils de Lyon, France,* [4] *Service de Neurologie A, Hôpital Neurologique, Hospices Civils de Lyon, France,* [5] *Cermep-Imagerie du Vivant, Université de Lyon, France.*

Multiple sclerosis (MS) is the most frequent disabling neurological disease in young adults. Clinical and imaging exams help to the diagnosis and to monitor the disease evolution. The MS patients are classified actually into 4 subtypes and the disability course and the risk for developing permanent disability are very

different from one patient to another. Today's neurologist challenge is to predict the individual disability evolution based on different type of markers. We proposed modeling individual disability course with latent class linear mixed model using longitudinal clinical, biological and imaging data.

80 MS patients divided in 4 subtypes were followed up with clinical and imaging exams every six months during the first 3 years and every year during the last 2 years. The disability course was modelled with the latent class linear mixed model which is able to take into account more than one mean profile of trajectories. The model was fitted on 4 latent class using lesion load, white matter (WM) and gray matter (GM) atrophy, age at onset, gender and subtype as well as an interaction between gender and subtype.

The results showed GM and WM atrophy had a significant effect on disability course as well as the effect of lesion load wasn't significantly different from 0. Furthermore, the males less affected from MS than females at disease onset. Moreover, EDSS was significantly lower among older patients at the disease onset.

Our results demonstrate the latent class linear mixed model proposed in this study was a good option for modeling the disability course using longitudinal clinical and imaging data; as MS patients are characterized by 4 different profiles of trajectory.

PC3 - M54: Comparing the fit of linear mixed models with different covariance structures based on Bayesian model comparison

Authors: Christos Thomadakis[1], Loukia Meligkotsidou[2], Nikos Pantazis[1], Giota Touloumi[1].
[1] *Department of Hygiene and Epidemiology, University of Athens, Greece,* [2] *Department of Mathematics, University of Athens, Greece.*

Likelihood-based methods ignoring the missingness mechanism are unbiased provided missingness is at random (MAR) and the correct model is used. This means that both the mean evolution and the covariance structure of a linear mixed model (LMM) should be modeled correctly. When modeling CD4 cells during untreated HIV infection, measurements taken after treatment initiation (ART) are excluded, leading to dropout. Pre-ART CD4 data are usually analyzed using an LMM with random slopes (RS) assuming MAR dropout. Two other approaches are (a) addition of a stochastic process e.g. Brownian motion (BM) on top of the RS and (b) use of splines in the design matrix of the random effects. We propose Bayesian model comparison based on the posterior model probabilities to discriminate between the two approaches. The proposed method requires the calculation of the marginal likelihoods, which was carried out using adaptive quadrature, after analytically integrating out both random and fixed effects and the within-subject variance. We compare the proposed method with the BIC criterion in a simulation study mimicking the CD4 data in the CASCADE study. Under the LMM with a BM process, the LMM with splines for the random effects overestimated the CD4 decline and both criteria always identified the true model. Under the LMM with splines, both models were nearly unbiased, and the proposed criterion identified the true model 66% of the time, whereas the BIC criterion only succeeded 7% of the time. Fitting the models to CASCADE CD4 data, the best model in terms of both criteria was the LMM with a BM process. As fixed effects estimates can be biased under MAR dropout and incorrect covariance specification, choosing the correct model is essential.

PC3 - M57: Comparison of regression methodology for analyzing health-related quality of life

Authors: Dal Ho Kim[1], E.J. Jang[2], J. Hwang[3], J. Lee[4].
[1] *Kyungpook National University, Daegu, Republic of Korea,* [2] *Andong National University, Andong, Republic of Korea,* [3] *Daegu University, Gyeongsan, Republic of Korea,* [4] *National Evidence-Based Healthcare Collaborating Agency, Seoul, Republic of Korea.*

The health-related quality of life (HRQOL) has become an important outcome measure in clinical trials, outcomes research and economic evaluations. The distribution of HRQOL such as EQ-5D tends to be skewed to the left, bounded at both ends (usually ranging between 0 and 1). Several regression models have been applied to estimate a relationship between patients' characteristics and HRQOL and to predict HRQOL using the estimated model. Beta distribution is very flexible for modeling bounded data, and beta regression is useful statistical technique in medical research for modeling of outcomes with ranged in (0,1), such as proportions and HRQOL. In this paper, we compared the regression methods including the ordinary least square regression model, censored regression model which is a generalization of the standard Tobit model and beta regression model for analyzing HRQOL. In addition, HRQOL data occasionally inflate at 1 or 0, and have heteroscedasticity property, but the beta regression model for mean is insufficient for handling these features. Therefore, we further considered the beta regression model for mean and precision and the zero-or-one-inflated beta regression model. We simulated data without inflation and with inflation at 0 or 1 and compared the regression methods using bias and root mean square error. For real data application, we used EQ-5D data from Korea Health Panel Survey data in 2013, and applied a cross-validation method to compare the predictive accuracy. Statistical analyses were performed using the package censReg, betareg, gamlss in the software R.

PC3 - M60: Estimating the economic costs of functional gastrointestinal tract disease using marginalized two-part model for semi-continuous data

Authors: Mohadese Shojaei Shahrokh Abadi[1], Anoshirvan Kazemnejad[1].
[1] *Tarbiat Modares University, Iran.*

Healthcare managers and policy makers need precise estimates of the future medical costs. Distribution of medical cost data is generally right-skewed and includes substantial number of zero values. A proper model must reflect these properties. Two-part models are proposed to manage the semi-continuous feature of cost data. These models will allow two separate models, one for zero values and another for positive values. Despite wide use of conventional two-part models, they are limited to conditional (on positive values) interpretation of regression coefficients from the second part; therefore, generalization of the results is only applicable to consumer population. To create such marginal inferences, marginal two-part models (MTP) are suggested.

Data of 5850 gastrointestinal tract patients was collected at the research center for gastroenterology and liver disease of Shahid Beheshti University of Medical Science, Tehran in a cross-sectional study (2010-2016). Themarginalized two-part model was adopted for modeling GT disease cost.

The average costs of GT diseases was estimated $85.99±12.48, by the MTP method. Minimum value of costs was estimated $3.80. While the actual average cost was $88.35±22.36, with a minimum of zero. MTP model of costs revealed that the number of days absent from work or reduced efficiency, number of times visiting a doctor and hospitalization are influential factors in both parts of the model.

The estimated average cost of MTP method is close to the true value. In addition, this model has performed well at estimating the true value of so-called zero costs and reducing the standard deviation. According to simulation results the MTP is preferred for modeling cost data.

PC3 - M63: Trends in CD4 cell counts for patients on antiretroviral therapy: a Kenyan clinic

Authors: Caroline Wanja Mugo[1], Ziv Shkedy[2], Samuel M. Mwalili[1], Roel Braekers[2], Christina W. Mwachari[3].
[1] *Jomo Kenyatta University of Agriculture and Technology, Kenya,* [2] *Hasselt University, Belgium,* [3] *Kenya Medical Research Institute, Kenya.*

For patients on combination antiretroviral therapy (ART) in resource-limited settings, changes in CD4 counts constitute an important component in patient monitoring and evaluation of treatment response as these patients do not have access to routine viral load testing. In this study, we quantified trends on CD4 counts in patients on combined ART in a comprehensive health care clinic in Kenya between 2005 and 2011 and evaluated the rate of change in the CD4 count in response to antiretroviral treatment. We also assessed factors that influenced time to treatment change focusing on baseline characteristics of the patients and the different drugs used. The study involved 2,036 naïve HIV patients that had at least two CD4 count measurement. The relationship between the CD4 cell count and time was modelled using a semi parametric mixed effects model. The Cox proportional hazards model was used to assess factors associated with the first regimen change. The results show that CD4 counts increased over time; these trends were similar regardless of the treatment used. Males were less likely to have their drug regimens changed (adjusted hazard ratio 0.7944, 95% CI: 0.67-0.95) compared to females. Stavudine based regimens had a higher drug substitution (aHR 2.067, 95% CI: 1.81-2.36) compared to Zidovudine. These findings suggest that CD4 count increases over time .Gender and the backbone used were found to be associated with regimen changes among the patients.Key words: ART;CD4 count; semiparametric mixed effects model.

Reference 1: Tadesse A.,Adetayo K.,Alemayehu W,et al(2016).Outcomes of first-line antiretroviral therapy and rate ofCD4 change among a cohort of adult HIV/ AIDS patientsin Ethiopia: Retrospective cohort study. PLOS.

PC3 - M66: Estimation of the odds ratios for continuous covariates using generalized additive neural networks

Authors: Ana Luisa Papoila[1], Carlos Bras-Geraldes[1], Patricia Xufre[1].
[1] *Nova Medical School|Faculdade de Ciências Médicas da Unl and Ceaul, Portugal.*

The application of artificial neural networks (ANNs) has been increasing in several areas of knowledge, mainly due to their flexibility, not only when modeling data from complex realities such as pattern and voice recognition, but also in simpler situations that include several independent variables and a response (dependent variable).

The Multilayer Perceptron (MLP) is a widely used ANN architecture, however its application in biomedical research is lower than that of Generalized Linear Models (GLMs) and Generalized Additive Models (GAMs). The main reason is the opacity of the MLP respecting to the analysis of the effects of each explanatory variable on the response.

Thus, the application of generalized additive neuronal networks[1] (GANNs: ANNs that mimetize GAMs) in the field of Medicine is much more promising as, beyond producing good estimates, their results are also interpretable.

For GAMs, the odds ratio function has already been proposed[2] for continuous covariates. In this study, we also incorporated in the GANN model the estimation of that function and corresponding confidence intervals.

This amelioration has inevitably increased the interpretability of the results of a GANN, resulting in a potential increase of the applicability of these ANNs in the biomedical area.

Reference 1: de Waal, D. A. e du Toit, J. V. (2011). Automation of generalized additive neural networks for predictive data mining. Applied Artificial Intelligence, 25(5):380–425.

Reference 2: Cadarso-Suárez, C., Roca-Pardiñas, J., Figueiras, A., e González-Manteiga, W. (2005). Non-parametric estimation of the odds ratios for continuous exposures using generalized additive models with an unknown link function. Statistics in medicine, 24(8):1169–1184.

11th July
Tuesday

Invited Session

IS4: Joint Modelling in Practice

Tuesday 11th July - 09.00-10.30 h. - Room: Auditorio
Chair: Carmen Cadarso
Organised by Carmen Cadarso, *University of Santiago de Compostela, Spain*

IS4-1: Semiparametric Copula-Based Regression Models

Giampiero Marra, *University College London, United Kingdom*

In this talk I will discuss a general class of bivariate copula regression models where the parameters of the marginal distributions and of the copula can be specified as functions of additive predictors allowing for several types of covariate effects (such as linear, non-linear, random and spatial effects). The estimation approach permits all model's parameters to be estimated simultaneously within a general penalized likelihood framework that uses a trust region algorithm with integrated automatic multiple smoothing parameter selection. The models can be easily used via the functions available in the R package SemiParBIVProbit. Many choices are available for the outcome distributions and dependence structure. The approach also allow to fit joint survival models. The modelling framework will be motivated and illustrated using some case studies.

References:

- Marra G, Radice R et al (in press), *A Simultaneous Equation Approach to Estimating HIV Prevalence with Non-Ignorable Missing Responses, Journal of the American Statistical Association*
- Marra G, Radice R et al (in press), *Bivariate Copula Additive Models for Location, Scale and Shape, Computational Statistics and Data Analysis*
- Radice R, Marra G, Wojtys M (2016), *Copula Regression Spline Models for Binary OutcomesStatistics and Computing, 26(5), 981-995.*

IS4-2: Joint Models with Multiple Longitudinal Outcomes and a Time-to-Event

Dimitris Rizopoulos, *Erasmus University Medical Center, Rotterdam, The Netherlands*

Joint models for longitudinal and survival data have gained a lot of attention the recent years. There have been extended to handle among others multivariate longitudinal data, competing risks and recurrent events, and nowadays there also exist several freely available software packages for their implementation. From the aforementioned extensions, the one that is most practically relevant is the multivariate longitudinal data one. Even though this extension is mathematically straightforward, from a computational viewpoint joint models with multiple longitudinal outcomes remain difficult to fit in practice due to the high number of random effects they require. This difficulty has also hampered to a degree their practical application. Here we present a novel approach that enables fitting such joint models in realistic computing times. The idea behind our approach is to split the estimation

in two steps, first to estimate a multivariate mixed model for the longitudinal outcomes, and then use the output of this model to fit the survival submodel. Such two-stage approaches have been previously proposed in the literature and have been shown to be biased. What is different in our approach is a correction we apply in the resulting estimates that transform them to the estimates we would expect to obtain if we would fit the multivariate joint model. This correction is based on importance sampling ideas. Simulation studies have shown that this corrected-two-stage approach works very satisfactorily also in difficult settings.

IS4-3: Joint Modelling Approaches in Diabetes Research

Francisco Gude, *University Hospital Complex of Santiago de Compostela, Spain*

Diabetes is a serious, chronic disease that occurs either when the pancreas does not produce enough insulin, or when the body cannot effectively use the insulin it produces.

A-Estrada Glycation and Inflammation Study (AEGIS), is a population-based study carried out in A-Estrada, Spain, which involved 1516 adults. From this project, two interesting case studies are presented, showing the benefits of using Joint Modelling approaches in diabetes research. In the first application, we investigated which factors influence the relationship between two glycated proteins: glycated hemoglobin and fructosamine, by using Bayesian Structured Additive Distributional Regression (Klein&Kneib, 2016) for bivariate continuous responses. In the second application, we will talk about the metabolic syndrome. To be diagnosed with metabolic syndrome a person must have obesity and insulin resistance. Copula Generalized Additive Models for Location, Scale and Shape (CGAMLSS, Marra&Radice, 2016) were used with the aim of obtaining a deeper knowledge about the association between obesity and insulin resistance, depending on clinical covariates.

AEGIS participants were also invited to undergo a 6-day period of continuous glucose monitoring procedures. Thus, our clinical research is now being focused on using patients' glucose profiles over time (i.e. functional data) to assessing its relationship not only with glycation markers, but also with developing diabetes and patient's survival. From a biostatistical point of view, the challenge posed is multi-faceted, requiring the development of new statistical extensions of CGAMLSS and Joint Models for Longitudinal and Time-to-Event data (Rizopoulos, 2012) to the context of functional data analysis.

Oral Contributed Sessions

OC18: Biostatistics for high dimensional data 2

Tuesday 11th July - 09.00-10.30 h. - Room: Sala Terra 2
Chair: Marianne Jonker

OC18-1: A tale of two networks

Authors: Wessel Van Wieringen[1], Carel Peeters[1], Renee de Menezes[1], Mark Van De Wiel[1].
[1] *Vumc, The Netherlands.*

The two-sample problem is addressed from the perspective of Gaussian graphical models (GGMs). It concentrates on the particular situation in which partial correlations (i.e. edge strength measures between node pairs in a GGM) are systematically smaller/larger (in an absolute sense) in one of the groups. Biologically, systematically weaker/stronger partial correlations represent an inactivate/active pathway.

Data in both groups are assumed to follow a GGM but their partial correlations are proportional, differing by a multiplier (common to all partial correlations). The multiplier reflects the overall strength of the conditional dependencies. Model parameters are estimated by means of penalized maximum likelihood, using a ridge-like penalty. A permutation scheme to test for the multiplier differing from zero is proposed.

A re-analysis of publicly available data (from six studies) on the Hedgehog pathway in normal and cancer prostate tissue shows its activation in the disease group. The analysis is accompanied by extensive diagnostics to assess the value of this conclusion.

Reference 1: Van Wieringen WN, Peeters CFW, de Menezes RX, and Van de Wiel MA (2017). Testingfor pathway (in)activation using Gaussian graphical models. submitted.

OC18-2: Joint variational inference for genetic association studies with multiple outcomes

Authors: Hélène Ruffieux[1,2], Anthony C. Davison[1], Jörg Hager[2], Irina Irincheeva[2].
[1] *Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland,* [2] *Nestlé Institute of Health Sciences, Lausanne, Switzerland.*

Combined inference for heterogeneous high-dimensional data is critical in modern biology, where clinical and various kinds of molecular data may be available from a single study. Classical genetic association studies regress a single clinical outcome on many genetic variants one by one, but there is an increasing demand for joint analysis of many molecular outcomes and genetic variants in order to unravel functional interactions. Unfortunately, most existing approaches to joint modelling are either too simplistic to be powerful or are impracticable for computational reasons. Inspired by Richardson et al. (2010, Bayesian Statistics 9), we consider a sparse multivariate regression model that allows simultaneous selection of predictors and associated responses. As Markov chain Monte

Carlo (MCMC) inference on such models can be prohibitively slow when the number of genetic variants exceeds a few thousand, we propose a variational inference approach which produces posterior information very close to that of MCMC inference, at a much reduced computational cost. Extensive numerical experiments show that our approach outperforms popular variable selection methods and tailored Bayesian procedures, dealing within hours with problems involving hundreds of thousands of genetic variants and tens to hundreds of clinical or molecular outcomes. Software is available at https://github.com/hruffieux/locus.

OC18-3: Type I error and false discovery rate control in RNA-seq differential analyses through a variance component score test

Authors: Boris Hejblum[1], Denis Agniel[1].
[1] *Rand Corporation, Harvard Medical School, U.S.A.*

Gene expression measurement technology is shifting from microarrays to sequencing, and since RNA-seq data are measured as counts, the statistical tools available for their analysis must be adapted. We model RNA-seq counts as continuous variables using nonparametric regression to account for their inherent heteroscedasticity, in a principled, model-free, and efficient manner for detecting differentially expressed genes from RNA-seq data. Our method can identify the genes whose expression is significantly associated with a factor or a group of factor, through a variance component score test, while accounting for both covariates and heteroscedasticity without assuming any specific parametric distribution for the (transformed) counts. Despite the presence of a nonparametric component, our test statistic has a simple form and limiting distribution, which can be computed quickly. A permutation version of the test is also derived for small sample sizes. Applied to both simulated data and real benchmark datasets, we show that our test has very good statistical properties, with an increase in stability and power when compared to state-of-the-art methods limma/voom, edgeR, and DESeq2. In particular, we show that those three methods can all fail to control the type I error and the False Discovery Rate under realistic settings while our method behaves as expected.

OC18-4: Overall assessment of statistical significance in the presence of selection bias

Authors: Donghwan Lee[1], Yudi Pawitan[2], Woojoo Lee[3].
[1] *Ewha Womans University, South Korea,* [2] *Karolinska Institutet, Sweden, Inha University,* [3] *South Korea.*

Validation studies have been proposed as a way to improve the generalizability of scientific findings, and the reproducibility of findings has been widely discussed in scientific experiments. We focused on situations in which many biomarkers are tested for their statistical significance in training study, and only a few markers among the significant findings are re-tested in a validation study. The standard FDR procedure based on the p-values of the validation study can be biased because all features of the validation are already significant in the training study. So, for overall assessment of the experiment, the information about the p-value or corresponding statistics in both the training and validation study

need to be integrated. In this work, we propose a novel method to assess the reproducibility and false positivity in the presence of validation study as well as training study. Furthermore, we show a link between the overall assessment and existing measures such as rediscovery rate and false discovery rate. Some simulation studies and two metabolomics datasets are considered to illustrate the application of overall assessment of statistical significance in high-throughput data analysis.

OC18-5: Testing the equality of a large number of populations: an application to genome-wide studies

Authors: Marta Cousido Rocha[1], Jacobo de Uña Álvarez[1,2], J. D. Hart[3].
[1]*Sidor Research Group, University of Vigo, Spain,* [2] *Department of Statistics and Operations Research & Centro de Investigaciones Biomédicas (CINBIO), University Of Vigo, Spain,*
[3] *Department of Statistics, Texas A&M University, U.S.A.*

The statistical analysis of data from genome-wide studies is non-trivial due to their complexity. In particular, such analysis often involves the comparison of a large number of correlated populations (e.g. gene expression levels) from a small number of individuals. In this work we introduce a goodness-of-fit method to test for the equality of densities in such a small sample, large dimension setting. The method extends the one in Zhan and Hart (2012) in that it allows for correlated outcomes across populations. The asymptotic null distribution of the test as the number of populations grows is derived, and its consistency under certain alternatives is established. Simulation studies are conducted. A combined application of the proposed method and multiple comparison procedures to the detection of genes differently expressed among two types of tumours in breast cancer is included.

Reference 1: Zhan, D., Hart, J. (2012). Testing equality of a large number of densities. Biometrika, 99, 1-17.

OC19: Design and analysis of clinical trials

Tuesday 11th July - 09.00-10.30 h. - Room: Sala Mar 2
Chair: Erik Cobo Valeri

OC19-1: Beyond mean modelling: bias due to misspecification of dispersion

Authors: Gillian Heller[1], Stephane Heritier[2], Dominique-Laurent Couturier[3].
[1] *Macquarie University, Australia,* [2] *Monash University, Australia,* [3] *University of Cambridge, UK.*

Traditionally in clinical trials the effect of a treatment on the mean of an outcome of interest is hypothesised. The underlying assumption in statistical models for the mean, is that the treatment effect on the response distribution is a location shift, with other aspects of the distribution (dispersion/shape/

variance) remaining the same. We consider data from a clinical trial for a treatment hypothesised to reduce the mean number of falls, in which it is apparent that the treatment not only reduces the mean number of falls, but also the variability in falls. As the response is overdispersed, potential statistical models include Poisson mixture distributions such as the negative binomial and Poisson-inverse Gaussian (PiG), which are typically parametrised in terms of a mean and dispersion parameter. For our clinical data, the PiG was found to provide a good fit. The conventional analysis, which hypothesises a treatment effect on the mean while assuming a constant dispersion parameter, yields a non-significant treatment effect. Mean and dispersion models allow linear models to be specified on the mean and dispersion parameter(s) of a broad range of distributions. We show that, on our data, if we model a treatment effect on both the mean and dispersion parameters, both effects are highly significant. In a simulation study we show that if a treatment effect exists on the dispersion parameter and is ignored in the modelling, estimation of the treatment effect on the mean can be severely biased. The question of what is meant by a treatment effect arises. This has implications in the planning of statistical analyses for clinical trials: should a treatment effect on the dispersion be prespecified.

OC19-2: Conceptual framework and development of CONSORT extension guidelines for reporting pilot and feasibility trials

Authors: Gillian Lancaster[1], Sandra Eldridge[2], Claire Chan[2], Mike Campbell[3], Lehana Thabane[4], Christine Bond[5], Sally Hopewell[6].
[1] *Keele University, UK,* [2] *Queen Mary University of London, UK,* [3] *University Of Sheffield, UK,* [4] *McMasters University, Canada,* [5] *University of Aberdeen, UK,* [6] *Oxford University, UK.*

Trials carried out in a health care setting typically involve complex interventions that require considerable planning if they are to be implemented successfully. Complex interventions are conventionally made up of several interacting components and present special problems related to the logistics of applying experimental methods in a health care setting. The UK Medical Research Council's guidance document on complex interventions emphasises the importance of thorough groundwork in designing and evaluating complex interventions and stresses the importance of conceptualising the problem at the development stage.

Pilot and feasibility studies are an essential part of trial preparation. Recent papers have shown that there was, and still is a dearth of pilot studies in the literature that state they are specifically in preparation for a randomised controlled trial, and that give a clear list of key objectives relating to the pilot phase. This talk will present (i) an overarching conceptual framework for defining pilot and feasibility studies conducted in preparation for a RCT that is testing the effectiveness of a complex intervention, and (ii) explain the rationale for and principles of the newly developed CONSORT extension to randomised pilot and feasibility trials.

Reference 1: Eldridge S., Chan C., Campbell M., Bond C., Hopewell S., Thabane L., Lancaster G.A. (2016) CONSORT Statement: extension to randomised pilot and feasibility trials. BMJ 355:i5239.

Reference 2: Eldridge S., Lancaster G.A., Campbell M., Thabane L., Hopewell S., Coleman C., Bond C. (2016) Defining feasibility and pilot studies in preparation for randomised controlled trials: using consensus methods and validation to develop a conceptual framework. PloS One, 11(3): e0150205.

OC19-3: Impact of carry-over effects in a classical cohort stepped wedge design under model misspecification

Authors: Ann Christina Foldenauer[1].
[1] *RWTH Aachen University, Department Of Medical Statistics, Germany.*

Background: The stepped wedge design (SWD) is a unidirectional cross-over design where units cross over from a control to an interventional treatment over time in a step-wise manner. The unidirectional assignment entails the risk to bias the treatment effect through carry-over effects. Unfortunately, carry-over effects are often neglected in the statistical models of affected SWD trials. We investigate the impact of carry-over on the treatment effect under model misspecification for a classical cohort SWD. As recommended by Cornu et al. (2013), we consider small sample trials where units are patients treated throughout all periods.

Methods: We investigate classical cohort SWDs in the presence of carry-over effects assuming correlated outcomes (AR(1)) within patients across periods for a continuous endpoint. The statistical base model is formulated similar to Hussey and Hughes' model for cluster randomized SWD trials, but evaluated on patient level. Carry-over bias will be calculated for the miss-specified model ignoring carry-over effects. Self- and between treatment carry-over effects are modelled as fixed effects following J. Kunert, and are analyzed via generalized least square theory.

Results and Conclusion: Due to the linear dependence within the classical SWD the GLS estimate of the treatment effect is biased by carry-over effects, showing opposing behavior for the between and intervention-self carry-over. For a large number of steps in SWD, the impact of intervention self-carry-over, respectively between treatment carry-over, becomes neglectable depending on the size of the autocorrelation factor. These results stress the importance of adequate study design and statistical model choices.

OC19-4: Mediation analysis in randomized trials: the direction of bias

Authors: Jordi Cortés Martínez[1], Erik Cobo Valeri[1], José Antonio González Alastrue[1].
[1] *Universitat Politècnica de Catalunya, Spain.*

Objective: To quantify how large the bias can be in mediation analysis when simple assumptions hold for relationships between outcomes and backdoor variables.

Design: We draw on data from the REVASCAT trial (1), which aimed to evaluate the effects of mechanical thrombectomy on modified Ranking Score (mRS) at 3 months, but it also evaluated mRS at 1 year after stroke. We analysed the additional effect of thrombectomy at 12 months when previous effects at 90 days were included as predictors in the model (2).

Results: The thrombectomy effect size of achieving independence at 12 months was reduced from an OR of 2.14 (95%CI, 1.15, 3.97) to an OR of 1.09 (95%CI, 0.40 to 2.98) once the already observed effect on mRS at 3 months (OR=2.46, 95%CI 1.33 to 4.57) was taken into account. This provides evidence that 89% [=1- ln(1.09)/ ln(2.14)] of the thrombectomy effect at 12 months is already present at 3 months post randomization.

Although those results rely on a randomized trial, adjusting for an intermediate variable opens the backdoor and may induce selection bias. We assume that the relationships between the hidden variables and both outcomes (at 3 and 12 months) are of the same magnitude; thus, we use simulation to approach the questions of: (a) how large the bias can be; and (b) how this bias can be moderated by adjusting for baseline minimisation factors.

Reference 1: Jovin TG, Chamorro A, Cobo E et al. Thrombectomy within 8 hours after symptom onset in ischemic stroke. N Engl J Med. 2015;372:2296-306.

Reference 2: Molenberghs G, Geys H, Buyse M. Evaluation of surrogate endpoints in randomized experiments with mixed discrete and continuous outcomes. Statist. Med. 2001; 20:3023–3038

OC19-5: Covariate adjustment and prediction of mean response in randomised trials

Authors: Jonathan Bartlett[1].
[1]*Astrazeneca, United Kingdom.*

A key quantity which is almost always reported from a randomised trial is the mean outcome in each randomised group. When baseline covariates are collected, these can be used to adjust these means to account for imbalance in the baseline covariates between groups, thereby resulting in a more precise estimate. Qu and Luo (2015) recently described an approach for estimating baseline adjusted treatment group means which, when the outcome model is non-linear (e.g. logistic regression), is more appropriate than the conventional approach which predicts the mean outcome for each treatment group, setting the baseline covariates to their mean values.

I will first describe how, for many commonly used outcome model types, the 'Qu and Luo' baseline adjusted outcome mean estimates are unbiased even when the outcome model is misspecified. Qu and Luo described how standard errors and confidence intervals can be calculated for these estimates, but treated the baseline covariates as fixed constants. When, as is usually the case in trials, the baseline covariates of patients would not be fixed in repeated sampling, I show that these standard errors are too small. I will describe a simple modification to their approach which provides valid standard errors and confidence intervals.

I will then discuss the impact of stratified randomisation or missing outcome data on the preceding results. The analytical results will be illustrated through simulations and application to a trial with recurrent events analysed using negative binomial regression.

Reference 1: Y Qu and J Luo. Estimation of group means when adjusting for covariates in generalized linear models. Pharmaceutical Statistics 2015; 14:56-62

OC20: Statistical methods in epidemiology 2

Tuesday 11th July - 09.00-10.30 h. - Room: Sala Terra 4
Chair: David Van Klaveren

OC20-1: Accounting for differential detection bias when evaluating race and body mass index association with prostate cancer in a prevention trial

Authors: Catherine Tangen[1], Phyllis Goodman[1], Jeannette Schenk[1], Cathee Till[1], Wendy Barrington[1], Scott Lucia[2], Ian Thompson[3].
[1] *Fred Hutchinson Cancer Center, U.S.A.,* [2] *University of Colorade Denver, U.S.A.,* [3] *Christus Santa Rosa, U.S.A.*

Prostate cancer is usually asymptomatic and detected through PSA screening and subsequent biopsy. Adoption of PSA screening and recommendations for and acceptance of prostate biopsy can depend on a number of demographic and health-related factors, resulting in biased and sometimes misleading estimates of the association of candidate risk factors or biomarkers with prostate cancer (PC) (ref 1). A recent publication using the United States National Cancer Institute- funded Selenium and Vitamin E (SELECT) cancer prevention trial (n=35,000 men), conducted in the U.S. and Canada, reported a strong positive correlation of BMI with risk of PC among African American men, but an inverse association of BMI among non-Hispanic white men (ref 2). We will report on an updated analysis strategy that seeks to minimize disease ascertainment bias by using imputed prostate cancer outcomes for those men who did not have a biopsy while on study. Comparison of original naïve estimation and improved imputed results, and strengths and limitations of these methods will be discussed along with implications for other disease settings.

Reference 1: Tangen et al. J Clin Oncol 34:4338-4344, 2016. Biases in recommendation for and acceptance of prostate biopsy significantly affect assessment of prostate cancer risk factors: results from two large randomized clinical trials

Reference 2: Barrington et al. JAMA Oncol 2015;1(3):342-349. Difference in association of obesity with prostate cancer risk between US African American and non-Hispanic white men in the Selenium and Vitamin E Cancer Prevention Trial.

OC20-2: Rethinking meta-analysis: addressing problems of non-collapsibility when combining treatment effects across patient population

Authors: Tat-Thang Vo[1], Stijn Vansteelandt[1].
[1] *Department of Applied Mathematics, Computer Science and Statistics, Gent University, Belgium.*

Introduction: One of the most overlooked statistical problems in meta-analysis is the non-collapsibility of effect measures across studies. According to this phenomenon, treatments that

are equally effective on patient subgroups, may appear to have different effectiveness on patient populations with different case mix. In view of this, it is important that meta-analyses be explicit for what patient population they describe the treatment effect.

Method: We developed novel meta-analysis approaches for randomized clinical trials, which use individual patient data from all trials to infer the treatment effect for the patient population in a given trial, either based on direct standardization or inverse probability weighting. Accompanying random-effect meta-analyses are developed, which enable disentangling heterogeneity in treatment effects due to non-collapsibility from that due to differential treatment effectiveness.

Result: We conducted simulation experiments in which treatment was equally effective in the three patient populations. Using the proposed methods, the false positive risk of the modified Cochrane Q's test of heterogeneity was well controlled at 5% (5.3%, 4.2% and 3.6% respectively), but inflated up to 31.5% when using the classical meta-analysis approach.

Conclusion: The new meta-analysis approach based on direct standardization was is more successful at correctly capturing treatment effect heterogeneity, and yields treatment effect estimates that express the effectiveness for well-defined patient populations.

OC20-3: Modeling of mean-covariance structures in marginal structural models

Authors: Chen Qu[1], Jianxin Pan[1].
[1] *University of Manchester, UK.*

In epidemiological studies, marginal structural models (MSMs) are used for properly estimating the causal effect of a time-dependent treatment, especially when confounders are present. Estimating the mean structure in the marginal structural model framework has been studied for a long time period, but there has been little research conducted on modelling of variance or covariance structures. According to the generalised estimating equations (GEE) approach of Zeger and Liang (1986), Hernan, Brumback and Robins (2000) suggested a selected covariance structure such as compound symmetry and AR(1) etc. However, questions arise whether the assumed covariance structure is indeed correct and what the consequences might be otherwise. In this research, we propose to use the inverse probability weighted generalized estimating equations (WGEE) approach to model the mean and covariance structures,simultaneously. These models allow for appropriate adjustment for confounding. The proposed WGEE approach yields unbiased estimators for both the mean and covariance parameters for longitudinal data with confounders. We demonstrate the use of the proposed approach in simulation studies and a real data analysis.

Reference 1: Robins James M, Heman Miguel A and Brumback Babette. Estimating the causale ect of zidovudine on CD4 count with a marginal structural model for repeatedmeasures.Statistics in medicine.2002;1689-1709

Reference 2: Pan Jianxin and Ye Huajun. Modelling of covariance structures in generalizedestimating equations for longitudinal data . Biometrika. 2006;927-941

OC20-4: Marginal structural model to evaluate the joint effect of socioeconomic exposures on the risk of developing end-stage renal disease in patients with type 1 diabetes: a longitudinal study based on data from the Swedish Childhood Diabetes Study Group

Authors: Laura Pazzagli[1], Anna Möllsten[2], Ingeborg Waernbaum[2].
[1] *Karolinska Institutet, Sweden,* [2] *Umeå University, Sweden.*

Purpose: Diabetic nephropathy is a severe complication of type 1 diabetes (T1D) that may lead to renal failure and end-stage renal disease (ESRD) demanding dialysis and transplantation. The aetiology of diabetic nephropathy is multifactorial and both genes and environmental and life style related factors are involved. In this study we investigate the effect of the socioeconomic exposures unemployment and receiving income support on the development of ESRD in T1D patients, using a marginal structural model in comparison with standard logistic regression.

Methods: This study is based on the Swedish Childhood Diabetes Register which started to register patients developing T1D before 15 years of age in 1977. In the analyses we include patients born between 1965 and 1979, developing diabetes between 1977 and 1994, followed until 2013 (n=4034). A marginal structural model (MSM) was fitted to adjust for both baseline and time-varying confounders.

Results: The main results of the analysis indicate that being unemployed for more than one year and receiving income support are risk factors for the development of ESRD. Multiple exposure over time to these risk factors increases the risk associated with the disease.

Conclusions: Using a MSM is an advanced method well suited to investigate the effect of exposures on the risk of complications of a chronic disease with longitudinal data. The results show that socioeconomic disadvantage increases the risk of developing ESRD in patients with type 1 diabetes.

Reference 1: Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550–60.

OC20-5: Automatic variable selection for exposure-driven propensity score matching with unmeasured confounders and complex correlation structures

Authors: Daniela Zöller[1], Leesa F. Wockner[2], Harald Binder[1].
[1] *University Medical Center Mainz, Mainz, Germany,* [2] *Qimr Berghofer Medical Research Institute, Brisbane, Australia.*

When investigating late effects of cancer, data from cancer registries provide an established way for contacting a representative sample of cancer survivors. Yet, reference data from a healthy population are required to judge the health status. While general population reference values are available,

comparisons to individual data allow for a fine-grade matching with respect to a larger number of confounders. Such a fine-grade matching approach can be performed with the help of propensity scores, which combine several characteristics into a score used for matching. However, multivariable model building for propensity score matching is challenging. The regression model for the propensity score may not only require variable selection, but it is still unclear to what extent effects on the exposure and the outcome should be required as a selection criterion. Unmeasured confounders, complex correlation structures, and non-normal covariate distributions further complicate matters. We consider the performance of different modeling strategies in a simulation design with complex but realistic data with effects on a binary outcome. Of the two main investigated strategies for variable selection, one focusses solely on the exposure, and the second requires association with both the exposure and the outcome. As a result, the strategies will be compared with respect to bias in estimated marginal exposure effects and increase in variance. When investigating the effect of unmeasured confounders on both, we distinguish between three types of covariates that might be missing (generally or strictly replaceable and bystander), and suggest tools based on resampling for potentially identifying the type of unmeasured confounder at hand.

OC21: Diagnostic tests

Tuesday 11th July - 09.00-10.30 h. - Room: Sala Mar 4
Chair: Inmaculada Arostegui

OC21-1: Binocular sensitivity and specificity of screening tests in cross-sectional diagnostic studies of paired organs

Authors: Joyce Raymond Punzalan[1], Yamuni Perera[2], Mingchen Ren[2], Christopher J. Rudnisky[3], Alexander R. de Leon[2].
[1] *University of The Philippines, Philippines,* [2] *University of Calgary, Canada,* [3] *University Of Alberta, Canada.*

We introduce new binocular accuracy measures as alternatives to conventional marginal measures that can be used to evaluate screening tests in diagnostic studies involving paired organs (e.g. eyes and ears). Specifically, we consider screening studies based on a cross-sectional design, where both diagnosis and disease status are determined after study enrolment or sampling, yielding paired binocular binary data described via two models, namely, the extended common correlation model and the Gaussian copula probit model. The first relies on the assumption of exchangeability of fellow organs, while the second is more flexible. Binocular versions of sensitivity and specificity are defined, respectively, as the probability of at least one correct positive diagnosis in patients with one or both organs truly diseased and the probability of two correct negative diagnoses for patients with both organs truly un-diseased. Comparisons between the conventional marginal and binocular sensitivities and specificities are illustrated for both models using data from a diabetic retinopathy study. We show that our methodology provides a viable alternative to conventional ways of assessing diagnostic accuracy of screening tests for paired organs. The binocular versions of sensitivity and specificity reflect the way screening tests are conducted in practice, and they overcome the shortcomings of conventional measures.

Reference 1: de Leon AR, Soo A, Bonzo D, Rudnisky CJ. Joint estimation of diagnostic accuracy measures for paired organs— application in ophthalmology. Biometrical Journal 2009; 51:837–850.

OC21-2: Individual participant data random-effects meta-analysis of the clinical utiltity of diagnostic tests and prediction models

Authors: Laure Wynants[1], Richard Riley[2], Dirk Timmerman[1], Ben Van Calster[1].
[1] *KU Leuven, Belgium,* [2] *Keele University, UK.*

Using patient data from multiple studies or centers to validate a diagnostic test or prediction model allows to investigate the predictive performance in relevant care settings and populations in which the model is intended to be used. Recently, meta-analytic techniques have been proposed to deal with heterogeneity in discrimination and calibration across populations (1). However, methods to meta-analyze the clinical utility of the model, as measured by the Net Benefit (2), are not available yet. In this study, we explain how to compute a summary measure of Net Benefit and the accompanying prediction interval at given risk thresholds, based on a trivariate random-effects meta-analysis of sensitivity, specificity and the prevalence, in a Bayesian framework. Further, we propose relevant probability statements on the likely Net Benefit in new populations. The proposed method is illustrated by case studies: one on the meta-analysis of published studies on ear thermometry to diagnose fever in children, and one on the validation of a multivariate risk prediction model for the diagnosis of ovarian cancer in a multicenter dataset. The clinical utility of the diagnostic tests was variable accross settings in both case studies, and, in addition, depended on the risk threshold that was assumed. These findings suggests that appropriate meta-analytic techniques should be used to appreciate the heterogeneity in the clinical utility across care settings.

Reference 1: Riley RD et al. Summarising and validating test accuracy results across multiple studies for use in clinical practice. Stat Med. 2015; 34.

Reference 2: Vickers AJ et al. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016; 352.

OC21-3: Likelihood ratio tests for non-inferiority of a new screening test in a paired design when the non-diseased population is a mixture

Authors: Johannes Berkhof [1], Matejka Rebolj[2], Peter Van De Ven[1].
[1] *VUMC, NL,* [2] *QMUL, UK.*

In a market where a standard screening test is already available, producers may develop new tests for a multitude of reasons. In this setting, test validation guidelines usually demand assessment of non-inferiority of the new test, rather than superiority. We consider the situation

where paired test results of subjects who are free of disease come from a mixture of two distributions. The first component corresponds to subjects without a disease-related condition, with positive test results reflecting noise of the tests. The second component corresponds to subjects in a pre-disease state. A difference in test performance between the screening tests observed in a specific population depends on the relative contribution of the two components. We present non-inferiority likelihood ratio tests with margins defined in terms of the mixture components. We present a local and a global likelihood ratio test where the latter test simultaneously assesses non-inferiority of both mixture components. Establishing global non-inferiority enables us to transfer non-inferiority to populations with a different prevalence of pre-disease. The performance of the non-inferiority tests is examined in a simulation study. A motivating example is human papillomavirus (HPV) testing in cervical cancer screening. HPV infection is a pre-disease state necessary for the development of cervical disease. In 2016, 16 million HPV tests from 125 producers were conducted to validate new screening tests by showing their non-inferiority to the FDA approved hybrid capture 2 test. We will apply the local and global non-inferiority tests to diagnostic studies in which the BDOnclarity test is compared to the hybrid capture 2 test.

OC21-4: Quantifying how test accuracy depends on threshold in a meta-analysis

Authors: Hayley Jones[1], Constantine A. Gatsonis[2], Thomas A. Trikalinos[3], A. E. Ades[1].
[1] *School of Social and Community Medicine, University of Bristol, UK,* [2] *Center for Statistical Sciences, School of Public Health, Brown University, USA,* [3] *Center for Evidence Synthesis in Health, School of Public Health, Brown University, USA.*

Many tests for disease produce an explicit continuous measure, e.g. the concentration of a biomarker in a blood sample. This is dichotomised at some threshold to call the result positive or negative. As the threshold is varied, the sensitivity and specificity of the test trade off against each other. Choice of the optimum threshold is a key question of clinical importance. For this to be based on a meta-analysis, we require summary estimates of test accuracy across all possible thresholds. Threshold can be included as a covariate in either of the two standard diagnostic test accuracy meta-analysis models. However, this implies strong – and not widely plausible – assumptions about the distributions of underlying continuous test measures in the diseased and disease-free populations.

Another common 'problem' is that of some studies reporting sensitivity and specificity at multiple thresholds. In fact, these extra data allow us to model the threshold effect much more flexibly, requiring weaker assumptions. We describe a model that uses such extra data and can be considered a generalised version of that recently described by Steinhauser et al. We assume multinomial likelihoods and parameterise in terms of the means and scale parameters of transformed test results in the two populations and a transformation parameter. We demonstrate using two case studies.

Using more data, where available, allows test accuracy to be summarised across all possible thresholds with minimal assumptions. This increases the potential clinical utility of the meta-analysis results.

Reference 1: Steinhauser S, Schumacher M. Rücker G, 2016. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. BMC Medical Research Methodology, 16:97.

OC21-5: Evaluating diagnostic accuracy in free-response detection-localization tasks using ROC tools

Authors: Andriy Bandos[1], Nancy Obuchowski[2].
[1] *University of Pittsburgh, USA,* [2] *Cleveland Clinic Foundation, USA.*

Diagnostic systems designed to detect possibly multiple lesions per patient (e.g., CT colonoscopy) are often evaluated in "free-response" studies that allow for diagnostic responses unconstrained in their number and locations. Analysis of the free-response studies requires extensions of traditional Receiver Operating Characteristic (ROC) analysis, which are termed free-response ROC (FROC) methodology. Despite substantial developments in this area, FROC tools and approaches are much more cumbersome to implement and interpret than traditional ROC methods. Alternative approaches that use well-known ROC tools (e.g., ROI-ROC) require defining regions of interest (ROI) and combining FROC data within ROIs. We propose an approach that allows analyzing FROC data using conventional ROC tools but without defining ROIs or reducing data. We describe how the intrinsic parameters of FROC study can be used to make FROC data analyzable using ROC tools. We calibrate the corresponding FROC and ROC curves on both conceptual and numerical levels and show that the differences in the overall performance indices for the nonparametric FROC and the new approach are asymptotically negligible and typically rather small in practice. The new approach is illustrated on the data from a large multi-reader study of colon cancer detection.

Poster Contributed Session

PC4: Survival analysis

Tuesday 11th July - 09.00-10.30 h. - Room: Hall
Chair: Ronald Geskus

PC4 - T1: Built-in clustering technique for time-to-event data through a discrete frailty term

Authors: Francesca Gasperoni[1], Francesca Ieva[1], Anna Maria Paganoni[1], Chris Jackson[2], Linda Sharples[3].
[1] *Politecnico di Milano, Italy,* [2] *MRC Biostatistics Unit Cambridge, United Kingdom,* [3] *London School Of Hygiene & Tropical Medicine, United Kingdom.*

In this work, we propose an innovative method to detect a hierarchical clustering structure for survival data. In particular, we focus on clustered time-to-event data. Times to hospital readmission for patients suffering from a chronic disease are a possible example of this kind of data. Statistical units are patients, while groups are hospitals.

In this framework, we are interested in investigating a second level of clustering structure, a hidden clustering structure. We want to find and characterize hidden populations among data, where each population is composed by a certain number of groups and each group belongs to only one population.

To address this research question, we propose a new model for the hazard function, starting from a semiparametric Cox model. Specifically, we add a random multiplicative factor to the Cox model and we set it as a discrete shared frailty term.

In this context, we have to estimate several parameters, i.e. the number of hidden populations, the frailty terms and the proportion associated to each population. Moreover, we have to estimate the classical elements of a semiparametric Cox model, i.e. the nonparametric baseline and the regressive parameters associated to the covariates. In order to estimate these quantities, we propose a tailored Expectation-Maximization algorithm.

To conclude, we show an application to a clinical administrative database. In this database, several information of patients suffering from Heart Failure is collected, like age, comorbidities, procedures etc. Applying our model, we are able to detect a hidden clustering structure among hospitals and this result can also be exploited by healthcare managers.

PC4 - T4: Flexible modeling of the effects of a sparsely measured continuous time-varying covariate in time-to-event analyses

Authors: Yishu Wang[1], Michal Abrahamowicz[2].
[1] *McGill University, Canada.*

Cox proportional hazards model implies that covariate effects are constant-over-time, and most applications assume linear effects of continuous covariates on the log of hazard. However, many prognostic factors have been found to have time-dependent (TD) and/or nonlinear (NL) effects. In contrast, these issues were little studied in the context of time-varying covariates (TVCs). Furthermore, when the TVC are sparsely measured, the estimated effects may not capture the true relationships between the current TVC and the hazard. To investigate these complexities, we extend the flexible modeling approach of Wynant & Abrahamowicz [Statistics in Medicine 33: 3318-37] to the modeling of the TVCs in survival analyses. To account for sparsely measured TVCs, we model the effect of time elapsed since last observation (TEL), which acts as an effect modifier for the TD/NL effects. Alternative conditional estimation is used to simultaneously estimate the TD, NL and TEL effects. We designed simulations based on the Framingham Heart Study (FHS) to evaluate the performance of the proposed model. Results indicate that the TD and NL effects of a TVC can be correctly estimated when the TVC is measured with high frequency. However, when the measurements of TVC are sparse, the estimates are biased, but the TEL estimate helps reduce the bias. We applied our model to re-analyze the FHS data. Most recent measures of systolic blood pressure (SBP) had a significant TD effect on the hazard of cardiovascular events, with the effect becoming weaker with aging of the cohort. The TEL effect was also significant and the estimate suggested the effect of SBP was lagged by about 8 months. This work is related to STRATOS Survival Analysis group (TG8).

PC4 - T7: Time-dependent effects in Cox regression models – do they matter?

Authors: Julia Krzykalla[1], Axel Benner[1].
[1] *German Cancer Research Center, Germany.*

The basic assumption for a Cox proportional hazards model is that the effect of every single covariate on the hazard is constant over time. Although this is rarely the case, an approximation by the average effect over time is often satisfactory. But, especially when investigating the effect of some (highly) variable covariate on the time-to-event outcome, a time-constant effect might not be reasonable. In this case, the Cox model can be extended by a time-dependent coefficient to be able to paint a more precise picture of the underlying association. This could be realized by creating a time-dependent covariate as the product of the variable of interest and a function of time. The crucial point is the selection of an appropriate time-transformation function. Various alternatives have been proposed, from step-functions or continuous parametric forms (linear, logarithmic, polynomials, ...) to non-parametric dependency terms such as splines. The appropriateness of the form of dependency can be tested or checked visually using Schoenfeld residuals. Tests for the goodness-of-fit of such models are also available but are very likely to lack power. In addition, the prediction performance of the final model needs to be evaluated, e.g. by using the time-dependent Brier score.

We evaluate different approaches of time-transformation functions to model time-varying effects by means of a practical example, mainly but not exclusively with respect to the time-dependent prediction error.

Reference 1: Therneau, T.M., Crowson, C. and Atkinson, E. (2016). Using time dependent covariates and time dependent coefficients in the Cox model [electronic resource]. https://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf

PC4 - T10: Gamma- generalized gamma mixture cure fraction model in survival data

Authors: Serifat Folorunso[1], Angela Chukwu[1].
[1] *Department of Statistics, University of Ibadan, Ibadan Nigeria.*

In this study, we use a methodology based on the Gamma Generated link function in the presence of mixture cure fraction models, considering that survival data are skewed in nature. The objective of the study is to propose a skewed family distribution on mixture cure model using a skewed Gamma link function which can handle heavily skewed survival data. The mathematical properties of this proposed model were explored and the inferences for the models are obtained. The proposed model called Gamma- generalized gamma mixture cure model (GGGMCM) will be validated in the presence of a real life survival data set.

PC4 - T13: Estimating risk of seizure recurrence for people with epilepsy

Authors: Laura Bonnett[1], Anthony G. Marson[1], Jane L. Hutton[2], Catrin Tudur Smith[1].
[1] *University of Liverpool, UK,* [2] *University of Warwick, UK.*

Prognostic models within epilepsy tend to model time to a specific event such as 12-month remission from seizures or first seizure after randomisation via Cox's proportional hazards model. Although such models are simple to fit and easy to interpret clinically, they fail to account for the recurrent nature of seizures which define epilepsy. The Prentice, Williams and Peterson - Counting Process (PWP-CP) model was therefore used to estimate risk of seizure recurrence including all seizures across the whole follow-up period, not just those occurring prior to the specified time point. This is the first time such a model has been applied to epilepsy data. Such a model can demonstrate improved power and enables a better understanding of treatment effects within a clinical trial.

The results from the PWP-CP model were similar to those for conventional Cox models for related fixed time point outcomes. In general, the direction of effect was consistent. However the confidence intervals obtained via the PWP-CP model tended to be narrower due to the increase in statistical power.

This initial exploration suggests that the PWP-CP model is a plausible one for seizure recurrence by accounting for the clustering of seizures experienced by patients with epilepsy. Further work is required to validate the model and demonstrate its increased statistical power in alternative data. Such a model may be worth considering when designing future clinical trials in medical conditions typified by recurrent events to ensure improve efficiency and statistical power.

PC4 - T16: Analysis of time to first viral reactivation with transient reactivations and intermittent observation times

Authors: Ian James[1], Elizabeth McKinnon[1].
[1] *Murdoch University, Australia.*

We consider analysis of the time to first occurrence of a particular state in a fluctuating two-state process, where the process is only observed intermittently. Our motivating example comes from assessment of factors associated with times from organ transplantation to the reactivation of latent cytomegalovirus (CMV). CMV viral load is determined post-transplant from samples taken at follow-up visits so it is possible that reactivation may occur and subside within an inter-visit interval. The time to first reactivation may thus be missed or at least be interval censored. We consider potential biases inherent in methods that are based simply on the time of the first observed reactivation, and explore some strategies to incorporate the observation uncertainties.

PC4 - T19: Handling informative cluster size in discrete survival analysis to model in-vitro fertilization data in multiple stages

Authors: Azadeh Ghaheri[1], Aliakbar Rasekhi[1], Reza Omani Samani[2], Ebrahim Hajizadeh[1].
[1] *Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran,* [2] *Department of Epidemiology and Reproductive Health, Reproductive Epidemiology Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran.*

In-vitro fertilization (IVF) cycles should be passed successfully through multiple stages during the procedure (implantation, clinical pregnancy, no abortion, and delivery) to achieve live birth. So the risk sets at each stage are defined conditional on success at the previous stage as in discrete survival models. Pregnancy outcomes are also believed to be correlated within a single woman. To model the factors associated with stage-specific probability of success/failure during IVF cycles, the model presented by Maity et al. (2014) was first used to deal with multiple cycles and multiple stages of each cycle. In IVF data the number of cycles an infertile woman undergoes is believed to be associated with the success probability (known as informative cluster size (ICS)). However, the model presented by Maity et al. does not take account of ICS. In data on 996 cycles of 511 infertile women, during April 2011 and March 2012 in a referral infertility center of Iran, the number of cycles each woman has undergone is reversely associated with success conditional on other predictors and cluster size is believed to be informative. Therefore, clustered-weighted generalized estimating equation was also considered in the model presented by Maity et al. to handle ICS. Receiving frozen embryo transfer was associated with higher odds of success compared to receiving fresh but, exploring the effect

of fresh/frozen embryo transfer on stage-specific odds of success showed that women receiving fresh embryo transfer did significantly better in clinical pregnancy stage, but women receiving frozen embryo transfer could continue the cycle as successfully as those receiving fresh embryos.

PC4 - T22: Modeling cardio-vascular disease related mortality in breast cancer survivors using large, registry based data from Norway

Authors: Milada Cvancarova Småstuen[1], Kristin Reinertsen[2], Sophie D. Fosså[2], Grethe S. Tell[3].
[1] *Hioa, Norway,* [2] *Oslo University Hospital, Norway,* [3] *University of Bergen, Norway.*

Background: Cancer survivors may develop side-effects of their treatment and often demonstrate increased risks for specific diseases. Cancer is often diagnosed in the elderly, so it is difficult to detangle if a selected risk is due to life style, genetics or previous cancer treatment. Further, a cancer survivor can have increased risks for multiple diseases but only one cause of death. We present our statistical approach to a complex clinical question using several large nation-wide registries in Norway. We examined the effect of breast cancer treatment on cardio-vascular disease (CVD) related mortality comparing women with left and right sided cancer and with the individuals from the general population. Materials and Methods: We identified 18614 females who were diagnosed with breast cancer between 1984 and 2002 and who survived at least 10 years after their diagnosis. Each cancer survivor was age-matched with 5 individuals selected at random from the whole Norwegian population. Each individual in Norway is at birth given a unique ID number which enables linkage with registry based data. All individuals in our sample were linked with data from the Death registry, Cancer registry and Statistics Norway. Cause specific mortality was modeled using Cox proportional hazards model stratified by matched set. Further, causes of death were categorized as CVD- related, cancer- related (breast cancer) and other causes. Cumulative incidences for all causes of death were modeled using competing risk approach and Fine and Grey regression. Conclusion: Large registry based data make it possible to answer complex clinical questions, however, advanced statistical methodology and close co-operation with medical experts is needed.

PC4 - T25: Statistical modelling of cure rates in cancer and their relationship to residual tumour

Authors: Crispin Musicha[1], David A. Cairns[1], Linda D. Sharples[1, 2].
[1] *Leeds Institute of Clinical Trials Research, University of Leeds, United Kingdom,* [2] *London School Of Hygiene And Tropical Medicine, United Kingdom.*

Background: Efficacious anti-cancer treatment has provoked an interest in estimating the proportion of 'cured' patients and how this relates to residual tumour. Two common approaches for estimating 'cure' have been proposed: considering relapse following treatment, where focus is on estimating the proportion that will not relapse; and considering relative survival where the 'cured' group regains the hazard of death of the general population and the 'non-cured' have an elevated hazard.

Methods: Models for estimating the cured proportion were investigated in randomized trials in multiple myeloma (MM) and chronic lymphocytic leukaemia (CLL). Parametric and semi-parametric mixture models for time to relapse and relative survival models incorporating population mortality data were considered. The effects of residual tumour and other factors were investigated by including covariates on the cured proportion parameter. Models were implemented using Bayesian Updating using the Gibbs Sampler (BUGS).

Results: The MM trial followed-up 427 patients for a maximum of 7 years. A clear plateau in time to relapse was observed; enabling robust cure proportion estimation. The proportion of patients who returned to population survival rates was also simple to estimate. There was a relationship between residual tumour and cure proportion. 415 patients with CLL were followed-up a maximum of 6 years. In contrast no plateau was observed in time to relapse and the cure proportion was difficult to identify, unless strong prior constraints were imposed.

Conclusion: It is possible to estimate the 'cured' proportion with sufficient follow-up and a clear plateau. Bayesian methods are useful for incorporating external information.

PC4 - T28: Small sample effects on large sample estimation methods for parametric lifetime models

Authors: Hasinur Rahaman Khan[1].
[1]*Associate Professor of Applied Statistics. Institute of Statistical Research and Training (ISRT). University of Dhaka. Bangladesh.*

If number of uncensored observations is small, the maximum likelihood based asymptotic normal approximation using pivotal quantity for the parameter under parametric lifetime models is not very accurate since the log-likelihood function tends to be asymmetric and not closely approximated by a quadratic. As alternative, the Sprott's parametrization and the likelihood ratio test approaches are used by the practitioners since the approaches provide more accurate results even for small samples. The study proposes a new parametrization approach that outperforms the Sprott's and likelihood ratio test approaches in most of the cases. The study also attempts to indicate the extent to which the large sample methods including the proposed method give appropriate confidence interval coverage in small samples for lifetime models under right censoring mechanism with several simulation studies. Effect of sample size and censoring rates are considered broadly to demonstrate the results. Real examples are provided for further illustration purpose. Both type of results reveal that all methods having equal confidence interval coverage if sample size is reasonably large but for any small sample generally smaller than 50--80 the method of maximum likelihood based asymptotic normal approximation having significantly much lower confidence interval coverage while the proposed method including the Sprott's parametrization and the likelihood ratio test methods having always impressive coverage irrespective of the level of censoring and the size of true parameter values.

PC4 - T31: United milestones survival analysis in single-arm trial

Authors: Isao Yokota[1], Satoshi Teramukai[1].
[1] *Kyoto Prefectural University Of Medicine, Japan.*

In cancer clinical trials, single-arm trials are conducted to explore new treatment strategy. Time-to-event outcome is often compared with a result estimated from historical study. MLE test imposes parametric survival distribution like exponential model, that may not be reliable. Recently phase II designs based on one sample log-rank test were proposed (Kwak and Jung 2014, Wu 2016). Since these nonparametric methods require hazard or survival function over a period under the null hypothesis, they seem to be applicable if individual participant data can be obtained. When we have the survival proportion at the fixed time point under the null hypothesis, one-sample test for the difference in survival proportion can only be used, which Chen (2015) called `milestone survival analysis'. The milestone time need not be single point because of statistical efficiency as the null hypothesis consists of several survival proportions. Furthermore, two or three milestone survivals have same clinical importance, for example short-term and long-term prognosis. For decision making, unified p-value must be required. In this talk, we propose one sample test for united survival proportions of several milestone times. The method is identical to one sample log-rank test when no censoring occurred. The performance of tests will be shown through Monte Carlo simulations.

Reference 1: Chen TT. Milestone survival: A potential intermediate endpoint for immune checkpoint inhibitors. JNCI 2015; 107: djv156.Kwak MJ, Jung SH. Phase II clinical trials with time-to-event endpoints: optimal two-stage designs with one-sample log-rank test. Stat Med 2014; 33: 2004-16. Wu J. Single-arm Phase II cancer survival trial designs. J Biopharm Stat 2016; 26: 644-56.

PC4 - T34: Modelling time to multiple event types and an extended model for recurrence with an application to the Drakenstein child health study

Authors: Jordache Ramjith[1], Landon Myer[1], Heather Zar[1], Francesca Little[1].
[1] *University of Cape Town, South Africa.*

Competing risks (CR) models are widely used in failure time analyses when individuals can experience one of several events of interest. Cause-specific hazards (CSH) and sub-distribution hazards have been used for CR. When individuals experience multiple events within follow-up, researchers make use of CR models and only consider the first failure but available data are not fully utilized and there may be biased regression estimates. A CSH model with an unrestricted risk set can allow multiple event types but assumes event independence. Wei et al.(1989) proposed allowing individuals to be at risk for all events throughout follow-up and then adjusting for event dependence by clustering on the individual's longitudinal profile (WLW model). A frailty approach uses a random effects term to capture dependence between events within a cluster. To test the utility of this approach, we applied these models to data from the Drakenstein Child Health Study and explored the effects of risk factors on time to first severe and first non-severe pneumonia. The WLW model and unrestricted CSH model performed similarly overall, but the WLW model may be superior with greater event dependence. The frailty model showed much larger effects, indicating bias in the model estimates. We then extended the WLW model by combining the model with the Prentice et al.(1981) conditional

model to include recurrence of each event type. The new model was applied to the study data and appeared to be fitting well in the residuals plot. We were able to distinguish the effect of different risk factors on different event types and identify whether these effects were on the first occurrence of these event types or carried through the recurrence of each event type.

PC4 - T37: Outcomes of blood pressure targets in clinical trial versus primary care setting

Authors: Lisanne Gitsels[1], Elena Kulinskaya[1], Ilyas Bakbergenuly[1], Nicholas Steel[1].
[1]*University Of East Anglia, England.*

Objective: The primary objective was to compare outcomes of different systolic blood pressure (SBP) targets in the US clinical setting and the UK primary care setting.

Methods: Data from the SPRINT randomised control trial and The Health Improvement Network (THIN) primary care database were used to develop survival models for longevity and adverse renal outcome (ARO, main adverse effect) at different SBP targets given treatment in people without diabetes and chronic kidney disease. The hazard of all-cause mortality or ARO associated with SBP targets was calculated by a multilevel Cox's proportional hazards regression, adjusted for sex, age, race, smoking, BMI, SBP, cardiovascular disease, number of antihypertensive agents at baseline, additional medication after trial entry, their interaction, and clinical site.

Results: Compared to SBP target of ≤120 mmHg, SBP target of ≤140 mmHg was associated with increased hazard of mortality of 1.42 (1.06-1.90) in SPRINT, but with decreased hazard of 0.70 (0.65-0.76) in THIN. Both in SPRINT and THIN, SBP target of ≤140 mmHg was associated with decreased hazard of ARO of 0.32 (0.22-0.46) and 0.87 (0.80-0.95), respectively. Additional antihypertensive agents (3+) were associated with increased hazards of both outcomes, with H. of 1.71-1.74 in SPRINT and 1.43-2.23 in THIN, yet being on 2 agents had survival benefits in SPRINT (H. 0.70-0.79).

Conclusions: Lower SBP target was associated with survival benefits in the clinical setting, but with increased hazard of mortality in the primary care setting. In both settings, polypharmacy patients tended to be worse off. An intensive control of SBP may benefit a selected subgroup of patients, but it appears harmful for the broader population.

PC4 - T40: Expression and clinicopathological correlations of retinoid acid receptor responder protein 1 in renal cell carcinomas

Authors: Aenne Glass[1], *G. Kundt*[1].
[1]*Institute for Biostatistics, University of Rostock, Germany.*

Aim: To evaluate the expression and prognostic value of RARRES1 at protein level in renal cell carcinoma (RCC).

Materials & Methods: Expression profile of RARRES1 was analyzed in 903 documented RCC followed by clinicopathological correlations and survival analysis.

Results: RARRES1 expression was seen in 72.5% of RCC. A stronger RARRES1 expression was seen in high grade compared with low grade RCC ($p < 0.001$). Logrank tests revealed shorter overall survival in RARRES1 positive RCC ($p = 0.006$) and in pT1/2 tumors with RARRES1 expression ($p = 0.002$).

Conclusion: The variable expression profile in low and high grade RCC may reflect and confirm the differences of previous gene expression analysis. There was a significant prognostic value of RARRES1 expression in patients with RCC, especially in pT1/2 tumors.

Reference 1: Expression and clinicopathological correlations of retinoid acid receptor responder protein 1 in renal cell carcinomas. Zimpfer A, Dammert F, Glass A, Zettl H, Kilic E, Maruschke M, Hakenberg OW, Erbersdobler A. Biomark Med. 2016 Jul;10(7):721-32. doi: 10.2217/bmm.16.12.

PC4 - T43: The impact of selection bias for the log rank test

Authors: Marcia Rückbeil[1], Ralf-Dieter Hilgers[1].
[1] *Rwth Aachen University, Department of Medical Statistics.*

Background: The presence of selection bias in clinical trials prevents a valid comparison of the treatment effects between two study groups. Selection bias may arise even in single- or double-blinded randomized trials if past treatment assignments are unmasked, for example due to characteristic side effects. For the modelling of selection bias in trials with a time-to-event outcome there exists only a parametric biasing policy for exponentially distributed outcome. In practice, however, the assumption of exponentially distributed outcome is usually too restrictive, especially since the most commonly used test methods are either nonparametric or semiparametric.

Methods: We propose a semiparametric biasing policy for the case of a time-to-event outcome variable. Based on this model we study the impact of third order selection bias on the type I error rate if the treatments are compared using a log rank test. Special attention is paid to the influence of censoring. We compare the observed type I error rates to the nominal significance level in a simulation study and give a theoretical approximation for the shifted test statistic in the presence of bias.

Results and Conclusions: Our simulation results show that the observed type I error rates always exceed the nominal significance level and that our approximation obtains reasonable estimates of the biased type I error probabilities. Censoring reduces the increase of type I error probability. Since the presented biasing policy depends on the random allocation of patients, our results can further be used to compare different randomization procedures with respect to their susceptibility to selection bias.

PC4 - T46: Development of an accurate prognosis model in practice

Authors: Iola Pinto[1], Ana Luísa Papoila[2], Karina Soto[3].
[1] *Isel and Cma Fct-Unl, Portugal,* [2] *Nova Medical School/Fcm Unl and Ceaul, Portugal,* [3] *Hospital Fernando Fonseca and Cedoc.*

The ability of making predictions on the prognosis of a disease became an important aid in the medical decision process. Physicians in their daily lives are confronted with doubts and uncertainty that may be supported by prediction models. These are multivariable models as the outcomes usually depend on several predictors. Still in medical research, a vast literature can be found about this subject such as on the development of biomarkers in several areas of Medicine. Nowadays these type of models are broadly applied, however, often not in a very correct way. Therefore, some guidelines emerged that rule their implementation and dissemination of results1, 2. Care must be taken with problem definition and data management, model specification and estimation, model performance, model validity (including penalized regression shrinkage), and model presentation. In this study we illustrate the use of the referred guidelines in nephrology. In a prospective cohort study, 616 patients were enrolled at admission to the emergency department and followed up for a median time of 62.1 months with the aim of investigating whether community-acquired acute kidney injury, encountered in a tertiary hospital emergency department setting, increased the risk of chronic kidney disease.

Reference 1: Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. European Heart Journal 2014; 35, 1925-1931.

Reference 2: Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015;162: W1-W73.

PC4 - T49: Flexible parametric models applied to the study of time to non-persistence in a chronic disease treatment

Authors: Ana Rita Godinho[1], Cristina Rocha[2], Zilda Mendes[1].
[1] *Centre for Health Evaluation & Research (CEFAR), National Association of Pharmacies (ANF), Portugal,* [2] *Faculty of Sciences University of Lisbon and Center of Statistics and Applications University of Lisbon, Portugal.*

Population ageing has led to an increase in the incidence of chronic diseases associated with painful conditions, that reduce patients' quality of life. Medication adherence and persistence are vital to improve health outcomes, thus being crucial to understand the factors that influence them. This work aimed to identify the factors that influence time to non-persistence in a particular chronic disease treatment, by fitting and comparing flexible parametric regression models [1]. Another goal was to identify the model that best described the hazard function of each group of patients and, thus, produced more precise estimates of the adjusted hazard ratios. Data were obtained from an observational, prospective cohort study with 360 patients. During patient recruitment, in the Portuguese community pharmacies, several sociodemographic and health variables were collected, but only three significantly influenced time to non-persistence: Age, Living alone and Treatment

(monthly or weekly). The comparison of the graphs of the estimated survival functions, obtained with the simple parametric models with one covariate and with the flexible models, shows that the introduction of a cubic spline with m internal knots greatly increases the parametric models' flexibility. An increase in the number of internal knots leads to more precise estimates and to a decrease in the AIC value of each model. Through flexible parametric modelling, it is expected to better understand the evolution of the patients' risk of non-persistence over time.

Reference 1: Royston P, Parmar MK (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects.

PC4 - T52: Comparison of predictive abilities under different approaches to treat transplantation in survival studies

Authors: Elisabet García[1], Alex Amorós[1], Carme Deulofeu[1], Ferran Aguilar[1], Marco Pavesi[1].
[1] *European Foundation for the Study of Chronic Liver Failure, Spain.*

Background: Survival studies consider transplantation (T) in different ways, including: 1. Exclusion; 2. Censoring at T time; 3. Using T as alive at analysis time-point; 4. Considering T as competing event; 5. Using T as death. OBJECTIVE: To compare the abilities of these approaches in predicting mortality. METHODS: The data came from a prospective observational study in 1,343 cirrhotic patients with decompensated cirrhosis (CANONIC study1), who have a 1-year transplant and mortality rates of 16% and 34%, respectively. Survival and predictive abilities for Model for End-stage Liver Disease (MELD) score were studied at 1-year after inclusion. Additionally, other time-points and scores for specific liver disease subpopulations were assessed. Harrel's C-index (HC) was obtained in all approaches. For #4, the database was appropriately transformed. The Area Under Receiver Operating Characteristic (AUROC) could be estimated only in #1, #3 and #5. The Competing Risk (CR,#4) was assumed as the most methodologically appropriated approach. RESULTS: One-year HC was 0.666 (0.640-0.692) under #4, 0.665 (0.639-0.691) under #3 and 0.700 (0.675-0.725) under #2. AUROC under #3 was 0.660 (0.628-0.691). HC and AUROC under #1 and #5 were significantly deviated from the results of the ideal approach (#4). Similar findings were observed for all time-points, scores and subpopulations. CONCLUSIONS: Approaches #3 and #4 showed identical predictive abilities with HC. Approach #3 with AUROC leads to similar estimates, although it is not methodologically adequated. Approach #2, which is commonly used, overestimates HC. T exclusion (#1) and composite event (#5) approaches lead to skewed results.

Reference 1: 1 Moreau R, et al. Gastroenterology 2013;144:1426–1437.

PC4 - T55: Modified weighted Kaplan-Meier cancer survival estimate for data with higher loss to follow-up

Authors: Jagathnath Krishna K. M.[1].
[1]*Regional Cancer Centre, Thiruvananthapuram.*

Background: Kaplan-Meier (K-M) method is used to estimate survival from life-time data and it assumes loss to follow-up (LFU) as random. But if the study outcome is influenced by the factors for loss, LFU can be non-random and it leads to over-estimated survival. Present study proposed a modified weighted Kaplan-Meier (MWKM) Method and validated it using realistic data.

Materials and Methods: Breast cancer (BC) patients treated in the Regional Cancer Centre, Trivandrum during 2009-2010 were used. The proposed MWKM method was based on weightage to the LFU and duration of the study. To validate the MWKM method, we estimated BC survival using standard K-M and the MWKM methods.

Results: Three-year BC survival by stage (>80% follow-up) showed almost similar survival estimates for standard K-M and the MWKM methods [3-year survival, K-M method- stage I: 97.7%, II: 94.8% & III: 84.7%; MWKM method- stage I: 96.6%, II: 93.5% & III: 81.9%]. However 4-year survival with <80% follow-up, showed slightly over estimated survival for all stages using standard K-M method [I: 96.2%, II: 92.7% & III: 80.0%] than MWKM method [I: 93.2%, II: 89.7% & III: 77.2%]. The validated results showed that MWKM method provided consistent estimates [4-year MWKM estimates by stage I: 94.3%, II: 89.8% & III: 76.8%].

Conclusion: When factors for follow-up loss influence the study outcome and the proportion of LFU is high, the proposed modified weighted K-M method provides more unbiased and consistent estimates.

Reference 1: Kaplan EL and Meier P. Nonparametric estimation from incomplete observations. Journal of American Statistical Association. 1958; 53 (282): 457–481

Invited Session

IS5: Complex Survival Data

Tuesday 11th July - 11.00-12.30 h. - Room: Auditorio
Chair: Ronald Geskus
Organised by Ronald Geskus, *Academic Medical Center, Amsterdam, the Netherlands*

IS5-1: Estimation of the Transition Probabilities in Non-Markov Models: New Contributions and Software Development

Luis Meira-Machado, *University of Minho, Portugal*

One major goal in clinical applications of multi-state models is the estimation of transition probabilities. The usual nonparametric estimator of the transition matrix for nonhomogeneous Markov processes is the Aalen-Johansen estimator. Several non-Markov estimators have been proposed in the recent literature and their superiority with respect to the Aalen-Johansen estimator has been proved in situations in which the Markov condition is strongly violated. However, most of the proposed estimators have the drawback of requiring that the support of the censoring distribution contains the support of the lifetime distribution, which is not often the case. Recently, landmark-based estimators were proposed to estimate these quantities which are consistent regardless the Markov condition and the referred assumption on the censoring support. An in-depth review of the subject will be given with a brief discussion on how to estimate these quantities conditionally on current or past covariate measures. The available software for implementing the proposed methods will be given using real medical data.

Reference 1: Aalen, O. and Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. Scandinavian Journal of Statistics 5, 141–150.

Reference 2: de Uña-Álvarez, J. and Meira-Machado, L. (2015). Nonparametric estimation of transition probabilities in the non-Markov illness-death model: a comparative study. Biometrics, 71(2), 364–375.

IS5-2: The Use of Tumor Dynamics and New Lesions to Predict Survival with Multivariate Joint Frailty Models

Virginie Rondeau, *INSERM, France*

Abstract: The RECIST criteria are used as standard guidelines for the clinical evaluation of cancer treatments. The assessment is based on the anatomical tumor burden: change in the sum of the longest diameters (SLD) of target lesions and evolution of non-target lesions (NTL): appearance of new lesions (NL) and unequivocal progression NTL determined before treatment. Despite indisputable advantages of this standard tool, RECIST are subject to some limitations such as categorization of continuous tumor size or negligence of its longitudinal trajectory. In particular, it is of interest to capture the tendency of short-term decrease and long-term re-growth of tumor size that is often present under an advanced cancer treatment and model it simultaneously with time to event for NTL

and OS. We propose a new joint model for longitudinal biomarker (SLD) and two types of survival data. In the model, the tumor size trajectory is described using an ordinary differential equation (ODE) that accounts for the natural growth and treatment induced decline. We apply the model to a real dataset of a phase III clinical trial for metastatic colorectal cancer. We compare this model with models that consider parametric functions for the SLD trajectory.

Reference 1: A. Krol, et al.. Joint model for left-censored longitudinal data, recurrent events and terminal event: Predictive abilities of tumor burden for cancer evolution with application to the cd 2000-05 trial. Biometrics, 2016.

Reference 2: A. Krol, et al. Tutorial in joint models and prediction: a statistical software for correlated longitudinal outcomes, recurrent events and a terminal outcome. JSS, 2016.

IS5-3: Robustness Properties of the (Landmark) Aalen-Johansen Estimator

Hein Putter, *Leiden University Medical Center, The Netherlands*

Multi-state models are useful extensions of classical survival and competing risks models to multiple types of events occurring sequentially in time. The Aalen-Johansen estimator plays a central role in multi-state models. Where transition hazards form the building blocks in multi-state models, the Aalen-Johansen formula could be seen as the cement that connects the transition hazards with the quantities that are of final interest, the state occupation and transition probabilities.

The relation between transition hazards and transition probabilities described by Aalen-Johansen is valid only under the Markov assumption. This assumption is often taken for granted, but many situations exist where it does not hold. A remarkable result from Datta & Satten (2001) states that in the absence of covariates the Aalen- Johansen estimator of the state occupation probabilities is consistent, even if the Markov assumption is violated.

In this presentation, we show how the result of Datta & Satten can be exploited to obtain robust estimators of transition probabilities based on landmarking. We will also study alternative representations of the Aalen-Johansen estimator and show how these alternative representations can be used to derive robust estimators of state occupation probabilities with covariates.

Reference 1: Aalen, O. O. & Johansen, S. (1978), `An empirical transition matrix for nonhomogeneous Markov chains based on censored observations', Scandinavian Journal of Statistics 5, 141-150.

Reference 2: Datta, S. & Satten, G. A. (2001), `Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson-Aalen estimators of integrated transition hazards for non-Markov models', Statistics and Probability Letters 55, 403-411.

Oral Contributed Sessions

OC22: Bayesian methods in clinical research 4

Tuesday 11th July - 11.00-12.30 h. - Room: Sala Terra 4
Chair: Lisa Hampson

OC22-1: Weighted pairwise estimation of multivariate longitudinal outcomes in a Bayesian joint modelling framework

Authors: Katya Mauff[1].
[1]*Erasmus Medical Centre, Rotterdam, The Netherlands.*

Using a random effects approach, mixed effect models for longitudinal outcomes may be easily extended to account for multivariate outcomes that must be jointly modelled. However, as the number of k (potentially non-guassian) outcomes increases, so too does the dimension of the joint covariance matrix of the random effects, which quickly becomes computationally prohibitive. Verbeke et al. (1) proposes a pairwise modelling approach, whereby all possible pairwise models are fitted, and inference is based on pseudo-likelihood theory. We now demonstrate the use of this approach within the Bayesian setting. Having obtained MCMC iterations for each parameter in each model, we combine those obtained for the same parameters, making use of Self-Normalizing Importance Sampling theory to weight the iterations prior to calculation of the mean and 95% credibility interval. In this way we aim to more closely approximate the estimated parameter values that would be obtained via estimation of the full multivariate model.

Reference 1: Fieuws S., Verbeke G. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. Biometrics 2006; 62: 424-31.

OC22-2: Dynamic Bayesian variable selection for discrete time-to-event data in the presence of competing events

Authors: Rachel Heyard[1], Jean-François Timsit[2], Leonhard Held[1].
[1] *University Of Zurich, Switzerland,* [2] *Université Paris Diderot, France.*

Developing good prediction models by selecting the most relevant variables is highly im-portant in clinical research. Techniques to perform objective Bayesian variable selection inthe linear model are well developed and have been extended to the generalized linear modelsetting [1] as well as to the Cox proportional hazards model [2]. These Bayesian approachesautomatically induce optimal shrinkage of the regression coefficients. We propose an ex-tension of this methodology to discrete time-to-event data with competing events. In ourspecific application to a large French database, the goal is to predict the daily risk of acquir-ing a ventilator associated pneumonia (VAP) attributed to P. aeruginosa (PA) in intensivecare units (ICUs). The competing events for a PA VAP are extubation and death. Baselinevariables are potentially important to predict the outcome at the start of ventilation, but maylose some of their predictive power after a certain time. Therefore, we use a landmark ap-proach for dynamic Bayesian variable selection where the set of relevant predictors dependson the time already spent at risk (Section 3.3 in [2]).

Reference 1: Leonhard Held, Daniel Sabanés Bové, and Isaac Gravestock. Approximate Bayesianmodel selection with the deviance statistic. Statistical Science, 30(2):242–257, 2015.

Reference 2: Leonhard Held, Isaac Gravestock, and Daniel Sabanés Bové. Objective Bayesian modelselection for Cox regression. Statistics in Medicine, 35(29):5376–5390, 2016.

OC22-3: Bayesian dose-finding with adaptive cohort sizes for combination drugs

Authors: Masakatsu Onoda[1], Seiichi Yasui[1].
[1] *Tokyo University Of Science, Japan.*

In Phase I cancer researches, the maximum tolerated dose (MTD) is estimated by appropriately changing dose levels while accumulating safety information. Recently, cancer clinical trials in which patients are treated with combination of two drugs have grown. We propose a new Bayesian dose-finding method with combination drugs of a discrete-dose agent (key drug) and a continuous-dose agent (support drug), in which cohort sizes are adaptively determined to reduce the number of cohorts in cancer phase I clinical trials.

Either dose escalation or de-escalation, and cohort sizes are based on Bayesian posterior estimates of the joint toxicity probabilities of combined doses. The clinical trial proceeds using univariate escalation with overdose control (EWOC). At each stage of the trial, we seek a dose of one agent using the current posterior distribution of the agent's MTD given the current dose of the other agent. This method searches for a target doses of key and support drug that achieve the target toxicity rate as close as possible. The dose of support drug can be controlled more strongly than dose of key drug. Hence, conditional EWOC is used for the method of dose determination of support drug. Thus, the number of cohorts is adjusted while controlling over dose of support drug. Also, in order to improve the accuracy of MTD estimation, toxicity data is accumulated during the trial, so the proportion of severely punishing excessive doses over underdoing will be relaxed as the trial progresses.

Simulation studies show that the proposed method reduced the number of patients treated with low doses (i.e. low therapeutic effect) while controlling the dose of support drug without violating the accuracy of MTD estimation.

OC22-4: Detecting microRNA-mRNA-interactions using a Bayesian hierarchical model

Authors: Eva-Maria Huessler[1], Martin Schäfer[1], Holger Schwender[1], Pablo Landgraf[2].
[1] *Heinrich Heine University Düsseldorf, Germany,* [2] *University Hospital Cologne, Germany.*

MicroRNAs are small non-coding RNAs that play an important role in gene regulation as they bind to target mRNA to initiate translational repression and mRNA destabilization. Such targeted

mRNA regions can be detected using PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation). The transition of T-to-C in the sequenced cDNA induced by this biotechnology helps to distinguish potential binding sites from noise. Nevertheless, detected T-to-C nucleotide exchanges can also be due to other reasons such as sequencing errors or single nucleotide polymorphisms (SNPs). Only few statistical methods have been developed to detect potential binding sites in PAR-CLIP data, most of these do not account for errors and SNPs.

We have developed a fully Bayesian hierarchical model that accounts also for other sources of nucleotide exchange than the PAR-CLIP induced ones. Moreover, this model is - to the best of our knowledge - the first method for the analysis of PAR-CLIP data that allows to incorporate additional information relevant for the biology of microRNA binding sites such as the mRNA region.

In order to evaluate and compare the performance of different versions of this hierarchical model, different types of prior information and different ways of accounting for noise and SNPs, we have set up an extensive simulation study in which the data sets were simulated based on real PAR-CLIP data.

In our talk we will present this hierarchical Bayesian procedure as well as the results of the simulation study and of the application of this method to a publicly available data set. We will also compare the method to other procedures developed for the analysis of PAR-CLIP data.

OC22-5: Distributional assumptions in random effects bivariate and network meta-analysis of surrogate endpoints

Authors: Sylwia Bujkiewicz[1], Keith R. Abrams[1], John R. Thompson[1].
[1] *University of Leicester, United Kingdom.*

We discuss how bivariate meta-analysis and network meta-analysis in a Bayesian framework can be used to model surrogate endpoints and how the choice of distributional assumptions, to capture uncertainty around all relevant parameters, may impact on the predictions of a treatment effect on a final clinical outcome from a treatment effect on a surrogate endpoint. Daniels and Hughes [StatMed1997;16:1965] proposed a Bayesian meta-analytic model for surrogate endpoints which allows for the full inclusion of uncertainty around all relevant parameters. A similar approach in the form of a bivariate random-effects meta-analysis was then proposed by Bujkiewicz et al. [Stat Meth Med Res; e-pub 13Aug2015] to allow borrowing of strength across studies when making predictions. A limitation of this approach was the assumption of a normal distribution for the random effects for treatment differences measured on surrogate endpoint, which was relaxed by the use of t-distribution to capture outlying observations. In this research we extend this work to allow a mixture normal distribution which may be a better fit when the distribution of true effects is clearly not normal, for example, bimodal. This work is further extended by use of network meta-analysis which assumes individual distributions for each treatment contrast whilst still allowing borrowing of strength across treatments by taking into account the network structure of the data. This can lead to more precise predictions of a treatment effect on a final clinical outcome from the effect on a surrogate endpoint. Methodology is demonstrated on illustrative example in multiple sclerosis.

OC23: Statistical methods for precision medicine 1

Tuesday 11th July - 11.00-12.30 h. - Room: Sala Mar 4
Chair: Kelley Kidwell

OC23-1: Monitoring futility and efficacy in phase II trials: criteria for futility and efficacy in interim and final analysis

Authors: Annette Kopp-Schneider[1], Manuel Wiesenfarth[1], Ruth Witt[2].
[1] *Dkfz Heidelberg,* [2] *Nct Heidelberg, Germany.*

A phase II trial is typically a small-scale study to determine whether an experimental treatment should continue further clinical evaluation. In this setting, interim analyses are commonly performed to allow for early stopping for futility and/or efficacy. The use of Bayesian posterior probability as decision rule for early stopping and for final analysis has been suggested in the last decade, especially in the context of biomarker-targeted therapies with small numbers of patients. In comparison to traditional hypothesis testing-based approaches the advantage is the flexibility with respect to number and timing of interim analyses as well as to the final number of patients included in the study.

The INFORM2 phase I/II trial series addresses individualized therapy for relapsed malignancies in childhood using next-generation diagnostics. It is currently in the planning stage and the first trials are expected to start in 2017. The trials are one-arm trials with a dichotomous endpoint and will include interim futility and/or efficacy evaluations. Identical evaluation criteria for interim and final analyses will be used. We discuss performance criteria such as the probability of correct decision and expected number of patients enrolled. The influence of the choice of the Bayesian model and the prior distribution on the performance of the design will be discussed. The final choice of stopping criterion will depend on the principal investigator's preference on the basis of the design's operating characteristics.

OC23-2: Estimating the quality of optimal treatment regimes

Authors: Aniek Sies[1], Iven Van Mechelen[1].
[1] *KU Leuven, Belgium.*

Often, multiple treatment alternatives are available for a certain disease. When this is the case, an obvious question is which of these alternatives is most effective. As it is well-known that the best treatment alternative commonly differs between patients, one may address this question by searching for an optimal treatment regime that assigns each individual to a treatment alternative based on the individual's baseline characteristics. When an optimal treatment regime has been estimated, of obvious interest is its quality in terms of the expected potential outcome if it would be used to assign all patients in the population to treatment. Estimating this quantity is a major challenge, as, on the one hand, often data sets are not large enough to split them into an independent training and test set; on the other hand, using the whole data set to first estimate a treatment regime and to

subsequently estimate its expected outcome will lead to positively biased estimates of the expected outcome. A common way out is to rely on resampling methods. Up to now however, the quality of the estimates of the expected outcome returned by resampling methods is not yet fully clear. To address this, we performed a simulation study in which we compared several variants of the bootstrap (viz., ordinary, out-of-bag, .632 and .632+ bootstrap) and cross validation (viz., 5-fold and 10-fold cross validation) on a wide range of data sets. This study showed that most methods have satisfactory performance, and that, in general, the .632+ bootstrap performs best. We will discuss these results and compare them to related findings regarding the estimation of the prediction error of prediction rules resulting from statistical learning approaches.

OC23-3: Analyzing multiple endpoints simultaneously in crossover trials using generalized pairwise comparisons and patient-defined thresholds

Authors: Joris Giai[1], Delphine Maucort-Boulch[1], Julien Péron[1], Jean-Luc Cracowski[2], Matthieu Roustit[2], Pascal Roy[1], Marc Buyse[3].
[1] *Université Claude-Bernard, Lyon, France,* [2] *Université Grenoble Alpes, Grenoble, France,* [3] *International Drug Development Institute (IDDI), San Francisco, Ca 94109, USA.*

Background: We sought to apply Generalized Pairwise Comparisons (GPC) as a proof of concept for assessing the individual benefit–risk balance in crossover trials with patient-defined thresholds of interest using data from a completed N-of-one trial evaluating the benefit of Sildenafil in Raynaud's Phenomenon (PROFIL study, NCT02050360).Methods:GPC were performed within each patient by using every possible combination of one prioritized daily outcome when treated versus under placebo. Only the low Sildenafil dose was considered. First to third prioritized outcomes were daily crises number ranging from zero to ten, Raynaud's Condition Score ranging from zero to ten, and daily sum of crises duration ranging from zero to 755 minutes with five minutes steps respectively. Analyses were performed by considering every possible threshold combination, thus allowing treatment clinical utility assessment on every patient self-defined relevant set of thresholds. P-values were computed by bootstrap.Results:A total of 847 sets of thresholds were tested for each of the 38 patients of this trial, by calculating the net chance of a better outcome as defined in the original GPC publication and their associated p-values testing the difference from zero. For example patient number 11 had all its 847 sets of thresholds statistically favoring Sildenafil over placebo.Conclusion:GPC provide a flexible way of incorporating several prioritized efficacy and safety outcomes with thresholds defined by the patient itself in crossover trials such as N-of-one trials. A clinician could use the results for a given patient to assess the clinical utility of the experimental drug after asking the patient for his set of thresholds of interest.

OC23-4: Deciding on the best biomarker-guided trial deisgn: a case study

Authors: Miranta Antoniou[1], Andrea L. Jorgensen[1], Ruwanthi Kolamunnage-Dona[1].
[1] *MRC North West Hub for Trials Methodology Research & Department of Biostatistics, Institute of Translational Medicine, University of Liverpool, Liverpool L69 3Gl, UK.*

Biomarker-guided clinical trial designs have drawn considerable attention in the era of stratified medicine, which focuses on the identification of subgroups of patients with specific characteristics who are most likely to benefit from a particular treatment. However, testing the effectiveness of a biomarker-guided approach to treatment yields challenges both in terms of trial design, analysis and practical applications. A number of biomarker-guided trial designs have been proposed in the past decade, including both biomarker-guided adaptive and non-adaptive trial designs which are thoroughly reviewed in our published papers Antoniou et al. (2016) and Antoniou et al. (2017) providing the research community with clarity in their definition, methodology and terminology.

We have re-considered the most appropriate design for a clinical trial we previously proposed for testing whether a genotype-guided treatment strategy results in reduced rate of relapse in alcohol-dependent patients. The design originally proposed was not plausible due to low prevalence of the genetic marker. Here, we explore the various biomarker-guided non-adaptive trial designs and apply a strategy to choose the most appropriate. We also apply statistical techniques to calculate the corresponding sample size, considering both binary and time-to-event outcomes. In addition, an adaptive version of the proposed design is explored, to address the uncertainty surrounding the estimated effect size.

OC23-5: Subgroups analysis of scoring systems – identifying hard to classify HCC patients

Authors: Irene Schmidtmann[1], Arndt Weinmann[1], Roman Kloeckner[1], Daniela Zöller[1], Harald Binder[1].
[1] *University Medical Center, Johannes Gutenberg Universität Mainz, Germany.*

Introduction: There is an ongoing debate which staging system should be used to classify hepatocellular carcinoma (HCC) patients and to determine initial treatment. Several staging systems are available for stratifying in patients with HCC; recent research indicates that performance may depend on etiology [1].

Methods: We compared the predictive ability of the staging systems BCLC, HKLC and MESH by applying them to data of 1109 patients from the Mainz HCC registry. We assessed predictive ability using the integrated Brier score (IBS) and variants of the c-index such as Harrell's C. We used pseudo value residuals [2] as a novel approach to determine subgroups of patients in which predictive ability is poor.

Results: In our data, survival time differed significantly between stages as defined by BCLC (median survival 55.0, 22.2, 10.7, 3.8 months respectively; p<0.001), HKLC (median survival 63.3, 35.5, 11.5, 5.1, 3.3 months respectively; $p < 0.001$) and MESH (103.0, 38.4, 25.2, 9.8, 4.8, 3.4, 2.6 months respectively; $p < 0.001$). The cases identified as influential by plotting pseudo residuals are shown.

Discussion: Both IBS and c-index indicate the same trends in discriminative ability of the three scores. Pseudo residuals allow identifying influential observations and hard to predict subgroups of patients, thus facilitating the choice of appropriate staging system and potential improvement.

Conclusion: Clinical registries provide opportunity to perform independent evaluation of staging systems. Different performance measures capture different aspects of predictive ability. Pseudo residuals provide a valuable tool to investigate and improve staging systems.

Reference 1: Liu P et al, Medicine 2015

Reference 2: Perme MP, Andersen PK, StatMed 2008

OC24: Design and analysis of clinical trials 3

Tuesday 11th July - 11.00-12.30 h. - Room: Sala Mar 2
Chair: John Whittaker

OC24-1: Incorporating data from patients with missing biomarker status into biomarker-based trial designs

Authors: Cornelia Ursula Kunz[1].
[1]*Boehringer Ingelheim Pharma GMBH & Co. Kg, Germany.*

The focus of modern medicine has shifted from broad spectrum treatments to targeted therapeutics leading to new challenges for the design and analysis of clinical trials. Developments in genomics are providing a biological basis for the heterogeneity of clinical course and response to treatment that have long been apparent to clinicians. The impact of genomic variability is often assessed in a biomarker-based design. Currently, such studies are restricted in the sense that only patients for whom the biomarker can be measured can be enrolled. Patients unable to provide adequate biosamples are thus excluded from the trial and in some cases the proportion of such patients can be substantial. A recent study in non-small lung cancer showed that the epidermal growth factor receptor (EGFR) could not be measured in about 30% of the patients [1]. Hence, results from trials including only patients whose biomarker can be assessed only provide evidence on how to treat these patients.

We show how current biomarker-based designs can be extended to also include patients with missing biomarker status. Analytical solutions for the test statistic of the interaction effect between the biomarker status and the received treatment are derived. We also show under what circumstances the data from these patients can be used to obtain a more powerful test of the interaction effect.

Reference 1: Jackman, Miller, Cioffredi, et al. (2009): Impact of Epidermal Growth Factor Receptor and KRAS Mutations on Clinical Outcomes in Previously Untreated Non-Small Cell Lung Cancer Patients: Results of an Online Tumor Registry of Clinical Trials. Clinical Cancer Research, 15:5267-5273.

OC24-2: Cancer phase I trial design using drug combinations when a fraction of dose limiting toxicities is attributable to one or more agents

Authors: José Jiménez[1], Mourad Tighiouart[2], Mauro Gasparini[1].
[1] *Politecnico Di Torino, Italy,* [2] *Cedars-Sinai Medical Center, USA.*

Drug combination trials are increasingly common nowadays in clinical research. However, very few methods have been developed to consider toxicity attributions in the dose escalation process. We are motivated by a trial in which the clinician is able to identify certain toxicities that can be attributed to one of the agents. We present a Bayesian adaptive design in which toxicity attributions are modeled via Copula regression and the maximum tolerated dose (MTD) curve is estimated as a function of model parameters. The dose escalation algorithm uses cohorts of two patients, following the continual reassessment method (CRM) scheme, where at each stage of the trial, we search for the dose of one agent given the current dose of the other agent. The performance of the design is studied by evaluating its operating characteristics when the underlying model is either correctly specified or misspecified. We show that this method can be extended to accommodate discrete dose combinations.

Reference 1: Yin, G., & Yuan, Y. (2009). A latent contingency table approach to dose finding for combinations of two agents. Biometrics, 65(3), 866-875.

Reference 2: Tighiouart, M., Li, Q., & Rogatko, A. (2017). A Bayesian adaptive design for estimating the maximum tolerated dose curve using drug combinations in cancer phase I clinical trials. Statistics in medicine, 36(2), 280-290.

OC24-3: Incorporating historical data into the design and analysis of a clinical trial with normally distributed outcome data

Authors: Maxine Bennett[1], Simon R. White[1], Adrian P. Mander[1].
[1] *MRC Biostatistics Unit, United Kingdom.*

A standard two-arm randomised controlled trial usually compares an intervention to a control treatment and the analysis is based only on patients within the current trial. Historical data are often used when designing trials, as a basis for sample size calculations and have only recently been considered for formal use in the analysis. When historical and current control data agree trials can be more efficient and have increased precision in treatment effect estimates. However, when the historical and current data are inconsistent, there is a potential for bias, inflated type I error and reduced power.

For normally distributed outcome data, a difference in means or variances between the historical and current controls could indicate that the data represent two different populations and that the historical data should be down-weighted in the current trial analysis.

We present a Bayesian design where the historical data are treated as additional information to increase the power of the current trial while maintaining the current study sample size. We introduce

two novel weights to assess agreement based on the joint posterior distribution of the current and historical control means and variances: A probability weight based on tail area probabilities; a weight based on mean and variance equivalence.

We compare the operating characteristics of a design using the proposed weights to methods proposed in the literature, power priors and robust meta-analytic predictive priors. An example illustrates that all methods discount the historical data when there is disagreement with the current controls but the methods differ in the rate at which the historical data are discounted and in their flexibility and implementation.

OC24-4: Multivariate individual bioequivalence

Authors: Rasheed Abdur[1], Juniad Saghir Siddiqi[1], Tasneem Ahmad[2].
[1] *Department of Statistics, University of Karachi, Karachi, Pakistan,* [2] *Pharmaceutics, University Of Karachi, Karachi, Pakistan.*

Bioequivalence (BE) studies are conducted to exhibit the similarity of two drug products with each other in terms of their efficacy and safety. The pharmacokinetic (PK) parameters generally used in BE studies are AUC, Cmax and Tmax. Extensive literature is available for the univariate average, population and individual (three types) of bioequivalences where the analysis is carried out using one PK parameter at a time. Contrary to this, only scarce material is found in multivariate situations for average and population bioequivalences where more than one PK parameters are considered simultaneously. We developed a criterion for Multivariate individual bioequivalence. Unlike to the previously developed criteria for multivariate bioequivalences, this criterion combines not only the comparison of means and variances but also the within-subject correlations into a single aggregate criterion. In this research, we identify the threshold value (theta) for multivariate individual bioequivalence (MIBE) criterion since there is no guidance related to theta which shows that whether it should be greater or not while assessing MIBE. Moreover, various combinations of within-subject correlations among test and reference PK parameters has also been used in calculation of theta. We also performed a simulation study to evaluate the distribution of developed multivariate criterion under various combinations of sample sizes, correlations and differences in means and variances.

Reference 1: Chervoneva, I, Hyslop, T & Hauck, W W (2007). A multivariate test for population bioequivalence. Stat Med, 26(6),1208-1223.

Reference 2: Hauschke, D, Steinijans, V & Pigeot, I. (2007). Bioequivalenec Studies in Drug Development Methods and Application: Jhon Willey & Sons.

OC24-5: How to address the heterogeneity in the design of phase II clinical trials in geriatric oncology?

Authors: Bastien Cabarrou[1], Patrick Sfumato[2], Loïc Mourey[1], Eve Leconte[3], Laurent Balardy[4], Jean-Pierre Delord[1], Jean-Marie Boher[2], Thomas Filleron[1].
[1] *Institut Claudius Regaud - Iuct-O, France,* [2] *Institut Paoli Calmettes, France,* [3] *TSE-R Université Toulouse 1, France,* [4] *Chu Toulouse, France.*

Cancer in the elderly is a major public health problem. However, this heterogeneous population is often excluded from clinical trials and the lack of prospective data makes difficult the management of these patients. As classical single-arm phase II designs do not take into account the heterogeneity, elderly specific phase II clinical trials are very uncommon and generally conducted in specific groups defined by geriatric criteria which increases the number of patients to be included and thus reduces the feasibility. Several sequential designs have been proposed in the literature to address the patients' heterogeneity in phase II trials for target therapy (Jones, 2007; Tournoux, 2011; Parashar, 2016). The main objective of this work is to present these adaptive designs and to compare their operational characteristics with classical designs in order to make recommendations on the methodology to be used in the design of phase II clinical trials in geriatric oncology.

Simulations results shows that the use of these adaptive designs could reduce the number of patients to be included while maintaining statistical power and type I error with a high heterogeneity detection rates at interim analysis (≥70%). Recruitment duration, that is a major cause of early stopping in elderly trials, can also be shortened.

To improve the efficiency of clinical research in geriatric oncology, it is essential to conduct elderly specific phase II clinical trials using appropriate methodologies by taking into account the heterogeneity of this population. The use of sequential adaptive designs allows the possibility to select a subpopulation that could benefit (or not) from the experimental treatment at the end of the first or the second stage.

OC25: Survival analysis 2

Tuesday 11th July - 11.00-12.30 h. - Room: Sala Terra 2
Chair: Andrew Titman

OC25-1: Identification issues and efficient estimation strategies in cure survival models under insufficiently long follow-up

Authors: Philippe Lambert[1, 2].
[1]*Université de Liège, Belgium,* [2]*Université Catholique de Louvain, Belgium.*

The promotion time cure rate model is a special case of cure survival models assumming that an unidentified proportion of subjects will never experience the event of interest whatever the study duration. The challenges raised by an insufficiently long follow-up will be discussed from a theoretical and practical point of view. In particular, when the same covariates enter the probability

to be cured and the time-to-event model for susceptible subjects, identification issues arise for the regression parameters of shared covariates. We investigate how, despite this, plausible values for these parameters can be obtained in a computationally efficient way. Properties of the derived estimators will be investigated by simulation, followed by an illustration on clinical data.

This project started from joint work with Vincent Bremhorst (Bremhorst & Lambert, 2016) and Oswaldo Gressani (Gressani & Lambert, 2016).

Reference 1: Bremhorst V. and Lambert P. (2016). Flexible estimation in cure survival models using Bayesian P-splines. Computational Statistics and Data Analysis, 93, 270-284.

Reference 2: Gressani O. and Lambert P. (2016). Fast Bayesian inference in semi-parametric P-spline cure survival models using Laplace approximations. Discussion paper DP2016/41, UCL, Belgium.

OC25-2: Impact of model misspecification in survival models with frailty terms

Authors: Alessandro Gasparini[1], Keith R. Abrams[1], Michael J. Crowther[1].
[1] *University of Leicester, United Kingdom.*

Survival models incorporating random terms to account for unmeasured heterogeneity are being used more and more in biostatistical research. Specifically, unmeasured covariates whose lack of inclusion in the model would lead to biased, inefficient results are commonly modelled by including a frailty term. Those frailty terms may be subject-specific or cluster/group-specific and are usually assumed to follow either a gamma or log-normal distribution. In the context of parametric frailty models, little is known about the impact of misspecifying the choice of parametric survival model, i.e. how we model the baseline hazard function. Hence, we aim to quantify the impact of misspecifying the baseline hazard function in a wide variety of clinically plausible scenarios. We simulate survival data under various baseline hazard functions, including those with turning points, and furthermore assess the impact of sample size, the effect size of covariates, and the frailty distribution. We estimate the model coefficients via maximum likelihood, using numerical integration to handle intractable terms. Models compared include standard distributions, such as the exponential, Weibull and Gompertz, and spline based approaches including the Royston-Parmar model, and penalized splines on the log hazard scale. Our simulations indicate the importance of assessing model fit with respect to the baseline hazard function, showing bias can occur when it is misspecified. We illustrate our conclusions on an example dataset in colon cancer recurrence.

OC25-3: Regression models in survival analysis with left-truncation using pseudo-observations

Authors: Mia Klinten Grand[1], Arthur Allignol[2], Hein Putter[3], Per Kragh Andersen[1].
[1] *Copenhagen University, Denmark,* [2] *Ulm University, Germany,* [3] *Leiden University Medical Center, The Netherlands.*

Pseudo-observations[1] was introduced as a way to perform regression analysis on an object related to a right-censored time-to-event outcome, such as the survival probability or the restricted mean survival time.

The pseudo-observations are jackknife estimates, which represents the subjects' contribution to the non-parametric estimator of the object of interest. Pseudo-observations are calculated for all subjects in the sample and the regression model parameters are then obtained by solving the corresponding generalised estimating equations using the pseudo-observations as outcome.

Since the introduction of the approach there have been several extensions from the original setting. However, the proper definition and performance of pseudo-observations under left-truncation has not yet been addressed. We explore this in a simulation study, where we also look at the performance of different versions.

Reference 1: Andersen PK, Klein JP, Rosthøj S. Generalised linear models for correlated pseudo-observations, with applications to multi-state models. Biometrika 2003; 90: 15–27.

OC25-4: Replicating individual patient data including covariate distributions to inform extrapolation of clinical trials using external data

Authors: Lynsey Chudleigh[1], Mark Rutherford[1], Keith Abrams[1].
[1] *University of Leicester, United Kingdom.*

Accurate extrapolation of data from clinical trials is an important area of research as these estimates are used to inform policy decisions on new treatments. A number of different approaches have been suggested for extrapolation, and more sophisticated approaches account for the fact that as time progresses, patients' age and thus their hazard of dying from causes other than the disease of interest also increase. Recent approaches in population-based cancer data have used flexible parametric models to extrapolate the relative comparison between a group of cancer patients and a matched cohort from the general population in order to make lifetime extrapolations[1]. However, this approach can be inhibited due to the fact that access to individual patient data is often restricted, and thus matching the study population to the general population group is not possible. Replication of IPD from published Kaplan-Meier curves is a solution to this problem, however approaches to do this do not generate appropriate covariate distributions meaning matching to external data would still not be possible. Here, we propose applying a modified algorithm to that proposed by Guyot et. al.[2] to replicate covariate distributions that are linked to the survival observed in the dataset. The replicated data is then extrapolated using relative survival[1]. We exemplify our approach using an example dataset in early breast cancer.

OC25-5: Adjusting the effect of integrating antiretroviral therapy and tuberculosis treatment on mortality for non-compliance: an instrumental variables analysis

Authors: Nonhlanhla Yende-Zuma[1].
[1] *Centre for the Aids Programme of Research in South Africa (Caprisa).*

Background: Using intent-to-treat (ITT) comparisons, it has been shown that the integration of antiretroviral therapy (ART) and tuberculosis (TB) treatment improves survival. Because the magnitude of the effect of ART initiation during TB treatment on mortality is less well understood due to non-compliance, we used instrumental variables (IV) analyses.

Methods: We studied 642 HIV-TB co-infected patients from Starting Antiretroviral Therapy at Three Points in Tuberculosis (SAPiT) trial in South Africa. Patients were assigned to start ART either early or late during TB treatment or after TB treatment completion. We utilised two-stage predictor substitution (2SPS) and two-stage residuals inclusion (2SRI) methods under additive and proportional hazards regressions with a time-fixed measure of compliance defined as the fraction of time on ART during TB treatment (i.e. months on ART/months on TB therapy). We moreover developed IV methods for additive hazards regressions with a time-varying measure of compliance defined as 1 at time t when the considered patient was on ART at or prior to time t, and 0 otherwise.

Results: ITT results from additive hazards models showed that patients in early and late integrated arms had a reduced hazard of -0.05 (95% confidence interval (CI): -0.09 to -0.01) and -0.06 (95% CI: -0.11 to -0.02) respectively when compared to the sequential arm. Adjustment for noncompliance produced an effect of -0.07 (95% CI: -0.12 to -0.01) of the fraction of time on ART during TB treatment, and a larger effect of -0.29 (95 % CI: -0.54 to -0.03) of the (time-varying) history of ART exposure.

Conclusion: IV analyses enable assessment of the effectiveness of TB and ART integration, corrected for non-compliance, and thereby enable a better public health evaluation of the potential impact of this intervention.

Poster Contributed Session

PC5: Statistical methods in epidemiology

Tuesday 11th July - 11.00-12.30 h. - Room: Hall
Chair: Carmen Cadarso

PC5 - T2: Two indices to measure disease activity in patients with systemic lupus erythematosus: a comparative study

Authors: Ana Matos[1, 2], Carla Henriques[3], Diogo Jesus[4].
[1] *School of Technology and Management, Polytechnic Institute of Viseu,* [2] *Centre for the Study of Education, Technologies and Health, Portugal,* [3] *School of Technology and Management, Polytechnic Institute of Viseu, Centre for The Study of Education, Technologies and Health, Centre for Mathematics of The University of Coimbra, Portugal.* [4] *Lupus Clinic, Rheumatology Department, Centro Hospitalar E Universitário De Coimbra, Coimbra.*

Activity evaluation of the disease Systemic Lupus Erythematosus (SLE), is a crucial matter in order to define the therapeutics to adopt and to monitor its effectiveness. SLEDAI - Systemic Lupus Erythematosus Disease Activity Index- is an index ranging from 0 and 105 points, frequently used in clinical practice for the assessment of SLE disease activity. Even so, the best considered way of evaluation of the disease is through the Physician Global Assessment (PGA), in which an experienced physician quantifies the overall disease activity between 0 and 30 points. Our study involves records of 279 patients, who were followed between January 2014 and December 2016, and aims to compare these two assessment measures. PGA and SLEDAI were studied at baseline and through a summarized measure - the adjusted means during the follow-up, and a relatively high correlation was found between the two scores. Still, the strength of agreement, assessed by weighted kappa, could not be considered good. In fact, the SLEDAI's sensitivity to changes in the activity of the disease has been questioned. To further explore this issue, taking the PGA as the gold standard, ROC curves were considered to measure the ability of SLEDAI to differentiate patients with improvements or worsening. A change of at least one point in SLEDAI could only detect about one half of the improvement cases and also about one half of the cases of worsening. This study supports the limited performance of SLEDAI in detecting a clinically relevant change in the activity of the disease, suggesting the need to optimize this index.

Reference 1: Jos W. R. Twisk. (2003) Applied Longitudinal Data Analysis for Epidemiology, Cambridge University Press

PC5 - T5: Sample size recommendations for binary logistic prediction models: beyond events per variable criteria

Authors: Maarten Van Smeden[1], Gary S. Collins[2], Douglas G. Altman[2], Marinus J. C. Eijkemans[1], Rolf H. H. Groenwold[1], Joris A.H. de Groot[1], Johannnes B. Reitsma[1], Karel G. M. Moons[1].
[1] *UMC Utrecht, Netherlands,* [2] *University of Oxford, UK.*

Binary logistic regression analysis is one of the most frequently used approaches for making biomedical prediction models for predicting future prognostic events and diagnostic outcomes.

Earlier simulation studies have considered the impact only of specific data characteristics, such as the number of events per variable, and analysis choices, such as regularization, on the predictive performance of these binary logistic prediction models. The combined effects of development data characteristics and analysis choices on predictive performance are not well understood. To identify the key development data characteristics and analyses choices that contribute to changes in the out-of-sample predictive performance an extensive full-factorial simulation study was conducted. In this talk I will show the importance of considering both the expected number of events and the total sample size when developing a prediction model. A new and simple-to-apply approach for determining minimal sample size when developing prediction models will be presented.

PC5 - T8: Excess zeros and estimating the true rate of work-related ill-health: the fatigued physician effect

Authors: Matthew Gittins[1], Roseanne McNamee[1], Fiona Holland[1], Lesley Carter[1].
[1] *University of Manchester, UK.*

Background: Accurately estimating incidence of work related ill-health (WRIH) is a goal of many surveillance systems. Zero reporting is recommended in order to remove ambiguity with non-responses, however reporters may increase respond with a false zero case report especially as membership fatigue increases, which in turn may underestimate the true incidence rate and overestimate any trends.

Methods: To look for evidence of excess zeros in reports submitted to three UK surveillance schemes and to determine their impact on the true WRIH incidence rates a set of Zero Inflated Negative Binomial Models with random effects adjustment were fitted to case reports during 1996-2012 to three of The Health and Occupational Reporting network (THOR) schemes. The Zero Inflated portion of the model contained 'membership year' and 'peak holiday period' as predictors, while the Negative Binomial portion contained 'calendar year' and 'first month' as a reporter and 'bank holiday' months. Reported annual WRIH cases were then adjusted for the excess zero rate, along with the standard response rate and the participation rate.

Results: Time since joining the scheme was associated with the odds of excess zero case reports for most schemes. Current estimated incidence rates (95% C.I) of 16 (7-25), 76 (56-96), 17 (11-22), per 100,000 pyrs, were approximately doubled to 30 (21-39), 137 (116-157), 33 (27-39), respectively and weaker trends were observed when excess zero rate adjustment was applied.

Conclusion: If we accept, as assumed here, that excess zeros are in reality non-response by busy reporters, then usual estimates of incidence and trends are likely to be significantly underestimated and potentially overestimated respectively.

PC5 - T11: The benefits of spatial models for analyzing public health data in Bangladesh lattice

Authors: Paritosh Kumar Roy[1], M. Shafiqur Rahman[1].
[1] *University Of Dhaka, Bangladesh.*

Collection of spatial data and their use in public health research become popular recently. Particularly, presenting results through mapping has drawn special attention because its feasibility of sketching the country's whole scenarios regarding the problem dealt with and to identify the vulnerable areas so that policy makers can design effective policies for urgent improvement. Although methods for analyzing and mapping spatial data has already embedded into commonly used statistical softwares, many public health studieshave analyzed the spatial data, where spatial correlation is expected, ignoring the spatial effect and conveyed findings to the policy makers through mapping. We reviewed the benefits of considering spatial models for analyzing spatial data by exploring the effect of different level of spatial dependency on estimation and prediction within Bangladesh lattice using simulation. For spatial referencing we used district level Bangladesh lattice and Geo-spatial locations of clusters in Bangladesh Demographic and Health Survey (BDHS) 2014. With combinations of both binary and continuous responses each with one covariate and spatial references, four simulation studies were described with similar application using real BDHS data. Bayesian methods of estimation and prediction were carried out using Integrated Nested Laplace Approximation approach. Results reveals that the use of spatial models to analyze spatial data within Bangladesh lattice is statistically beneficial even when spatial correlation is week. This leads to provide accurate inference on the association of covariates with response and correct estimation and prediction for small areas, which in turn provides a reliable map for the policy planners.

PC5 - T14: Adjustment for the length of follow-up while assessing the performance of a clinical risk score

Authors: Kristi Läll[1, 2], Aet Saar[3], Maris Alver[1], Krista Fischer[1].
[1] *Estonian Genome Center, University of Tartu, Estonia,* [2] *Institute of Mathematics and Statistics, University of Tartu, Estonia,* [3] *Department Of Cardiology, University Of Tartu, Tartu, Estonia.*

Several phenotypic risk scores (PRS) have been proposed to guide clinical decision making in the primary prevention of common complex diseases. Usually those risk scores are estimates of the 10-year risk for certain outcomes. Before actual implementation of a PRS, it is common to evaluate the performance of the score in the target population, especially if the score is initially derived for a different population. We discuss some methodological problems that arise when the follow-up time in the validation cohort is shorter than 10 years and compare alternative approaches to adjust for the follow-up.

Predictive ability of PRS is usually described through calibration and discrimination. Usually the original PRS is modified to correspond to the length of follow-up in the validation cohort using an exponential survival function. To account for censoring, one of the two options is used: a) the PRS is modified to correspond to the length of follow-up regardless of the event status of the individual

and b) while modifying the PRS, both the follow-up and the event status are accounted for. We use standardized incidence ratio and Harrell's c-statistic to illustrate the impact of modification option.

Simulations are done to demonstrate that scaling PRS under option a) tends to underestimate both the true values of Harrell´s c-statistic and expected number of events. To better communicate the importance of such findings, we also compare performances of cardiovascular PRS such as Pooled Cohort Equations (PCE), QRISK2 and Systematic COronary Risk Estimation (SCORE) in the Estonian Biobank data using both modifications.

PC5 - T17: Estimating the prevalence of treated attention deficit hyperactivity disorder in Ontario, Canada using Bayesian hierarchical modelling

Authors: Yu Luo[1], David A. Stephens[1], David L. Buckeridge[1].
[1] *McGill University, Canada.*

Attention deficit hyperactivity disorder (ADHD) can interfere with functioning atschool or work and is usually treated with medication. The prevalence of ADHD is usuallymeasured by survey, and little is known about the prevalence of ADHD in Canada as only a few,small surveys have been conducted and these are subject to errors arising from small sample sizes. One potential strategy for routine measurement of ADHD is to estimate prevalence from dispensing records for drugs used to treat ADHD. Using prescription data obtained from Health Canada from January 1st to December 31st, 2014, we developed a Bayesian hierarchical model to predict the indication for dispensed medications that may be used to treat ADHD. Prior information was parametrized using an external dataset [Medical Office for the Twenty First Century (MOXXI)] containing demographic information, drug indications, and drug dispensations. The number of prescriptions for ADHD was predicted for each drug class in a Forward Sortation Area (FSA) in Ontario, Canada, and then aggregated to obtain the the prevalence of treated ADHD for each FSA in Ontario. We estimated the prevalence of treated ADHD in Ontario to be 2.30% with 95%credible interval (2.05%, 2.38%). Geographically, prevalence was higher in southeastern part of Ontario, most notably in the Toronto area.

PC5 - T20: Outlier classification performance of risk adjustment methods when profiling multiple providers

Authors: Timo Brakenhoff[1], K.C.B. Roes[1], K.G.M. Moons[1], R.H.H. Groenwold[1].
[1] *UMC Utrecht, The Netherlands.*

Background: When profiling multiple health care providers, adjustment for case-mix is essential to accurately classify the quality of providers. Unfortunately, misclassification of provider performance is not uncommon and can have grave implications. Propensity score (PS) methods have been proposed as viable alternatives to conventional multivariable regression methods to adjust for case-mix differences. Objective: To assess the outlier classification performance of risk adjustment

methods when profiling multiple providers. Methods: In a simulation study based on empirical data, the classification performance of logistic regression (fixed and random effects), PS adjustment, and three PS weighting methods was evaluated when varying parameters such as the number of providers (10-50), the average incidence of the outcome (3%-20%), and the percentage of outliers (8%-40%). Traditional classification accuracy measures were considered, including sensitivity and specificity. Results: Fixed effects logistic regression consistently had the highest sensitivity and negative predictive value, yet a low specificity and positive predictive value. Of the random effects methods, PS adjustment and random effects logistic regression performed equally well or better than all the remaining PS methods for all classification accuracy measures across the studied scenarios. Conclusion: Of the evaluated PS methods, only PS adjustment can be considered a viable alternative to random effects logistic regression when profiling multiple providers in different scenarios.

PC5 - T23: Selection of optimal cut points in a logistic regression model with interactions

Authors: Irantzu Barrio[1, 2], María Durbán[3], Cristóbal Esteban[2, 4], María Xosé Rodríguez-Álvarez[5, 6], José María Quintana[2, 7], Inmaculada Arostegui[1, 2, 5].
[1] *Universidad del País Vasco UPV/EHU,* [2] *Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Spain,* [3] *Universidad Carlos III De Madrid, Spain,* [4] *Servicio de Neumología, Hospital Galdakao-Usansolo,* [5] *BCAM - Basque Center for Applied Mathematics,* [6] *Ikerbasque, Basque Foundation for Science, Spain,* [7] *Unidad De Epidemiología Clínica, Hospital De Galdakao-Usansolo.*

In the development of clinical prediction models, continuous parameters are often categorized. For a binary response variable and in the context of a logistic regression model a methodology has been proposed to select the optimal cut points based on the maximal AUC (Barrio et al, 2015).

Let Y be a dichotomous response variable and X a continuous covariate which we want to categorize in the presence of a dichotomous predictor Z. The methodology proposed by Barrio et al, selects the optimal cut points for X considering them to be equal for all the categories of Z. In this work we propose an extension of that methodology allowing selecting different optimal cut points for each of the categories of Z. We evaluated the empirical performance of the proposed methodology by means of a simulation study in which we considered different relationships between X and the logit of the expected Y for each of the categories of Z, with and without an interaction between X and Z. Simulation results suggest that the methodology performed successfully.

We applied this methodology to a cohort study of patients with an exacerbated COPD. From a clinical point of view, the relationship between the predictor variable PCO2 and the response variable (admission to ICU or IRCU) could be different depending on whether the pH indicated acidosis or not. Hence, we looked for the optimal cut points of the PCO2 in both categories of the pH, which turned out to be 48 for a normal pH and 65 for altered pH, which make sense from a clinical point of view.

Reference 1: Barrio I, Arostegui I, Rodríguez-Álvarez MX, Quintana JM. "A new approach to categorising continuous variables in prediction models: Proposal and validation". Statistical methods in medical research. in-press.

PC5 - T26: Development and population-specific calibration of a mortality risk prediction model in 123,697 middle-aged and older Australian men

Authors: Grace Joshy[1], Emily Banks[1,2], Anthony Lowe[3,4], Rory Wolfe[5], Leonie Tickle[6], Bruce Armstrong[7], Mark Clements[8].
[1]*Australian National University, Australia,* [2]*The Sax Institute, Australia,* [3]*Prostate Cancer Foundation Of Australia, Australia,* [4]*Griffith University, Australia,* [5]*Monash University, Australia,* [6]*Macquarie University, Australia,* [7]*University of Sydney, Australia,* [8]*Karolinska Institutet, Sweden.*

Evidence-based guidelines to target preventive services or treatment often require knowledge of a person's predicted risk of death. However, robust prediction models for middle-aged and older men, are lacking. Further, the methods for calibrating predictions and transferring results from cohort studies to populations are under-developed. We aimed to develop methods and validate a prediction score for 5-year adult male mortality. Using the "45 and Up" cohort study questionnaire data was linked to the State Death Register and associations of all-cause mortality with 40 health measures were assessed. Multiple imputations by chained equations were used to accommodate individuals with missing data on some of these variables. Prediction models were validated internally using 10-fold cross-validation and survival predictions were calibrated to the Australian population using Australian Health Survey (AHS) data. For re-calibration, we compared (i) the marginal predicted risk integrated across population subgroups defined by predictor variables with (ii) the observed age-specific all-cause mortality. Of 123,697 men aged ≥45 years at baseline, 12,160 died during a median follow-up of 5.9 years. Following age-adjustment, self-reported health was the strongest predictor of all-cause mortality (C-index=0·827, 95%CI 0.824-0.831). Three prediction models for all-cause mortality were validated, with predictors: (i) age group and self-rated health; (ii) variables common to the 45 and Up Study and the AHS; and (iii) all variables selected using stepwise regression. Final predictions calibrated well with observed all-cause mortality rates. Calibrated estimates with life tables could be used to predict mortality risks.

PC5 - T29: Use of a bias-adjusted three-step approaches to assess relationships in dietary pattern analysis

Authors: Anna Schritz[1], Nicolas Sauvageot[1], Sonia Leite[2], Ala'a Alkerwi[1], Saverio Stranges[3], Faiez Zannad[4], Sylvie Streel[5], Axelle Hoge[5], Anne- Francoise Donneau[5], Michèle Guillaume[5].
[1]*Luxembourg Institute of Health, Luxembourg,* [2]*Ministry of Health, Luxembourg,* [3]*Schulich School of Medicine & Dentistry, Canada,* [4]*Centre Hospitalier Universitaire, France,* [5]*University of Liège, Belgium.*

Mixture model analysis can be used to define groups of individuals sharing similar dietary patterns by classifying each individual to a group with a given probability. Further, when analysing relationships between dietary patterns and external variables, researchers often recommend the standard 3-step approach in which individuals' dietary patterns are defined by their most probable group:

- Step 1: Estimation of mixture model
- Step 2: Assignment of individuals to their most probable groups
- Step 3: Assessment of relationships between groups assignment and external variables

The assignments made in step 2, however, yield classification errors which lead to biased associations between external variables and dietary patterns. The bias-adjusted three-step approaches suggested by Vermunt (2010) use estimates of classification errors to correct for bias in step 3.

We illustrate and apply these approaches on nutritional database collected during the NESCaV (Nutrition, environment and cardiovascular health) survey (Alkerwi et al., 2010) by analysing the associations between dietary patterns and cardiovascular risk factors.

Reference 1: Alkerwi, A. a., Guillaume, M., Zannad, F., Laufs, U., & Lair, M.-L. (2010). Nutrition, environment and cardiovascular health (NESCAV): protocol of an inter-regional cross-sectional study. BMC Public Health, 10(1), 698. doi: 10.1186/1471-2458-10-698

Reference 2: Vermunt, J. K. (2010). Latent Class Modeling with Covariates: Two Improved Three-Step Approaches. Political Analysis, 18(4), 450-469. doi: 10.1093/pan/mpq025

PC5 - T32: Impact of neighborhood size when examining contextual effects of school and neighborhood on adolescent BMI in cross-classified models: a simulation study

Authors: Carly Milliren[1], Clare R. Evans[2], Tracy K. Richmond[1], Erin C. Dunn[3].
[1] *Boston Children's Hospital, Boston, Ma, USA,* [2] *University of Oregon, Eugene, Or, USA,*
[3] *Massachusetts General Hospital, Boston, Ma, USA.*

Recent advances in multilevel modeling allow for modeling non-hierarchical levels (e.g., students in non-nested schools and neighborhoods, patients across multiple providers) using cross-classified models (CCMM). Current practice is to cluster samples from only one context (e.g., schools) and utilize the observations however they are distributed across the second context (e.g., neighborhoods). It is unknown whether an uneven distribution of sample size across the first and second context leads to incorrect estimates of random effects.

Using the school and neighborhood data structure in the National Longitudinal Study of Adolescent to Adult Health (Add Health), we examined the effect of neighborhood sample size imbalance on the estimation of variance parameters. By simulating students from each of the Add Health schools and assigning them to neighborhoods using three scenarios of (im) balance, we examine how often the true variance parameters were recaptured at the school and neighborhood levels. 1,000 random datasets were simulated across five combinations of school-variance and neighborhood-variance values for a total of 15 combinations of neighborhood size/balance and variance. For each simulation, we calculated the 95% CI for each variance parameter to determine whether the true simulated variance fell within the interval.

Simulation results indicated that estimates of random effects (across a range of possible variance values) were robust to the distribution of students within neighborhoods. Regardless of neighborhood size, both neighborhood and school variances were successfully estimated in 93-96% of simulations suggesting that there is no bias in the ability of CCMM to capture the true variance parameters.

PC5 - T35: Multilevel logistic modelling: issues when working with large datasets

Authors: Scott Harris[1], Edith M. Y. Cheng[2].
[1] *University of Southampton, United Kingdom,* [2] *The Chinese University Of Hong Kong, Hong Kong.*

In public health research it is becoming increasingly common for studies to combine data that is collected at the individual level with higher-level aggregated data which could come from hospitals, specialist treatment centres or GP practices. Multilevel models are often used to analyse such datasets as they allow the hierarchical structure of the data to be taken into account. With large datasets this can result in computational issues depending on the software package being used, the computing environment and the complexity of model being fitted. For this comparison we focused on a two-level model, which is the most commonly seen in practice.

An example dataset containing in excess of 8 million cases and a higher level term with over 32,000 levels was used for the statistical modelling. Random subsets of varying size were taken from this dataset and models with differing levels of complexity were applied to them all. We investigated models with random intercepts as well as more complicated models which included random slopes. Variations on the number of classes in the higher-level variable were examined and the impacts on model fitting noted.

The GLIMMIX and HPLMIXED procedures in SAS 9.4 were used to fit the models. Other software packages were examined but they struggled with the size of the dataset. Each scenario was run 3 times, with the mean computation time reported. Computation times varied from 1.21 to 2072.4 seconds, although actual model run times were longer. The number of levels in the higher level term is the strongest driver to computation time in SAS PROC GLIMMIX, although there appears to be an interaction between the number of cases and the number of levels when looking at random slope models.

PC5 - T38: Selection of reference groups in cohort studies

Authors: Benjamin French[1], John B. Cologne[1], Ritsu Sakata[1], Mai Utada[1], Dale L. Preston[2].
[1] *Radiation Effects Research Foundation, Japan,* [2] *Hirosoft International Corporation, United States.*

In cohort studies, unbiased estimation of exposure-outcome associations requires selection of an appropriate reference group of unexposed individuals. We illustrate strategies for analyzing cohort data with multiple potential reference groups. We analyzed the association between radiation exposure and incidence of first primary solid cancer among 105,444 participants of the Life Span Study (Hiroshima and Nagasaki, Japan, 1958–2009). Potential reference groups included zero-dose survivors at different ground distances from the hypocenter (internal) and city residents who were not in either city at the time of the bombings (external). DS02R1 weighted absorbed colon doses were estimated by the DS02 dosimetry system. Piecewise constant hazard models estimated excess relative risks of first primary solid cancer. We focused on sex-averaged excess relative risks and the shape of the dose-response curve. Incidence rates differed among the potential reference groups. Selection of not-in-city residents as the reference group resulted in an overstated excess relative risk and reduced evidence of a curvilinear dose response among males. These differences were particularly apparent at weighted absorbed colon doses less than 1 Gray. In cohort studies, selection of an appropriate reference group requires understanding of the nature of unmeasured confounding to which the results could be sensitive.

PC5 - T41: Shared and unshared exposure uncertainty in proportional hazards models: effects on risk estimation and shape of the exposure-response curve

Authors: Sabine Hoffmann[1], Chantal Guihenneuc[2], Sophie Ancelet[1].
[1] *Institut de Radioprotection et de Sûreté Nucléaire, France,* [2] *Université Paris Descartes, France.*

In randomised clinical trials (RCT), comparative effectiveness of therapies measured on a dichotomous outcome are often reported as a risk difference (RD), risk ratio (RR) or odds ratio (OR). In cost-effectiveness (CE) evaluations in the context of health technology assessment (HTA), it is not uncommon for evaluations to consider consistency of results across an absolute difference (RD) and a relative measure (RR and/or OR). We explore the implications of using this approach via an indirect treatment comparison (ITC), a comparison of experimental treatments from two different RCTs via a common comparator present in both trials, conducted in the absence of direct head-to-head evidence. A drawback of an ITC is the increase in estimation error resulting from the contribution each trial makes to the overall estimation error. This increase in variance can lead to vague hierarchies of therapeutic relativity in disease indications where multiple treatments are available, but for which direct head-to-head evidence is not available. Establishing unequivocal therapeutic relativity between treatments can be advantageous to sponsor companies in price negotiations within the context of CE analyses in HTA. We explore the implications of the lack of symmetry of the RR measure on statistical efficiency in the context of an ITC. As the statistical efficiency of the RR is dependent on the response rates, we conclude that under certain circumstances, efficiency of an ITC is increased by considering a RR based on non-response, rather than response. An ITC of two treatments for severe chronic plaque psoriasis which is refractory to treatment with non-biological disease modifying anti-rheumatic drugs is used to demonstrate our findings.

PC5 - T44: Getting the balance right: comparing propensity score matching algorithms using real-world data from the US National Immunization Survey

Authors: Bryony Langford[1], Aimee Hall[1], Chris McDonald[1], Lucy Eddowes[1].
[1] *Costello Medical Consulting LTD, United Kingdom.*

Objectives: Propensity score matching (PSM) models are often tested with synthetic data. We compared the speed and consistency of various PSM models using the real-world National Immunization Survey (NIS) 2015 dataset to determine which are most appropriate in health-related biostatistics.

Methods: For southern US NIS respondents, we compared household data (number of children [NUM], income quartile [INQ]), child characteristics (SEX, age group [AGE], ethnicity [ETH], firstborn status [FB]) and response language (LAN) between those with insurance (INS; n=2762) and without (NOINS; n=2490) with PSM models. We used the R packages MatchIt (standard PSM: matchit) and Matching (standard PSM: Match; automatic optimisation: genetic matching [GenMatch]) to balance means in the INS and NOINS subsets, assessed using a two-sided Student's t-test.

Results: Prior to PSM, all variable means other than SEX (p=0.2) and AGE (p=0.01) were significantly different between INS and NOINS (p<<0.001). Matchit (runtime t<1min), Match (t<1min) and GenMatch (t=90min) matched 767–793 (31–32%) NOINS to 2762 INS subjects. Matchit improved balance meaningfully for only three means (FB: p=0.04, NUM: p=0.08, SEX: p=0.4). All means were better balanced after Match (LAN: p=0.03, FB: p=0.05, AGE: p=0.07, ETH: p=0.08, NUM: p=0.9, INQ: p=0.9, SEX: p=0.9) but GenMatch generally balanced the means best (p>0.3 except for AGE [p=0.2]).

Conclusions: When applied to the NIS dataset, the PSM model Match was better than matchit. While GenMatch was computationally intensive, it was the only model to well-balance all variables. Our results demonstrate the value of comparing standard and genetic PSM models using real-world data as opposed to synthetic data alone.

PC5 - T47: Dichotomization of continuous variables with maximum OR values method in logistic regression

Authors: Jialing Huang[1], Xianying He[2], Jinxin Zhang[1].
[1] *Sun Yat-Sen University, Guangzhou, China,* [2] *The First Affiliated Hospital Of Zhengzhou University, Zhengzhou, China.*

Logistic regression is a widely used multivariable method in medical research, which requires the relationship between logitϖ and continuous covariates should comply with the linearity. It is possible to obtain a false relationship between the impact factors and outcome using continuous independent variables to fit logistic regression directly when the linearity hypothesis is violated. To make a rational classification of the continuous variables to solve the situation is a research hotpot[1]. However, there are no judgment on the linearity hypothesis and the point selection is too casual in common classification methods. Thus, we developed maximum OR values method to dichotomize continuous variables, which firstly determines whether the linear relationship is met, and choose the independent values as final cut-off point to fit logistic regression when the corresponding OR value reaches its maximum value. The result of simulation studies and real data analysis showed, maximum OR values method presented better fitting effect with smaller AIC, maximum Nagelkerke R2 and larger overall accuracy and performed better in screening out the impact factors related to outcome, compared with continuous variables method, median split method and one cut-off point minimum P values method which used commonly in processing independent variables in logistic regression. It is recommended that utilizing maximum OR values method to discretize continuous covariates to fit logistic regression model if these variables and logitϖ have non-monotonic relationship.

Reference 1: Turner E L, Dobson J E, Pocock S J. Categorisation of continuous risk factors in epidemiological publications: a survey of current practice. Epidemiol Perspect Innov, 2010,7:9.

PC5 - T50: A comparison of statistical methods to estimate the proportion of (non-specific) acute gastroenteritis burden attributable to rotavirus

Authors: Nicole Mealing[1], Anthony T. Newall[2], Andrew Hayen[1].
[1] *University of Technology Sydney, Australia,* [2] *University Of New South Wales, Australia.*

Background: Rotavirus disease cases may not be recorded as such in routine healthcare databases. Examination of acute gastroenteritis (AGE) encounters (i.e. non-specific disease) and pathology testing can help to establish those encounters that are rotavirus-attributable. The statistical methods used in the literature to do this vary and it is unclear which method is most appropriate. Method: We compare statistical methods to estimate the mean weekly rate of AGE Emergency Department (ED) presentations that are attributed to rotavirus in Results: The estimated mean weekly rate of rotavirus-attributable ED presentations ranged from 8.2-14.4 per 100,000 children in the pre-vaccination era and 3.4-6.0 per 100,000 population in the vaccination era. Most methods assumed that the ratio of positive rotavirus laboratory tests to ED presentations was constant over time. This restricted the reduction in the mean weekly rate of rotavirus-attributable ED presentations from the pre-vaccination to the vaccination era to be 58%, which was the observed reduction in the mean weekly number of positive rotavirus laboratory tests. The process to account for seasonality had less of an impact on the final results. Conclusion: Statistical methods that allowed the ratio of positive rotavirus laboratory tests to ED presentations to vary over time were found to be the more appropriate in determining rotavirus-attributable AGE ED presentations.

PC5 - T53: Digital spiral analysis metrics for discrimination of pathological tremor subtypes

Authors: Roberto Melotti[1], Isabel Wurster[2], Yuri D'elia[1], Cristian Pattaro[1], Peter P. Pramstaller[1], Daniela Berg[3].
[1] *Center for Biomedicine, Eurac Research, Bolzano, Italy,* [2] *Center of Neurology, Department of Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany,* [3] *Clinic for Neurology, Department of Neurology, Christian-Albrechts-University, Kiel, Germany.*

Pathologic tremor represents an involuntary oscillatory and rhythmic movement disorder, which impairs daily activities. Resting tremor is typical of Parkinson's disease (PD), whilst action and postural tremor are typical of essential tremor (ET). However, motor symptoms overlap frequently and are difficult to discern in the early stages.

Digital Spiral Analysis (DSA) represents a cost-effective aiding tool for the objective assessment of action tremor. Aims were to generate ad-hoc metrics from DSA for different features of tremor, and gauge repeatability and discriminant performance on a sample of tremor affected patients and controls.

We assessed four distinct nosologic entities: tremor dominant PD (n=20), ET (n=33), ET and idiopathic PD (n=10), and healthy controls (n=19). On a random sequence of the 2 hands, participants drew 6 consecutive guided Archimedean spiral tracings on a digital tablet. Fully automated data cleaning

and processing in Python yielded metrics for drawing amplitude, acceleration, speed and pressure. Linear mixed models on transformed metrics revealed high repeatability of each metric (ICC range: 0.76-0.81). Little learning (time-trend) systematic effects were detected, none for amplitude. Clinical scores of tremor severity correlated strongly with individually averaged tremor amplitude (r=0.87) and acceleration (r=0.75). Quadratic discriminant analysis, including differences of speed and weight between hands, discriminated among the 4 nosologic groups comparably better (reclassification k=0.74) than using scores of visually rated spirals (k=0.55).

Ad-hoc metrics derived from DSA appear suitable to discriminate among different types of pathologic tremor and for large-scale applications.

PC5 - T56: A multivariate approach to investigating impact of obesity on iron status

Authors: Elizabeth McKinnon[1], Anita Chua[2], Debbie Trinder[2], Leon Adams[2], John Olynyk[3].
[1] *Murdoch University, Australia,* [2] *University of Western Australia, Australia,* [3] *Fiona Stanley Hospital, Australia.*

Clinical practice is frequently guided by biomarker cut-offs for disease diagnosis or classification of risk profiles. The simplicity of many limits entrenched in clinical guidelines – usually adjusted only for broad demographic categories – often belies the complexity of the biological system contributing to both the mechanism underlying the condition of interest and markers used for its diagnosis. Adiposity is a known risk factor for both iron deficiency and dysregulated iron metabolism but the underlying mechanisms, particularly their dependence on increased fatty deposits in the liver, are not well-understood. In this study we consider how obesity, a burgeoning clinical and public health problem in industrialized countries, impacts iron status in a large longitudinal cohort of young adults. We demonstrate how the diagnostic value of serum ferritin as a marker of low body iron stores may be compromised by obesity, and apply multivariate techniques to build a useful picture of the interplay between factors contributing to heterogeneity of important indices of iron status.

PC5 - T58: Multiple imputation and inverse probability weighting for estimating pulmonary disease prevalence in survey data with non-response

Authors: Omololu Stephen Aluko[1].
[1]*University of Kwazulu-Natal, South Africa.*

Incomplete data is a frequent occurrence in many research areas especially cross sectional survey data in epidemiology, health and social sciences research. Complete case analysis is the most used technique in survey data analysis, where it simply takes the form of a listwise deletion of observations with missing rates. This can induce bias of parameter estimates, loss of statistical information and loss of distributional relationships between the predictors, if data is not missing completely at random. In this paper, the effect of missing observations were accounted for by using multiple imputation

(MI) and inverse probability weighting (IPW) methods. Generally, multiple imputation has the ability to draw multiple values from palusible predictive distribution for the missing values. However, under the inverse probability weighting procedure the weights are the inverse of the predicted probabilities of response estimated from the missingness models of incomplete variables. A simulation study is conducted to compare methods and demonstrate that a cross sectional survey data can be used to mitigate bias induced by missing data. The simulation results show the benefit of the IPW compared with the MI. The former performs well but not as the latter. The findings were based on an application to a real dataset on chronic obstructive pulmonary disease.

PC5 - T59: Has the impact of risk factors on hearing loss in infants changed over 10 years? Polish Universal Neonatal Hearing Screening Program

Authors: Kinga Salapa[1], Grazyna Greczka[2], Maciej Wrobel[2], Witold Szyfter[2], Malgorzata Wierzbicka[2].
[1] *Department of Bioinformatics and Telemedicne, Jagiellonian University Medical College, Poland,* [2] *Department of Otolaryngology and Oncological Laryngology, Poznan University of Medical Sciences, Poland.*

Objective: The study was conducted to evaluate the change of influence of risk factors and outcome of the transient evoked otoacoustic emission test (TEOAE) on hearing loss (HL) in infants registered in the Polish Universal Neonatal Hearing Screening Program between 2003 and 2015.

Materials and methods: The selected population included newborns exposed to at least one of 13 risk factors (RFs) as indicated by the Joint Committee of Infant Hearing (JCIH), with failed TEOAE test or selected children who passed TEOAE test without any RF. Total of 242,393 infants were examined, whereof 10,248 were diagnosed as HL (4.2%). Generalized Linear Model was applied to model the probability of HL. Information about program duration (1-10 years) was included into the model as the confounder.

Results: The results showed significant paired association between four RFs connected with bad condition of premature infants. As a consequence the new RF was created based on these four (1 - neonate exposed to at least one of four risk factors, 0 - otherwise). One of RF (TORCH infection) was removed from the analysis due to changes of its definition in 2011. There were no significant interactions between program duration and RFs. In the homogeneity situation the associations can be summarized by the common odds ratios adjusted for program duration. Almost all RFs, apart from hyperbilirubinemia, were significant predictors of hearing loss in infants. The most important ones were TEOAE outcome (ORadj=7.8), the craniofacial anomalies (ORadj=7.2), complex congenital anomalies (ORadj=5.1) and familial HL (ORadj=2.2).

Conclusions: The impact of each RF remained the same over 10 years of Program.

PC5 - T61: An epidemiological model to find out factors associated with nodal involvement among Indian oral cancer patients and its validation

Authors: Vishwajeet Singh[1], S. N. Dwivedi[1], S. V. S. Deo[1], M. A. Khan[1].
[1] *India Institute of Medical Sciences, India*

Background: Oral cancer is the most common cancer among Indian men.Treatment of these patients can be tailored with the knowledge of nodal involvement.Hence, understanding of its associated factors may provide clues to the clinician for better management.Accordingly,objective of this study was to develop a prediction model to find out factors associated with nodal involvement and validate it.Materials and Methods:The data on oral cancer patients available with the Dept. of Surgical Oncology,BRAIRCH,AIIMS,New Delhi,was used.All histopathologically proven oral cancer patients,who underwent for surgery including neck dissection,were included.For stepwise multivariable analysis,a subset of covariates was selected on its clinical relevance and/or significance (p=0.25) under univariate.The results were considered significant at 5% level of significant.The results in the form of odds ratio (corresponding 95% CI) were considered.Results:Pathologically nodes were positive in 39.8% of patients.Patient with pain [1.34(1.02- 1.77)] and clinical nodes involvement [2.38(1.69-3.35)] at the time of presentation;having tumor other than well differentiated [1.41(1.05-1.89)] and tongue as compared with buccal mucosa [1.62(1.08-2.46)] were more likely to experience positive nodes.Non significant HL test for goodness of fit showed that developed model describes the distribution satisfactorily.The proportion of agreement was obtained as 65.61% in data used for model development & 63.73% in external data set.Conclusion:The developed model described the distribution satisfactorily.Pain at the time of presentation,presence of clinical node,involved oral site & degree of differentiation were most probable associated factors with nodal involvement.

PC5 - T62: Estimating case fatality risk and reproduction number for Ebola disease using daily notification data

Authors: Zheng Chen[1], Zihang Lu[2], Huilin Chen[1], Lixian Li[1], Yawen Hou[3].
[1] *Southern Medical University, China,* [2] *University Of Toronto, Canada,* [3] *Jinan University, China.*

When an emerging influenza virus appears in humans, one of the early concerns is whether the virus has the potential to be fatal and widely spread in a substantial fraction of infected individuals. Case fatality risk (CFR) and reproduction number (R0) are two most important epidemiological parameters used to quantify the mortality and morbidity of the epidemic. This study aims to compare the availability and accuracy of different approaches to estimate timely CFR and R0, and to aid in further planning disease control and intervention strategies. Daily notification data were available for Ebola virus disease (EVD) in 1976, 1995 and 2014 outbreaks. For each of these outbreaks, we compared the estimates of CFR and R0 based on multiple approaches. Where appropriate, recovery-death hazard ratio plot and Lam test were used to evaluate the constancy of timely CFR. Our finding suggested that while the CFR of previous outbreak (1976 and 1995) were constant, the CFR of 2014 outbreak in both Sierra Leone and Guinea was decreasing over time. Although less severe as compared to the two historical outbreaks, the 2014 outbreak of EVD remained a significant concern and highly fatal, with CFR in Guinea higher than that in Sierra Leone.

Moreover, the estimates of R0 (1.4 for Sierra Leone and 1.3 for Guinea) implied that infected cases were continuous to grow and the outbreak has not yet brought under control on 2014. This study was supported by NSFC (81673268, 81202288).

Reference 1: Lu Z, Chen Z. Timely Case-Fatality Risk Estimation. Epidemiology.2015.26(2):e26-7.

Reference 2: Chen Z, Akazawa K, Nakamura T. Estimating the case fatality rate using a constant cure-death hazard ratio. Lifetime Data Analysis. 2009.15:316-329.

PC5 - T65: Analysis of spatially clustered data on schistosoma haematobium infection in southwestern Tanzania

Authors: Kirsi Manz[1], Petra Clowes[2], Inge Kroidl[3], Dickens O. Kowuor[2], Christof Geldmacher[1], Nyanda E. Ntinginya[2], Leonard Maboko[2], Michael Hoelscher[1], Elmar Saathoff[1].
[1] *LMU Munich, Germany,* [2] *Nimr-Mbeya Medical Research Center (MMRC), Tanzania.*

Background: Urinary schistosomiasis is a worm infection which is common in the tropics and caused by the human parasitic trematode Schistosoma haematobium. Its transmission involves Bulinus spp. freshwater snails as intermediate host. Since the presence of susceptible snail species strongly depends on environmental conditions, knowledge about the factors favoring transmission might help to control infections.

Methods: We analyzed data from a cross-sectional study conducted in Mbeya Region/Southwestern Tanzania to assess the spatial pattern of S. haematobium infection and its potential association with various ecological factors. The analysis included data on 17,310 individuals from 4,199 households and nine study-sites. Since schistosome infection was clustered within households and study-sites, we used mixed-effects Poisson regression models with study-site and household as random effects to calculate prevalence ratios (PRs). This approach allows for comparing PRs across the different households and study sites, because effects of clustering on both levels are accounted for.

Results: The multivariable mixed-effects models showed a significant association of S. haematobium infection with individual (e.g. HIV status; PR = 0.59, 95% CI 0.39-0.89) and environmental factors (e.g. distance to water course; PR = 1.26, 95% CI 1.15-1.37).

Conclusion: Mixed-effects modeling is useful to account for clustering, which is common in epidemiological data. Furthermore, our findings regarding the association of S. haematobium infection with various factors could aid in the design and implementation of effective control measures.

Oral Contributed Sessions

OC26: STRengthening Analytical Thinking for Observational Studies (STRATOS) 2

Tuesday 11th July - 12.36-13.30 h. - Room: Auditorio
Chair: Willi Sauerbrei

OC26-1: Comparison of methods for adjusting for time-varying confounding

Authors: Philip Clare[1], Timothy Dobbins[1], Raimondo Bruno[2], Richard P. Mattick[1].
[1] *UNSW Australia, Australia,* [2] *University of Tasmania, Australia.*

Adjusting for time-varying confounding is important when estimating causal associations from non-randomised longitudinal data. A range of statistical methods for adjusting for time-varying confounding have been proposed, including Marginal Structural Models (MSMs) and Targeted Maximum Likelihood Estimation (TMLE), however, there has been little comparison of the available methods. We conducted Monte Carlo simulations comparing a range of methods, including MSMs, G-computation and TMLE, under various scenarios of correct and incorrect specification, such as omission of interaction or polynomial effects. In addition, different methods for estimating the models were also compared, including generalised linear models (GLM), mixed-effect models (GLMM), and gradient boosted models (GBM). As expected,when there was no misspecification, the methods performed better than naïve analysis, with around 95% coverage of confidence intervals, and relatively little bias in estimates. Performance reduced under misspecification, with reasonable coverage but greater bias in estimates. TMLE did not perform as well as expected, with generally comparable bias to MSMs, but relatively poor coverage even when estimates were unbiased, primarily due to relatively small standard error estimates. Even in the presence of misspecification, GBM did not perform substantially better than GLMs/GLMMs. While there are definite theoretical advantages to more complex methods such as TMLE and other doubly robust methods for adjusting for time-varying confounding, simpler methods display comparable performance, even when there is misspecification in the exposure or outcome models. However, further study is required for more complex scenarios.

OC26-2: Measurement error and timing of predictor values used in prediction model research: a systematic review of current practice and reporting

Authors: Rebecca Whittle[1], George Peat[1], John Belcher[1], Richard Riley[1].
[1] *Keele University, England.*

Prediction models are important for clinical decision making, and utilise values of multiple predictors to enable individualised risk prediction. However, when developing such models, measurement error may affect the observed predictor values, which could potentially lead to biased estimates of

predictor effects, and thus misleading model predictions. In this presentation, we provide a systematic review of recent articles that developed a diagnostic or prognostic prediction model, and examine the potential for measurement error in the predictors used and how it was acknowledged and/or accounted for in the analysis. We show that many of the prediction models proposed included predictors that were potentially measured with error, and yet this was rarely acknowledged or accounted for in the statistical modelling. We also identify that most of the articles included did not explicitly state the exact timing that the model is intended to be used in clinical practice, or exactly when the predictors used in the modelling development were measured. These issues raise a concern that many developed models may be sub-optimal and pose challenges for clinicians wishing to implement the models in practice. Recommendations for improvement are provided.

OC26-3: Is kidney transplantation best preceded by dialysis? On avoiding potential biases through G-estimation

Authors: Camila Olarte Parra[1], Els Goetghebeur[1].
[1] *Ghent University, Belgium.*

Estimating the impact of initial dialysis vs immediate kidney transplantation on outcomes after transplantation is a typical question which registry-based studies seek to answer. For conclusions drawn from such non-randomized studies, the recently published ROBINS-I tool highlights 7 domains of potential bias(1). We review each of them and examine how adapted G-estimation methods can improve on the methodology used thus far.

Confounders available for adjustment to avoid bias (domain 1, D1) tend to depend on the registry. Registries entering patients upon inclusion on the transplant waiting list, delay entry for patients referred to transplantation after starting dialysis which leads to bias in the selection of study participants (D2). The dynamic nature of dialysis can lead to bias in classification error and deviations from intended interventions (D3,D4). Registries are not mandatory and many suffer from missing data(D5).

When patients are lost to follow-up after returning to dialysis, informative censoring generates non-representative outcome measurements (D6). And finally, papers cutting follow-up at different times for different types of failure raise questions on selective reporting (D7).

We review the potential impact of these sources of bias in some key papers using realistic simulation. We then build on work from Vock et al(2) to adapt G-estimation to answer the new causal question on the value of pre-emptive treatment in one particular registry under the no-unmeasured confounders assumption.

Reference 1: Sterne JAC. BMJ. 2016;355(7040):i4919

Reference 2: Vock DM. Biometrics. 2013;69(4):820-9

OC27: Joint modelling in practice 2

Tuesday 11th July - 12.36-13.30 h. - Room: Sala Mar 2
Chair: Francisco Gude

OC27-1: Integrating Latent Classes in the Bayesian Shared Parameter Joint Model of Longitudinal and Survival Outcomes

Authors: Eleni Rosalina Andrinopoulou[1], Kazem Nasserinejad[1], Dimitris Rizopoulos[1].
[1] *Erasmus MC, The Netherlands.*

When patients are monitored after a kidney transplantation, it is of clinical interest to investigate the association between the longitudinal biomarker protein to creatinine ratio and time-to kidney failure. A feature of this data set is that different sub-populations exhibit different longitudinal profiles. Patients can be categorized in several sup-groups (latent classes) with different trajectories. There fore, to better model the association between the longitudinal and the survival out come these latent classes should be taken into account.

The joint model of longitudinal and survival data constitutes a popular frame work to analyze such data sets. In particular, two paradigms within this framework are the shared parameter joint models and the joint latent class models. The former paradigm allows to quantify the strength of the association between the longitudinal and survival outcomes but does not allow for latent sub-populations. On the other hand, the latter paradigm explicitly postulates the existence of sub-populations but does not directly quantify the strength of the association.

To answer our motivating research question we propose to integrate latent classes in the shared parameter joint model in a fully Bayesian approach. Specifically, the model allows us to investigate the association between protein-to-creatinine ratio and time-to kidney failure within each latent class. We, furthermore, focus on the selection of the true number of latent classes.

OC27-2: Penalized likelihood estimation of a trivariate additive binary model

Authors: Panagiota Filippou[1], Giampiero Marra[1], Rosalba Radice[2].
[1] *University College London, UK,* [2] *Birkbeck, University Of London, UK.*

This work proposes a penalized likelihood method to estimate a trivariate system of binary regressions,which accounts for several types of covariate effects (such as linear, nonlinear, random and spatial effects) through the use of semi-parametric predictors, as well as error correlations. The dependence structure of the three binary regressions is specified through the use of copula functions with arbitrary margins. The link functions considered are the probit, logit and complementary log-log. The proposed approach also addresses the difficulty in estimating accurately the correlation coefficients, which characterize the dependence of binary responses conditional on covariates.

We found that this is not an unusual occurrence for trivariate binary models and as far as we know such a limitation has been neither discussed nor addressed in the literature. The parameters of the model are estimated within a penalized likelihood framework based on a carefully structured trust region algorithm with integrated automatic multiple smoothing parameter selection. The relevant numerical computation can be easily carried out using the SemiParTRIV() function in a freely available R package. The proposed method is illustrated through a case study which uses data from the North Carolina State Center for Health Statistics whose aim is to model jointly adverse birth binary outcomes in North Carolina.

OC27-3: Using copula GAMLSS to study the relationship between the age at menarche and reproductive lifespan of women in Central Portugal breast cancer screening program

Authors: Elisa Duarte[1], Bruno De Sousa[2], Carmen Cadarso-Suárez[1], Giampiero Marra[3], Vítor Rodrigues[2].
[1] *Universidad de Santiago de Compostela, Spain,* [2] *Universidade de Coimbra, Portugal,* [3] *University College London, UK.*

The time of exposure of a woman to hormones is considered as a breast cancer risk factor. The reproductive lifespan is a woman's life period when she is exposed to endogenous hormones responsible to ensure the functioning of her reproductive system. Therefore the age at menarche and age at menopause set a woman's reproductive lifespan. The aim of this study is to assess the influence of the year of birth in the upward trend of a woman's reproductive lifespan and if it is conditioned by the downward trend of the age of menarche, ascertained by the literature. We develop a bivariate copula model to relate the age at menarche and the woman's reproductive lifespan and quantify the effect the woman's year of birth in this relationship. In addition, the equations include a woman's residence municipality which allow us to take into account the spatial correlations and hence explore any possible effect of the geographic location of a woman's residence in the behavior of age at menarche and menopause. For this analysis we employ Bivariate Copula Additive Models for Location, Scale and Shape. Such models extend the scope of univariate GAMLSS by binding two equations with binary, discrete or continuous responses. The equations can be flexibly specified using smoothers with single or multiples penalties, hence allowing for several types of covariate effects. The copula dependence parameter can also specified as a function of flexible covariate effects. All the model's parameters are estimated simultaneously. The inference is carried out using the R package SemiParBIVProbit.

Reference 1: Marra, G., & Radice, R. (2016). A Bivariate Copula Additive Model for Location, Scale and Shape. arXiv preprint arXiv:1605.07521.

OC28: Survival analysis 3

Tuesday 11th July - 12.36-13.30 h. - Room: Sala Terra 2
Chair: Luis Meira-Machado

OC28-1: A shared frailty model using Gamma shape mixtures and the EM algorithm for interval-censored data

Authors: Aysun Cetinyurek-Yavuz[1, 2], Philippe Lambert[2, 3].
[1] *Danone Nutricia Research, Utrecht, Netherlands,* [2] *Institut Des Sciences Humaines et Sociales, Université de Liège, Belgium,* [3] *Institut De Statistique, Biostatistique Et Sciences Actuarielles,Université Catholique De Louvain, Belgium.*

Recently, there has been an increasing interest in statistical analysis of interval-censored time-to-event data. This type of data is quite usual for clinical trials or longitudinal studies especially in practical settings of AIDS and cancer research where the individuals have pre-scheduled visits but the event of interest occurs between the visits. Moreover, in clinical trials, the units may be collected in clusters and they share some observed or unobserved characteristics, i.e. patients from multiple centres, teeth of multiple subjects. In dependent right-censored data setting, numerous extensions of familiar survival procedures have been proposed. A popular model accounting for the correlation among the observations is the shared frailty model. In this approach, it is assumed that there exists an unobserved latent variable shared by all individuals in the same cluster and named the frailty and is assigned a parametric distribution, typically, a gamma distribution due to its conjugacy in the Cox PH model. However, in the case of interval-censored time-to-event data, the inclusion of gamma frailties results in complicated intractable likelihoods. The use of the EM algorithm solves this issue because gamma distribution becomes conjugate for complete data likelihood.

The EM algorithm also facilitates the use of more flexible distributions such as gamma shape mixtures (GSM) for the frailty. The combination of the EM algorithm and of a GSM distribution for the frailty provides a more flexible modelling framework for clustered interval-censored data. Therefore, we aim to present a shared frailty proportional hazards model with a flexible form for the time-to-event distribution and the GSM distribution for the frailty.

OC28-2: A frailty model for the estimation of the excess mortality hazard in case of mismatched lifetables

Authors: Francisco Javier Rubio[1], Bernard Rachet[1], Roch Giorgi[2], Aurelien Belot[1], Censur Working Survival Group[3].
[1] *Cancer Survival Group, Faculty of Epidemiology and Population Health, Department of Non-Communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, WC1E 7Ht, UK,* [2] *Aix Marseille University, Inserm, IRD, Sesstim, Sciences Economiques & Sociales de la Sante & Traitement de L\'Information Medicale, Marseille, France,* [3] *Censur Working Survival Group.*

In cancer epidemiology using population-based data, excess mortality hazard is a useful method to estimate cancer survival (Graffeo et al., 2012). However, a limitation of this method in practice is that cancer patients need to be matched to groups of the population, sharing their available characteristics, in order to obtain their corresponding background mortality rates from the lifetables. Some characteristics such as deprivation, ethnicity, and comorbidities, are not available in a number of countries, which may affect the mortality rates assigned to each patient. It has been shown that not including relevant information for matching the mortality rates of patients induces a bias in the estimation of the parameters of the excess hazard (Graffeo et al., 2012). In this work, we present a frailty model that accounts for misspecification of the background mortality rate, generalising a previously proposed approach (Zahl, 1997). The proposed model has a closed form expression, which facilitates its implementation. We evaluate the statistical performance of this approach in a simulation study, compare it to alternative models, and finally illustrate its use on real data. We detail the conditions and limitations of this method to apply, and give some recommendations of its use in practice, mainly if some key assumptions are in doubt.

Reference 1: P.H. Zahl. Frailty modelling for the excess hazard. Statistics in Medicine, 16(14):1573–1585, 1997.

Reference 2: N. Graffeo, V. Jooste, and R. Giorgi. The impact of additional life-table variables on excess mortality estimates. Statistics in Medicine, 31(30):4219–4230, 2012.

OC28-3: Non-proportional hazards or unobserved heterogeneity in clustered survival data: can we tell the difference?

Authors: Theodor Adrian Balan[1], Hein Putter[1].
[1] *Leiden University Medical Center, The Netherlands.*

In survival analysis, shared frailty (random effect) models are often used to model heterogeneous survival data, such as clustered failures or recurrent events. In Hougaard (2000) a large family of infinitely divisible frailty distributions were proposed, including the positive stable and compound Poisson distributions with a probability mass at 0. The estimation of semiparametric frailty models with these distributions has proved challenging.

The assumption of proportional hazards, usually made conditional on the frailty, does not carry over to the marginal model for most random effect distributions. It has been shown that, for univariate

data, this makes it impossible to distinguish between the presence of a frailty or marginal non-proportional hazards. Difficulties also arise when the data consists of small sized clusters.

We discuss the results of a simulation study carried out in the situation where the clusters have a small size or individuals have few recurrent events. For a large number of frailty distributions, we analyze the behaviour of test statistics for the presence of the frailty and for the proportional hazards assumption. With a novel software implementation for estimating semiparametric shared frailty models, we discuss the situations when the unobserved heterogeneity can be distinguished from the non-proportional hazards. The practical implications are illustrated in real-world data analysis examples.

Reference 1: Hougaard P (2000). Analysis of Multivariate Survival Data. New York: Springer-Verlag.

Reference 2: Yashin, AI, Iachine, IA, Begun, AZ and Vaupel, JW (2001). Hidden frailty: myths and reality. Research report, Odense University.

OC29: Topics in biostatistics 1

Tuesday 11th July - 12.36-13.30 h. - Room: Sala Terra 4
Chair: María del Carmen Iglesias Pérez

OC29-1: Impact of standardization in L1-penalized regression

Authors: Michael Kammer[1], Georg Heinze[1].
[1] *Medical University of Vienna, Austria.*

The estimation process of L1-penalized regression (LASSO) involves the absolute norm of the regression coefficients. Hence, it is essential to bring variables to a common scale. Otherwise, variables with small coefficients due to their natural scaling will be favored unduly.

However, what exactly constitutes this common scale? Discussions about how to properly standardize variables have a long history in statistics, mainly in the context of relative importance of variables. We will investigate the impact of some of the proposed methods in the setting of LASSO estimation, especially with regard to categorical predictors which are very common in medical applications.

We will investigate how the standardization process influences shrinkage of regression coefficients, selection probabilities and prediction performance when the pool of candidate predictors includes balanced and unbalanced categorical variables. Furthermore, we question if categorical and continuous variables are treated similarly by standardization.There are indications that differences in variability affect the LASSO estimates even after simple standardization in certain situations. We will present simple considerations, complemented by a simulation study for more complex scenarios. Our focus will be on low-dimensional problems with more observations than variables. However, we will also explore the high-dimensional case, where the LASSO is a common method to analyze e.g. SNP data. Real data will be used to demonstrate our recommendations.

OC29-2: Model selection based on combined penalties for biomarker identification

Authors: Eleni Vradi[1], Werner Brannath[2], Thomas Jaki[3], Richardus Vonk[1].
[1] *Bayer Ag, Germany,* [2] *Institute for Statistics and Competence Center for Clinical Trials, Bremen University, Germany,* [3] *Department of Mathematics and Statistics, Lancaster University, UK.*

The growing role of targeted medicine has led to an increased focus on the development of actionable biomarkers. Current penalized selection methods that are used to identify biomarker panels for classification in high dimensional data, however, often result in highly complex panels that need careful pruning for practical use. In the framework of regularization methods a penalty that is a weighted sum of the L0 and L1 norm has been proposed to account for the complexity of the resulting model. In practice, the limitation of this penalty is that the objective function is non-convex, non-smooth, the optimization is computationally intensive and the application to high-dimensional settings is challenging. We propose a stepwise forward variable selection method which combines the L0 with L1 or L2 norms. The penalized likelihood criterion that is used in the stepwise selection procedure results in more parsimonious models, keeping only the most relevant features. Simulation results and a real application show that our approach exhibits a comparable performance with common selection methods with respect to the prediction performance whilst minimizing the number of variables in the selected model resulting in a more parsimonious model as desired.

Reference 1: 1. Liu Y, Wu Y. Variable selection via a combination of the L0 and L1 penalties. Journal of Computational and Graphical Statistics. 2007,16(4):782-798.

OC29-3: A comparison of model selection methods in the presence of multiply imputed data

Authors: Thao Le Thi Phuong[1], Guy Thwaites[2], Ronald Geskus[2].
[1] *Oxford University of Clinical Research Unit, Vietnam,* [2] *Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.*

In prognostic studies, many approaches to model selection have been proposed. In the presence of multiply imputed (MI) data, the number of options increases even further. Despite some significant contributions to elucidate the problem, no method prevails as uniformly best.

We conducted a logistic regression simulation study to compare performance (AUC and Brier score) of two variable selection methods and possible variations with MI data:

BF: Model selection on Bootstrap (BS) data, using backward elimination based on AIC or LASSO; fit model based on most Frequently (e.g. $\geq$ 50%) selected variables over all MI and BS data sets

LO: Apply LASSO to Original MI data. We used a suboptimal penalty based on the 1-se rule to obtain a more parsimonious model. The final model averages estimates over MI data sets. An alternative is to restrict to variables that are selected in >50% of the MI data. We considered recalibration of over-shrinkage, both of the linear predictor and of all individual variables. A variation that selects variables

uniformly over all MI data sets is to first stack the data before using LASSO. We also considered stacked LASSO with individual weights as determined by the fraction of missingness. We will apply the methods on a real dataset of 951 tuberculosis meningitis adult patients to predict mortality at 9 months.

Results: In the complete data, the LO method with recalibration of all selected variables performed best. The same was seen with MI data; stacking, with and without weights, had equal model performance.

Conclusion: The OL selection method with recalibration of all variables is a valid model selection strategy, which is easily extended to MI data.

OC30: Methods for handling missing data 2

Tuesday 11th July - 12.36-13.30 h. - Room: Sala Mar 4
Chair: María Helena Gonçalves

OC30-1: Imputation of missing covariates: when standard methods may fail

Authors: Nicole Erler[1], Dimitris Rizopoulos[1], Oscar H. Franco[1], Emmanuel M.E.H. Lesaffre[2].
[1] *Erasmus Medical Center, Rotterdam, The Netherlands,* [2] *L-Biostat, KU Leuven, Belgium.*

Our work is motivated by examples from two large cohort studies, the Generation R Study and the Rotterdam Study, in which the analysis models of interest involved non-linear effects, interaction terms or had a longitudinal outcome. As is the case for most observational datasets, missing values in multiple variables complicated the analyses. The most popular method to deal with missing values is multiple imputation using the fully conditional specification (FCS). In settings like our motivating examples, however, the analysis and imputation models specified by FCS are incompatible, which violates an important assumption of FCS and may result in severely biased estimates. Even though many applied researchers have to deal with incomplete data, often they are not aware of the assumptions that are required to obtain valid results from the imputation methods implemented in standard software or the bias that may result from violations.

In our present work, we briefly review assumptions of FCS and the potential effects of violations thereof. We discuss previously proposed extensions aiming to reduce bias due to incompatibility and contrast them to recent approaches that specify imputation models that assure compatibility utilizing the Bayesian framework. Focusing on methods that are available in existing or newly developed R packages, the application of these methods will be illustrated for generalized linear as well as linear mixed models that involve non-linear effects or interaction terms, using relevant data from several recent studies.

OC30-2: Propensity score analysis incorporating missingness patterns for partially observed confounders

Authors: Helen Blake[1], Clémence Leyrat[1], Kathryn Mansfield[1], James Carpenter[1, 2], Elizabeth Williamson[1, 3].
[1] *London School of Hygiene and Tropical Medicine, UK,* [2] *MRC Clinical Trials Unit at Ucl, UK,* [3] *Farr Institute Of Health Informatics, UK.*

Electronic Health Records (EH.) are a valuable data source for investigating health related questions”,” and propensity score analysis has become an increasingly popular approach to address confounding bias in such investigations. However”,” because EHR data are typically routinely recorded as part of standard clinical care”,” there are often missing values”,” particularly for potential confounders. In our case study – investigating the effect of renin-angiotensin system blockers on the risk of acute kidney injury – two key confounders”,” ethnicity and chronic kidney disease stage”,” have 59% and 53% of missing data respectively.

The missingness pattern approach (MPA) has been proposed as a method for handling partially observed confounders in propensity score analysis. In the MPA”,” propensity scores are estimated separately for each missingness pattern present in the data”,” and then collected into one variable. Although the assumptions underlying the validity of the MPA are stated in the literature”,” the way they are stated means that it can be difficult for analysts to assess their plausibility in specific clinical applications.

By using causal diagrams to examine the underlying assumptions of the MPA”,” we develop a framework to guide researchers considering whether the assumptions will hold. We apply the framework to our case study”,” showing that the underlying assumptions of the MPA appear reasonable. We provide practical guidance on when the MPA will yield valid inferences”,” finding that the MPA can often apply where more commonly-used missing confounder methods”,” such as complete case analysis and multiple imputation”,” may not be appropriate.

OC30-3: A causal approach to handling missing data in multiple variables: coming to grips with missing not at random

Authors: Margarita Moreno-Betancur[1, 2], Finbarr P. Leacy[3], Ian R. White[4], John B. Carlin[1].
[1] *Murdoch Childrens Research Institute, Australia,* [2] *University of Melbourne, Australia,* [3] *Royal College of Surgeons In Ireland, Ireland,* [4] *MRC Biostatistics Unit and MRC Clinical Trials Unit, United Kingdom.*

In studies with multiple incomplete variables, it is widely understood that if the data are missing at random (MAR) then unbiased estimation is possible with appropriate methods. While the need to assess the plausibility of this assumption has been emphasised, the practical difficulty of this task and the stringency of MAR in the context of multivariable missingness are rarely acknowledged. Further, while MAR is sufficient, it is certainly not necessary: in a wide range of missing not at random (MNAR) scenarios unbiased estimation of certain parameters is possible. Recent developments in the computer science literature suggest that a causal reframing of missing data problems could

prove more natural for stating assumptions, and could provide a more useful guide to the treatment of missing data, beyond the MAR-MNAR dichotomy. We build on that work to develop a causal approach to handling missing data in the context of a typical point-exposure epidemiological study with incomplete exposure, outcome and confounders. Specifically, we use directed acyclic graphs to depict missingness assumptions, and consider a counterfactual approach to elucidating the conditions required for non-parametric identification, or recoverability, of a target parameter. Further to providing a strategy for tackling the complexities of MNAR, this paradigm suggests a novel estimation method akin to g-computation for recoverable parameters, based on tailored record deletion. In a simulation study, this approach performed better in terms of bias than multiple imputation. We outline implications of the strategy for the conduct of sensitivity analyses for non-recoverable parameters and use the Longitudinal Study of Australian Children for illustration.

12th July
Wednesday

Invited Session

IS6: Biostatistics for High Dimensional Data

Wednesday 12th July - 09.00-10.30 h. - Room: Auditorio
Chair: Jeanine Houwing-Duistermaat
Organised by Jeanine Houwing-Duistermaat, *University of Leeds, United Kingdom*

IS6-1: An Integrative Network-Based Analysis of Multi-Omics for Survival Cancer Data

Pietro Lio, *University of Cambridge, United Kingdom*

Multi-omics studies are believed to provide a more comprehensive picture of a complex biological system than traditional studies with one omics data source. From a statistical perspective, the most important challenge in integrating multi-omic analyses is the high-dimensionality of the data.

In particular, taking more levels into account increases the dimensionality of the problem including unknown parameters, which are often difficult to estimate, thereby making the overall inference weaker. The aims of our study is to (i) integrate different data types identifying the best approach for combining multiple matrices that include data from different scales in a biologicallymeaningfulway; (ii) compare the performance of genomic integration with the individual data types making use of network penalized methods; (iii) test the predictive power of our signatures on independent datasets; (iv) derive the network interactions and associated risk pathways based on combined omics data.

In particular, different online databases and bioinformatic infrastructure available for integrated omics data in relationto time-to-event data are used to provide a more comprehensive picture of breast, ovarian and lungcancer.

Overall our analysis shows that the integrative network-based analyses are expected to play an increasingly important role in the interpretation of high-throughput omics cancer data.

IS6-2: Latent Variable Modelling for Multiple Omics Data

Hae Won Uh, *Leiden University Medical Center, The Netherlands*

The recent advances of life science technologies have enabled the collection of a virtually unlimited quantity of data from multiple sources. The 'systems biology' approach involves parallel investigation of, for example, transcriptomic, proteomic, or metabolomic datasets. It generally consists of two steps: each of the data sources are analysed in parallel, and each of the results are summarized to outline novel findings or hypotheses based on prior knowledge of existing pathways.

These 'omics' datasets are noisy, heterogeneous, highly correlated within each dataset and among the datasets, and often high-dimensional (p>>n), with some underlying hierarchical structure of the biological system. To unwrap systematic and biologically relevant information, a simultaneous approach provides meaningful insights. Therefore, latent variable modelling (e.g. using common

factor analysis) is considered. This well-established but computationally intensive method, however, is not equipped to handle high-dimensional data. Dimension reduction and data integration can be achieved using a Partial Least Squares regression (PLS) variant. A drawback is the lack of a probabilistic model that leads to statistical inference and hypothesis testing.

A unified approach of latent variable modelling that has merits of PLS is developed for integration of multiple omics datasets. Various real world applications will be shown.

IS6-3: Genomics and Drug Discovery

John Whittaker, *Glaxo Smith Kline Pharmaceuticals, United Kingdom*

I will discuss the use of genomics and data science in drug discovery at GSK, and particularly in the selection and validation of drug targets. I'll highlight the interplay between experimental science and informatics, and will discuss recently published work highlighting the value of genetic information in selecting drug targets/indications. Motivated by this, I will also include a sketch of work ongoing and planned in this area at GSK, particularly with respect to the potential to integrate electronic health record and genomics information. I will also describe work at Open Targets (https://www.targetvalidation.org/), including both experimental and informatics aspects. I will focus on the key concepts and avoid technical detail, both regarding genomics and data science.

Oral Contributed Sessions

OC31: Design and analysis of clinical trials 4

Wednesday 12th July - 09.00-10.30 h. - Room: Sala Mar 2
Chair: Frank Bretz

OC31-1: Sequential trials in the context of competing risks – a case study

Authors: Corine Baayen[1], Paul Blanche[2].
[1] *Capionis, France,* [2] *University of South Brittany, France.*

Group sequential trials allow for early stopping of a trial in case of a clear treatment effect, or a clear lack thereof. Advantages of such trials include that individuals are not exposed to unsafe, inferior or ineffective treatments unnecessarily and that less recourses may be required.

Analysis results of sequential trials should take into account that multiple tests are being performed. Therefore, critical values of interim and final test statistics are often based on their joint distribution. For many tests, this joint distribution follows a canonical joint distribution, which is at the basis of most standard group sequential methods.

Recently it was shown that Gray's test statistic [1] and the Wald test statistic of the Fine and Gray model [2] also follow a canonical joint distribution and can therefore be used in sequential trials. These tests have been developed to evaluate competing risks endpoints for survival data.

Recently the authors of this paper were involved in a sequential trial in the context of competing risks in prematurely born infants. They found very limited available literature in this context and no applied examples. In this presentation they share their experience with performing such a trial. The steps and choices to be made are illustrated using the real trial as example and remaining open issues are highlighted.

Reference 1: Logan, B. R., & Zhang, M. J. (2013). The use of group sequential designs with common competing risks tests. Statistics in medicine, 32(6), 899-913.

Reference 2: Martens & Logan (2016). A Group Sequential Test of a Competing Risk Endpoint for Treatment Effect Based on the Fine-Gray Model. In JSM 2016, Chigago.

OC31-2: Test-compatible inference for adaptive two-stage single-arm designs with binary endpoint

Authors: Kevin Kunzmann[1], Meinhard Kieser[1].
[1] *Heidelberg University, Germany.*

Inference after two-stage single-arm designs with binary endpoint is challenging due to the non-unique ordering of the sampling space in multi-stage designs. Naïve confidence intervals and p values might indicate a different conclusion than the design's test decision. In this talk, the issue of test-compatible p values and confidence intervals for designs with varying stage-two sample size is presented. P values and point estimators are closely connected and test-compatible p values can be obtained from estimators which are optimized under the additional restriction that their corresponding natural p values be test-compatible [1]. Here, the natural p value of an estimator is defined as the p value resulting from the ordering of the sampling space that is induced by the estimates. Consequently, Clopper-Pearson confidence intervals based on these p values must also be test-compatible. The resulting confidence intervals tend to be conservative but assure the nominal coverage probability. As an alternative, we pursue a direct optimization approach minimizing the expected mean width of the confidence intervals under the Jeffreys prior of the design. We investigate the operating characteristics of a test-compatible estimator that minimizes expected mean square error with respect to the Jeffreys prior, the corresponding p values and Clopper-Pearson confidence intervals, as well as the test compatible confidence intervals obtained from direct minimization of the expected mean width of the intervals.

Reference 1: Kunzmann, K., and Kieser, M. (2017) Point estimation and p-values in phase II adaptive two-stage designs with a binary endpoint. Statist. Med., 36: 971–984.

OC31-3: Optimal planning of phase II/III programs for clinical trials with multiple endpoints

Authors: Meinhard Kieser[1], Marietta Kirchner[1], Eva Doelger[1], Heiko Goette[2].
[1] *Institute of Medical Biometry and Informatics, University of Heidelberg, Germany,* [2] *Merck Kgaa, Darmstadt, Germany.*

Due to increased costs and competition pressure, drug development becomes more and more challenging. Therefore, there is a strong need for improving efficiency of clinical research by developing and applying methods for quantitative decision making. In this context, the integrated planning for phase II/III programs plays an important role as numerous quantities can be varied that are crucial for cost, benefit, and program success. Within a utility-based approach, an optimal planning of phase II/III programs can be achieved that puts the choice of decision boundaries and phase II sample sizes on a quantitative basis [1]. We propose a procedure for optimal phase II/III planning in the situation of clinical trials with multiple endpoints and (asymptotically) normally distributed test statistics. Optimal phase II sample sizes and go/no-go decision rules are derived both for the "all-or-none" and the "at-least-one" win criterion. Application of the method is illustrated by drug development programs in the fields of Alzheimer's disease and oncology.

Reference 1: Kirchner M, Kieser M, Götte H, Schüler A. Utility-based optimization of phase II/III programs. Statistics in Medicine 2016; 35:305–316.

OC31-4: Covariate-adjusted response-adaptive designs in clinical trials where the survival responses of patients follow a semi-parametric model

Authors: Ayon Mukherjee[1].
[1] *Queen Mary, University of London, United Kingdom.*

Covariate-adjusted response-adaptive (CARA) designs use the available responses to skew the treatment allocation in a clinical trial in favor of the treatment found at an interim stage of the trial to be best for a given patient's covariate profile.

There has recently been extensive research on diverse aspects of CARA design with the patient responses assumed to follow a theoretical model. However the range of application for such designs becomes limited in real life clinical trials where the responses of the patients infrequently fit a certain parametric form. On the other hand, the parametric assumption yields robust estimates about the covariate adjusted treatment effect. To balance these two requirements, designs are hereby developed without any distributional assumption about the survival responses, relying only on the assumption of proportional hazards of patients between the two treatment arms.

The proposed designs have been developed in two ways, namely, by deriving two variants of optimum allocation, and also by using an appropriate link-function. The optimal designs are based on the doubly-adaptive biased coin design (DBCD) in one case, and the efficient randomized adaptive design (ERADE) in the other. The derived treatment allocation proportions for these designs converge towards the expected targeted values. The design based on the link function is derived using the distribution function of a Gumbel model. The comparative merits of the proposed designs have been elaborated, their preferred application has been discussed in detail, and their operating characteristics have also been established through extensive simulation study. An existing clinical trial has been redesigned by applying the proposed methods.

OC31-5: A new design for trials for finding optimal treatment duration

Authors: Matteo Quartagno[1], James Carpenter[1], Mahesh Parmar[2], Patrick Philips[2], Sarah Walker[2].
[1] *London School of Hygiene and Tropical Medicine, United Kingdom,* [2] *MRC Clinical Trials Unit At UCL, United Kingdom.*

Introduction: There is substantial interest in investigating the use of shorter durations of antibiotic treatment for bacterial infections to counter the global threat of antimicrobial resistance (AMR). Classic designs compare two arbitrarily chosen durations in a non-inferiority trial. Given their important limitations (including arbitrary non-inferiority margins), our aim was therefore to develop new alternative designs to optimise duration of treatment.

Methods: We recast the problem of determining an optimal duration as estimating the shape of the `Duration-Response Curve' (DRC). Patients are randomised to one of multiple treatment durations within a pre-specified range. We explore fractional polynomials and spline-based methods to estimate

the unknown DRC. We begin with equidistant treatment durations designs, and evaluate MSPE and area under the fitted prediction curve to investigate the absolute and relative bias in DRC estimation under a range of possible scenarios. We then look at the sensitivity of results to changes in design parameters including different numbers and durations of arms.

Results: Our proposed design estimates the DRC within 5% relative error across a wide range of possible plausible simulation scenarios, representing our uncertainty on the real shape of the DRC. We show that the best flexible regression technique in this situation is fractional polynomial regression and that a moderate number of equidistant duration arms (5 or 7) is enough to adequately model the DRC. These results suggest that our design allows for robust identification of the shortest effective duration as an alternative to the classic 2-arm non-inferiority design.

OC32: Multistate survival analysis and dynamic prediction 2

Wednesday 12th July - 09.00-10.30 h. - Room: Sala Terra 2
Chair: Klaus Langohr

OC32-1: Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model

Authors: Krithika Suresh[1], Jeremy M.G. Taylor[1], Alexander Tsodikov[1].
[1] *University of Michigan, USA.*

Dynamic prediction incorporates time-dependent marker information accrued during follow-up to improve personalized survival prediction probabilities. At any follow-up, or "landmark", time, the residual time distribution for an individual, conditional on their updated marker values, can be used to produce a dynamic prediction. To satisfy a consistency condition that links dynamic predictions at different time points, the residual time distribution must follow from a prediction function that models the joint distribution of the marker process and time to failure, such as a joint model. To circumvent the assumptions and computational burden associated with a joint model, approximate methods for dynamic prediction have been proposed. One such method is landmarking, which fits a Cox model at a sequence of landmark times, and thus is not a comprehensive probability model of the marker process and the event time. Considering an illness-death model, we derive the residual time distribution and demonstrate that the structure of the Cox model baseline hazard and covariate effects under the landmarking approach do not have simple form. We suggest some extensions of the landmark Cox model that should provide a better approximation. We compare the performance of the landmark models with joint models using simulation studies. We examine the predicted probabilities produced under both methods using data from a prostate cancer study, where metastatic clinical failure is a time-dependent covariate for predicting death following radiation therapy.

OC32-2: Predicting treatment outcome using intermediate state information based on a multistate landmark model including pre-landmark co-variables

Authors: Ulrich Beyer[1], Francesca Michielin[1].
[1] *F. Hoffmann-La Roche LTD.*

Clinical development programs in Oncology are often gated based on intermediate endpoints. For example if a new treatment shows a certain increase in Recist response the development of the drug will continue as a survival benefit is expected. However the correlation between response and survival is often poor and the prediction of the survival endpoint should be based on a more complete assessment.

A good method is provided with multistate models, where the disease progression is described by the four states of stable disease (SD), response (RESP), progression (PD) and the absorbing state death (D). At a predefined gating landmark t, transition probabilities from each of the 3 states SD, RESP and PD into death will be calculated. If these transition probabilities are estimated based on a reference trial, the prediction for a new investigational compound can be made assuming that the transition probabilities stay constant and only the percentage of patients being in stable disease and response will change. As a new treatment might not only change the SD or response rates, but also e.g. the duration of response or SD, pre-landmark co-variables (e.g. tumor burden) can be included into the model which will have a direct impact on these durations.

The presentation will provide a brief description of the method and show a retrospective analysis, where this method was applied to data from a phase III breast cancer combination trial to predict the combination survival rates based on post landmark transition probabilities derived from monotherapy data only. The predicted combination survival curve fitted the observed data very well and demonstrates that this method can improve predictions based on response rates only.

OC32-3: A dynamic prediction approach to analyse the population-attributable fraction of ventilator associated pneumonia

Authors: Maja Katharina Von Cube[1], Martin Wolkewitz[1], Martin Schumacher[1], Jan Beyersmann[2], Jean-Francois Timsit[3, 4], Wafa Essaied[3, 4], Michael Darmon[5], Maité Garrouste-Orgeas[6], Elie Azoulay[7], Yves Cohen[8].
[1] *Institute for Medical Biometry and Statistics, Germany,* [2] *University of Ulm, Germany,* [3] *Université Paris Diderot,* [4] *Hopital Bichat, France,* [5] *Saint-Etienne Teaching Hospital, France,* [6] *Service de Médecine Intensive et de Réanimation, France,* [7] *Saint Louis Teaching Hospital, France,* [8] *Avicenne Teaching Hospital, France.*

A main interest in hospital epidemiology is to analyse the impact and consequences of hospital -acquired infections (HAIs). One issue is the estimation of mortality due to infections caused by specific pathogens. A way to analyse this quantity is the concept of population-attributable fraction (PAF) which measures the benefit in regard to overall mortality if the infection could be eliminated.

Difficulties arise due to the fact that the occurrence of HAIs is a time- dynamic process and that discharge is a competing event to hospital-death. Moreover, adjustment for confounders is essential to obtain an unbiased estimator.

Little literature is available accommodating an estimation of the PAF accounting for confounders in the case where both exposure and outcome are time-dependent.

We propose a landmark approach for dynamic prediction of the adjusted PAF. At each landmark we use a logistic regression model to account for the current exposure state and further covariates. With a simulation study we show how the method is able to capture the effect of a time-dependent exposure on a time-dependent outcome. Moreover, the method is applied to a large French database of intensive care unit patients (n=7319) to estimate the adjusted PAF of ventilator-associated pneumonia caused by the pathogen Pseudomonas aeruginosa.

Reference 1: van Houwelingen, Hans, and Hein Putter. Dynamic prediction in clinical survival analysis. CRC Press, 2011.

Reference 2: Nicolaie, M. A., et al. "Dynamic Pseudo-Observations: A Robust Approach to Dynamic Prediction in Competing Risks." Biometrics 69.4 (2013): 1043-1052.

OC32-4: Multi-state net survival model

Authors: Leyla Azarang[1].
[1]*Aix Marseille University, France.*

Net survival, is the probability of surviving cancer under study in the absence of other causes of death. The concept arises from missing or unreliable records of the cause of death in many data sources. This seemingly too hypothetical concept plays a pivotal role in comparison of specific cancer related mortality among countries. It is, also, important to identify prognostic factors for cancer related mortality. The excess hazard regression model proposed by Esteve et al. (1990) addresses this issue by assuming proportionality of the hazards. However, in many cancer studies some nonfatal events might be observed during the follow-up which increase the mortality risk. This kind of datasets are usually analyzed using multi-state models.

The focus of our work is a three states multi-state model, called progressive illness-death model, where the final state is death; thus, the challenge of the uncertainty about cause of death still exists. Considering net survival, we estimate the effect of some prognostic factors on the probability of occurrence of events in the progressive illness-death model, by employing direct binomial approach (Azarang et al. 2017). The flexibility of our approach is its applicability in non-Markov and non-proportional settings. The performance of the proposed method is evaluated in a simulation study, also it is applied to a real dataset.

Reference 1: Azarang L, Scheike T, de Uña-Álvarez J. (2017). Direct modeling of regression effects for transition probabilities in the progressive illness-death model. Statistics in Medicine. In Press.

Reference 2: Estève J, Benhamou E, Croasdale M. and Raymond L. (1990). Relative survival and the estimation of net survival: elements for further discussion. Statistics in Medicine,9, 529-538.

OC32-5: Assessing treatment effects on registry data in presence of competing risks

Authors: Brice Ozenne[1], Christian Torp-Pedersen[2], Thomas Alexander Gerds[1].
[1] *University of Copenhagen, Copenhagen,* [2] *Aalborg University, Aalborg.*

Registry data offer the opportunity to estimate the incidence of a disease ata national scale and to compare treatment effects. Estimation technics, however,have to be corrected for the confounders that will arise due to the absenceof randomization and have to handle the occurrence of competing events (e.g.death). Here we investigate the use of cause-specific Cox models to estimatethe cumulative incidence of a disease over time and the use of the G-formula tocompare this incidence between treatment modalities.We show how analytic expressions for the confidence interval of the cumulativeincidence, and their difference, can be derived using the influence formula. Thevalidity of the G-formula estimates relies, among other hypothesis, on a correctmodeling of the relationship between exposure, confounders and outcome.We thus investigate an alternative modeling strategy where we would reject theinitial Cox model if a model that relax one of the Cox model assumption givesignificantly different treatment estimate.We illustrate our work on data from the Danish national registries comparingthe effect of preventing treatments on the risk of recurrent stroke. Theproposed estimators for the cumulative incidence and their confidence intervalare implemented in the riskRegression package that is available on github(https://github.com/tagteam/riskRegression/).

OC33: Heterogeneity in biomedical studies and meta-analysis

Wednesday 12th July - 09.00-10.30 h. - Room: Sala Terra 4
Chair: Jarek Harezlak

OC33-1: Guidance for deriving and presenting percentage study weights in meta-analysis of test accuracy studies

Authors: Danielle Burke[1], Joie Ensor[1], Kym I.E. Snell[1], Danielle Van Der Windt[1], Richard D. Riley[1].
[1] *Keele University, United Kingdom.*

Percentage study weights in meta-analysis reveal the contribution of each study toward the overall summary results, and are especially important when some studies are considered outliers or at high risk of bias. In meta-analyses of test accuracy reviews, such as bivariate meta-analyses of sensitivity and specificity, percentage study weights are usually expressed relative to the study sample size or to the standard error of study-specific estimates of logit sensitivity and logit specificity. In this presentation, we explain why these approaches give incorrect study weights, and adopt an alternative method based on a decomposition of Fisher's information matrix [1]. This method also generalises to a bivariate meta-regression, so that percentage study weights can be derived for estimates of study-level modifiers of test accuracy, as well as for overall summary sensitivity and specificity results. We illustrate the method with two meta-analyses examining test accuracy: one of ear temperature for diagnosis of fever in children; and the other of positron emission tomography for diagnosis of

Alzheimer's disease. These highlight that the percentage study weights based solely on sample size can be hugely inaccurate, and should no longer be used. We suggest that the proposed percentage weights should be presented routinely on forest and ROC plots for sensitivity and specificity, to provide transparency of the contribution of each study in test accuracy meta-analyses.

Reference 1: Riley R, et al., Deriving percentage study weights in multi-parameter meta-analysis models: with application to meta-regression, network meta-analysis, and one-stage IPD models. Stat Meth Med Res, 2016, in press.

OC33-2: A semi-parametric approach for meta-analysis of survival curves from orthopedic registries

Authors: Samprit Banerjee[1], Art Sedrakyan[1].
[1] *Weill Cornell Medical College, United States of America.*

The increasing demand for timely, yet evidence-based decisions in health care and public health, synthesis of evidence or meta-analysis, has gained popularity over the last decade. While there are numerous methods/models for performing meta-analysis on various types of outcomes, in this talk I will focus on meta-analysis of survival outcomes. For survival outcomes (e.g. time to event) the standard method employed in meta-analysis studies involves combining log-hazard ratios in a linear, fixed or random effects, model. However, such methods are not suitable for meta-analysis of multiple time-dependent survival estimates (generally, survival curves). Our motivating example is a study of comparative effectiveness of medical devices used in hip-replacement and knee-replacement surgeries. The United States Food and Drug Administration funded an International Consortium of Orthopedic Registries (ICOR), headed by Cornell University's medical college, which comprises of world-wide registries and teams of researchers. In this study, time to revision surgery, post hip-replacement surgery, is being examined as a survival outcome and data from 7+ US and international registries has been aggregated to conduct inference. Here, I present a multivariate random-effects model to combine reported survival probabilities at multiple time-points for various covariates. I generalize this model to include non-parametric estimation of the cumulative hazard function via a stochastic process in a flexible linear mixed model framework. I demonstrate the advantages of this approach via simulations and apply this method to the ICOR data.

OC33-3: Population-adjusted treatment comparisons: estimates based on matching-adjusted indirect comparison and simulated treatment comparison

Authors: David Phillippo[1], A. E. Ades[1], Sofia Dias[1], Stephen Palmer[2], Keith R. Abrams[3], Nicky J. Welton[1].
[1] *University of Bristol, UK,* [2] *University of York, UK,* [3] *University of Leicester, UK*

We present the findings and recommendations of a recent NICE Technical Support Document (available from www.nicedsu.org.uk) regarding the use of population-adjusted treatment comparisons in health technology appraisal.

Standard methods for indirect comparisons and network meta-analysis are based on aggregate data, with the key assumption that there is no difference between trials in the distribution of effect-modifying variables. Two methods which relax this assumption, Matching-Adjusted Indirect Comparison (MAIC) and Simulated Treatment Comparison (STC), are becoming increasingly common in industry-sponsored treatment comparisons, where a company has access to individual patient data (IPD) from its own trials but only aggregate data from competitor trials. Both methods use IPD to adjust for between-trial differences in covariate distributions. Despite their increasing popularity, there is a distinct lack of clarity about how and when these methods should be applied. We review the properties of these methods, and identify the key assumptions. Notably, there is a fundamental distinction between "anchored" and "unanchored" forms of indirect comparison, where a common comparator arm is or is not utilised to control for between-trial differences in prognostic variables, with the unanchored comparison making assumptions that are infeasibly strong. Furthermore, both MAIC and STC as currently applied can only produce estimates that are valid for the populations in the competitor trials, which do not necessarily represent the decision population. We provide recommendations on how and when population adjustment methods should be used to provide statistically valid, clinically meaningful, transparent and consistent results.

OC33-4: Longitudinal meta-analysis of percentile ratios for survival data

Authors: Fotios Siannis[1].
[1]*Department of Mathematics University of Athens Panepistimioupolis Athens, Greece.*

The meta-analysis of survival data is of interest and is based on log hazard ratios, since the proportional hazards model is the most prevalent way of analyzing survival data. However, the proportionality assumption becomes a rather restrictive assumption when it is applied to a number of studies. For that reason, the benefits of a meta-analysis based on percentile ratios is of interest. This measure has a useful interpretation that can be appealing to clinicians and others who would like to utilize such approach. The main complication is that percentiles are by construction associated and such association needs to be accounted in the analysis. Multivariate meta-analysis could provide the needed tools in order to account for this association. In this work, we propose an approach of meta-analyzing associated data that have a natural ordering. Percentile ratios can be seen as longitudinal data measured in "percentile time", that goes from zero to one. Thus, we have a number of independent studies providing a number of observations over time, exactly in the

same way longitudinal data do. In this way we can utilize known methodology for the analysis of longitudinal data in order to build the necessary modeling structures needed in order to summarize the information that comes from several studies and at the same time account for the association between observations. This can be seen as a two stage meta-analysis, where in the first stage data from each study can come from parametric, semi-parametric or non-parametric approaches. The second stage can be seen as a standard longitudinal data analysis.

Reference 1: F. Siannis et.al. (2010) "One-stage parametric meta-analysis of time-to-event outcomes", SiM, 29, 3030-3045.

OC33-5: Individual patient data meta-analysis of time-to-event data: a review of the methodology

Authors: Valentijn de Jong[1], Thomas P. A. Debray[1], Marinus J. C. Eijkemans[1], Karel G.M. Moons[1].
[1] *UMC Utrecht Julius Center, The Netherlands.*

Relative treatment effects are frequently estimated using individual participant data (IPD) with time-to-event outcomes. When combining IPD from multiple studies and performing a so-called IPD meta-analysis, several methods have been proposed to summarize treatment effects and investigate sources of heterogeneity.

We performed a literature review to provide an overview of methods for conducting an IPD-MA with a time-to-event outcome. Hereby, we focused on assessing good practice for modelling frailty of patients across studies, for choosing appropriate effect measures, for dealing with (differences in) censoring, and for addressing time-varying treatment effects.

We discuss several parametric and semi-parametric methods for modelling time-to-event data, and describe how to implement these in a one-stage or two-stage meta-analysis framework. Further, we discuss how these models might be affected by heterogeneity due to the choice of effect measure, the presence of censoring and the presence of time-varying treatment effects or effect modification.

We will illustrate several key methods in an empirical IPD-MA example where clinical trial data from 2793 patients from six countries were combined to compare the effects of online haemodiafiltration with those of haemodialysis on mortality.

OC34: Statistical challenges of EHR (eHealth Records) analysis

Wednesday 12th July - 09.00-10.30 h. - Room: Sala Mar 4
Chair: María Durbán

OC34-1: How should missing time-varying confounders be handled to estimate marginal effects from electronic health records?

Authors: Clemence Leyrat[1], Ruth Farmer[1], Krishnan Bhaskaran[1], James Carpenter[1, 2], Elizabeth Williamson[1, 3].
[1] *London School of Hygiene and Tropical Medicine, UK,* [2] *MRC Clinical Trials Unit at UCL, London, UK,* [3] *Farr Institute of Health Informatics, London, UK.*

Electronic health records (EHR) are useful for addressing health-related questions, such as estimating the marginal effect of a treatment over a long period. In practice, a patient's treatment exposure may not be constant over time but updated as their medical history evolves. In our motivating example using data from the Clinical Practice Research Datalink to compare the effect of first-line drugs for diabetes on mortality, 73% of the 98080 participants change exposure during the follow-up. Key covariates (HbA1c or BMI) are updated every 7 months, so longitudinal data may be missing depending on the follow-up interval.

Marginal structural models (MSMs) have been proposed to estimate marginal effects with time-varying confounders, based on inverse-probability-of-treatment-weighting (IPTW). A major issue when applying this method is the presence of missing data among confounders. For a single time-point, we showed that multiple imputation (MI) of the confounders performs well for IPTW estimators, but little attention has been paid to extending this approach to multiple time points.

Using our motivating example, we present the different missingness patterns observed in EHR data and how they relate to Rubin's taxonomy. These patterns informed the design of our simulation study comparing MI to ad hoc methods often used to handle missing time-varying confounders: complete case, mean imputation, missing indicator approaches and last observation carried forward. We present our simulation results and provide recommendations about how to address missing confounders in MSMs. Finally, we discuss the assumptions required to obtain a consistent marginal estimate in this context and how likely they are to hold in EHR data.

OC34-2: The role of electronic health records in the design of trials of complex interventions: a view from surgery

Authors: Olympia Papachristofi[1], Linda Sharples[1].
[1] *London School of Hygiene and Tropical Medicine, United Kingdom.*

Surgical interventions are an indispensable part of modern healthcare; however, their multi-component nature renders formal evaluation in RCTs more challenging than for medical treatments.

Understanding component effects and their interactions on outcomes is imperative for designing adequately powered trials. Our study focuses on how Electronic Health Records (EHR) can be used to identify components of variation and design efficient studies.

We advocate hierarchical models with cross-classifications to establish effects and interactions between components of surgery that are not necessarily in strict hierarchy. For instance, surgery is delivered by multidisciplinary teams with varying outcomes induced by differences in operator skill, the environment in which surgery occurs and patient-specific traits. We illustrate methods by exploring variation in binary and continuous outcomes between two key team components, the surgeon and anaesthetist, adjusting for patient heterogeneity. We use cross-classifications to access surgeons operating with multiple anaesthetists and estimate composite surgeon-anaesthetist effects. An extension to multicentre trials tackles further variation due to centre infrastructure or policy differences, and employs random coefficients to capture potential drivers of between-centre variation.

Methods are applied on >100,000 cardiac surgery patients from 10 centres. We discuss the challenge of applying statistical methods in EHR, using an example where inefficient methods led to false-positive signals. Finally, we use EHR to estimate parameters for various trial designs, considering differing intervention components per arm (surgery vs surgery or drug), interaction effects and unit of randomisation.

OC34-3: A quasi-experimental approach to account for unmeasured confounders in the assessment of side effects using observational data

Authors: Lauren Rodgers[1], John Dennis[1], Beverley Shields[1], Andrew Hattersley[1], William Henley[1].
[1] *University of Exeter Medical School, UK.*

A major challenge to the usefulness of electronic health records in research are measured and unmeasured confounders. The lack of randomisation can lead to bias in treatment effect estimates due to these hidden confounders. The prior event rate ratio is a novel quasi-experimental design which aims to reduce this bias through the use of data from a period prior to treatment. In this method the ratio of a period before and after treatment yields an estimate which reflects the combined effects of known and unknown confounders. Previous work has shown that bias is reduced using this method. We apply this quasi-experimental method to evaluate side effects on treatment using observational data.

The MRC APBI Stratification and Extreme Response Mechanism In Diabetes project investigates patient response to Type 2 diabetes treatment. One arm of the project is the use of electronic health records from the Clinical Practice Research Datalink to identify which patients respond best to which treatments. We compare side effects on two common treatments, thiazolidinediones and sulphonylureas. Thiazolidinediones have known associations with weight gain and oedema. Incidence of the side effect on metformin, a first line treatment for Type 2 diabetes, is used as the prior period. To validate the method we compare the traditional survival model and the prior event rate ratio method to two clinical trials.

OC34-4: Genetic variation in the Estonian population: a pharmacogenomic study of adverse drug reactions using electronic health records

Authors: Kristi Krebs[1], Tõnis Tasa[1, 2], Mart Kals[1], Reedik Mägi[1], Tõnu Esko[1], Andres Metspalu[1], Jaak Vilo[2], Lili Milani[1].
[1] *Estonian Genome Center, University of Tartu, Estonia,* [2] *Institute Of Computer Science, University of Tartu, Estonia,*

Advances in next generation sequencing (NGS) technologies coupled with electronic health records (EHR) provide new opportunities for the interpretation of the role of genetic variation in different diseases and traits. Pharmacogenomics applies NGS methods to document genetic variants important in drug response with the goal of reducing the negative effects of variability in drug response, including adverse drug reactions (ADR).

Here we performed a large-scale study of genes important in drug response of Estonian biobank participants by using e-health databases. We studied the use of ICD10 system to identify ADR diagnoses, and determined drug intake by searching the Biobank participants EHR. We sequenced the whole genomes of 2,240 participants at the Broad Institute, and to increase the sample size included all variants imputed for 14,219 genotyped participants in the biobank. To identify association between genetic variants and the occurrence of ADRs, we used Fisher exact test and logistic regression methods. We faced the restrictions of the low frequency of prescriptions for certain drugs, and the use of ICD10 codes for ADR detection is limited and additional validation of adverse reactions would be necessary for stronger statistical power. Nevertheless, we were able to validate several previously documented genetic variants associated with drug induced ADRs and additionally found new associations.

Population studies of drug response using EH. enable the use of greater sample sizes and faster data collection. By overcoming the limitations, EH. together with genotype information and additional thorough validation of ADRs propose a way to identify individuals potentially at risk for unexpected drug response.

OC34-5: Efficient and robust semi-supervised estimation of treatment effects in electronic medical records data

Authors: David Cheng[1], Tianxi Cai[1].
[1] *Harvard T. H. Chan School of Public Health, USA*

There is strong interest in conducting comparative effectiveness research in electronic medical records (EMR) data to assess treatment strategies in real-world patients. A primary challenge of conducting such analyses in EMR data is the lack of direct observation on a pre-specified true outcome, prompting the need for phenotyping algorithms that impute the outcome given available data. It is often unclear whether such imputations are adequate when used to estimate the treatment effect. We frame the problem of estimating average treatment effects (ATE) in this situation in a semi-supervised learning setting, where a small set of observations are labeled with the true outcome and a

large set of observations are unlabeled. We develop an estimator for the ATE that imputes outcomes in a way such that the final estimator is robust to mis-specification of the imputation model. As a result, information from surrogate variables that predict the outcome in the unlabeled data can be safely leveraged to improve the efficiency in estimating the ATE. The estimator is also doubly-robust in that it will be consistent under correct specification of either an initial propensity score model or a baseline outcome model. It is more efficient than complete-case estimators that neglect the unlabeled data and also more efficient than related missing data/causal inference estimators that we adapt to this setting to make use of the unlabeled data. Simulations exhibit the benefits of this estimator in finite samples. We illustrate the method by comparing rates of treatment response to two anti-TNF therapies for treatment of inflammatory bowel disease.

Poster Contributed Session

PC6: Bayesian methods and joint modelling

Wednesday 12th July - 09.00-10.30 h. - Room: Hall
Chair: Carmen Armero

PC6 - W1: Bayesian central statistical monitoring based on adverse event incidence for risk based monitoring approach: a practical usage of historical data on central statistical monitoring

Authors: Tomoyoshi Hatayama[1].
[1]*Biostatistics Department A2 Healthcare Corporation, Department of Biostatistics Graduate School of Medicine Yokohama City University, Japan.*

Aims: There is growing interest in Risk-Based Monitoring (RBM) as an effective way to ensure data quality in clinical trials. Central monitoring (CM) that monitors data in across all sites has an important role in RBM. However, statistical methods are not sufficiently utilized in CM so far. In this study, we propose a Bayesian central statistical monitoring (CSM) method to detect sites whose tendency of adverse event (AE) occurrence differs from the others as atypical sites. In addition, we propose the method to use prior information efficiently to carry out Bayesian CSM more effectively.Methods: We use Bayesian Poisson regression model to model AE counts by sites and derive Bayesian posterior predictive distribution of AE counts of each site. In our proposed method, sites are detected as atypical when the AE counts of relevant sites are observed as out of pre-specified ranges in posterior predictive distribution. In using prior information, loss of detection performance is a concern in the case of using wrong prior information. Therefore, in the method, we adjusted amount of information used in prior distribution based on posterior precisions of the parameter used in the model to prevent prior distribution from significantly impacting on detecting performance. Results: Intensive simulation showed that the detection performance was improved by using appropriately discounted prior distribution even in the case that the information of likelihood and prior information are inconsistent. Conclusion: We proposed a Bayesian CSM method detecting atypical sites based on the incidence of AEs. We also proposed a practical usage of prior information in CSM. We believe that the proposed method is useful in practical use of RBM.

PC6 - W4: A preventive program for university instructors based on acoustic features automatically extracted from voice recordings

Authors: María Jesús Rufo Bazaga[1], J. Martín[1], C. J. Pérez[1], Y. Campos-Roca[1], A. Moreno[1].
[1] *University of Extremadura, Spain.*

Vocal fold nodules are recognized as an occupational disease for people whose professions require the continued use of voice. This recognition demands the detection of associated risks in a preventive program.

The feasibility of using acoustic features extracted from voice recordings has been previously analyzed to assess the risk of vocal fold nodules in university instructors. A database composed of 90 instructors was collected. A protocol, which consisted of a physical examination, a survey and voice recordings, was applied to all these instructors within the Health Prevention Program of the University of Extremadura. Based on 19 patients suffering from nodules and 53 healthy individuals from the MEEI database (Massachusetts Eye and Ear Infirmary), probability of having vocal fold nodules was calculated to the instructors, classifying them as healthy or suffering from nodules.

The aim of this work is to provide a methodology to classify the previous instructors into one of three risk groups: low, medium and high. The identification of medium-risk patients will lead them to a speech therapy program, whereas those allocated to the high-risk group will be candidates for a more advanced treatment that could include surgery.

A Bayesian decision analysis is proposed for this objective. Firstly, the posterior probabilities of being healthy are calculated for each individual based on a probit regression model. Next, a multicriteria additive utility function is considered. Those criteria are the cost of the corresponding treatment and the utility of recovery. This information is provided by doctors. Finally, a classification approach for each patient based on the maximization of the expected utility is obtained.

PC6 - W7: A simple way to unify Multi-Criteria Decision Analysis (MCDA) and Stochastic Multi-criteria Acceptability Analysis (SMAA) using a Dirichlet distribution in benefit-risk assessment

Authors: Gaelle Saint-Hilary[1], Stephanie Cadour[2], Veronique Robert[2], Mauro Gasparini[1].
[1] *Dipartimento Di Scienze Matematiche (DISMA), Giuseppe Luigi Lagrange, Politecnico Di Torino, Torino, Italy,* [2] *Department of Biostatistics, Institut de Recherches Internationales Servier (IRIS), Suresnes, France.*

Quantitative methodologies have been proposed to support decision-making in drug development and monitoring. In particular, Multi-Criteria Decision Analysis (MCDA) and Stochastic Multicriteria Acceptability Analysis (SMAA) are useful tools to assess the benefit-risk ratio of medicines according to the performances of the treatments on several criteria, accounting for the preferences of the decision-makers regarding the relative importance of these criteria. However, even in its probabilistic form, MCDA requires the exact elicitations of the weights of the criteria by the decision-makers, which may be difficult to achieve in practice. SMAA allows for more flexibility and can be used with unknown or partially known preferences, but it is less popular due to its increased complexity and the high degree of uncertainty in its results. We propose a simple model as a generalization of MCDA and SMAA, by applying a Dirichlet distribution to the weights of the criteria and by making its parameters vary. This unique model permits to fit both MCDA and SMAA, and allows for a more extended exploration of the benefit-risk assessment of treatments. The precision of its results depends on the precision parameter of the Dirichlet distribution, which could be naturally interpreted as the strength of confidence of the decision-makers in their elicitation of preferences.

PC6 - W10: Design of phase I clinical trial based on Bayesian tests

Authors: Chia-Ru Chung[1], Yuh-Ing Chen[1].
[1] *National Central University, Taiwan.*

The purpose of phase I clinical trials is to estimate the maximum tolerated dose (MTD) of the drug under study and to protect patients from the risk of overdose in the dose-escalation procedure. Note that the MTD is the highest dose producing an acceptable toxicity, that is, the probability of experiencing the dose-limiting toxicity (DLT) is less than the pre-specified target toxicity probability (TTP). Therefore, we consider sequentially using Bayesian tests (BT) for the acceptable toxicity based on the available response in the dose-escalation procedure. When the maximum number of patients is reached, the mode of the posterior distribution of the MTD is then recommended for the further phase II trial. Hence, the proposed design is denoted by BTMD. A simulation study is then conducted to investigate the performances of the BTMD and competitive designs of CRM and EWOC on the toxicity probability and MTD estimation under different dose-toxicity curves. A sensitivity study is also implemented to explore the performances to miss-specified dose-toxicity models. The results indicate that the BTMD design generally produces reasonably toxicity probability and gives an more accurate MTD estimation comparing with the competitive trial designs. Moreover, the performance of the BTMD design is relatively robust to the miss-specified dose-toxicity curves.

Reference 1: O'Quigley J, Pape M and Fisher L (1990). "Continual Reassessment Method: A Practical Design for Phase 1 Clinical Trials in Cancer", Biometrics, Vol 46, No. 1, pp. 33-48

Reference 2: Babb J, Rogatko A and Zack S (1998). "Cancer phase I clinical trials: efficient dose escalation with overdose control", Stat Med, 17, 1103-1120.

PC6 - W13: Trialr: Bayesian clinical trial designs in Stan, R and beyond

Authors: Kristian Brock[1].
[1]*Cancer Research UK Clinical Trials Unit, University of Birmingham, UK.*

Trialr is a new software package implementing Bayesian clinical trial designs in Stan. Stan is a probabilistic programming language that performs no-U-turn MCMC sampling to draw from the Bayesian posterior distribution. There are implementations of Stan in R, Stata, Python, Matlab and Julia. Models in Stan are specified using a proprietary modelling language, similar to that used in WinBUGS. This allows analysts to specify rich models without having to handle the potentially arduous prior-to-posterior calculations, making Stan particularly useful for clinical trial designs.

Flexibility is one of the key motivations for using Bayesian analysis. trialr seeks to be a cookbook for Bayesian methods in clinical trials. We provide implementations of the EffTox dose-finding design, Thall's hierarchical Bayesian model for analysing treatment effects of a drug in a disease with many subtypes, the BEBOP design for including predictive information to analyse co-primary efficacy and toxicity outcomes at phase II; and others. Of particular importance is Stan's ability to use non-standard probability functions. This allows it to be used in dose-finding designs like CRM and

EffTox. Users can implement the models as provided. More likely is that users will tailor the provided examples to their own clinical scenarios. We will continue to add further trial designs to trialr.

Trialr is an R-package. The probability models are written in Stan and helper functions are written in R. However, the clinical trial designs can be used in any of the Stan implementations, making trialr a valuable resource to Stata, Python and Matlab users too.

Full source code for trialr is available on GitHub at https://github.com/brockk/trialr and on CRAN.

PC6 - W16: Changing the trial design in an ongoing dose-finding haematological trial: from 3+3 to the Continual Reassessment Method (CRM)

Authors: Samuel Muñoz Vicente[1], Mark Drummond[2], Leyre Navarro-Nuñez[1], Christina Yap[1].
[1] *Cancer Research UK Clinical Trials Unit, Institute of Cancer and Genomic Sciences, University of Birmingham, UK,* [2] *Beatson West of Scotland Cancer Centre, Glasgow, UK.*

It is well known that accrual in trials often suffers difficulties. PHAZAR, a 5-dose 3+3 dose-finding trial followed by an expansion phase, encountered such complications. The initial design meant that recruitment was halted between the completion of each cohort and safety assessment, resulting in a loss of potential patients. This may result in an increased duration leading to excess costs and the potential of stopping early before obtaining sufficient evidence, rendering any statistical analysis and trial conclusions inconclusive.

For this reason, the initial 3+3 design was discarded after completion of two cohorts, and a 2-stage Bayesian CRM design implemented. Our task was 3-fold: facilitate a smooth transition between the two designs, avoid the loss of potential patients and increase the probability of correctly declaring the Maximum Tolerated Dose. Our model allows flexible cohort sizes, with a consequent gain of information at each dose. Prior probabilities of Dose Limiting Toxicity (DLT) were assigned to each dose and a target DLT rate was elicited through clinical discussion.

Stage 1 adheres to an initial design prior to observing DLT, which is similar to a3+3 but with cohorts of up to 5 patients, allowing for an additional 2 patients whilst waiting for the next dose update. If a DLT is observed, stage 2 begins and the posterior DLT rate at each dose is determined in a Bayesian framework. Each dose decision will then depend on a holistic assessment of the model's recommendation, opinions of the Clinical Investigators and the Trials Steering Committee. We discuss how the trial progressed from the original to its current state, the feasibility of changing the trial design as well as the increased efficiency.

PC6 - W19: Combination of prior distributions elicited from expert opinions previous to Bayesian inference. Application to precision medicine

Authors: Fabien Subtil[1, 2], Angély Loubert[1, 2], Marie Laure Delignette-Muller[3, 4, 5], Muriel Rabilloud[1, 2].
[1] *Hospices Civils de Lyon, Service de Biostatistique, Université de Lyon,* [2] *CNRS, Umr5558, Laboratoire de Biométrie et Biologie Evolutive, Equipe Biostatistique-Santé, Lyon, France,* [3] *Université De Lyon,* [4] *CNRS, Umr5558, Laboratoire de Biométrie et Biologie Évolutive, Villeurbanne, France,* [5] *Vetagro Sup Campus Vétérinaire de Lyon, Marcy L'etoile, France.*

In clinical trials in rare diseases, it may be difficult to enroll enough subjects to conclude efficiently on the tested treatment efficiency. One possibility is to bring additional information through elicitation of expert opinions about the primary endpoint value for each arm of the trial. The information is combined with the trial results as a prior distribution in a Bayesian inference approach. The problem is finding a way of combining several distributions of the endpoint value from different experts to build a valid prior distribution. The work compares two ways: i) Mathematical combination methods using either linear or logarithmic pooling with potentially unequal weights, or quantiles aggregation; ii) Modelling with a supra-Bayesian approach. The latter approach assumes that the distribution of each expert is one form stemming from a single general distribution and takes into account the uncertainty of each expert, as well as a potential variability between experts (or experts group). The prior distributions obtained by the different methods are compared. Because of the lack of trial results at this stage, different scenarios of results are simulated to compare the hypothetical impact of each method on the conclusion of the clinical trial. This work uses a combination of eight distributions elicited from 8 physicians regarding the proportion of patients in non-progression after a six-week treatment (A vs. B) for an Ear-Nose-Throat epidermoid carcinoma.

PC6 - W22: Addressing measurement errors in multivariate continuous monotone disease processes

Authors: Lizbeth Naranjo[1], Emmanuel Lesaffre[2], Timothy Mutsvari[3], Carlos J. Pérez[4].
[1] *Department of Mathematics, Faculty of Sciences, Universidad Nacional Autónoma de México (UNAM), 04510 México D. F., México,* [2] *L-Biostat, School of Public Health, KU Leuven, 3000 Leuven, Belgium,* [3] *Arlenda S.A., Chaussée Verte 93, 4470 Saint-Georges-Sur-Meuse, Belgium,* [4] *Department Of Mathematics, Faculty Of Veterinary, Universidad De Extremadura, 10003 Cáceres, Spain.*

Motivated by a longitudinal oral health study, the Signal-Tandmobiel® study (Vanobbergen et al., 2000), an inhomogeneous mixed hidden Markov model with continuous state-space is proposed and discussed. In this Bayesian approach the true unobserved correlated response variables are subject to an unconstrained measurement error process and have a monotone non-decreasing behaviour. The proposed approach is different from the one proposed by García-Zattera et al. (2012), who modelled a multivariate monotone disease processes in the presence of misclassification assuming conditional independence on the manifest level. In our proposal, latent continuous variables are defined to model the caries experience process, and these are related to the covariates through a continuous monotone non-decreasing process. In addition, random effects are included to model

the relationship among teeth for each subject. The measurement error assumption is defined on these latent continuous variables by using the classical additive measurement error model. Besides, conditional independence assumptions are defined on the latent level. Finally, the binary observed responses, which are subject to misclassification, are related to these last latent continuous variables.

Reference 1: García-Zattera, M. J., Jara, A., Lesaffre, E., Marshall, G. (2012). Modeling of multivariate monotone disease processes in presence of misclassification. Journal of the American Statistical Association, 107(499):976-989.

Reference 2: Vanobbergen, J., Martens, L., Lesaffre, E., Declerck, D. (2000). The Signal-Tandmobiel® project, a longitudinal intervention health promotion study in Flanders (Belgium): baseline and first year results. European Journal of Paediatric Dentistry, 2:87-96, 2000.

PC6 - W25: Incorporating individual historical control and aggregated treatment effect into Bayesian survival clinical trial: a simulation study

Authors: Caroline Brard[1], Lisa V. Hampson[2, 3], Marie-Cécile Le Deley[1], Gwénaël Le Teuff[1].
[1]*Université Paris-Saclay, Université Paris-Sud, UVSQ, Cesp, Inserm, Gustave Roussy, France,* [2]*Medical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Fylde College, Lancaster University,* [3]*Statistical Innovation Group, Advanced Analytics Centre, Astrazeneca, United Kingdom.*

When historical data is available, Bayesian trial can be a good alternative or complement to the frequentist approach to augment the trial data, in particular in rare diseases. We designed a new randomised Phase 2 trial with a censored endpoint in high-risk osteosarcoma, with a planned sample size of 105 patients; historical data are available both regarding the control arm (individual patient data) and relative treatment effect (published hazard ratio, HR). We proposed an approach combining these historical data, using a non-adaptive power prior and a mixture prior, respectively. The operating characteristics of this approach were evaluated by extensive simulations assuming Weibull or piecewise exponential distributions for the future trial data. Different degrees of exchangeability of future versus historical data (anticipated, identical, disappointing scenarios) for (i) control arm and (ii) HR were investigated as well as the impact of a misspecification of survival modelling. Incorporation of historical data on the HR entails a large increase in power in all cases, but with no bias only in the 'identical' scenario. The additional gain in power brought by the incorporation of historical data on control arm is small and very dependent on the exchangeability of data. In case of misspecification of the survival modelling, the incorporation of historical data on control arm may lead to an important bias. In the context of rare disease, the use of a piecewise model to analyse data is a better option than a more simplistic model as it seems to better resists to misspecification.

PC6 - W28: Sequential Monte Carlo methods in Bayesian joint models for longitudinal and time-to-event data

Authors: Danilo Alvares[1], Carmen Armero[1], Anabel Forte[1], Nicolas Chopin[2].
[1]*University of Valencia, Spain,* [2]*Crest-Ensae And Hec Paris, France.*

In recent decades, the statistical models have greatly contributed to the comprehension of the risk of various diseases as well as to the connection between them and their behaviours/symptoms. A particular case is the joint models of longitudinal and time-to-event data, where both a repeatedly measured biomarker and the elapsed time to the event of interest are collected on each individual of the study.

A key characteristic of the learning process in this type of models is its dynamic nature. This is clear in biomedical studies where data usually come from individual follow- up over time: when new information of a given patient is collected, physicians are interested in updating the relevant estimated and/or predicted outcomes.

Focusing on this context, we propose a dynamic procedure for Bayesian joint models of longitudinal and time-to-event data. We rely on sequential Monte Carlo methods, in which our primary interest is to reduce the processing time of the inferential update after obtaining new data. Hence, this dynamic mechanism updates the posterior distributions when new data is available, taking into account all the previous information. Our approach is illustrated by means of a joint model with competing risk events in ICU patients receiving mechanical ventilation.

PC6 - W31: Joint modeling for longitudinal variations of blood pressure and circulatory disease incidence in association with radiation exposure

Authors: Kyoji Furukawa[1], Misumi Munechika[1], Ikuno Takahashi[1].
[1]*Radiation Effects Research Foundation, Japan.*

With increasing chances of exposure to radiation for medical purposes, health effects of radiation exposure have been of great public concern. Despite evidence being accumulated from epidemiological and clinical studies, the nature of radiation effects on circulatory diseases, the leading cause of death in many countries, is less clear compared to that on cancers. In particular, a hypothesis of our primary interest is that the increase in the radiation-associated circulatory diseases might be related to longitudinal variations of blood pressure, which are known to be important risk factors to circulatory diseases and also to be associated with radiation exposure. To investigate this hypothesis, the long-term clinical follow-up data of Japanese atomic-bomb survivors, one of the most informative sources to study late effects of radiation exposure, is particularly useful. In this study, we consider a joint modeling of longitudinal data of blood pressure and survival outcome of stroke incidence, which have been repeatedly measured through biennial clinical checkups for more than 50 years since 1950s. Jointly analysing the associations among radiation, longitudinal patterns of blood pressure, and incidence of stroke, we expect to elucidate new clinical implications about the mechanism about how radiation exposure affects blood pressure and the risk of circulatory diseases.

PC6 - W34: Joint modelling for flexible multivariate longitudinal and survival data: a two-stage model based proposal

Authors: Francisco Gude[3], Ipek Guler Caamaño[1], Christel Faes[2], Carmen Cadarso-Suárez[1].
[1]*Center for Research in Molecular Medicine and Chronic Diseases (CIMUS), University of Santiago de Compostela, Santiago de Compostela, A Coruna, Spain,* [2]*I-Biostat, Hasselt University, Diepenbeek, Belgium,* [3]*Clinical Epidemiology Unit. Hospital Clinico Santiago de Compostela, Spain.*

Many follow-up studies in biomedical research typically produce different types of outcomes including both longitudinal biomarkers and time-to-event outcomes. Often, the interest is on assessing the relationship between the longitudinal and the time-to-event processes. Joint modelling approaches of longitudinal and survival data have gained an increasing attention recently. Various authors have proposed joint modelling for a single longitudinal marker and a time-to-event process. However, in many studies several longitudinal biomarkers are of interest and instead of selecting one single biomarker, the relationships between all these outcomes and their association with survival needs to be investigated. Furthermore, flexible regression techniques may be necessary for the non-linear longitudinal trends. Existing joint modelling approaches are difficult to implement when the number of longitudinal biomarkers is large and the longitudinal profiles are non-linear due to the high dimensional complexities. We propose a two-stage based modelling approach for flexible modelling of multivariate longitudinal and survival data. The model proposal is applied to a real data example on Orthotopic Liver Transplantation (OLT) study. For the model validation, a simulation study was implemented followed by the case study.

PC6 - W37: Validation of the French version of Conners' Parent Rating Scale Revised, Short Version (CPRS-R:S). Scale measurement invariance by sex and age

Authors: Pascal Roy[1], Catherine Mercier[1], Sylvain Roche[1], Pierre Fumeaux[2], Jean Iwaz[1], Michel Bader[3], Philippe Stephan[4], René Ecochard[1], Olivier Revol[1].
[1]*Hospices Civils de Lyon, France,* [2]*Cabinet De Pédopsychiatrie Et Neuropsychologie, Switzerland,* [3]*Centre Hospitalier Universitaire Vaudois, Switzerland,* [4]*Université de Lausanne, Switzerland*

Objective: Attention deficit hyperactivity disorder is one of the most frequent neurodevelopmental disorders. In addition to clinical assessment, its diagnostic requires the use of validated and reliable behaviour questionnaires such as the Conners' Parent Rating Scale-Revised: Short Form (CPRS-R:S). Though various French versions of the CPRS-R:S have been already put to use in clinical practice and research, only a few have undergo a stringent validation process. After a previous validation of the factorial structure of Lausanne French version, we sought here for the analysis of its invariance across sex and age.

Method: This validation step was carried out in a rather homogeneous French population of 365 boys and 374 girls from a single school. Two age classes were considered: children (boys and girls) aged 9 to 11 years (n=258) and adolescents aged 12 to 15 years (n=481).

Results: Regarding age, dimension Oppositional showed a strong invariance whereas dimensions Hyperactivity and Cognitive problems/Inattention showed a partially strong invariance. Regarding sex, dimensions Oppositional and Hyperactivity showed a partially strong invariance whereas dimension Cognitive problems/Inattention showed a partially weak invariance. The distribution of the CPRS-R:S scores is given by sex and age-class.

Conclusions: The Lausanne French version of the CPRS-R:S, already validated regarding its factorial structure, internal consistency, and reliablility, is here validated regarding its invariance across sex and age. Caution should be taken in using dimension Cognitive problems/Inattention in comparisons between boys and girls.

PC6 - W40: Joint regression modeling of radiation effects on longitudinal changes in red blood cell measurements and mortality

Authors: Munechika Misumi[1], Kengo Yoshida[1], Michiko Yamada[1], Yoichiro Kusunoki[1].
[1] *Radiation Effects Research Foundation, Japan.*

Data from atomic bomb survivors are important resource for investigating the health effects of radiation exposure. Participants in the Adult Health Study of atomic bomb survivors (Hiroshima and Nagasaki, Japan, since 1958) have contributed blood samples and clinical information at biennial health examinations. Their radiation dose was estimated by the DS02 dosimetry system. Previous research has suggested that radiation exposure has deleterious effects on blood cell measurements. Red blood cell distribution width (RDW) is a blood cell measure assessing variability in the size of circulating red blood cells, and its increase is known to be associated with elevated mortality and morbidity of several diseases such as ischemic stroke, cancers, anemia, etc., in the general population. However, association between radiation exposure and longitudinal RDW change is yet to be evaluated, and the investigation requires careful consideration of the fact that radiation exposure is associated with mortality and other diseases. Also, the increase of RDW associated with some disease should be taken into consideration. In this study, we apply joint regression models to estimate the association between radiation dose and longitudinal changes in RDW, while accounting for informative censoring due to death. In addition, possible effect modification of association between radiation dose and mortality by RDW change is investigated to elucidate the possible involvement of RDW in the development of radiation-associated diseases.

PC6 - W43: Joint model for longitudinal marker and time-to-event in presence of competing risks: application on cystic fibrosis

Authors: Lionelle Nkam[1], Mounia N. Hocine[1].
[1]*Conservatoire National des Arts et Métiers, France.*

Forced Expiratory Volume in on second (FEV1) is an important factor in assessing Cystic Fibrosis (CF) severity. The analysis of this marker, which dynamically changes over time, allows us to describe the progression of the disease and anticipate the occurrence of clinical events.

The aim of this study was to develop a model which provides dynamic predictions of cause-specific hazard (CSH) of lung transplantation (LT) and CSH of death taking into account the evolution of FEV1.

We developed a joint latent class model which assumes that the population can be divided into G latent homogeneous subgroups. We choose the model with 3 latent classes based on BIC criteria and computed subject-specific predictions at various landmark times with a prediction window of 3 years.

A total of 1625 adults from the French CF Registry were included in the study, over the period 2007-2013. A total of 86 (5.3%) patients died and 277 (17%) patients received a lung transplant during this period. The developed joint model identified three evolution profiles: a "low risk", a "moderate risk" and a "high risk" profile representing respectively 10%, 33% and 57% of the subjects. The estimated AUCs corresponding to the prediction model were high (ranged from 0.7 to 0.9 for LT and ranged from 0.6 to 0.8 for death). The estimated Brier scores were low (values less than 0.1) for predictions of CSH of LT and CSH of death.

The developed joint model showed good predictive accuracy in terms of discrimination and calibration. The dynamic individual predictions obtained may be useful for clinicians in identifying patients eligible for LT, which is proposed to CF patients with terminal respiratory failure.

PC6 - W46: Joint longitudinal and survival-cure models to improve the assessment of being cured in oncology studies

Authors: Antoine Barbieri[1], Catherine Legrand[1].
[1] *Université Catholique de Louvain, Belgium.*

In the cancer framework, a number of studies such as certain relating to breast and prostate cancers actually include two kinds of patients: those who will not experience the event of interest (e.g. clinical recurrence, time to progression of cancer...) and are said to be "cured", and those who will develop the event, and are said to be "susceptible". However, the cure status is unobserved in (right-)censored patients. While most of the work on cure models focus on the time-to-event for the uncured patients (latency) or on the baseline probability to cure or not (incidence), we focus in this research on the conditional probability of being cured after a medical intervention (surgery, chemotherapy, radiotherapy...) given that a patient or a profile of patients survived until a certain time. An important feature of the cancer studies is the availability of longitudinal measurements collected over the follow-up such as, biomarkers, size of tumor, or quality of life. We therefore consider a joint model for the longitudinal and survival data given a cure fraction. This model includes a mixed model to fit the trajectory of longitudinal measurements and a mixture cure model. Different share latent structures to link both models are compared in order to improve the analysis and the computation of the conditional probability of interest.

PC6 - W49: Joint modelling of longitudinal and repeated event data using nonlinear mixed effect models- an application in evaluation of treatment of Gaucher disease

Authors: Marie Vigan[1], Jerome Stirnemann[2].
[1] *Debrc, GH Hupnvs, France Mentré (Inserm, Iame, Umr 1137, F-75018 Paris, France. Université Paris Diderot, Iame, Umr 1137, Sorbonne Paris Cité, F-75018 Paris, France),* [2] *Médecine Interne Générale, Hôpitaux Universitaires de Genève, Rue Gabrielle-Perret-Gentil 4, Ch-1211 Genève 14, Suisse.*

Gaucher disease (GD) is a rare recessively inherited disorder due to the deficiency of lysosomal enzyme glucocerebrosidase. Several biomarkers (ferritin, chitotriosidase, hemoglobin, platelets) are significantly increased during this disease. Bone events (BE: Avascular necrosis, bone infarct or pathological fracture) are important complications of GD. GD can be treated by enzyme replacement therapy (ERT), imiglucerase. The aim was to analyse the link between evolution of biomarker concentrations and repeated BE in treated patients.

We analysed patients from the French Registry of GD, treated by ERT. All measurements from initiation of ERT to discontinuation or the end of follow-up were used. We used a nonlinear mixed effects model to analyse longitudinal biomarkers concentration and a frailty model to analyse repeated BE. We developed a joint model, to analyse link between biomarkers and BE. Several covariates were tested including age at initiation of ERT, splenectomy and sex. Estimations were performed with the SAEM algorithm implemented in MONOLIX. Likelihood was evaluated by importance sampling.

Data were available for 156, 169, 210 and 211 patients with a total number of 241 patients and a total number of observations of 859, 762, 2197, and 2251 for ferritin, chitotriosidase, hemoglobin and platelets, respectively. Forty-five patients had a BE with a total number of 65. Median time of follow-up was 9 [1-19] years. Estimated half-life of normalization of biomarkers was 0.8 years. Having BE before ERT increased three-fold the risk under ERT. If the concentration of chitotriosidase is doubled, the risk of BE is multiplied by 1.1.

This is the first study who study the link of biomarkers and repeated BE in GD.

PC6 - W52: A Bayesian joint model of longitudinal and competing risks data for assessing cardiovascular risk in chronic kidney disease

Authors: Carles Forné[1], Montse Rué[1, 2], Danilo Alvares[3], Elvira Fernández[4], José M. Valdivielso[4].
[1] *University of Lleida-Irblleida, Department of Basic Medical Sciences - Biostatistics Unit, Av. Rovira Roure 80, 25198-Lleida, Spain,* [2] *Health Services Research Network in Chronic Diseases (REDISSEC), Spain,* [3] *University of Valencia, Department of Statistics and Operational Research, Doctor Moliner 50, 46100-Burjassot, Spain,* [4] *Vascular and Renal Translational Research Group, Irblleida, Lleida, Spain.*

Joint models of longitudinal and time-to-event data can be used for assessing the association between the trajectories of longitudinal markers and the risk of events. In the present study we propose a Bayesian joint model with a longitudinal non-negative outcome and three competing

events. In particular, we consider a mixed effects gamma regression model for the longitudinal outcome and a cause-specific hazards model for the time-to-event outcomes.

The model is applied to assess the relationship between changes in subclinical atheromatosis disease and cardiovascular risk in 2445 patients with chronic kidney disease, aged 18-75 years enrolled from 81 Spanish hospitals between October 2009 and June 2011. Competing events to the event of interest are non-cardiovascular death and kidney transplant. The longitudinal marker is the number of atheromatous plaques detected by ultrasound in ten arterial territories (carotid and femoral arteries). Markov Chain Monte Carlo methods are used to approximate the posterior distribution by means of the R and the JAGS software.

PC6 - W55: Longitudinal discriminant analysis for predicting preeclampsia using a joint mixed effects model

Authors: Malihe Nasiri[1], Farid Zayeri[2], Noorosadat Kariman[3], Nastaran Safavi Ardebili[4], Massume Khosravi[5].
[1] *Faculty of Nursing and Midwifery, Shahid Beheshti University of Medical Sciences, Tehran, Iran,* [2] *Department of Biostatistics, Faculty of Paramedical, Shahid Beheshti University of Medical Sciences, Tehran, Iran,* [3] *Department of Midwifery, Faculty of Nursing and Midwifery, Shahid Beheshti University of Medical Sciences, Tehran, Iran,* [4] *Department of Midwifery, Faculty of Medical Sciences, Islamic Azad University, Ardebil, Iran,* [5] *Shiraz University Of Medical Sciences, Shiraz, Iran.*

In diagnostic medical research, discriminant analysis is usually applied to evaluate the ability of biomarkers to separate patients into pre-defined groups. In our data set, the hemoglobin and hematocrit levels (two markers) and the diabetes were measured for 650 women in three trimesters and preeclampsia was the main outcome under study. We employed two models for classifying the women into groups (with or without preeclampsia). In the first model, a random subject effect and the Kroneker product of the correlation matrix of the repeated measures and the correlation matrix between two markers were assumed for the variance-covariance pattern. Also the Kroneker product of matrix of time dependent covariates and the matrix of time trend was supposed for the mean. In this model the diabetes was assumed as a time dependent covariate. In the second model, the responses that we analyze are the vectors of time varying hemoglobin, hematocrit and diabetes measurements. We proposed a linear mixed and a nonlinear mixed model to analyze the evaluation of hemoglobin, hematocrit and diabetes responses, respectively. Three response trajectories are tied together through a joint distribution for the random effects. In this model, the diabetes was assumed as a separate binary response that is related to preeclampsia and two markers, not a time dependent covariate. The present study aimed to predict preeclampsia by the hemoglobin and hematocrit profiles and the diabetes, through longitudinal discriminant analysis and comparing the error rate of discrimination in two models. Statistical analyses were performed using the SAS software version 9.2.

PC6 - W58: Population pharmacokinetics of cyclosporine using nonlinear mixed-effects modelling and Bayesian estimation in renal transplant patients

Authors: M. Salomé Cabral[1], A. Sofia Cardoso[2], A. Paula Carrondo[2,3], José Guerra[4].
[1] *Ceaul and Deio, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal,* [2] *Serviço de Gestão Técnico-Farmacêutica, Chln-Hospital de Santa Maria, Lisboa, Portugal,* [3] *Faculdade de Farmácia da Universidade de Lisboa, Lisboa, Portugal,* [4] *Serviço De Nefrologia E Transplantação Renal, Chln-Hospital De Santa Maria, Lisboa, Portugal.*

Background: Cyclosporine (CyA) is an immunosuppressive drug with a narrow therapeutic window and a large pharmacokinetic (PK) variability that makes therapeutic drug monitoring (TDM) indispensable.

Purpose: To develop a population PK model for cyclosporine, using nonlinear mixed-effects models and Bayesian approaches, to be applied in clinical practice, using late renal transplant Portuguese patients.

Methods: The routine monitoring data of CyA were retrospectively collected from 104 renal transplant patients (682 steady-state blood concentrations). Routine data consisted on sampling time at predose, 1, 2, 3 and 4h after oral CyA administration. A nonlinear mixed-effects modelling was performed on 543 concentrations (82 patients) with a one compartment open model and absorption constant rate (Ka) fixed. The nlme function in S-Plus software was used. The model was validated by internal (data splitting and jack-knife) and external methods (22 patients, 139 concentrations).

Results: The population PK parameters estimated were oral clearance (Cl/F)=34.5 l/h and oral volume of distribution (V/F)=81.3 l with 22.2% and 41.5% of interindividual variability, respectively. Covariate screening reveals that Cl/F increased with body surface area (BSA), and decreased with age and delayed sampling time (Spt). Validation demonstrated that model was robust and could predict CyA concentrations with a mean prediction error of 21.3 ng/ml.

Conclusions: This population PK model showed a reasonable predictability and can be applied to CyA dosing optimization in late stage renal transplant recipients by using conventional TDM data. Clinical model evaluation and further research in early stage of transplant are still required.

PC6 - W61: Trajectories of eGFR 6 months to 7 years after hospital discharge in critically ill patients: group-based trajectory modeling analysis

Authors: Salma Ayis[1], Johna Powell-Tuck[1], Hugh Leonard[1], Ryan Haines[1], Marlies Ostermannn[1].
[1] *King's College London, UK.*

Objective: Worsening of estimated glomerular filtration rate (eGFR) is associated with higher risk of end stage renal disease (ESRD) and mortality. Little is known about the development of eGFR over time, particularly in critically ill patients with acute kidney injury (AKI). We aim to detect eGFR growth patterns in these patients, long term after hospital discharge.

Methods: Data on 1,223 critically ill patients admitted to an ICU in a University Hospital between 2004-2008 were collected and follow-up data obtained up to June 2016. The eGFR was assessed at 6 months, then annually up to 7 years after discharge. The primary analysis comprised 427 (63% males) patients who have at least 4 measures of eGFR. Group based trajectory methods (GBTM) based on the software STATA (14.0) were used.

Results: Four trajectories of eGFR were identified: Persistently low (PL) 10.1% of patients; Rapidly increasing (RI) 3.98%; Persistently moderate (PM), 79.16% and Rapidly dropping (RD) 6.79%. These differ significantly in demographic and clinical characteristics. PL and RD, have higher risk of ESRD, 69.77% and 34.48% respectively. The corresponding proportions for RI and PM, were none and 0.59% respectively. The prevalence of AKI stage III in the PL and RD trajectories, were, 53.49% and 31.03%, and in the RI, and PM these were 17.65% and 25.15% respectively. Sensitivity analyses of different inclusion criteria were done to assess the robustness of the findings. The 4 trajectories' solution provided good fit in all situations.

Conclusions: Clinically meaningful trajectories of eGRF were identified. Such information may help to guide the implementation of cost effective interventions and the long term management of critically ill patients.

Plenary Speaker

Keynote Speaker

Wednesday 12th July - 11.00-12.00 h. - Room: Auditorio
Chair: Guadalupe Gómez Melis

Model Uncertainty and Covariate Selection in Causal Inference

Francesca Dominici, *Harvard University, USA*

Dr. Francesca Dominici is Professor of Biostatistics, Senior Associate Dean for Research, and Associate Dean of Information Technology at the Harvard T.H. Chan School of Public Health (USA). Her research focuses on the development of statistical methods for the analysis of large and complex data. She leads several interdisciplinary groups of scientists with the ultimate goal of addressing important questions in environmental health science, climate change, comparative effectiveness research, and health policy.

Abstract: Researchers are being challenged with decisions on how to control for a high dimensional set of potential confounders in the context of a single binary treatment (e.g, drug) and in the context of a multivariate exposure vector with continuous agents and their interactions (e,g, exposure to mixtures). Typically, for a binary treatment, a propensity score model is used to adjust for confounding, while the uncertainty surrounding the procedure to arrive at this propensity score model is often ignored. Failure to include even one important confounder will results in bias. We discuss how to overcome issue of confounding selection and model uncertainty in causal inference. Specifically, we introduce the model averaged double robust (MA-DR) estimator, which accounts for model uncertainty in both the propensity score and outcome model through the use of model averaging. We also consider estimating the effect of a multivariate exposure that includes several continuous agents and their interactions when the true confounding variables are an unknown subset of a potentially large (relative to sample size) set of measured variables. We develop a new approach rooted in the ideas of bayesian model averaging to prioritize confounders among a high-dimensional set of measured covariates. We introduce a data-driven, informative prior that assigns to likely confounders a higher probability of being included into a regression model for effect estimation. We illustrate the performance of these estimators and applications to comparative effectiveness research and environmental problems.

Invited Session

IS7: Bayesian Methods in Clinical Research

Wednesday 12th July - 14.30-16.00 h. - Room: Auditorio
Chair: Carmen Armero
Organised by Carmen Armero, *University of Valencia, Spain*

IS7-1: Bayesian Evidence Synthesis for Extrapolation in Clinical Research

Heinz Schmidli, *Novartis Pharma AG, Switzerland*

Extrapolation from a source population to a target population has recently received much interest in clinical research. Examples include extrapolation of treatment effects from adults to children, from one ethnic group to another, or from early to late disease stage. Bayesian approaches seem particularly attractive in this context, as data from the source population can be summarized and used as prior information to be combined with data from the target population. For a robust extrapolation, the possibility that the source information may not be relevant has to be taken into account. Examples from various settings in clinical research will be used to illustrate the framework.

Reference 1: Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics 2014 70(4):1023-32.

Reference 2: Schmidli H, Wandel S, Neuenschwander B. The network meta-analytic-predictive approach to non-inferiority trials. Statistical Methods in Medical Research 2013, 22(2):219-240.

IS7-2: A Population-Finding Design with Covariate-Dependent Random Partitions

Peter Müller, *University of Texas at Austin, USA*

Targeted therapies on the basis of genomic aberrations analysis of thetumor have become a mainstream direction of cancer prognosis andtreatment. Regardless of cancer type, trials that match patients totargeted therapies for their particular genomic aberrations, are wellmotivated. Therefore, finding the subpopulation of patients who can most benefitfrom an aberration-specific targeted therapy across multiple cancertypes is important. We propose an adaptive Bayesian clinical trialdesign for patient allocation and subpopulation identification. We startwith a decision theoretic approach, including a utility functionand a probability model across all possible subpopulation models. The mainfeatures of the proposed design and population finding methods arethat we allow for variable sets of covariates to be recorded bydifferent patients, adjust for missing data, allow high orderinteractions of covariates, and the adaptive allocation of each patient totreatment arms using the posterior predictive probability ofwhich arm is best for each patient. The new method is demonstrated via extensive simulation studies.

IS7-3: Use of Bayesian Approaches for Extrapolating from Adult Efficacy Data to Design and Interpret Confirmatory Trials in Children

Lisa Hampson, *Lancaster University, United Kingdom*

New medicines for children should be subject to rigorous examination whilst taking steps to avoid unnecessary experimentation. Extrapolating from adult data can reduce uncertainty about a drug's effects in younger patients meaning smaller trials may suffice.

We consider how to design a confirmatory trial in children intended to compare the efficacy of a new drug, E, against control. Assuming that conduct of this trial is conditional on having demonstrated a significant beneficial effect in adults, we adopt a Bayesian approach to incorporate these adult data into the design and analysis of the paediatric trial. At each stage, inferences are made using all available data to update a Bayesian mixture model for prior opinion on the degree of similarities between adults and children. Using this framework, we propose designs for the paediatric trial which are specified by calibrating the sample size and final decision rule to: a) achieve a high frequentist power and high minimum (or average) Bayesian positive predictive value of a significant result in children; or b) ensure that a final decision to adopt (abandon) drug E in children is always associated with a minimum positive (negative) predictive value. Operating characteristics of our Bayesian designs are evaluated and compared with those of a recently proposed hybrid approach (Hlavin et al. Statistics in Medicine 2016; 35: 2117) where the sample size and significance level of a frequentist confirmatory trial in children are set to achieve a high frequentist power and high average positive predictive value of a significant result in children.

Oral Contributed Sessions

OC35: Complex survival data 2

Wednesday 12th July -14.30-16.00 h. - Room: Sala Terra 2
Chair: Hans Van Houwelingen

OC35-1: Flexible modelling of personalised dynamic prediction curves using landmarking, with a case-study in cystic fibrosis

Authors: Ruth Keogh[1], Jessica Barrett[2], Shaun Seaman[3], Rhonda Szczesniak[4], Angela Wood[2].
[1] *LSHTM, UK,* [2] *University of Cambridge, UK,* [3] *MRC Biostatistics Unit, UK,* [4] *Cincinnati Children's Hospital Medical Center, US.*

In 'dynamic' prediction of survival we make updated predictions of individuals' survival over time as new information becomes available about their health status via longitudinal measures. Landmarking is an attractive and flexible method for making dynamic predictions. Applications of landmarking have typically focused on estimating the probability of survival to time t+w given survival to time t (the 'landmark'), using a person's health status data up to t. To obtain such predictions, Cox models are fitted from time t, with censoring enforced at t+w. However, interest often lies in the entire survival curve conditional on survival to time t, rather than in survival to a single time point t+w. We show how 'dynamic prediction curves' can be used in landmarking. Use of Cox models and Royston-Parmar flexible parametric survival models will be discussed, including use of time-varying effects of predictors. Estimation is based on a 2-stage procedure. First, mixed models are fitted to the longitudinal data up to each landmark t. Second, predicted values from the mixed model are used as predictors in a flexible survival model conditional on survival to time t. We also show how the models for dynamic prediction curves can be combined across landmark times and how to obtain confidence intervals for the curves.

The methods will be discussed with reference to a case-study in cystic fibrosis (CF). The aim was to obtain personalised dynamic prediction curves for survival in CF from a series of landmark ages using longitudinal health-status data from the UK CF Registry. It will also be shown how we address some of the obstacles arising in this study, including occurrence of intermediate events and missing data on predictors.

OC35-2: Joint modelling of progression-free and overall survival

Authors: Kaspar Rufibach[1], Matthias Meller[2,3], Jan Beyersmann[4].
[1] *Department of Biostatistics, F. Hoffmann-La Roche, Basel,* [2] *Ams Advanced Medical Services, Germany,* [3] *F. Hoffmann-La Roche, Basel,* [4] *Institue of Statistics, University of Ulm, Germany.*

Progression-free survival (PFS) is a commonly used surrogate endpoint in oncology trials. Quantities of interest in this context are the correlation coefficient between PFS and OS as well as the survival function for OS. Fleischer et al. (2009) and Li and Zhang (2015) use a latent-time illness-death model without recovery to jointly model PFS and OS and based on this model, derive parametric point estimates for the correlation between PFS and OS and the survival function of OS. They either assume exponential or Weibull transition hazards with a common shape parameter. We generalize their approach by showing that the latent time assumption is not necessary, derive parametric and

nonparametric point estimates as well as inference methods for the transition hazards, the correlation between PFS and OS, as well as the survival function of OS. We do this by relaxing the equal shape parameter assumption and under various assumptions on the stochastic process underpinning the multistate model, namely time-homogeneous and non-homogeneous Markov as well as non-Markov. Our results shed light on the implicit assumptions in Fleischer et al (2009) and Li and Zhang (2015). The methods are illustrated using a large Phase 3 oncology clinical trial.

Reference 1: Fleischer, F., Gaschler-Markefski, B. and Bluhmki, E. (2009). A statistical model for the dependence between progression-free survival and overall survival. Stat. Med. 28 2669-2686.

Reference 2: Li, Y. and Zhang, Q. (2015). A weibull multi-state model for the dependence of progression-free survival and overall survival. Statistics in medicine 34 2497-2513.

OC35-3: A ROC curve approach for cure status prediction from survival data

Authors: Mailis Amico[1], Ingrid Van Keilegom[1].
[1] *KU Leuven, Belgium.*

In medical studies (as in other fields), the classical survival analysis assumption standing that the event of interest will eventually be observed for all observations if the follow-up will be infinite is often not realistic. When interest lies in the time until a relapse from breast cancer or the time until the occurrence of a certain disease for example, a fraction of the patients may never experience the event of interest, and the survival data can contain a 'cure' fraction associated with infinite survival times. A common approach to analyse this type of data consists in using cure models. Two types of information can then be drawn from the data: the survival for the uncured observations and the cure status, both possibly modelled as functions of the covariates.

The cure status is often of interest for medical practitioners, and one is usually interested in predicting it based on markers. ROC curves are one way to evaluate the predicting performance of cure models for the cure status. However, the classical ROC curve approach is not appropriate because the cure status is only partially observed due to censoring. In this research, we propose a ROC curve methodology to evaluate the classification performance of cure models to predict the cure status. We develop an estimator by means of Bayes' theorem where the quantities of interest are estimated by means of weighted empirical distribution functions, where the weights are expressed using the mixture cure model. Based on simulations, we demonstrate the performance of the proposed method and compare it with the estimator obtained when the cure status is fully observed. Finally, we illustrate the methodology on a cancer data set.

OC35-4: Comparison of net survival with non-proportional hazard functions: a test based on restricted net survival time

Authors: Anna Wolski[1], Nathalie Graffféo[2, 3], Leyla Azarang[1], Roch Giorgi[1].
[1] *Aix Marseille University, Inserm, Ird, Sesstim, Sciences Economiques & Sociales de la Santé & Traitement de L'Information Médicale, Marseille, France,* [2] *Inserm U1153, Statistic and Epidemiologic Research Center Sorbonne Paris Cité (CRESS), Ecstra Team, Saint-Louis Hospital, 75010 Paris, France,* [3] *Paris Diderot University, Paris 7, Spc, Paris, France.*

Net survival, the hypothetical survival that would be observed if a particular disease were the only cause of death, is a proper measure for assessing excess mortality in population-based studies. Because it is independent of the general mortality, net survival enables comparisons through time and space and provides a useful tool in the evaluation of health policies. Only recently, a consistent estimator of net survival – the Pohar Perme estimator – has been proposed in the non-parametric framework, followed closely by the derivation of a suited log-rank type test.However, it is widely accepted that log-rank tests, though optimal under proportional hazards, perform poorly in extreme cases of non-proportionality, e.g. when the hazard functions cross. The hazard ratio is then no longer an appropriate measure of the difference, and an advocated alternative approach, in the general survival setting, is to estimate the restricted mean survival benefit. The latter measure is of particular interest since it is evaluated geometrically as the area between two survival curves, and is also a relevant clinical indicator of the life gain over a specified period of one group versus the other.Remaining in the non-parametric setting, we use as an initial approach between-group permutations to extend the idea of a restricted mean survival-based test to net survival. We generalize the test to more than two groups. Through extensive simulation of datasets where the proportional hazard assumption does not hold at various degrees, we assess its power and efficiency, and compare its performance to the previously derived log-rank approach. Finally, we propose an illustration on real data.

OC35-5: Improving the testing of treatment effect in clinical trials with time to event outcomes

Authors: Song Yang[1].
[1]*National Heart, Lung, And Blood Institute, Nih. USA.*

The log-rank test may be the most widely used tool for testing treatment effect with survival data, due to its optimality under proportional hazards alternatives and its robustness under mild departures from the proportional hazards assumption. However, it can have substantial power loss in non-proportional situations.

There are a variety of situations when the hazards are highly non-proportional: the treatment may bring a short-term benefit, and gradually lose its effect as time goes on; an aggressive treatment may provide a long term benefit while giving higher mortality early on due to toxicity or complications. In general, the longer the follow-up period is, the more likely it is for various non-proportional scenarios to develop.

In the literature, most of the alternative tests proposed so far more or less focus on certain specific alternative hypotheses. They may be more powerful than the log-rank test under those situations considered, but are no longer optimal under the proportional hazards assumption. Is it possible to get the best of both worlds? We discuss an approach that improves the log-rank test using weighted log-rank statistics with adaptive weights. The resulting test remains optimal under the proportional hazards assumption. To control the size while maintaining good power, we use a resampling method that leads to more power than the log-rank test in most cases in a wide range of non-proportional hazards scenarios. We also consider the issues related to use of the new test for interim looks during the process of monitoring the trials. When applied to clinical trial data, the proposed methods often lead to earlier stopping of the trial or a smaller p value.

OC36: Statistical methods for precision medicine 2

Wednesday 12th July -14.30-16.00 h. - Room: Sala Mar 4
Chair: Tim Friede

OC36-1: Point and interval estimation for predicting individual treatment effect based on randomized clinical trial data

Authors: Kukatharmini Tharmaratnam[1], Thomas Jaki[1].
[1] *Lancaster University, United Kingdom.*

Individuals within a population are typically heterogeneous. Characteristics such as gender, genetic variations, disease etiology and severity vary between individuals. These different patient characteristics potentially affect the response to treatment. Treatment effectiveness is typically assessed using the average treatment effect or (occasionally) the treatment effect within (prespecified) subgroups. Recently developed approaches allow researchers to predict individual's response to treatment allowing individualized approaches to medical treatment instead of relying on averages from a group or subgroup. In this paper we estimate the Predicted Individual Treatment Effects (PITE) utilizing multiple imputation and investigate different variable selection procedures to obtain a parsimonious PITE model that is robust in predicting treatment response of individuals within and outside the original study. We subsequently derive prediction intervals for individual patients PITE. Simulations show that the prediction intervals have adequate coverage when either the BIC or Lasso is used for variable selection. Our proposed approach to find the prediction interval for PITE performs well in the simulation studies and data example from ALS clinical trials. We could use other type of response variables and covariates in the model to estimate PITE and also we may use interaction models and more complex mixed models to get PITE.

Reference 1: Lamont, A. E., Lyons, M. D., Jaki, T. F., Stuart, E. A., Feaster, D., Tharmaratnam, K., Oberski, D., Ishwaran, H., Wilson, D. K., and Horn, M. V. (2016). Identification of predicted individual treatment effects (PITE) in Randomized clinical trials. Statistical Methods in Medical Research, E-pub ahead of print.

OC36-2: Approaches to incorporate drug-drug similarity in multi-response penalized likelihood methods for predicting drug response based on multi-omics data

Authors: Zhi Zhao[1], Manuela Zucknick[1], Arnoldo Frigessi[1].
[1] *University of Oslo, Norway.*

It is important to find efficient molecular targets for personalized cancer therapy based on large-scale DNA sequencing data and other molecular data sources (e.g., gene expression, methylation, mutation). To model multivariate pharmacological sensitivity of compounds and distinguish different sources of molecular data, we use or extend six current multi-response penalized methods for predicting response of multiple drugs. In simulation studies, standard multivariate lasso (and elastic-net) regressions cannot predict drug response well, since they cannot account for the drug-drug similarity or for different sources of features. Here, we focus on the lasso penalty, but all approaches can be extended to other penalties. Although Integrative LASSO with Penalty Factors (IPF-LASSO; Boulesteix et al., 2015) assigns different penalties to different sources of features, the dependence between drugs is not considered. So we combine the IPF technique with tree-lasso (Kim & Xing, 2012) which can capture hierarchical structure of drugs, as well as with spatial regression (Lam & Souza, 2014) that models spatial dependencies between drugs by regarding drug responses as covariates. We demonstrate that Spatial-IPF-LASSO outperforms other methods in predicting drug response under correlated multivariate responses in our simulation, and illustrate it with an application to the Cancer Cell Line Encyclopedia data.

Reference 1: Boulesteix AL, De Bin R, Jiang X, et al. (2015). IPF-LASSO: integrative L1-penalized regression with penalty factors for prediction based on multi-omics data. Technical Report 187, Department of Statistics, LMU.

Reference 2: Lam C, Souza P (2014). Regularization for spatial panel time series using the adaptive lasso. Econometrics Paper Series.

OC36-3: Development of treatment-selection markers for studies with failure-time outcomes

Authors: Erik Van Werkhoven[1], Parvin Tajik[2], Patrick Bossuyt[2], Koos Zwinderman[2].
[1] *The Netherlands Cancer Institute,* [2] *University of Amsterdam, The Netherlands.*

Biomarkers can be used to predict treatment benefit for individual patients, which may facilitate treatment selection. Methods for developing and evaluating such predictive biomarkers typically use data from randomized trials with a binary endpoint. We propose a method for developing predictive biomarkers from non-randomized studies in which the outcome of interest is a failure time.We suggest to use proportional hazards regression models with a cross-validated penalty to select predictive markers from the candidate variables. We introduce summary measures to evaluate the benefit of the (multi)marker-based strategy compared to the strategy of treating all patients with the treatment which is best for an average patient. The benefit can be measured by the increase in restricted mean survival time (RMST) and the number of events prevented. The RMST was obtained

from the models using the Breslow estimator of the baseline hazard. The expected number of events prevented was estimated using the cumulative hazard. Inverse probability of treatment weighting (IPTW) can be used for non-randomized treatment comparisons, and multiple imputation to take care of missing data. As an illustration, the procedures were applied to a large non-randomized study. The validity of the methods was evaluated using a simulated dataset, closely representing this study.

OC36-4: Evaluating treatment effect modification on the additive scale to investigate predictors of differential treatment response

Authors: Antonia Marsden[1], Graham Dunn[1], William Dixon[2], Richard Emsley[1].
[1] *Centre for Biostatistics, School of Health Sciences, The University of Manchester, Manchester Academic Health Science Centre, UK,* [2] *Arthritis Research UK Centre for Epidemiology, Centre for Musculoskeletal Research, Institute for Inflammation and Repair, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK.*

The success of personalised medicine relies on the existence of predictive markers, defined as factors associated with differential response to treatment. Differential response is most appropriately assessed by evaluating treatment effect modification on the additive measurement scale; this corresponds to a comparison of absolute treatment effects across patient subgroups. However, it is not uncommon for treatment effect modification to be only assessed and presented on the multiplicative measurement scale when a multiplicative regression model has been used for a statistical analysis. A multiplicative regression model is typically used for analysis of binary, count and time-to-event outcomes.

This work proposes and summarises practical guidance regarding the evaluation of treatment effect modification on the additive measurement scale relating to data where multiplicative regression models are most commonly fitted.

For binary and count outcomes, the additive interaction effect measure between treatment and a potential moderator cannot be directly estimated from the logistic or Poisson regression model when additional model covariates are included in the model, but one can average over the effects of any additional variables to obtain an estimate of the average marginal additive interaction effect. The assessment of effect modification on the additive scale in survival data is more complex due to the dependency on time. We propose the use of an original measure, the ratio of absolute effects, which can be calculated as a function of the regression coefficients from the Cox proportional hazards regression model. The use of additive regression models for binary, count and survival data will also be discussed.

OC36-5: Pathway-based approach to NGS data with multiple phenotypes

Authors: Taesung Park[1], Sungyoung Lee[1], Sungkyoung Choi[1], Seungyeoun Lee[2], Heungsun Hwang[3].
[1] *Seoul National University, Korea,* [2] *Sejong University, Korea,* [3] *McGill University, Canada.*

As one of possible solutions to solve the "missing heritability" problem, many methods have been proposed to address the pathway-based analysis for next generation sequencing (NGS) data. However, there is no publicly available method for pathway-based NGS data analysis of multiple phenotypes. Hereby we present two powerful pathway-based approaches based on the general structural component analysis (GSCA) to investigate the association between rare variants and multiple phenotypes. By reflecting the natural hierarchy of biological behavior and considering the correlation among both pathways and phenotypes, the proposed method is capable of analyzing multiple phenotypes and multiple pathways simultaneously. The simulation studies successfully demonstrated advantages of multivariate analysis, compared to univariate analysis. Moreover, the real data analysis of five type 2 diabetes-related traits using 1,000 whole exome sequencing dataset identified significant pathways that were not identified in the univariate analysis. A strong relationship between the identified pathways and the target phenotypes was validated via literature search.

OC37: Design and analysis of clinical trials 5

Wednesday 12th July -14.30-16.00 h. - Room: Sala Mar 2
Chair: Sandra Lee

OC37-1: Graphical displays for subgroup analysis in clinical trials

Authors: Yida Chiu[1], Thomas Jaki[1], Franz Koenig[2], Martin Posch[2].
[1] *Lancaster University, UK,* [2] *Medical University of Vienna, Austria.*

Investigating target populations potentially beneficial to an innovative intervention is challengeable because various issues are needed to address. Subgroup analyses as investigative measures have received extensive attention in recent clinical research for the development of stratified medicine. Graphical approaches are routinely employed in subgroup analyses, typically for depicting effect sizes of subgroups. Such visualisation encapsulates subgroup information and boosts the clinical decision-making process. However, existing approaches still have inherent drawbacks and their use may lead to misinterpretations to subgroup effect sizes. It is therefore crucial to correctly depict the effect sizes and essential subgroup information. In this talk, we present several developed effective visualisation approaches. Nine techniques (including level plots, bar plots, tree plot, matrix plots, forest plots and so on) were applied to exhibit certain subgroup information through synthetic data with a continuous endpoint. We assessed the graphical approaches based on sensible criteria and present some improved displays by mitigating their original demerits. In the end, we summarise the assessment and features of all the improved approaches. Remarks on their practical usefulness and implications in clinical trials are made as well.

Reference 1: Alosh M, Huque M, Bretz F, et al. (2016) Tutorial on statistical considerations on subgroup analysis in confirmatory clinical trials, Statistics in Medicine, 7(8), 889–894.

Reference 2: Ondra T, Dmitrienko A, Friede T, et al. (2016) Methods for identification and confirmation of targeted subgroups in clinical trials: a systematic review., Journal of Biopharmaceutical Statistics, 26(1), 99–119.

OC37-2: A correlation test for central statistical monitoring of multicenter clinical trials

Authors: Lieven Desmet[1], David Venet[2], Tomasz Burzykowski[3], Laura Trotta[4].
[1] *Université Catholique de Louvain, Belgium,* [2] *Université Libre de Bruxelles, Belgium,* [3] *Universiteit Hasselt, Belgium,* [4] *Cluepoints, Louvain-La-Neuve, Belgium.*

In multicenter clinical trials, central statistical monitoring aims at detecting atypical patterns in the data by comparing each center's data to the distribution obtained from all centers (1).

A number of tests have been proposed in the univariate framework, taking into account the nature of measurements and accounting for inter-center variability, e.g., with a random center effect (2). However, more subtle data quality issues may be detectable only at the multivariate level. For instance, an issue with fabricated blood pressure data may only show up because of an atypical correlation between diastolic and systolic pressure.

In this context we propose a correlation test based on a flexible multicenter modeling of the data where bivariate normality is assumed at the center level. The correlation itself is modeled according to a normal law for its Fisher z-transform, in line with Fisher's result on the variability of the Pearson correlation coefficient in bivariate normal data. The test is designed to detect a small proportion of atypical centers in a multicenter trial such that power increases with the extent of inter-center discrepancy while keeping specificity at high level. Performance is demonstrated in a simulation study covering a large range of scenarios of interest, including unbalanced setups. Applications to real clinical data are also given.

Reference 1: (1) Venet D, Doffagne E, Burzykowski T, Buyse M, et al. A statistical approach to central monitoring of data quality in clinical trials. Clin Trials 2012; 9(6): 705–713.

Reference 2: (2) Desmet L, Venet D, Doffagne E, Timmermans C, Burzykowski T, Legrand C, Buyse M. Linear mixed-effects models for central statistical monitoring of multicenter trials. Stat Med 2014; 33(30): 5265–527.

OC37-3: Assessing the similarity of dose response and target doses in two non-overlapping populations

Authors: Frank Bretz[1].
[1]*Novartis, Switzerland.*

We consider two problems of increasing importance in clinical dose finding studies. First, we assess the similarity of two non-linear regression models for two non-overlapping groups of patients over a restricted covariate space. To this end, we derive a confidence interval for the maximum difference between the two given models. If this confidence interval excludes the equivalence margins, similarity of dose response can be claimed. Second, we address the problem of demonstrating the similarity of two target doses for two non-overlapping populations, using again a confidence interval based approach. We illustrate the proposed methods with a real case study and investigate their operating characteristics via simulation.

OC37-4: Use of ordinal scale endpoint in therapeutic influenza clinical study

Authors: Yonghong Gao[1].
[1]*Department Of Health And Human Services.*

A global priority is to develop novel therapeutics for serious influenza illnesses in addition to neuraminidase inhibitors (NAI). Clinical studies of novel agents require clinical endpoints that can reliably assess clinically meaningful outcomes. Ordinal scale endpoint for influenza therapeutic study has been extensive discussed by a US multi-sector working group to better capture the clinical status of patients at each day across some time range. We explore some analysis methods on the ordinal scale endpoint with one real clinical study dataset, and we also evaluate the statistical performance of the ordinal scale endpoint in comparing against the traditional time-to-event endpoint analysis approach through simulation study.

OC37-5: Correlates of risk and protection of CYD-TDV, the first licensed dengue vaccine in endemic countries

Authors: Zoe Moodie[1], Michal Juraska[1], Ying Huang[1], Yingying Zhuang[2], Youyi Fong[1], Steven G. Self[1], Laurent Chambonneau[3], Robert Small[4], Nicholas Jackson[3], Fernando Noriega[4], Peter B. Gilbert[1].
[1] *Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, USA,*
[2] *Department of Biostatistics, University of Washington, USA,* [3] *Sanofi Pasteur, France,*
[4] *Sanofi Pasteur, USA.*

Identifying a correlate of protection is a major goal of vaccine research as it can then predict vaccine efficacy in a new setting without the need to follow participants for the efficacy endpoint. Correlates provide an efficient path to vaccine development, evaluation and deployment; however, very few correlates have been identified recently.

In the placebo-controlled Phase 3 efficacy trials of CYD-TDV in 2-14 year olds in Asia and in 9-16 year olds in Latin America, estimated vaccine efficacy (VE) against dengue between Months 13 and 25 post first vaccination was 57% and 61%. Case-cohort analyses assessed how dengue risk and vaccine efficacy (VE) vary with anti-dengue neutralizing antibody titers to the four serotypes in the vaccine one month post-vaccination (Month 13) in 2848 vaccine and 1574 placebo recipients. Month 13 titers were assessed as correlates of risk using Cox regression accounting for the case-cohort sampling design. Correlates of protection were assessed using the principal stratification framework to study how VE varies over vaccinated subgroups defined by Month 13 titers. These analyses used structural hinge logistic or logistic linear risk models with nonparametric kernel density estimation of Month 13 titers under assignment to vaccine conditional on baseline titers. VE-by-Month 13 titer curves were also estimated under different assumptions with a pseudo-score estimator.

For vaccine recipients in each trial, dengue risk of each serotype significantly decreased with increasing Month 13 homologous serotype titer. In both trials, Month 13 neutralizing antibody titers correlated positively with VE overall, for each serotype and across all age groups. Other factors may influence vaccine efficacy.

OC38: Statistical methods in epidemiology 3

Wednesday 12th July -14.30-16.00 h. - Room: Sala Terra 4
Chair: Ana Luisa Papoila

OC38-1: Using indices to estimate the association of >1 exposures with quantitative, binary and time-to event health outcomes

Authors: Evangelia Christodoulou[1], Christina Bamia[2].
[1] *KU Leuven-University of Leuven, Belgium,* [2] *National And Kapodistrian University Of Athens, Medical School, Department Of Hygiene, Epidemiology And Medical Statistics.*

In studies (e.g. Nutrition or Genetic) evaluating the association of many risk factors with health outcomes”,” the use of indices that represent the overall presence of >1 risk factors has gained popularity due to the appealing simplicity of this approach. Indices can be defined a-priori or through multivariable analysis (e.g. principal components) - they are subsequently used as single-exposures to model the association of the combination of the risk factors with a health outcome. We consider indices calculated as linear combinations of risk factors. First”,” we describe the properties of these indices in relation to the properties of the included risk factors. Second”,” we decompose the estimated association with quantitative outcomes of such indices and we show that the overall association is a weighted average of the risk factors associations. The weights depend on the correlation matric of the risk factors. Third”,” using simulations for a variety of scenarios regarding the distributions of the individual risk factors we show that the same formulae can be used for binary or time-to-event outcomes. Hence”,” we highlight informative results that need to be presented in the respective publications when indices are used. Finally we make a brief review on the respective literature in Nutritional Epidemiology and we show that presentation of dietary indices”,” the so-called dietary patterns”,” is often poor and not informative. We conclude with a number of issues that should be addressed when indices of >1 exposures are used in multi-exposure/disease associations.

OC38-2: Modeling of cumulative effects of time-varying drug exposures on changes in a longitudinal biomarker

Authors: Coraline Danieli[1], Michal Abrahamowicz[1].
[1] *McGill University Health Center, Canada.*

An accurate assessment of the safety or effectiveness of drugs in pharmaco-epidemiological studies requires defining an etiologically correct time-varying exposure model, which specifies how previous drug use affects the health outcome of interest. To address this issue, we develop, and validate in simulations, a new approach for flexible modeling of the cumulative effects of time-varying exposures on repeated measures of a normally distributed response variable, such as a quantitative surrogate outcome or a biomarker. In particular, we extend the linear mixed effects model approach to allow estimation of how past and recent drug exposure affects the way individual values of the outcome change throughout the duration of the study. To account for the dosage, duration and timing of past exposures, we rely on a flexible weighed cumulative exposure (WCE) methodology to model the cumulative effects of past drug use, as the weighted sum of past doses, with weights that reflect relative importance of doses taken at different times in the past. The weights are modeled with cubic regression B-splines. In simulations, we evaluate the performance of the model under different assumptions concerning (i) whether the patients have a different number of observations or not, (ii) whether their outcomes are measured at the same times, (iii) whether their observations are equidistant in time, (iv) strength of the autocorrelation between consecutive measures of the biomarker, and (v) true shape of the weight function. Results demonstrate accuracy of the estimates of both: the weight function and the between and within patients variances. We also illustrate this method using real data on a longitudinal marker of glycemic control.

OC38-3: Weighted two-stage calibration for adjustment on unobserved confounders with non-representative validation sample

Authors: Helene Jacqmin-Gadda[1], Bernard Silenou[1], Catherine Helmer[1], Antoine Pariente[1].
[1] *Inserm, France.*

Pharmacoepidemiology studies using health administrative databases may lead to biased results since information on potential confounders are often missing. Two-stage calibration based on the propensity score (TSC)1 has been proposed to avoid confounding bias by integrating confounder data from a validation sample. The validation sample may be a subsample from the main one or, more frequently, an external cohort which is assumed to be representative of the main population. Using a simulation study, we highlighted the bias of TSC when the validation sample is not representative, with a selection probability possibly depending on the outcome and/or the exposure. We then extended TSC by using weighted regressions to correct for this bias. The weighted approach was validated by simulation and then compared to the unweighted TSC in a study of the association between Benzodiazepine use and fractures in the elderly using the general sample of French health insurance beneficiaries (EGB). Two French cohorts of elderly were used as validation samples to emphasize on the potential impact of selection biases.

Reference 1: 1Lin HW, Chen YH. Adjustment for missing confounders in studies based on observational databases: 2-stage calibration combining propensity scores from primary and validation data. Am J Epidemiol. 2014;180(3):308-317.

OC38-4: The calibrated model-based concordance improved assessment of discriminative ability in patient clusters of limited sample size

Authors: David Van Klaveren[1], Ewout W. Steyerberg[1], Mithat Gönen[2], Yvonne Vergouwe[3].
[1] *Leiden University Medical Center, The Netherlands,* [2] *Memorial Sloan Kettering Cancer Center, New York, USA,* [3] *Erasmus University Medical Center, Rotterdam, The Netherlands.*

Discriminative ability is an important aspect of prediction model performance, but challenging to assess in clustered (e.g. multicenter) data. Concordance (c)-indexes may be too extreme within small clusters. We aimed to define a new approach for the assessment of discriminative ability in clustered data.

We assessed discriminative ability of a prediction model for mortality after traumatic brain injury within centers of the CRASH trial. With multilevel regression analysis we estimated cluster-specific calibration slopes which we used to obtain the recently proposed calibrated model-based concordance (c-mbc) within each cluster. We compared the c-mbc with the naïve c-index in centers of the CRASH trial, and in simulations of clusters with varying calibration slopes.

The c-mbc was less extreme in distribution than the c-index in 19 European centers (internal validation; n=1,716) and 36 non-European centers (external validation; n=3,135) of the CRASH trial. In simulations the c-mbc was biased but less variable than the naïve c-index, resulting in lower root mean squared errors.

The c-mbc, based on multilevel regression analysis of the calibration slope, is an attractive alternative to the c-index as a measure of discriminative ability in multicenter studies with patient clusters of limited sample size.

OC38-5: A graphical guideline for the selection of reliability and agreement methods for clinical studies

Authors: Pedro Sa-Couto[1, 2], Maria José Santiago[1], Andreia Hall[1, 2].
[1] *Center for Research and Development in Mathematics and Applications (CIDMA), University Of Aveiro, Portugal,* [2] *Department of Mathematics (DMAT), University Of Aveiro, Portugal*

Introduction: Validity is defined as the ability of a given instrument to be well-founded from a theoretical point of view. Reliability refers to the ability to discriminate subjects or objects while agreement refers to the degree to which scores are identical. OBJECTIVE: To propose a graphical guideline to indentifiy the most comon used reliability and agreement methods. METHODS: Methods based on the correction chance (e.g. Cohen's Kappa), their weighted kappas versions, methods

based on rankings (e.g. Kendall W), and methods based on the ratio of the variances (e.g. intraclass correlation coefficient) are presented and discussed. For the latter, a discussion between two different approachs and their implications are also addressed [1,2]. RESULTS: By presenting flowcharts, we are able to select the appropriate reliability/agreement methods depending on certain conditions such as: type of measure used, number of raters, or study design (inter-rater or intra-rater), among others. Their calculation and respective statistical inference (including sample size calculation) are illustrated through case studies using the software R. DISCUSSION: The unreliability poses problems about the validity of an instrument and therefore an instrument that is not reliable cannot be valid. On the other hand, the existence of reliability does not imply validity. This work regroups and completes the information available in the literature, and contributes to a more correct application of these methods for the development or adaptation of measurement instruments applied in clinical studies.

Reference 1: Eliasziw M et al. (1994). Physical Therapy, 74(8), 777–88

Reference 2: Shrout P et al. (1979). Psychological Bulletin, 86(2), 420–428

Poster Contributed Session

PC7: Observational studies

Wednesday 12th July -14.30-16.00 h. - Room: Hall
Chair: Els Goetghebeur

PC7 - W2: How beneficial is individual patient data in a mixed treatment comparison?

Authors: Joy Leahy[1], Cathal Walsh[2].
[1] *Trinity College Dublin, Ireland,* [2] *University Of Limerick, Ireland.*

Evaluating various treatment regimens is an important aspect of decision making in a clinical setting. Individual Patient Data (IPD) from Randomised Control Trials (RCTs) are considered the gold standard of evidence. IPD enables a more in depth analysis, such as investigating the effect of covariates, than can be carried out with aggregate data (AD) alone. However, as the majority of studies do not report IPD, most Mixed Treatment Comparisons (MTCs) are carried out using aggregate data (AD), or sometimes using IPD where possible, along with AD. Our work investigates the benefit of including varying proportions of IPD studies. We simulate trials where the effects are known, and while keeping the number of trials constant, we vary the proportion using IPD. We can then assess the accuracy of the estimate of both the treatment effect and the covariate effect. Under certain assumptions including more IPD studies will increase the accuracy of the covariate effect estimates. However, in well conducted RCTs IPD does not produce more accurate treatment estimates, as their randomised approach allows us to estimate the treatment effect accurately without an accurate estimate of the covariate effect.

Reference 1: Donegan, Sarah, et al. "Combining individual patient data and aggregate data in mixed treatment comparison meta analysis: Individual patient data may be beneficial if only for a subset of trials." Statistics in medicine 32.6 (2013): 914-930.

PC7 - W5: An empirical investigation of the impact of different methods for synthesising evidence in a network meta-analysis

Authors: Emily Karahalios[1], Georgia Salanti[2], Simon Turner[1], G. Peter Herbison[3], Ian R. White[4], Areti Angeliki Veroniki[5], Adriani Nikolakopoulou[2], Joanne E. McKenzie[1].
[1] *Monash University, Australia,* [2] *ISPM, University of Bern, Switzerland,* [3] *University of Otago, Dunedin, New Zealand,* [4] *MRC Biostatistics Unit, Cambridge, UK,* [5] *Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada.*

Network meta-analysis (NMA) is a method to synthesise evidence from multiple treatments. Two broad approaches are available to synthesize data across networks: arm-based and contrast-based, with a range of models that can be fitted within each. It is unclear how the two approaches compare and there has been limited empirical evaluation comparing results from different network meta-analysis methods applied to a large number of networks.

We re-analysed a subset of 158 networks from a cohort of 456 published networks of randomised trials. The subset of networks included those where the primary outcome was binary, the number of events and participants were reported for each direct comparison, and there was no evidence of inconsistency in the network. We re-analysed the networks using five methods, three of which are contrast-based and two of which are arm-based models. We compared the estimated treatment effects, their standard errors, treatment ranks, and the metric on which the ranks are based, and the between-trial heterogeneity variance, across the network meta-analysis methods. We investigated if differences in the results are modified by network characteristics.

Preliminary results show good agreement between the contrast-based, Bayesian and frequentist methods in terms of effect estimates and treatment ranks. However, differences are apparent in the effect estimates and ranks when comparing the arm-based method to the contrast-based methods.

PC7 - W8: Comparative meta-analysis of diagnostic studies: a review and comparison of currently proposed approaches

Authors: Junfeng Wang[1], Mariska Leeflang[1].
[1] *University of Amsterdam, The Netherlands.*

There is increasing interest in comparative meta-analysis of diagnostic studies which evaluates the accuracy of a test relative to an alternative test. Researchers have attempted to develop a network meta-analysis (NMA) like approach for evaluating two or more tests simultaneously.

Several approaches have been proposed in recent years (e.g. Trikalinos 2014, Ma 2015, Menten 2015, Dimou 2016 and Hoyer 2016) . They share some common features but there are also conceptual and statistical differences among them.

In this study, we aim to give a comprehensive review of the currently proposed approaches for comparative meta-analysis of DTA studies. We compared the following features with respect to conducting a comparative meta-analysis: (i) input data requested from primary studies; (ii) assumption of common threshold; (iii) focus on absolute or relative accuracy; (iv) inclusion of indirect evidence; (v) detecting inconsistency; and (vi) statistical package of implementation.

We also investigated and compared the performance of these approaches using simulation. The simulation is done by (i) generating summary test results data (2×2×2 or 2×2 tables) based on assumptions of these approaches; (ii) generating original test results from individual patients, which is more close to real world but in this case the assumptions on the distributions of sensitivity and specificity may not be valid.

PC7 - W11: Undertaking meta-analyses when only very few studies are available

Authors: Ralf Bender[1], Tim Friede[2], Armin Koch[3], Oliver Kuß[4], Peter Schlattmann[5], Guido Schwarzer[6], Guido Skipka[1].
[1] *IQWIG, Germany,* [2] *UMG, Germany,* [3] *Mhh, Germany,* [4] *DDZ, Germany,* [5] *IMSID, Germany,* [6] *Imbi, Germany.*

The DerSimonian-Laird method has been the standard for random effects meta-analysis for several decades. However, unfavorable statistical properties, especially in the case of few studies, have been highlighted and critically discussed for some time now. A working group of the Cochrane Collaboration recommended the Knapp-Hartung method as new standard approach, which in contrast to the DerSimonian-Laird method accounts for uncertainty in estimating the between-study heterogeneity. With very few studies (say 2-5), however, the Knapp-Hartung method can result in very wide confidence intervals, even if all studies are statistically significant in the same direction. Besides classical meta-analysis approaches, a number of alternative approaches are available including generalized mixed effects models and Bayesian methods incorporating weakly informative priors for the between-study heterogeneity. Basic features of these approaches are summarized and required conditions for practical applications are discussed. Methods are illustrated by various examples. Currently, none of the available approaches can be considered the uniformly best method. Besides classical approaches, the use of alternative methods such as generalized mixed effects models seems to be useful. Although meta-analyses with very few studies are very common, performing meta-analyses in the case of very few studies remains challenging. Currently no clear guidance exists on how to best proceed in these challenging scenarios. Further research in this field is required.

PC7 - W14: Evaluation of practical methods for estimating means for meta-analysis of continuous, non-normally-distributed outcomes

Authors: Valentina Assi[1], Peter Langhorne[2], Steff C. Lewis[1], Marian C. Brady[3], Gordon D. Murray[1], Chris J. Weir[1].
[1] *Edinburgh Clinical Trials Unit, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, UK,* [2] *University of Glasgow, UK,* [3] *Glasgow Caledonian University, UK.*

Background: Effects on continuous non-normally-distributed outcomes can be arduous to pool in meta-analyses as the mean is often not reported. We assessed three methods that estimate the missing mean using other summary statistics (Hozo 2005: median, minimum, maximum, sample size; Bland 2014: median, minimum, maximum, sample size, upper and lower quartile; Wan 2014: median, upper and lower quartile, sample size).

Methods: We assessed their performance on a skewed outcome (hospital length of stay), comparing their impact on the overall meta-analysis results. Using individual patient data from two sources (the GALA carotid surgery trial and a published meta-analysis of early supported discharge (ESD) after stroke) we simulated a wide range of scenarios, varying size and number of the included studies and size and number of studies not reporting a mean.

Results: The Wan approach most closely estimated the actual treatment effect, mean difference (SD) 0.07 (0.11) days in GALA and 0.38 (0.45) days in ESD, with little effect on the confidence interval (CI) width, mean CI ratio 0.97 (0.08) in GALA and 0.98 (0.03) in ESD. The Hozo and Bland approaches led to larger differences in effect (respectively 0.91 (2.41) and 0.40 (1.08) days in GALA and 1.58 (1.76) and 0.73 (0.83) days in ESD) and much wider CIs (ratios 2.05 (1.87) and 1.49 (0.89) in GALA and 1.14 (0.22) and 1.06 (0.09) in ESD).

Conclusions: The Wan method performed best and is also more practical, since it uses the median and quartiles which are more often available than other summary statistics, especially for non-normally distributed outcomes. As the Hozo and Bland formulae include also the minimum and maximum, their estimates are more susceptible to outliers.

PC7 - W17: From eQTL associations to suggested mechanisms

Authors: Silva Kasela[1], Kai Kisand[2], Krista Fischer[1], Pärt Peterson[2], Lili Milani[1].
[1] *Estonian Genome Center, University of Tartu, Estonia,* [2] *Institute of Biomedicine and Translational Medicine, University of Tartu, Estonia.*

Studies on disease-associated genetic polymorphisms have highlighted the importance of the biological context for many regulatory variants. In particular, analysing the effect of single nucleotide polymorphisms (SNPs) on gene expression levels, referred to as expression quantitative trait loci (eQTL), in specific cell types can be used to gain insight into specific mechanisms of disease.

By eQTL mapping in essential elements of adaptive immune response, purified CD4 and CD8 T cells of 313 healthy individuals, we have found an interesting association triangle between a missense SNP, IRF1 and STAT1 expression levels. Both of the genes have notable roles in immune regulation and the functional effect of the SNP is confirmed by additional experiments. Next, we use different statistical methods to shed light into the associations. We apply Bayesian test for colocalisation to link the SNP with type 1 diabetes susceptibility. Then we use structural equation modeling and mediation analysis to determine whether the SNP affects both genes independently or via each other. Overall, the best-fitting scenario suggests that IRF1 mediates the SNP and STAT1 relationship. This finding is further supported by a simulation experiment.

In conclusion, statistical modelling suggests a mechanism for the effect of the SNP on IRF1 and STAT1 expression in the context of type 1 diabetes. This illustrates the strength of the combination of advanced statistical methods and eQTL studies to establish specific mechanisms and pathways involved in diseases progression and create the basis for future explorations and drug interventions.

PC7 - W20: Applying the causal inference framework in policy evaluation: a case study in the evaluation of primary care reforms in Ontario, Canada

Authors: Nadia Sourial[1], Isabelle Vedel[1], Susan Bronskill[2], Tibor Schuster[1].
[1] *McGill University,* [2] *University of Toronto.*

Objectives: Estimating population-wide effects of health policies presents several challenges including lack of randomization and time-varying exposures. This study assesses the potential application of the potential outcomes framework to evaluate primary care reforms on health system use for persons with dementia in Ontario, Canada.

Methods: A series of primary care models (PCMs) were rolled out in Ontario between 2001 and 2008 with voluntary enrollment for physicians and patients and potential transitions from one model to another. A causal framework will be applied to determine which PCM leads to improved care and reduced health system use for persons with dementia. The common assumptions required for causal inference (consistency, positivity, exchangeability, no interference) will be assessed in the context of this observational policy evaluation study. A directed acyclic graph will be mapped and validated with subject experts.

Results: In the proposed analysis, a marginal structural model (MSM) will be used to address the specific time-varying nature of the PCMs as well as the clustering of persons within PCMs. Sensitivity analyses exploring the robustness of results with regard to unmeasured confounding and measurement error will also be conducted.

Conclusion: Recent advances in causal inference present an opportunity to improve the analysis of data from observational studies including health policy evaluation studies. If required conditions under the causal framework are met, inference on the impact of health policy interventions on desired outcomes can be strengthened to represent population-wide causal effects.

PC7 - W23: Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy

Authors: Jessica Rees[1], Angela Wood[1], Stephen Burgess[1].
[1] *University of Cambridge, UK.*

Methods have been developed for Mendelian randomization that can obtain consistent causal estimates while relaxing the instrumental variable assumptions. These include multivariable Mendelian randomization, in which a genetic variant may be associated with multiple risk factors so long as any association with the outcome is via the measured risk factors (measured pleiotropy), and the MR-Egger (Mendelian randomization-Egger) method, in which a genetic variant may be directly associated with the outcome not via the risk factor of interest, so long as the direct effects of the variants on the outcome are uncorrelated with their associations with the risk factor (unmeasured pleiotropy). We have extended the MR-Egger method to a multivariable setting to

correct for both measured and unmeasured pleiotropy. We have shown, through theoretical arguments and a simulation study, that the multivariable MR-Egger method has advantages over its univariable counterpart in terms of plausibility of the assumption needed for consistent causal estimation, and power to detect a causal effect when this assumption is satisfied. The methods were compared in an applied analysis to investigate the causal effect of high-density lipoprotein cholesterol on coronary heart disease risk. The multivariable MR-Egger method will be useful to analyse high-dimensional data in situations where the risk factors are highly related and it is difficult to find genetic variants specifically associated with the risk factor of interest (multivariable by design), and as a sensitivity analysis when the genetic variants are known to have pleiotropic effects on measured risk factors.

PC7 - W26: Screening the effects of lipidomic alterations on atrial depolarization

Authors: Fabiola Del Greco M.[1], Luisa Foco[1], Alexander Teumer[2], Niek Verweij[3], Roberto Melotti[1], Viviana Meraviglia[1], Giuseppe Paglia[1], Peter K. Joshi[4], Alan F. Wright[5], Harry Campbell[4], Marcus Dörr[6], Harold Snieder[7], James F. Wilson[4], Peter P. Pramstaller[1], Alessandra Rossini[1], Cristian Pattaro[1].
[1] *Institute for Biomedicine, EURAC research, Bolzano, Italy (affiliated to the University of Lübeck, Lübeck, Germany),* [2] *Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany,* [3] *Department of Cardiology, University of Groningen, University Medical Center Groningen, The Netherlands,* [4] *Centre for Population Health Sciences, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, Scotland, United Kingdom,* [5] *MRC Human Genetics Unit, University of Edinburgh, Edinburgh, Scotland, United Kingdom,* [6] *Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany,* [7] *Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands.*

Alterations of cardiac atrial conduction are associated with an increased risk of cardiovascular diseases and arrhythmias, like atrial fibrillation, which affects more than 1% of the adult population in industrialized countries. Altered lipid levels may affect atrial conduction, but the specific biological mechanisms are unclear. We performed a screening of 151 sphingo- and phospholipids for association with atrial depolarization, measured by P wave length (PWL) from standard electrocardiogram. To identify causal associations between lipid species and PWL, we established a workflow based on cross-sectional association and mediation analyses on individual-level data followed by 2-sample Mendelian randomization (MR) on summary genome-wide association study (GWAS) data. We tested association between lipidomics and PWL in 839 general population subjects from the Microisolates in South Tyrol study (Italy) by means of linear mixed effect models, to account for relatedness. Multiple-testing corrected significant associations were tested for direction-consistent effects in an independent population sample of 951 subjects from the Orkney Complex Disease Study (Scotland), confirming association of one phosphatidylcholine (PC) with PWL (pooled effect 2.6 ms/mol, 95% confidence interval: 1.3-3.9). Screening for potential mediators revealed body mass index (BMI) as a complete mediator of the observed association. MR analysis on summary GWAS data from studies on lipidomics (N=3254), BMI (N~230,000), and PWL (N=4338) highlighted causal association between PC and BMI (P-value=0.03) and between BMI and PWL (P-value=0.05). Lipidomic alterations may trigger a novel causal pathway in the biological mechanism of atrial depolarization.

PC7 - W29: Multiple imputation for handling missing values in a time-dependent covariate with a nonlinear trajectory over time: a simulation study

Authors: Anurika de Silva[1], Margarita Moreno-Betancur[2, 3], Alysha M. de Livera[1], Katherine J. Lee[2, 4], Julie A. Simpson[1].
[1] *Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia,* [2] *Clinical Epidemiology and Biostatistics Unit, Murdoch Childrens Research Institute, Royal Children's Hospital, Melbourne, Victoria, Australia,* [3] *Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Victoria, Australia,* [4] *Department of Paediatrics, University of Melbourne, Melbourne, Victoria, Australia.*

Missing data are a major problem in longitudinal studies due to multiple waves of data collection. Multiple imputation (MI) is increasingly being used for handling missing data in these studies, but standard methods, fully conditional specification (FCS) and multivariate normal imputation (MVNI), treat repeated measurements of the same time dependent variable as just another 'distinct' variable for imputation, ignoring the longitudinal structure of the data. The two-fold fully conditional specification (two-fold FCS) algorithm imputes a time-dependent variable using information from only the specified and adjacent times. We compared the performance of two-fold FCS with standard MI methods for handling missing values in a time-varying covariate with a nonlinear trajectory over time, a common scenario for health measures. We designed a simulation study based on the Longitudinal Study of Australian Children (LSAC), to compare the performance of FCS, MVNI, and two-fold FCS, where up to 50% of data in a continuous time-dependent exposure variable with a nonlinear trajectory over time were missing completely at random (MCAR) or missing at random (MAR). The standard two-fold FCS, which only uses information from specified and immediately adjacent time points, produced slightly more biased and less precise estimates than FCS and MVNI. We observed slight improvements in bias and precision when we extended the time window width of the two-fold FCS to include two adjacent time points. We recommend the use of FCS or MVNI in a similar longitudinal setting, and if convergence issues are encountered due to a large number of time points or variables with missing values, the two-fold FCS with exploration of a suitable time window.

PC7 - W32: A statistical significance test for the binormal ROC curve with measurements below the limit of detection

Authors: Alba María Franco Pereira[1], Christos T. Nakas[2], Alexander B. Leichtle[3], M. Carmen Pardo[1].
[1] *Universidad Complutense de Madrid, Madrid, Spain,* [2] *University of Thessaly, Volos, Greece,* [3] *University of Bern, Bern, Switzerland.*

The receiver operating characteristic (ROC) curve is commonly used to evaluate biomarker utility in clinical diagnosis of a disease. Some biomarkers are often measured subject to limits of detection (LoD) imposed by the laboratory analytical system precision. This inability to accurately determine values of biomarkers introduces bias in the analysis of data from such experiments. In this work we propose a parametric approach based on the statistic proposed in Metz et al. (1983) in the LoD context. This approach makes direct use of the binormal ROC curve parameters. This test is compared via simulations to other existing tests and an illustrative example of patients with colorectal cancer is also presented.

Reference 1: Metz, C. E., Wang,P.-L. and Kronman, H. B. (1983). A new approach for testing the significance of differences between ROC curves measured from correlated data. In Information processing in medical imaging (F Deconinck, ed.). The Hague: Nijhoff, 43.

PC7 - W35: Parameter estimation for normally distributed data with observations below multiple lower limits of quantification

Authors: Tanja Berger[1], Ralf-Dieter Hilgers[1], Nicole Heussen[1].
[1] *RWTH Aachen University, Germany.*

Background: The lack of perfect measurement sensitivity in clinical laboratories results in the frequently encountered problem of left-censored data. Due to this lack, some observations cannot be measured exactly because they are located below the lower limit of quantification (BLOQ). Multiple lower limits of quantification (MLOQ) result if more than one laboratory is involved in the analysis of concentration data. The most established methods to handle BLOQ and MLOQ data are simple imputation methods although their use is not recommended by the scientific community. Other methods such as the multiple regression method by Helsel (2012) and the maximum likelihood based methods by Beal (2001) have been proposed. However, Beal's methods are only applicable for BLOQ data and not in the case of MLOQ.

Methods: We propose two maximum likelihood methods to estimate mean and variance in the presence of observations below MLOQ for normally distributed data. We investigate the case of two and four MLOQ, different proportions of censored data and various sample sizes via a simulation study. Special interest is given to small sample sizes and large proportions of censored data. We assess the performance of our methods by comparing them to Helsel's and simple imputation methods with respect to bias and mean squared error.

Results and Conclusion: Our research findings serve to estimate the distribution parameters mean and variance in the context of censored data due to MLOQ. We show that the often used simple imputation methods can lead to serious errors in the estimation and that the newly presented methods are clearly superior in estimating the considered parameters in every investigated scenario.

PC7 - W38: Intention-to-treat analyses when missing outcomes depend on post-randomisation factors

Authors: Elizabeth Howarth[1], Richard Emsley[1].
[1] *University of Manchester, UK.*

Intention-to-Treat (ITT) analysis, with outcomes for all participants analysed according to their randomised group and irrespective of subsequent adherence to the assigned intervention, is the recommended approach for a pragmatic comparison of alternative treatments. To exclude participants with missing outcome data from an analysis is not strictly compatible with the ITT

principle; to include them using a likelihood-based analysis requires the assumption that the missing values are missing at random (MAR) after conditioning on observed data.

In cases where outcome missingness is likely to be associated with post-randomisation factors such as treatment adherence, inclusion of such data is necessary to justify the MAR assumption, even though post-randomization factors are not included in the substantive ITT analysis model. Multiple imputation may provide valid estimates under MAR if such factors are included as auxiliary variables in the imputation models. Alternatively, factor specific weights could be derived and used as inverse probability weights.

We present the results of a Monte Carlo simulation study comparing whether the following missing data methods produce valid ITT estimates: complete case analysis, full information maximum likelihood, inverse probability weighting and multiple imputation, using imputation models with and without compliance as an auxiliary variable. A range of data generation models are used, based on parameters from real datasets, and the methods are illustrated with analysis from a randomised trial of a psychological intervention in mental health. The results show that post-randomisation factors should be accounted for in order to estimate an unbiased ITT effect.

PC7 - W41: Sensitivity analysis within multiple imputation framework using delta-adjustment: application to Longitudinal Study of Australian Children

Authors: Julie A. Simpson[1], Panteha Hayati-Rezvan[1], Katherine J. Lee[2].
[1] *University of Melbourne, Australia,* [2] *Murdoch Childrens Research Institute, Australia.*

Multiple imputation (MI) is a well-recognised statistical technique for handling missing data. As usually implemented in standard statistical software, MI assumes that data are 'Missing at random' (MAR); an assumption that cannot be tested using the data available and in many settings may be implausible. The delta-adjustment method, implemented within the MI framework, can be used to perform sensitivity analyses that assess departures from the MAR assumption. This method requires specification of an unknown sensitivity parameter (delta).

We illustrate the application of the delta-adjustment method using data from the Longitudinal Study of Australian Children, where the epidemiological question is to estimate the association between exposure to maternal emotional distress at age 4-5 years and total (social, emotional, and behavioural) difficulties at age 8-9 years. We elicited unknown sensitivity parameters for the outcome (y) and exposure variables (x) from a panel of experts. The elicited quantile judgements from each expert were converted into a suitable parametric probability distribution and combined using the linear pooling method. We present results from sensitivity analyses that used different percentile values of the pooled distributions for the delta parameters for y and x, and demonstrate that two-fold increases in the magnitude of the association between maternal distress and total difficulties are only observed for large departures from MAR.

PC7 - W44: Observed vs. expected information matrix in linear mixed models under MAR dropout

Authors: Nikos Pantazis[1], Christos Thomadakis[1], Giota Touloumi[1].
[1] *National and Kapodistrian University of Athens, Medical School, Greece.*

It is known that likelihood inferences ignoring MAR missingness are valid. However, missingness should be taken into account when deriving the expected information matrix (EIM) otherwise precision estimates will be inconsistent (naïve EIM) [1]. The way statistical packages derive the fixed effects' variance is equivalent to the naïve EIM approach. Alternatively, the observed information matrix (OIM) can be used as it yields consistent precision estimates, ignoring the MAR mechanism. We implement the calculation of the OIM [2] and compare the resulting 95% coverage probabilities (CPs) with those obtained by the R lme function, through simulations based on the CD4 cell count decline in untreated HIV+ individuals. Initiation of antiretroviral treatment (ART) acts as a MAR mechanism since it is administered once CD4 drop below certain levels with subsequent measurements excluded from such analyses. With MCAR dropout, both approaches yielded nominal CPs. Increasing MAR dropout led to lower CPs for the CD4 slope when using the naïve EIM, (CP=0.694 with ART initiation at 500 CD4 cells/μL). CPs of slope differences between two groups were minimally affected except for cases with substantial differences in the MAR dropout between groups (CP=0.916). OIM-based CPs were always close to 0.95. Application of the proposed method to real data led to similar conclusions. The OIM is easily calculated and should be preferred for frequentist inference in linear mixed models under MAR dropout.

Reference 1: Kenward MG, Molenberghs G. Likelihood based frequentist inference when data are missing at random. Statist. Sci. 1998

Reference 2: Lindstrom M, Bates D. Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. JASA 1988

PC7 - W47: A comparison of multiple imputation methods for incomplete longitudinal binary data

Authors: Yusuke Yamaguchi[1], Toshihiro Misumi[1], Kazushi Maruo[2].
[1] *Astellas Pharma Inc., Japan,* [2] *National Center Of Neurology And Psychiatry, Japan.*

Longitudinal binary data are commonly encountered in clinical trials, where a binary variable is repeatedly measured per subject at several time points. Multiple imputation (MI) is an approach for getting a valid estimation of treatment effects under an assumption of missing at random mechanism and is flexible enough to be applied for any kinds of endpoints. Although there are a variety of options of the MI methods for the longitudinal binary data, a limited number of researches have reported on relative performances of the methods. Moreover, when focusing on the treatment effect throughout a period that has often been used in clinical evaluations of specific disease areas, no definite investigations comparing the MI methods have been available. We conducted an extensive simulation study to examine comparative performances of the MI methods for the longitudinal binary data under a wide range of scenarios regarding complete-data sampling models and missing

data generation, where two endpoints of a responder rate at a specified time point and a responder rate throughout a period were assessed. The simulation study suggested that results from naïve approaches of a single imputation with non-responders and a complete case analysis could be very sensitive against the missing data. The MI methods using a monotone method and a full conditional specification with a logistic regression as its imputation model were recommended for obtaining unbiased and robust estimations for the two endpoints.

PC7 - W50: Quantile regression for complex survey data

Authors: Jianhua Wu[1], Li Su[2], Ivonne Solis-Trapla[3].
[1] *University of Leeds, UK,* [2] *MRC Biostatistic Unit, UK,* [3] *Keele University, UK.*

The Diet and Nutrition Survey of Infants and Young Children, 2011, provide currentanthropometric measurement data of UK children aged 4–18 months [1]. Examination ofthe distribution of these measurements helps to monitor growth. We present a novel approachto quantile regression for positive continuous data from complex survey designs andapply the methods to describe body weight and length, and head circumference of youngchildren. We assume a generalised gamma distribution for the data collected using a multistagesampling design involving stratification, clustering of individuals and unequal samplingprobabilities. A pseudolikelihood approach with sampling weights is used for estimation. Wedevelop a sandwich estimator, bootstrap and a new estimating function bootstrap procedure,to estimate standard errors that account for clustering and stratification. Likelihood ratio tests,Akaike information criterion, Bayesian information criterion and model diagnostics based onquantile residuals are also provided. Our findings show that UK young children are overgrowncompared to the UK World Health Organisation growth standards for their age and sex.

Reference 1: Lennox A, Sommerville J, Ong K, Henderson H, Allen R. Policy paper. Diet andnutrition survey of infants and young children, 2011 [Internet]. 2013 [cited 2015 June13]. Available from: https://www.gov.uk.

PC7 - W53: A multilevel logistic model of poor quality staff-inpatient interactions accounting for variability across patients and observation periods

Authors: Inés Mesa Eguiagaray[1], E. Oliver[1], P. Griffiths[1], J. Bridges[1], L. Gould[1], *R. M. Pickering*[1].
[1] *University of Southampton, UK.*

Study objectives and design: an observational study was carried out to examine the impact of staffing levels on poor quality care in two acute hospitals. Each interaction occurring between designated patients and nurses or health care assistants (HCA) was observed during daytime, 2-hour long periods covering a variety of wards and times of day when staff were more or less busy. Interactions were rated according to the QuIS (Quality of Interactions Schedule) into 5 categories: positive social, positive care, neutral, negative protective and negative restrictive. Levels of nurse and HCA staffing were recorded.

Methods: multilevel logistic regression was used to investigate the effect of staffing levels on the likelihood of having a negatively QuIS rated interaction. Limited information on patients (n=269) and observation periods (n=120) was available and since we anticipated variation across these factors, they were included as random effects. Staffing levels along with ward, available patient characteristics and other contextual factors were included as fixed effects.

Results: 2,220 interactions were observed of which approx 10% were rated negatively. The variance component for observation period was greater than that for patient possibly because more patient characteristics were available to include as explanatory variables. Low staffing levels (either nurse or HCA) were associated with increased odds of a negative interaction. When nurse staffing levels were low, providing a greater number of HCAs not only did not improve the quality of patient care, but may have had a detrimental effect.

Conclusion: multilevel models were successfully used to better take account of the variability in ward conditions.

PC7 - W56: Ras inhibitor and risk of diabetic retinopathy using national claim data in Korea

Authors: Jo Aj[1], Rhee J.[1], Kang Sh[1], Lee Hi[1], Ahn Sy[2], Park Sj[3].
[1] *National Evidence-Based Healthcare Collaborating Agency, Seoul, South Korea,* [2] *Department of Nephrology and Hypertension, Korea Iniversity Guro Hospital, Seoul, South Korea,* [3] *Department of Ophthalmology, Seoul National University Bundang Hospital, Gyeonggi-Do, South Korea.*

Aims: This study evaluated the effect of Renin Angiotensin System(RAS) inhibitor use on diabetic retinopathy in type 2 diabetes mellitus(DM) with hypertension.

Methods: We identified 626,236 patients who were eligible for RAS inhibitor therapy among National Health Insurance Service of Korea between January 1, 2005, and December 31, 2011. We evaluated the clinical impact of RAS inhibitor use on clinical outcome. The primary outcome was diabetic retinopathy and the secondary outcomes included cataract, glaucoma, occlusion of retinal vessels. We performed a retrospective analysis by Cox proportional hazards with standardized mortality ratio(SMR) weight consist of estimated propensity score.

Results: A total of 52,446 patients met the inclusion criteria, of whom 20,916 (39.9%) used RAS inhibitor for hypertension. The SMR was estimated by propensity-score with the consideration of baseline demographic, clinical characteristics, and concomitant cardiovascular medication factors. Weighted analysis showed that risk of diabetic retinopathy and other outcomes depending on anti-hypertensive medications. The group with other anti-hypertensive medications had a significantly higher risk of diabetic retinopathy(hazard ratio: 1.15; 95% CI: 1.11-1.19), compared to the group with RAS inhibitor. Similarly, risk of secondary outcomes in the group with other anti-hypertensive medications was significantly higher.

Conclusions: This large-scale, population-based observational study lined with national claim data will provides a critical appraisal of effectiveness on diabetic retinopathy of RAS inhibitor use. Therefore, it needs to be emphasized that clinicians should routinely address adherence issues for patients with type 2 DM.

PC7 - W59: Prognostic models for death in operated acute pancreatitis

Authors: Carla Henriques[1,2,3], Ana Cristina Matos[1,3], Jorge Pereira[4], Carlos Daniel[4], Tercio de Campos[5].
[1] *Polytechnic Institute of Viseu, Portugal* [2] *Centre for Mathematics of the University of Coimbra (CMUC), Portugal,* [3] *Centre for Studies in Education, Technology and Health of Viseu (Ci&Dets), Portugal,* [4] *Centro Hospitalar Tondela-Viseu, Portugal,* [5] *Hospital da Santa Casa de Misericórdia – São Paulo, Brazil.*

Acute pancreatitis (AP) is a relatively common disease, usually without severity, but in some patients complications can occur that may endanger their life. Therefore, in patients with AP, an early assessment of the severity of the disease is crucial, but this evaluation is not straightforward, despite the existence of several scores of severity and current knowledge about prognostic variables (Gravante et al., 2009). The sample consisted of 66 patients surgically treated for AP, of whom 26 died. Aiming at assessing the risk of death as early as possible, logistic regression models were constructed and studied in order to enhance the predictive ability of two clinical scores usually considered in the evaluation of the severity of the disease. ROC curves and the area under these curves (AUC) were used to evaluate the discriminative ability of the models, comparing it with the discriminative ability of the scores alone. Furthermore, to recalibrate and improve predictive logistic regression models, penalized maximum likelihood estimation was applied. Validation of these models with bootstrapping revealed quite good performance measures (AUC, Nagelkerke's R-square, Brier Score and calibration slope) reinforcing their predictive value.

Reference 1: Gravante, G., Garcea, G., Ong, S. L., Metcalfe, M. S., Berry, D. P., Lloyd, D. M., & Dennison, A. R. (2009) Prediction of Mortality in Acute Pancreatitis : A Systematic Review of the Published Evidence. Pancreatology, 9, 601–614.

PC7 - W62: Relationship of statin use with the risk of type 2 diabetes mellitus and the benefit of cardiovascular disease using a nationwide administrative database

Authors: Min Jung Ko[1], Ae Jung Jo[1], Songhee Cho[1], Yun Jung Kim[1], Shin Hee Kang[1], Duk Woo Park[2].
[1] *National Evidence-Based Healthcare Collaborating Agency, Korea,* [2] *Asan Medical Center, University of Ulsan College of Medicine, Korea.*

The effects of cardiovascular benefit and diabetes risk of statin use in real world population remain unknown. This study aims to determine whether statin use is related with an increased risk of type 2 diabetes (T2D) and decreased risk of cardiovascular disease (CVD) in real world population who are eligible for statin therapy. The eligible population included patients aged 40 years or older who had no previous history of CVD within the preceding 3 years before cohort entry and had elevated level of total cholesterol over 240 mg/dL. We defined the ever use of statin as having filled at least 2 prescriptions for statin within a six-month period. The primary outcome was T2D based on the diagnosis and a new prescription for an oral antidiabetic medication or insulin. The secondary outcomes included a composite of cardiovascular death, nonfatal myocardial infarction, or nonfatal ischemic stroke. Propensity-score matching was used to assemble a cohort of patients

with similar baseline characteristics. Among 2,162,119 patients, 638,625 patients (29.5%) used statin and 1,523,494(70.5%) did not take any statin. In the propensity-score matched cohort (479,002 pairs), statin use was associated with a higher risk of T2D compared with statin non-use (13.5 vs 7.0 events per 1000 person-years, respectively; hazard ratio [HR], 1.89; 95% CI, 1.85–1.93; P<0.001). The use of statin was found to be associated with a reduced risk of the CVD (HR, 0.86; 95% CI, 0.78–0.95; P<0.001) compared to statin non-use. This study suggests the harm and benefit of statin use for T2D and CVD among Asian population whose risk for cardiovascular disease is relatively low.

PC7 - W65: Relationship of entecavir adherence with clinical outcomes in Korean patients with chronic hepatitis B

Authors: Hyo Jeong Kim[1], Minjung Ko[1], Songhee Cho[1], Jayoun Lee[1], Youngsuk Lim[2].
[1] *National Evidence-Based Healthcare Collaborating Agency, South Korea,* [2] *Asan Medical Center, South Korea.*

Aims: Medication adherence is an important factor for the success of treatment for chronic hepatitis B (CHB). Though, Poor adherence is a frequent occurrence in treatment for chronic disease. This study aimed to examine the association between medication adherence and clinical outcomes in patients with CHB.

Methods: The claims data was provided by National Health Insurance Service of Korea. Study participants were adult patients with CHD, treated with Entecavir from January 2005 through June 2015. We performed a retrospective analysis by multivariable Cox proportional hazards model. Multivariable analyses showed that risk of death or transplantation and HCC depending on level of proportion of days covered (PDC).

Results: A total of 49,276 patients with CHB were included. During the study period, 2,785 patients (5.7%) died or received a liver transplant, and 6,297 (12.78%) developed hepatocellular carcinoma (HCC). The group with PDC of 90% or higher had a significantly lower risk of developing HCC (hazard ratio [HR], 0.41; 95% confidence interval [CI], 0.36–0.48), compared to the group with PDC of less than 50%. A similar result was appeared for risk of death or liver transplantation (HR, 0.80; 95% CI, 0.71-0.90).

Conclusions: In a retrospective study of patients with chronic hepatitis B virus infection, excellent medication adherence was associated with a significantly lower risk of death or liver transplantation as well as risk of developing HCC. Therefore, it needs to be emphasized that clinicians should routinely address adherence issues for patients with CHB.

Invited Session

IS8: Clinical Trial Simulations: the When, Where and What

Wednesday 12th July - 17.00-18.30 h. - Room: Auditorio
Chair: Mouna Akacha
Organised by Geraldine Rauch, *University of Heidelberg, Germany*

IS8-1: Parametric and Nonparametric Bootstrap Methods for General MANOVA

Frank Konietschke, *University of Dallas, USA*

In this talk, we discuss parametric and nonparametric bootstrap methods for multi-factor multivariate data, without assuming normality, and allowing for covariance matrices that are heterogeneous between groups. The newly proposed, general procedure includes several situations as special cases, such as the multivariate Behrens-Fisher problem, the multivariate one-way layout, as well as crossed and hierarchically nested two-way layouts. We derive the asymptotic distribution of the bootstrap tests for general factorial designs and evaluate their performance in an extensive comparative simulation study. For moderate sample sizes, the bootstrap approach provides an improvement to existing methods in particular for situations with nonnormal data and heterogeneous covariance matrices in unbalanced designs. For balanced designs, less computationally intensive alternatives based on approximate sampling distributions of multivariate tests can be recommended.

Reference 1: Konietschke, F., Bathke, A., Harrar, S.W., Pauly, M. (2015). Parametric and Nonparametric Bootstrap Approaches for General MANOVA. Journal of Multivariate Analysis, 140, 291 -- 301.

IS8-2: A Framework for Simulations in Clinical Research with Applications in Small Populations and Rare Diseases

Tim Friede, *Georg-August University of Göttingen, Germany (Co-author: Norbert Benda, BfArM, Bonn)*

With rising pressure on resources for clinical trials and shifting patient populations there is an increasing demand for efficient and yet robust clinical trials. As a consequence the way clinical trials are planned, conducted and analysed is changing with a move to more complex designs and analysis methods, which in turn leads to more frequent use of Monte Carlo simulations. The so-called clinical scenario evaluation (CSE) framework was developed as an approach to structure such simulation studies and their reporting (Benda et al, 2010; Friede et al, 2010). In this presentation we start by introducing the elements of the CSE framework. Then we will discuss the application of the CSE framework to small populations and rare diseases. Examples for complex clinical plans in rare diseases in small populations include the use of external controls by means of power priors or (Bayesian) hierarchical models, extrapolation from larger population to a smaller one (e.g. adults to children) using PK/PD modelling or evidence synthesis methods formally combining treatment effect estimates from various sources.

Acknowledgement: EU's FP7 grant agreement number FP HEALTH 2013 – 602144 „Innovative methodology for small populations research" (InSPiRe).

Reference 1: Benda N, Branson M, Maurer W, Friede T (2010) Aspects of modernizing drug development using scenario planning and evaluation. Drug Information Journal 44: 299-315.

Reference 2: Friede T, Nicholas R, Stallard N, Todd S, Parsons N, Valdés-Márquez E, Chataway J (2010) Refinement of the clinical scenario evaluation framework for assessment of competing development strategies with an application to multiple sclerosis. Drug Information Journal 44: 713-718.

IS8-3: Simulating Recurrent Event Data with a Calendar Time Scale

Antje Jahn, *University of Mainz, Germany*

Simulation techniques are an important tool for investigating the use of statistical models for particular clinical applications. We specifically consider simulation of recurrent events in the presence of a competing terminal event. Such data are frequently collected in cardiovascular clinical trials, where treatment effects both on morbidity (assessed as recurrent non-fatal hospitalizations) and mortality (assessed as cardiovascular death) are to be evaluated. First, we consider hazard functions that vary over calendar time and thus reflect risks that vary with the time under treatment. In contrast to a gap time model, inter-event-times in calendar time models are not independent conditionally on the covariates. They are to be simulated recursively using their distribution conditional on the times of previous events. Second, we allow transition hazards for the recurrent and terminal events to be associated. This is important to reflect that an increased risk for hospital admissions can be associated with an increased risk for mortality and that events causing hospitalization can affect the risk for further recurrent or terminal events. We find that only complex simulation models can show important differences in treatment effect estimates that are derived from statistical models proposed for heart failure trials (composite endpoints, multistate models, joint frailty models). These findings can improve the interpretation of clinical trial results in heart failure disease.

Reference 1: Jahn-Eimermacher A, Ingel K, Ozga AK, Preussler S, Binder H: Simulating recurrent event data with hazard functions defined on a total time scale. BMC Medical Research Methodology 2015; 15:16.

Oral Contributed Sessions

OC39: Bayesian methods in clinical research 5

Wednesday 12th July -16.30-18.00 h. - Room: Sala Mar 2
Chair: Peter Müller

OC39-1: The hierarchical meta-regression approach and learning from clinical evidence

Authors: Pablo Emilio Verde[1].
[1]*Coordination Center for Clinical Trials, University of Düsseldorf, Germany.*

The Hierarchical Meta-Regression (HMR) approach (Verde et al. 2015) is a multi-parameters Bayesian approach for meta-analysis, which generalizes the standard random-effects meta-regression model. The HMR allows to investigate the potential external validity of the historical data as well as to assess the internal validity of the studies included in a systematic review. The HMR automatically identifies studies presenting conflictive evidence and it down-weights their influence in the conclusions of the meta-analysis. Under this approach we can perform Cross-Evidence-Synthesis, which combines aggregated results from randomize controlled trials (RCTs) with individual participant data (IPD) from observational studies. In this way, we can gain new insights from RCT's results which cannot be seen using only a meta-analysis of RCTs. We applied the HMR approach to a case study where RCTs' results, investigating efficacy in treatment of diabetic foot problems, are extrapolated to groups of patients treated in medical routine and who were enrolled in a prospective cohort study.

Reference 1: Verde, P.E, Ohmann, C., Icks, A. and Morbach, S. (2015) Bayesian evidence synthesis and combining randomized and nonrandomized results: a case study in diabetes. Statistics in Medicine. Volume 35, Issue 10, 10 May 2016, Pages: 1654–1675.

OC39-2: A Bayesian multivariate latent t-model for assessing the effects of intracranial radiotherapy and corticosteroid therapy on cardiometabolic risk factors in survivors of chilhood acute lymphoblastic leukemia

Authors: Miguel Caubet Fernández[1], Mariia Samoilenko[2], Simon Drouin[2], Emile Levy[2], Caroline Laverdière[2], Maja Krajinovic[2], Daniel Sinnett[2], Valérie Marcil[2], Geneviève Lefebvre[1].
[1]*Université du Québec À Montréal, Department of Mathematics, Montréal, Canada,* [2]*Centre Hospitalier Universitaire Sainte-Justine, Montréal, Canada*

Many childhood acute lymphoblastic leukemia (cALL) survivors experience chronic adverse effects often not clinically apparent until decades after treatment. Our objective was to investigate the association between treatment and cardiometabolic risk factors in a young cALL survivor cohort (n=180). We assessed the combined effects of cranial radiotherapy (CRT) and cumulative corticosteroid

dose (low/high dose) received on the prevalence of obesity, insulin resistance, (pre-)hypertension, and dyslipidemia. The prevalence of these cardiometabolic risk factors ranged from 10% to 39% in our cohort.

Treatment effect estimation was performed using the multivariate model for correlated binary outcomes of O'Brien and Dunson (2004). An MCMC algorithm was devised to sample from the posterior distribution of the model's parameters. Risk differences for each of the 16 possible combinations of the four binary metabolic outcomes were computed as in Hund et al. (2015).

The adjusted odds ratio for dyslipidemia associated with high dose of corticosteroids and CRT, as opposed to low dose without CRT, was 2.88 (95% IC: 1.25-5.72). cALL survivors that received CRT and high dose of corticosteroids had an increased risk of developing dyslipidemia alone, or dyslipidemia with (pre-)hypertension.

In summary, our analyses reveal that treatment could be an important contributing factor in dyslipidemia prevalence amongst cALL survivors, but not for insulin resistance or obesity.

Reference 1: O'Brien SM, Dunson DB (2004). Bayesian multivariate logistic regression. Biometrics 60: 739-746.

Reference 2: Hund L et al. (2015). A Bayesian framework for estimating disease risk due to exposure to uranium mine and mill waste on the Navajo Nation. JRSS-A 178:1069-1091.

OC39-3: Bayesian clinical trial design for rare diseases; prior opinion elicitation for chronic rheumatoid multifocal osteomyelitis

Authors: Despina Vasileiou[1], Lisa Hampson[1], Thomas Jaki[1].
[1] *Lancaster University, UK.*

Chronic Rheumatoid Multifocal Osteomyelitis (CRMO) is a rare painful bony condition that occurs in multiple sites on the body and causes loss of or impaired function of the associated areas, primarily affecting children and adolescents. However, the best treatment for it remains unknown, which results to differences in clinical practice amongst different hospitals and countries. It is therefore of interest to investigate the efficacy of the two most commonly adopted approaches for treatment and pain management. Despite the lack of published data on both pamidronate and adalimumab it is believed that there is expert prior knowledge on the treatment effects. Furthermore, the rarity of the disease prohibits the use of a frequentist design in a clinical trial, as the desired properties in power would necessitate the inclusion of an infeasibly large number of patients. Thus, a Bayesian design approach, that encompasses expert prior opinion, was adopted with a pragmatic total sample size of 40 patients. The design uses a conjugate normal-gamma model for the modelling of the change in pain scores from baseline. Prior opinion is elicited using bespoke software created via R-shiny by matching expert responses to the best fitting predictive prior distribution for the pain score at 26 weeks and subsequently mapping it to the mean change in pain score after 26-weeks for both treatments and their difference. We present the methods for the analysis behind the program created for the prior elicitation meeting and the resulting findings. Finally, we conclude with the presentation of the results of a subsequently conducted simulation analysis on the impact of the realisation of possible trial scenarios on key design features.

OC39-4: Extrapolation in meta-analysis

Authors: Christian Röver[1], Beat Neuenschwander[2], Simon Wandel[2], Tim Friede[1].
[1] *University Medical Center Göttingen, Germany,* [2] *Novartis Pharma Ag, Switzerland*

When data are sparse, extrapolation is a promising approach to making use of related external information [1]. In the context of meta-analysis, one is commonly faced with a small number of studies, while potentially relevant additional information may also be available. We describe a simple extrapolation strategy using heavy-tailed mixture priors [2] for effect estimation in a meta-analysis. The simple model setup is easily interpretable and leads to robust inference. We illustrate the method using examples of extrapolation from adults to children, and utilizing the "bayesmeta" R package.

Reference 1: European Medicines Agency (EMA). Reflection paper on extrapolation of efficacy and safety in pediatric medicine development, April 2016. EMA/199678/2016.

Reference 2: H. Schmidli, S. Gsteiger, S. Roychoudhuri, A. O'Hagan, D. Spiegelhalter, B. Neuenschwander. Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics, 70(4);1023-1032, 2014.

OC39-5: Dealing with baseline hazard functions in Bayesian joint models

Authors: Elena Lázaro[1], Carmen Armero[1], Danilo Alvares[1], Montse Rué[2].
[1] *Universitat de València, Department of Statistics and Operational Research, Spain,*
[2] *Universitat de Lleida-Irblleida, Department of Basic Medical Sciences, Spain.*

Joint modelling for longitudinal and time-to-event data focuses on the relationship between both processes. In studies with interest on survival, the most commonly proposed structure uses a longitudinal mixed effects model that estimates the trajectory of the time-dependent predictor, and incorporates this information into a Cox regression model. We focus on the baseline hazard function and discuss various parametric and non-parametric proposals that come from the Bayesian survival approach and apply to the Bayesian joint models. In particular, we consider a Weibull distribution as a parametric choice, and piecewise constant functions and B-splines basis functions as non-parametric approaches.

We illustrate those proposals by means of a Bayesian joint model for a survival study for patients at intensive care units (ICUs). We use a longitudinal mixed effects submodel and a survival competing risk submodel to assess the relationship between a severity index and the risk of death in the ICU or alive discharge from the ICU. Markov chain Monte Carlo methods are used to approximate the subsequent posterior distribution by means of the WinBUGS software. Model comparison is evaluated in terms of the Deviance Information and the Log Pseudo-Marginal Likelihood criteria.

Reference 1: Armero, C., Forné, C., Rué, M., Forte, A., Perpiñán, H., et al. (2016). Bayesian joint ordinal and survival modelling for breast cancer risk assessment. Statistics in Medicine 35(28): 5267-5282.

Reference 2: Ibrahim, J. G., Chen, M. H., & Sinha, D. (2005). Bayesian survival analysis. John Wiley & Sons, Ltd.

OC40: Joint modelling in practice 3

Wednesday 12th July -16.30-18.00 h. - Room: Sala Mar 4
Chair: Giota Touloumi

OC40-1: Dynamic predictions from joint models for multiple longitudinal markers correlated to interval-censored competing events

Authors: Anaïs Rouanet[1], Hélène Jacqmin-Gadda[1].
[1] *Inserm, Umr 1219, Univ. Bordeaux, Isped, Bordeaux, France.*

Analyses on predictive factors of dementia are often based on repeated cognitive tests and time-to-dementia onset data, which are correlated. Moreover, in elderly cohorts on dementia, the time-to-dementia is interval-censored as subjects can be diagnosed at visit times only. As subjects with dementia are at higher risk to die, they may die rapidly, without being diagnosed. Besides, the literature highlighted a high heterogeneity in cognitive declines. The joint latent class model developed by Rouanet et al. (2016) handles longitudinal data correlated to interval-censored competing events. It also captures heterogeneity by considering subgroups of the population with homogeneous longitudinal trajectories and risks of event. Based on the repeated measurements collected up to a landmark time, predictions of the risk of dementia at a given horizon are computed for each subject, and updated after each new measurement. These dynamic predictions can be evaluated by the dynamic AUC and Brier Score. This work presents a predictive model built and validated on two independent cohorts, providing dynamic predictions of the risk of dementia at a 5-year horizon, derived from a single or multiple cognitive tests. We quantify the gain in terms of predictive accuracy of the combination of MMSE, ISAACS and Benton tests to predict dementia occurrence. This type of models allows to detect earlier high risk individuals, but would also be useful for precision medicine, to help clinicians make informed decisions based on the partially observed cognitive trajectory of a patient.

Reference 1: Rouanet et al. (2016). Joint latent class model for longitudinal data and interval-censored semi-competing events: Application to dementia, Biometrics, 72(4): 1123-1135.

OC40-2: Joint models for longitudinal and time-to-event data in a case-cohort design

Authors: Sara Baart[1], Dimitris Rizopoulos[1], Eric Boersma[1].
[1] *Erasmus MC, The Netherlands.*

Longitudinal measurements are becoming increasingly more popular in clinical research, because estimated temporal patterns can be used to improve prediction of outcomes. A popular approach in combining longitudinal and time-to-event data is the joint modelling approach. Often a set of multiple biomarkers is measured to discover new biomarkers predictive for the outcome. Costs associated with assessing all biomarker values, however, can become exceedingly high. If, in addition, event

rates in the study are low and most information is to be expected from the patients experiencing the event (cases), it may be more cost efficient not to assess all biomarkers. For this research we are motivated by the BIOMArCS study, where patients admitted for acute coronary syndrome (ACS) are followed for one year to study the relation between temporal patterns of multiple biomarkers and recurring ACS. The BIOMArCS study follows a case-cohort design in which a random subcohort of patients is selected and supplemented with the other cases outside the subcohort. In standard survival models, weighting schemes have been proposed to account for the overrepresentation of cases in such designs. In the framework of joint modelling different approaches are needed. We propose to include survival information and any potential baseline covariate information of all patients in the analysis. The controls outside the subcohort will have missing values for the biomarker measurements. However, since the subcohort was chosen at random, the missingness mechanism is MAR, and hence results obtained from the joint model fitted in the constructed data set will remain valid. We evaluate this procedure with simulations and illustrate its use in the BIOMArCS study.

OC40-3: Bayesian joint models for multiple longitudinal biomarkers and a time-to-event outcome: software development and a melanoma case study

Authors: Sam Brilleman[1], Michael J Crowther[2], Margarita Moreno-Betancur[3], Serigne Lo[4], Rory Wolfe[1].
[1] *Monash University, Australia,* [2] *University of Leicester, UK,* [3] *Murdoch Childrens Research Institute, Australia,* [4] *The University of Sydney, Australia.*

Methodological developments in the joint modelling of longitudinal and time-to-event data abound. Implementations of various versions of this methodology now enable researchers to fit joint models using standard statistical software packages. Yet the software options available to users remain limited in several respects. In particular, situations in which there are multiple longitudinal markers are not well accommodated. The modelling of longitudinal biomarker data in patients diagnosed with melanoma, and their association with the risk of clinical events such as death or disease progression, is one situation in which such extensions would prove useful. A framework for fitting a multivariate joint model can be specified as follows. The longitudinal biomarkers are each modelled using a generalized linear mixed model, whilst the time-to-event is modelled through a parametric proportional hazards regression model. Dependence between the multiple biomarkers within a subject is captured through shared subject-specific random effects following a multivariate normal distribution. Dependence between each biomarker and the time-to-event can be parameterised in a variety of ways. This project motivated the development of general purpose software that allows users to fit these multivariate joint models under a Bayesian approach, now implemented as part of the rstanarm R package. Using data from the Melanoma Institute Australia, I will demonstrate how these models can be used to provide insights into the relationships between patient-specific longitudinal trajectories for biomarkers such as LDH, neutrophils and lymphocyte counts, and the risk of either death or disease progression.

OC40-4: Practical methods to pool multi-study joint longitudinal and time to event data

Authors: Maria Sudell[1], Dr. Ruwanthi Kolamunnage-Dona[1], Dr. Catrin Tudur Smith[1].
[1] *University of Liverpool, UK.*

Background: Joint longitudinal and time-to-event data models have been established in a single study case as beneficial compared to separate longitudinal or time-to-event analyses in a range of cases, including data with study dropout, time-to-event models with longitudinal covariates measured with error, or cases when the relationship between longitudinal and time-to-event outcomes is of interest. However the methodology available for multi-study cases such as meta-analyses is limited. Aims: To investigate different approaches of modelling of multi-study joint longitudinal and time-to-event outcome data. Methods: Several methods are examined to account for between study heterogeneity, including as one stage methods that can include random effects at the study level, stratification of baseline hazard by study and use of fixed study indicator terms and their interactions with treatment assignment, or approaches for two stage pooling of joint model fits. These methods are applied to a real data example and further investigated in a simulation study. Software have been developed in R to allow these methods to be easily applied in future investigations, which will be available in a package alongside joineR collaboration. Results: The results from the real data example and simulation study will be presented at conference.

OC40-5: Enhancing regression nomograms to provide a graphical representation of disease risk prediction regression models

Authors: Roger Marshall[1].
[1]*The University Of Auckland, New Zealand.*

Nomograms have, in recent years, been suggested as a way to calculate risk from a regression model [eg. 1]. However, computers and apps have made their use as calculators questionable. Regression nomograms may, however, be useful graphical representations of regression models to show the relative contribution of variables to a regression outcome. But they are also limited in this respect [2] unless the distribution of the predictor variables is somehow also represented. This presentation will show how, by adding the data distribution of predictor variables to the nomogram, it is very easy to immediately "see" how each variable contributes to the outcome. Box-plots, violin, or bean plots are suggested to represent the distribution, or scaled squares for factor variables. The method will be illustrated by different examples, based on ordinary, logistic and Cox survival models. The method is available in the author's R package – regplot. Further, the nomogram can be enhanced by superimposing an actual observation on the plot. The resulting graphic can be animated to immediately show how, the predicted outcome is affected by clicking a change in one or more predictors.

OC41: Topics in biostatistics 2

Wednesday 12th July -16.30-18.00 h. - Room: Sala Terra 4
Chair: Montse Rué

OC41-1: Simultaneous inference for multiple marginal GEE models with small sample sizes

Authors: Robin Ristl[1], Ludwig Hothorn[2], Christian Ritz[3], Martin Posch[1].
[1] *Medical University of Vienna, Austria,* [2] *Leibniz Universität Hannover, Germany,*
[3] *University of Copenhagen, Denmark.*

Motivated from two small-sample studies in ophthalmology and dermatology we study the problem of simultaneous inference for multiple endpoints in the presence of dependent observations. We propose a framework in which a generalized estimating equation model is fit for each endpoint marginally, taking into account dependencies within the same subject. Based on the work by Pipper et al. [1], the asymptotic joint normality of the stacked vector of marginal estimating equations is used to derive Wald-type simultaneous confidence intervals and hypothesis tests for linear contrasts of regression coefficients of the multiple marginal models. The small sample performance of this approach is improved by adapting the bias correction proposed by Mancl and DeRouen [2] to the estimate of the joint covariance matrix of the regression coefficients from multiple models. As a further improvement a multivariate t-distribution with appropriate degrees of freedom is specified as reference distribution. Alternatively, a generalized score test based on the stacked estimating equations is derived. Simulation results show control of the type I error rate for these methods even with small sample sizes and increased power compared to a Bonferroni multiplicity adjustment. Thus, the proposed methods are suitable to efficiently use the information from dependent observations of multiple endpoints in small-sample studies.

Reference 1: C.B. Pipper, C.Ritz, H. Bisgaard (2012) A versatile method for confimatory evaluation of the effects of a covariate in multiple models. JRSS C-Applied Statistics, 61:315-326, 2012.

Reference 2: L.A. Mancl, T.A. DeRouen (2001) A covariance estimator for GEE with improved small sample properties. Biometrics, 57(1), 126-134

OC41-2: Elicitation of the knowledge of several experts to derive informative prior distributions on biological parameters in radiation epidemiology

Authors: Sabine Hoffmann[1], Sophie Ancelet[1], Chantal Guihenneuc[2].
[1] *Institut de Radioprotection et de Sûreté Nucléaire, France,* [2] *Université Paris Descartes.*

In radiation epidemiology, a Bayesian hierarchical model may be used to account for exposure measurement error and dose uncertainty in the estimation of lung cancer risk due to the inhalation

of radon and its short-lived progeny in uranium miners. In this hierarchical model, some unknown input parameters will be well informed by the data, leading to informative posterior distributions even if vague priors are used. Other parameters are only poorly informed by the data. For these latter parameters, it seems indispensable to consider other sources of information and to elicit prior distributions that realistically reflect prior knowledge. We conducted standardized interviews with three experts on the exposure conditions in the French uranium mines to derive informative prior distributions on specific input parameters, such as average breathing rate. In these interviews, we chose a process of indirect elicitation, where experts were confronted with binary choices to derive probability distributions that accurately reflect their prior knowledge. Moreover, we combined this method with the possibility of direct feedback to rate two alternative probability distributions, which were in accordance with the information given by the experts. To combine the information of several experts, we used a hierarchical and probabilistic approach. We compare this approach with pooling and averaging, which are two alternative approaches to combine information from several experts. Finally, we propose a framework that would allow to combine the information obtained in the elicitation process with information available in the scientific literature.

OC41-3: Getting relevant groups of bacteria in microbiome analysis

Authors: Javier Rivera[1], Marc Noguera-Julian[1], Vera Pawlowsky-Glahn[2], J.J. Egozcue[3], Roger Paredes[1], M. Luz Calle[4].
[1] *Irsicaixa, Spain,* [2] *Universitat de Girona, Spain,* [3] *Universitat Politècnica de Catalunya, Spain,* [4] *Universitat De Vic - UCC, Spain.*

The development of high-throughput sequencing technologies has revolutionized the field of microbiome research by allowing a more precise quantification of microbiome abundances in a given environment.

However, microbiome data analysis poses important challenges. Mathematically, a microbiome data set consists of a matrix of counts corresponding to the number of DNA sequence reads assigned to each microbiome taxonomic unit. Because the different number of reads for each sample is related to the sequencing experiment and not to the total microbiome abundance, microbiome data is intrinsically compositional, carrying only relative information among the features. Classical statistical tools that ignore the compositional nature of microbiome data may provide invalid or inconsistent results.

One of the most important notions coming from the mathematical foundations of compositional data analysis is the concept of balance between groups of parts of a composition. In this work we transfer this mathematical concept to the field of microbiome research where it acquires a special ecology meaning.

As an alternative to common approaches that search for individual differentially abundant microbiome species, we propose an algorithm to find groups of microbiome species whose balance is differentially abundant or associated to a response variable or phenotype of interest.

OC41-4: Alfa coefficient: an alternative measure of categorical agreement among many raters

Authors: María Álvarez Hernández[1], Antonio Martín Andrés[2].
[1] *University of Vigo, Spain,* [2] *University of Granada, Spain.*

The most common measure of agreement for categorical data is the Kappa coefficient. This coefficient presents various problems, one of which is its poor performance when the marginal distributions are strongly unbalanced. For this reason several coefficients have been proposed in the literature, being the Delta coefficient one of the most recent, effective and cited. The coefficient is valid in any circumstance, but has the difficulty of being defined assuming that one of the raters is a standard; this causes that the parameters of the support model cannot be directly related to what is intended to be measured. The new Alfa coefficient proposed here, which is not based on the presence of a standard, produces the same results of the Delta coefficient. Additionally the model is extended to the case with more than two raters.

OC41-5: Robust inference under the beta regression model with application to health care studies

Authors: Abhik Ghosh[1].
[1]*Indian Statistical Institute.*

Data on rates, percentages, proportions arise frequently in many applied sciences like medical biology, health care, psychology & many others. In order to better understand the underlying data-generating mechanism and to predict such responses taking values in (0,1) using some related explanatory variables, a beta regression model has become very useful; but the existing inference procedures are based on the classical maximum likelihood approach. Although asymptotically optimum, they have serious lack of robustness against outliers in data and generate drastically different (erroneous) inference in presence of data contamination. Since such outliers are not uncommon in real-life datasets, we need to be very cautious before using them in practice.

In this paper, we develop a robust inference procedure for the beta regression model that can automatically take care of the outliers to yield stable inference. We follow the minimum divergence approach using the density power divergence; the resulting estimator has become very popular due to its high asymptotic efficiency along with strong robustness properties. Here, we develop the robust minimum density power divergence estimator for the beta regression model along with several applications. We derive the asymptotic properties of the proposed estimator and describe its robustness properties theoretically through the influence function analysis. Finite sample performance of the proposed estimator is examined through suitable simulation studies and several real data applications in the context of health care & psychology. Finally the applications of the proposed robust estimator in constructing robust tests of hypotheses under the beta regression model are briefly indicated.

OC42: Survival analysis 4

Wednesday 12th July -16.30-18.00 h. - Room: Sala Terra 2
Chair: Virginie Rondeau

OC42-1: Measuring the effectiveness of diagnostic markers in the presence of measurement error through the use of time-dependent ROC curve methodology

Authors: Adina Najwa Kamarudin[1], Ruwanthi Kolamunnage-Dona[1], Trevor Cox[1].
[1] *University of Liverpool, UK.*

The standard approach of ROC (receiver operating characteristic) curve analysis considers event (disease) status and marker value for an individual as fixed over time, however in practice, both the disease status and marker value can change over time. Individuals who are disease-free earlier may develop the disease later due to longer study follow-up. Based on our comprehensive review on time-dependent ROC curve analysis (Kamarudin et al., 2017), we found that all current estimation methods directly use the observed marker measurements by ignoring the presence of possible measurement error of the marker. In this work, we investigate the effectiveness of the observed marker, true marker and adjusted marker values by using time-dependent ROC curve methodology. We describe the relationship of those markers in an additive error model and estimate the adjusted marker by using a regression calibration method. Generally, the area under the ROC curve (AUC) for the observed marker is always overestimated which can be seriously misleading and useful markers might be overlooked. We compare the behaviour of this new method with the current estimation methods in a simulation study and using a real dataset.

Keywords: time-dependent ROC, biomarker, event-time, measurement error

Reference 1: Kamarudin AN, Kolamunnage-Dona R, Cox T: Time-dependent ROC curve analysis in medical research: current methods and applications. BMC Medical Research Methodology 2017, submitted.

OC42-2: Flexible parametric accelerated failure model

Authors: Steve Su[1].
[1]*Covance Pty Ltd, Australia.*

Due to the recent development of flexible parametric linear models using generalised lambda distributions (GLDs, Su 2015), the extension of these flexible models to accommodate log survival times with censoring is presented in this article. Unlike standard statistical distributions, the inverse quantile distribution of GLDs poses a different set of challenges in statistical modelling compared to standard statistical distributions. However, the rich shapes of GLDs make them very attractive tools for statistical modelling and the robustness of these regression models to outliers is demonstrated. Simulation studies demonstrating the consistency of these regression models for right skewed, left skewed and symmetric error distributions are presented along with comparisons to well-known Accelerated Failure models in the literature.

Reference 1: Su, S.(2015) "Flexible parametric quantile regression model" Statistics and Computing 25: 635. doi:10.1007/s11222-014-9457-

OC42-3: Quantifying the association between progression-free survival and overall survival in cancer trials

Authors: Enya Weber[1], Andrew Titman[1].
[1] *Lancaster University, UK.*

In oncology trials progression-free survival is often used as a surrogate endpoint for the traditional endpoint of overall survival. Defining the progression-free survival as the primary endpoint in a trial is efficient in terms of costs and time because potentially long follow-up periods after disease progression can be avoided. However, the extent to which the progression-free survival rate is effective is determined by the correlation between progression-free and overall survival. A review of existing approaches to estimating this correlation will be presented, including parametric model based approaches [1], semi-parametric copula models [2] and non-parametric methods based on inverse probability of censoring weights. The performance of these methods, in terms of bias and efficiency will be investigated through simulation and also illustrated using data from a clinical trial of treatments for colon cancer.

Reference 1: Li Y. & Zhang Q. (2015). A Weibull multi-state model for the dependence of progression-free survival and overall survival. Statist. Med., 34, 2497-2513.

Reference 2: Burzykowski T. & Molenberghs G. (2001). Validation of surrogate end points in multiple randomized clinical trials with failure time end points. Appl. Stat., 50(4), 405-422.

OC42-4: A probabilistic record linkage model for left truncated time-to-event data with competing risks

Authors: Michel Hof[1], Aeilko H. Zwinderman[1].
[1] *University of Amsterdam, The Netherlands.*

In absence of an unique identifier, combining information from multiple files relies on partially identifying variables (PIVs), such as gender or initials. With a record linkage procedure, these variables are used to distinguish record pairs that belong together (matches) from record pairs that do not belong together (non-matches). Generally, the combined strength of the PIVs is too low causing imperfect linkage; some true matches are identified as non-match and vice versa. To avoid bias in further analyses, it is necessary to correct for imperfect linkage.

For this study, we had access to a file with first deliveries (n=80000, gathered between 1-1-1999 and 31-12-1999) and a file with second deliveries (m=500000, gathered between 1-1-2002 and 31-12-2009) from the Perinatal Registry of the Netherlands. We were interested in the associations between the characteristics from the first delivery on the time and type (i.e. stillbirth, iatrogenic delivery, and

spontaneous birth) of second delivery. The three outcome types were considered to be competing risks for each other. In addition, the second delivery times were left truncated; all second deliveries before 1-1-2002 were not observed.

Due to privacy regulations, we had to rely on PIVs to determine which pregnancies belonged to the same mother. To deal with imperfect linkage in left truncated time-to-event data with competing risks, we developed a joint model in which the time-to-event submodel and the record linkage submodel are estimated simultaneously. We considered a marginal pseudo-likelihood approach in which the matching statuses are treated as missing data. If both submodels are correctly specified, our joint modeling approach gives unbiased parameter estimates.

OC42-5: Flexible accelerated failure time model for right censored data in survival analysis

Authors: Menglan Pang[1], Michal Abrahamowicz[1, 2], Robert W. Platt[3, 4].
[1] *Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada,* [2] *Division of Clinical Epidemiology, McGill University Health Centre, Montreal, Quebec, Canada,* [3] *Department of Epidemiology, Biostatistics and Occupational Health, and Department of Pediatrics, McGill University, Montreal, Quebec, Canada,* [4] *The Research Institute of The McGill University Health Centre, Montreal, Quebec, Canada.*

Background: The Cox proportional hazards (PH) model is the most widely used method to analyze time-to-event data in medical research, while the accelerated failure time (AFT) model has been suggested as an interesting alternative. However, a parametric AFT model requires the specification of an appropriate distribution for the event time, which is often difficult to identify in real-life studies and, thus, may limit applications.

Methods: We develop a flexible semiparametric AFT model for right censored survival data. The baseline hazard was modeled by regression B-splines. In comprehensive simulations, we validate the performance of our approach in terms of effect estimates and survival probabilities, and compare with the results of parametric AFT models and the Cox model with the Breslow estimator. Goodness of fit of alternative models is compared using AIC. Furthermore, a real-life epidemiological study will illustrate the application of this method.

Results: The parametric AFT model yielded the optimal results when the distribution of the event time was correctly specified, and the Cox model provided unbiased estimates and survival curves when the event times were generated so that the PH assumption was satisfied. However, when the distribution was miss-specified in the parametric AFT model and when the PH assumption was violated, the estimated covariate effects were biased and the estimated survival curves deviated from the truth. In contrast, the proposed flexible AFT model always provided approximately unbiased effects and survival curves.

Conclusion: Our flexible AFT model provides a useful approach to analyze survival data, and can provide insights regarding how a prognostic factor affects survival.

Poster Contributed Session

PC8: Biomedical studies 2

Wednesday 12th July -16.30-18.00 h. - Room:Hall
Chair: Jeanine Houwing-Duistermaat

PC8 - W3: Unemployment, obesity and economic crisis

Authors: Pilar Magdalena[1], M. Antelo[1], J. C. Reboredo[1].
[1] *University of Santiago de Compostela, Spain.*

We analysed the impact of unemployment on the prevalence of adult obesity in Spain andinvestigated whether this impact differed significantly in boom times versus recession periods.We used Spanish National Health Survey data for 2006 and 2011 to represent periods of economic upturn and downturn, respectively. We modelled the data using propensity score matching techniques to measure the effect of unemployment on obesity in adults aged 16-65 years and using difference-in-difference techniques to study whether the economic crisis changed the impact of unemployment's effect on obesity. The results reveal significant differences in obesity prevalence depending on the labour-market status of the individual in booth economic growth and recession periods; this difference was heightened for the longer-term unemployed. In times of economic recession, unemployment in a crisis period increases obesity prevalence by 2.7 points of probability compared to unemployment in a boom period; this figure is 4.1 for individuals unemployed for 6-12 months. Lower education levels and physical inactivity significantly increased adult obesity prevalence. An economic crisis combined with unemployment consequently has an aggravated negative impact on health.

PC8 - W6: An evaluation of methods for combining dependent p-values in genome-wide association studies

Authors: Ozan Cinar[1], Wolfgang Viechtbauer[1], Ruud Van Winkel[2].
[1] *Maastricht University, Netherlands,* [2] *Katholike Universiteit Leuven, Belgium.*

Examining the contribution of genetic factors to complex diseases is a common task in genome-wide association studies. However, due to the high number of single-nucleotide polymorphisms (SNPs) examined, the number of simultaneous tests and therefore the Type I error rate can easily inflate in association analyses. Furthermore, common methods for multiple testing correction, such as the Bonferroni method, can lead to overly conservative results. One possible solution to overcome this problem is shifting the focus of the analysis to higher structural levels, such as genes, to decrease the number of simultaneous tests. This process requires the combination of the p-values of the SNPs that belong to a gene to obtain a single p-value to test the joint null hypothesis at the gene level. There are common methods for combining p-values, such as Fisher's method, that assume independence among the tests. However, this assumption is known to be violated due to linkage disequilibrium (LD), the non-random association of alleles for different SNPs, which creates dependence among the p-values obtained from the SNPs. Several adjustments have been proposed in the literature to consider LD while combining the p-values.

In our study, we evaluated several adjustments to methods for combining p-values that consider LD. These adjustments include methods that estimate the effective number of tests, methods based on empirically-derived distributions, and two adjustments that directly account for the dependence (Brown's method and a generalized Stouffer's method). We examined Type I error rates and power of these methods on simulated data based on real LD values obtained from the TSI population in the HapMap phase 3 data set.

PC8 - W9: Evaluation of methods for gene selection in melanoma cell lines

Authors: Linda Chaba[1], John Odhiambo[1], Bernard Omolo[2].
[1] *Strathmore University, Kenya,* [2] *University of South Carolina-Upstate, USA.*

A major objective in microarray experiments is to identify a panel of genes that are associated with a disease outcome or trait. Many statistical methods have been proposed for gene selection within the last fifteen years. While the comparison of some of these methods has been done, most of them concentrated on finding gene signatures based on two groups. This study evaluates four gene selection methods when the outcome of interested is continuous in nature. We provide a comparative review of four methods: the Statistical Analysis of Microarrays (SAM), the Linear Models for Microarray Analysis (LIMMA), the Lassoed Principal Components (LPC), and the Quantitative Trait Analysis (QTA). Comparison is based on the power to identify differentially expressed genes, the predictive ability of the genelists for a continuous outcome(G2 checkpoint function), and the prognostic properties of the genelists for distant metastasis-free survival. A simulated dataset and a publicly available melanoma cell lines dataset are used for simulations and validation, respectively. A primary melanoma dataset is used for assessment of prognosis. No common genes were found among the genelists from the four methods. While the SAM was generally the best in terms power, the QTA genelist performed the best in prediction of the G2 checkpoint function. Identification of genelists depends on the choice of the gene selection method. The QTA method would be preferred over the other approaches in predicting a quantitative outcome in melanoma research. We recommend the development of more robust statistical methods for differential gene expression analysis.

PC8 - W12: A two-tailed hypothesis testing with a margin

Authors: Cheikh Ndour[1].
[1]*Université Catholique De Louvain, Belgique.*

This paper deal with a two-tailed test where under the null hypothesis the parameter of interest lies within an interval of small irrelevant values.A problem of classical two-tailed tests is that, with a sufficiently large sample, a statistical test will almost always demonstrate a significant difference, unless the effect size is exactly zero. It follows that very small differences, even if meaningless, can be significant. To solve this problem we propose a two-tailed hypothesis testing with a margin. Our proposal is explained, evaluated by simulations and applied to a published micro-array dataset.

Keywords: big data; large sample size; high-dimensional data; 2-Sided Equality test.

Reference 1: Sullivan G.M., Feinn R. (2012). Using E ect Size-or Why the P Value Is Not Enough.J Grad Med Educ. 4(3) 279-282. doi: 10.4300/JGME-D-12-00156.1.

Reference 2: Van der Laan M.J. and Rose S. (2010) Statistics Ready for a Revolution: NextGeneration of Statisticians Must Build Tools for Massive Data Sets. Amstat NewsSep. 2010.

PC8 - W15: Are single nucleotides the right scale for genome-wide association studies?

Authors: Florent Guinot[1], Christophe Ambroise[2], Marie Szafranski[3], Nathalie Jourdan[4].
[1] *Ueve, France,* [2] *CNRS, France,* [3] *Ensiie, France,* [4] *Bioptimize, France.*

Genome-Wide Association Studies (GWAS) aim to identify causal genomic variants implied in the expression of rare human diseases. From a statistical point of view, detecting these variants imply to perform hypothesis tests on the population subject to the rare disease (cases) against an healthy population (controls) at several locus on the genome. One individual's genome being caracterized by hundreds of thousands of SNPs, the type I error, resulting from a large number of hypothesis tests, can dramatically increase and lead to wrong conclusions about genetic associations with the disease.

Dimension-reduction methods are a way to improve the detection of true genetic associations by reducing the number of hypothesis to test. We thus propose a new dimension-reduction approach which can be applied in the context of GWAS by taking benefit of the structure in haplotype of the human genome.

For this purpose, we first cluster the SNPs with an hierarchical clustering algorithm using the linkage disequilibrium as a measure of dissimilarity. We then propose to :

1. Compress each cluster in the hierarchy into a unique variable which is built to reflect the number of mutation of each SNPs for each individuals compared to an average (most frequent) genotype.
2. Choose an appropriate cut into the tree with a ridge regression procedure. Replacing the initial genomic matrix with a lower-dimension predictors matrix allow us to perform fewer hypothesis tests and detect associations between the phenotype and clusters of SNPs with a much lower type I error.

PC8 - W18: Validation of classification and data reduction methods based on gene expression data

Authors: Soheila Khodakarim[1], Seyyd Mohammad Tabatabaei[1], Maryam Rafiei[2].
[1] *Shahid Beheshti University of Medical Sciences, Iran,* [2] *Arak University of Medical Sciences, Iran.*

The microarray technology has provided the monitoring of gene expression levels of thousands of genes, simultaneously [Harrington, 2000]. The analysis of these data sets is a problem in the century of bioinformatics revolution. The classifier methods such as data mining, machine learning

and regression have been applied to differentiate between normal and cancer samples in gene expression datasets, redundantly. This study aims to compare the accuracy of classifier methods ran on reduced gene expression data sets.

In order to compare the accuracy of some methods of classifying normal and abnormal samples, we used Support Vector Machine (SVM), Least Square Support Vector Machine (LSSVM), Radial Base Function Neural Network (RBFNN), Bayesian Probit Regression (BPR) and Bayesian Logistic Regression (BLR) as the classification methods on their two gene expression datasets and three reduced dimension sets. Data reduction was implemented with three approaches: Multivariate Median Gene Set Analysis (MMGSA), Principal Component with Karhunen–Loeve Transform (PCA_KL) and auto-encoder networks.

The accuracy of the SVM and LSSVM Methods does not depend on the kind of kernel. The accuracy of the SVM is 90% in PCA_KL and total data sets. The accuracy of the BKPR by Gaussian kernel is 94% in total data sets. The accuracy of the RBF is 90% in MMGSA and PCA_KL.

By using feature selection instead of feature exploration, MMGSA has the more simple interpretation for data reduction with sufficient accuracy. In addition, SVM and BKPR work better than the others in classification.

PC8 - W21: Ranking based variable selection for censored data using AFT models

Authors: Hasinur Rahaman Khan[1], Marzan Akhter[1].
[1] *University of Dhaka, Bangladesh.*

Variable selection for high dimensional data has become very attractive field of inquiry in modern statistical sciences. In case of high dimensional data, the interpretation of parsimonious models becomes more reliable in many fields including survival analysis because of the existence of censoring. There are many variable selection techniques that are developed for complete data but when censoring is present, the variable selection technique to be used for complete data must be modified because of the complex structure of the likelihood function. However, there is a number of variable selection approaches for two popular models-- Cox model and Accelerated failure time (AFT) model used for censored data. In this study, we modify the Ranking Based Variable Selection (RBVS) algorithm as proposed by Baranowski and Fryzlewicz (2014) as a variable selection technique for AFT models. In high dimensional-data, there are some subsets that contain irrelevant covariates and they may appear to have high impact on response variable. But the fact is that they are not consistently related to the response where the truly important variables will be related to the response consistently. The RBVS will overcome this problem using subsampling procedure. A real data example called Mantle Cell Lymphoma is used to identify the set of genes appearing non-spuriously at the top of a chosen variable ranking. A number of simulated examples have been conducted to demonstrate the performance of the revised RBVS method and to compare with the existing approaches in literature.

Reference 1: Baranowski, R. and Fryzlewicz, P. (2014). Ranking-based subset selection for high-dimensional data. Working Paper: URL http://personal.lse.ac.uk/baranows/rbvs.html

PC8 - W24: Functional data analysis of event-related potentials in assessing neural processing underlying mental rotation in time domain

Authors: Stanislav Katina[1, 2], Zdenka Gerslova[1], Igor Riecansky[2, 3].
[1] *Masaryk University, Czech Republic,* [2] *Slovak Academy of Sciences, Slovakia,*
[3] *University of Vienna, Austria.*

EEG is a method of recording electrical activity of the brain. Event-related potential(ERP) technique is an EEG method to assess brain activity with respect to a specific event or stimulus. We used ERPs to assess the neural processing underlying mental rotation(MR) of visual stimuli. Stimuli included rotated letters displayed in either canonical or mirror-reversed format. The task was to judge whether the stimulus was presented in one of the two formats. Stimuli were presented in random order and there were 80 trials in totals. EEG was recorded in 28 healthy volunteersfrom 61 equidistant scalp sites using electrodes mounted on an elastic cap. Data processing and analysis was carried out in the EEGLAB toolbox, MATLAB and R. After removing coarse artifacts, ICA was performed and components that separated artifactual signals(eye movements, eye blinks, ECG, muscle activity) were eliminated from the data. Each epoch was then baseline-corrected to the mean activity within 300 ms preceding stimulus onset. Trials in which response time was shorter than 300 ms and longer than 3000 ms were discarded. Epochs were truncated at the time of the onset of the probe stimulus. ERPs were calculated across all trials using methods of functional data analysis with signals aligned to the onset of the target stimulus and to the response time, and superimposed by Procrustes method of curve registration and spline smoothing by stimulus-response time. In this paper, these three alignment approaches are discussed with respect to signal-to-noise ratio, variability, and cognitive relevance. Supported by GACR project Nr. GA15–06991S.

PC8 - W27: The comparative performance of logistic regression and random forest in propensity score methods: a simulation study

Authors: Sara Khalid[1], M. Sanni Ali[1], Gary S Collins[1], Daniel Prieto-Alhambra[1].
[1] *Centre for Statistics in Medicine, Ndorms, University of Oxford, UK.*

Background: Propensity scores (PS) are typically estimated using logistic regression (LR). Machine learning techniques such as random forests (RF) are alternatives, but data are scarce on their comparative performance.

Objective: To compare LR and RF methods for PS estimation in terms of covariate balance, bias, and precision.

Methods: Simulation studies were conducted of binary covariates, treatment and outcome data. In several scenarios, different sample sizes, matching callipers and covariates were created. Treatment effect estimates were calculated after PS matching and inverse probability weighting (IPTW) using Poisson regression; covariate balance was checked using absolute standardized differences (ASD) pre and post-matching. Percentage bias (PB) was calculated for the proposed models, as were mean squared errors (MSE).

Results: LR was comparable to RF in terms of balance and bias when LR models included interactions and non-linear terms: 5% versus 4% ASD and 3% versus 1% PB (n=5000, calliper=0.05) respectively. Simpler LR models (including main terms only) resulted though in suboptimal balance (ASD=10%) and bias (PB=5%). PB increased in such LR models with stronger association between treatment/outcome, interactions, and non-linear terms. Both methods resulted in similar MSE for PS matching and IPTW when LR included interactions and non-linear terms.

Conclusions: PS estimation using LR should consider including interactions and non-linear terms in the model until an acceptable balance on covariates is achieved. Machine learning methods such as RF are useful alternatives.

PC8 - W30: Evaluation of potential bias of methods for selection of variables into propensity scores – a simulation study

Authors: Pär Karlsson[1].
[1]*Centre For Pharmacoepidemiology, Karolinska Institutet, Sweden.*

In pharmacoepidiologic studies there are a big concern about potential confounding. In a typical register study you could have in the order of 5000 potential confounding variables. One solution that is popular today is based High dimensional propensity scores. The high dimensional propensity scores are based on a subset of all potential confounding variables. Various selection methods exist for selecting the variables that should be included in this subset. Most of these method select variables based on the apparent dependence between a covariate and the outcome and/or the exposure.

This simulation study investigate the statistical properties of some of the selection methods. In the simulations 1000 confounding variables are generated together with an exposure variable and an outcome variable. All variables are dichotomous. Propensity scores are calculated for varying number of selected covariates, so the behavior of the method can be studied when the number of included variables are increasing.

The setting is such that the unadjusted estimate is biased, increasing the number of variables in the propensity score model will move the estimate closer to the true value, but if the variables are selected based on the dependence with the outcome or exposure, the estimate will cross the true value and then the bias will start to increase.

The conclusion of the simulation is that selection of covariates based on the apparent dependence on the outcome or exposure might give rise to biased estimates.

PC8 - W33: Multiblock sparse PLS for high dimensional data

Authors: Attila Csala[1], Michel H.P. Hof[1], Aeilko H. Zwinderman[1].
[1] *Dept. Clinical Epidemiology, Biostatistics and Bioinformatics Academic Medical Center, University of Amsterdam Amsterdam, The Netherlands.*

The goal of this study was to extract biological pathways (i.e. explanatory variables) from high-dimensional omics data which are related to phenotypes (i.e. response variables). A well-known technique to extract directional relationships from multiple sets of variables is Partial Least Squares path modelling (PLS).

In the PLS framework, a linear combination of the explanatory variables is found by iteratively regressing (a weighted combination of) the response variables on the current linear combination of the explanatory variables. Originally, PLS yields for the maximum correlation between the linear combinations of the explanatory and response variables, which indicates a non-directional relationship between omics data sets. One technique to extract directionality from multiple sets of variables is to find linear combinations of the explanatory variables that maximally correlate with all response variables. In this study, we aim to use multiblock PLS in omics data.

Because omics data is often characterised by high-dimensionality (i.e. number of explanatory variables is larger than the number of observations), standard PLS techniques cannot be used due to multicollinearity issues. As a solution, we proposed to incorporate elastic net (or other types) of regularization into PLS (by penalizing the regression steps), which we refer to as sparse Partial Least Squares path modelling (sPLS). By using the elastic net, which also contains a lasso type of regularization, it is possible to select a small set of explanatory variables to have easily interpretable results.

We used 3 data sets containing 485512, 18424, and 47 variables from 55 Marfan patients to identify genotypes that could explain particular phenotypes.

PC8 - W36: Goodness-of-fit and multiple comparison procedures for disorder detection in NGS experiments

Authors: Norman Jiménez Otero[1], Jacobo De Uña Álvarez[1].
[1] *Universidad de Vigo, España.*

Next-generation sequencing experiments (NGS) are often performed in biomedical research nowadays, leading to methodological challenges related to the high-dimensional and complex nature of the recorded data (Sathirapongsasuti et al. 2011). In this work we review some of the issues which arise in disorder detection from NGS experiments, that is, when the focus is the disorder (insertion/deletion) detection for homozygosity and heterozygosity in DNA sequencing. A statistical model to cope with documented sampling biases in such a setting (Benjamini and Speed 2012) is proposed, and a goodness-of-fit procedure for disorder detection is derived. The method combines the proper evaluation of local p-values (one for each DNA basis) with suitable corrections for multiple comparisons. The performance of the introduced procedure is investigated through simulations.

Reference 1: Sathirapongsasuti, J. F., Lee, H., Horst, B. A. J., Brunner, G., Cochran, A. J., Binder, S., ... Nelson, S. F. (2011). Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. Bioinformatics, 27(19), 2648–2654.

Reference 2: Benjamini, Y., & Speed, T. P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Research, 40(10), e72.

PC8 - W39: Fixed size confidence regions for the parameters of the linear mixed effects logistic model

Authors: Taoufik Zoubeidi[1].
[1]*UAE University, UAE.*

We develop fixed size confidence regions for estimating the fixed and random effects parameters of the mixed effects logistic regression model. This model applies to, among others, the study of the effects of covariates on a dichotomous response variable when subjects are sampled in clusters. Two sequential procedures are developed to estimate with a prescribed accuracy (confidence level) and fixed precision the set of fixed and random effects parameters and linear transformations of these parameters, respectively. We show that the two procedures are asymptotically consistent (i.e., the coverage probability converges to the nominal confidence level) and asymptotically efficient (i.e., the ratio of the expected random sample size to the unknown best fixed sample size converges to 1) as the width of the confidence region converges to 0. Suggestions to improve the performance of the procedures are provided based on Monte Carlo simulation and illustrated through a longitudinal clinical trial data.

PC8 - W42: Measurement error in exposure and confounder status in medical research: a systematic review of practices in high impact journals

Authors: Timo Brakenhoff[1], Marian Mitroiu[1], K.G.M. Moons[1], R.H.H. Groenwold[1], M. Van Smeden[1].
[1]*Umc Utrecht, The Netherlands.*

In medical research, measurements of exposure and confounder variables are often contaminated with some degree of measurement error. It is generally well accepted that prevalent measurement error in exposure or confounder measurements can harm inferences by introducing bias and/or imprecision in exposure-outcome associations. However, it is thus far unclear to what extent such issues are considered in research practice. Therefore, a systematic review was performed using full-text searching strategies of publications in high impact general medicine and epidemiology journals, to study common practices when confronted with measurement error. Results will be presented at the conference and include: (1) the percentage of articles that mention any form of the concept of measurement error in the manuscript; (2) the percentage of articles that make qualitative statements concerning the possible contamination of exposure or confounder status by measurement error; (3) the percentage of articles that applied one or more correction methods to account for measurement

error in the analyses; (4) an assessment whether assumptions that underlie the applied measurement error correction methods are adequately addressed and appropriately reported. This review aims to quantify the attention given to measurement error and its consequences in original applied research in high impact journals. In addition, practical guidance when dealing with possible measurement error in exposure or confounder variables will be provided.

PC8 - W45: Leveraging free text notes in the electronic medical record: an example identifying peripheral IV infiltrates using natural language processing

Authors: Carly Milliren[1], Victoria Mkpanam[1], Amir A. Kimia[1], Assaf Landschaft[1], Al Ozonoff[1].
[1] *Boston Children's Hospital, Boston, Ma, USA.*

The clinical narrative or text is often the sole source of critical data and may be more accurate than structured data. Text notes within the electronic medical record (EMR) contain data to improve patient care via quality improvement or clinical decision support. Useful data may be hard to extract from unstructured text without time-intensive manual chart review. Natural Language Processing (NLP) methods allow text mining for quality improvement, research and operational applications. We describe the use of NLP in a quality improvement project to identify peripheral intravenous infiltration events in a pediatric population, an application of NLP to safety surveillance in the inpatient setting.

We extracted over 21,000 free text inpatient nursing and intravenous team consult notes from the EMR system at Boston Children's Hospital from Jan–Jun 2016. We used a suite of NLP tools (DrT) to build a document classifier. We manually reviewed over 700 notes and identified regular expressions indicative of events (e.g. "Vv infiltrated") or non-events (e.g. "no infiltration or swelling"). We classified notes as events, non-events, or possible events and input the labeled dataset to a classifier that offered newly identified cases. We used a 70-30 train-test split sample validation to assess classifier performance. Combining both methods, we manually reviewed 775 notes and identified 96 events of which 45% were previously unidentified with our existing event reporting system.

This project illustrates how NLP methods can be applied to a novel source of clinical data which is otherwise inaccessible using other methods. Lay users can adopt intuitive tools for basic NLP methodology without a need for robust computing infrastructure.

PC8 - W48: Modifying the ggtree function to visualize IgG recognition strength of peptide variants in the context of their phylogenetic relationship

Authors: Verena Hoffmann[1, 2], Georgios Pollakis[3], Mohamed I.M. Ahmed[1, 2], Christof Geldmacher[1, 2].
[1] *Medical Center of the Ludwig Maximilian University of Munich, Munich, Germany,* [2] *German Center for Infection Research (DZIF), Partner Site Munich, Germany,* [3] *University of Liverpool, Liverpool, United Kingdom.*

Improved tools to visualize immune recognition of highly variable antigens, such as the HIV Envelope variable region 3 (V3) in the context of their phylogenetic relationship help to better understand the parameters influencing antigen variant recognition by IgG and hence contribute to improved design of HIV vaccines.

We studied HIV Envelope-specific IgG responses using a peptide array approach that included 36 V3-tip peptide variants (amino acids 304-319 in reference stran HxB2) using plasma samples from HIV vaccine recipients.

The R package GGTREE by Yu et al.(1) provides a general solution for visualization of phylogenetic trees and annotation using data of different types and from various sources. We then created a supplementary function that assesses the frequency of variant peptides sequences within the >40.000 global HIV primary isolate sequences deposited in the HIV sequence database (www.hiv.lanl.gov) and links this variable to the icon size of a given peptide variant sequence. Additionally, the magnitude of IgG recognition for a given peptide variant is visualized by heat map colour coding.

The function hence is a new tool for comprehensive visualization of the immune recognition of variant antigens in the context of their phylogenetic relationship and the respective frequency of the variant amino acid sequence in the global HIV epidemic in one graph.

Reference 1: Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. 2017;8(1):28-36.

PC8 - W51: Modelling the spread of gonorrhoea in an MSM population

Authors: Jozefien Buyze[1], Chris Kenyon[1].
[1]*Institute of Tropical Medicine, Belgium.*

There is considerable uncertainty as to the effectiveness and optimal timing of Neisseria Gonorrhoea(NG) screening in Men who have Sex with Men (MSM). NG has evolved resistance to a wide range of antibiotics, which makes it particularly important to ensure that NG screening in this population does not lead to excessive consumption of antibiotics. Our goal is to evaluate the effectiveness of different NG screening strategies on NG prevalence in an MSM population. Separable Temporal Exponential Random Graph Models are used to model the sexual relationships network with main and casual partners in MSM. Next

the transmission of NG is simulated on this dynamic network. We have adapted the standard model in the R package 'EpiModel' to include different infection statuses per person for the pharynx, urethra and rectum. Accordingly, different possible transmission routes (anal sex, oral sex and rimming, both active and passive) with their own act and transmission rate have been implemented. Furthermore a different recovery rate for symptomatic and asymptomatic infections was specified. The model was used to compare different screening programmes in terms of NG prevalence. Our models simulate day-by-day evolution of a population of 10000 MSM for 10 years. If one half of MSM is screened once a year, the prevalence of NG infection decreases from 13% to 10% (pharyngeal), 8% to 6% (urethral), and 16% to 12% (rectal), as compared to no screening. When only one third is screened, prevalence decreases to 11%, 7% and 13% respectively. The achieved prevalence reduction might not outweigh the larger risk of development of antibiotic resistance.

PC8 - W54: Method of interaction on site*subject level for central statistical monitoring

Authors: Andrey Myslivets[1], Tatiana Volodina[1], Aleksei Kiselev[1].
[1] *Data Matrix Ltd., Russia.*

Effective data cleaning strategies in clinical research is a really important question of budget - quality combination. Central Statistical Monitoring (CSM) based on advanced statistical methods is a cheaper and more efficient alternative to on-site data monitoring of all sites. There are many methods for checking and verifying data on different levels (Subject level, Site level, Country level and "Manually entered" data sources analysis). This paper presents method of interaction on Site*Subject level using the Set theory and Combinatorics for detecting suspicious and fraud data. Unique algorithm which uses the vectors of subjects clinical data allows to find the acceptable level of similarity based on the union of all study centres data and to compare subsets of subjects in each centre for solving the CSM problem. As an example, the case of events (0=No, 1=Yes) registered within a certain period of time is examined. This paper presents examples with endpoints based on binary data and the possibility of extending the algorithm to other types of data.

PC8 - W57: Estimating marginal proportions and intraclass correlations with clustered binary data

Authors: Josep Carrasco[1], Lluis Jover[1], Yi Pan[2], Rosa Abellana[1].
[1] *University of Barcelona, Spain,* [2] *Center for Disease Control and Prevention, USA.*

A logistic regression with random effects model is commonly applied to analyze clusteredbinary data, and every cluster is assumed to have a different proportion of success. However, it could be of interest to obtain the proportion of success over clusters, (i.e. the marginal proportion of success). Furthermore, the degree of correlation among data of the same cluster (intraclass correlation) is also a relevant concept to assess, but when using logistic regression with random effects it is not possible to get an analytical expression of the estimators for marginal proportion and intraclass correlation. Here, we assess and compare approaches based on different kinds of approximations, including estimators from generalized linear mixed modes and generalized estimating equations. The comparisons are completed by using a real data example and a simulation study.

PC8 - W60: Guidances for statistics in regulatory affairs in Wikipedia: initiative of the ISCB Statistics in Regulatory Affairs Subcommittee (SiRA SC)

Authors: Harbajan Chadha-Boreham[1], Christopher J. Weir[2], Nicole C. Close[3], Martin Schumacher [4], Tim Friede[5], Jonathan Siegel[6], Christoph Gerlinger[7], Ralf Bender[8], Christos T. Nakas[9], Jorgen Seldrup[10], Stanislav Katina[11], Juan V. Torres[12].
[1] *Clinical Biostatistics Consultancy, France,* [2] *University Of Edinburgh, Edinburgh, Eh8 9Yl Uk,* [3] *Empiristat, Inc., Mount Airy, Md 21771-3900 USA,* [4] *Institute for Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Freiburg, 79104 Germany,* [5] *Dept. Medical Statistics, University Medical Center Göttingen, Göttingen, 37073 Germany,* [6] *Bayer Healthcare Pharmaceuticals Inc., Whippany, Nj 07981 USA,* [7] *Bayer Ag, Berlin, 13353 Germany,* [8] *Institute for Quality and Efficiency in Health Care (IQWIG), Cologne, 50670 Germany,* [9] *Institute for Clinical Chemistry, Inselspital, Bern University Hospital, University of Bern, Bern, 3010 Switzerland,* [10] *Jstatconsult, Barr, 67140 France,* [11] *Institute of Mathematics and Statistics, Faculty of Science, Masaryk University, Brno, 61137 Czech Republic,* [12] *Syntax for Science S.L., Palma de Mallorca, 07121 Spain.*

Statistics in regulatory affairs guidance for the Pharmaceutical and Medical Devices Industry has been issued for more than 25 years world-wide. However, there is no unique shared free open-access depository that offers a synthetic view of the international-, regional- and country-level regulatory guidance. At the ISCB Utrecht conference (2015) the SiRA SC decided to fill this gap via Wikipedia, which is a free Internet encyclopaedia, ranked among the ten most popular websites worldwide.

The objective of SiRA SC was to initiate a comprehensive source of references for statistical regulatory affairs guidance documents and related articles in Wikipedia [1]. A suitable location in Wikipedia for the SiRA guidances was identified as the category "Biostatistics", classified under "Statistics", comprising various relevant topics (e.g. Medical Statistics, Clinical Trials and Epidemiology). The SiRA SC initiated a Wikipedia "List" of guidance documents in July-2016 and a related Wikipedia "Page" in January-2017 [2].

The poster presentation will include a description of the Wikipedia Project implementation (Wikipedia List and Wikipedia Page), an update to the Wikipedia User Guide [1] for editing/ reviewing the content in Wikipedia and a plan for dissemination of the Wikipedia Project to the wider statistical community.

Reference 1: Torres JV, Weir CJ, Close NC, Schumacher M, Friede T, Siegel J, Gerlinger C, Bender R, Nakas C, Seldrup J and Chadha-Boreham H. (2015) Statistics in Regulatory Affairs in www.wikipedia.org: Initiative of the ISCB SiRA SC. ISCB Conf. Poster.

Reference 2: https://en.wikipedia.org/wiki/List_of_Guidances_for_Statistics_in_Regulatory_Affairs; https://en.wikipedia.org/wiki/Draft:Guidances_for_statistics_in_regulatory_affairs

PC8 - W63: Sequential analysis in the context of hospital benchmarking

Authors: Lena Schneiderheinze[1], Ulrich Mansmann[1], Nicholas Lack[2].
[1] *University of Munich, Germany,* [2] *Bavarian Institute for Quality Assurance (BAQ), Germany.*

Sequential analysis of hospital benchmarking (SAHB) using control charts is a common instrument for quality control [1].The theoretical background is well founded [2], yet reports on the statistical performance of the procedure are sparse: What are optimal decision rules? What is the false signal rate? What is the probability to overlook quality deficits?

This work is motivated by the fact that SAHB is not part of the German external hospital quality assurance (EQA), where performance is evaluated annually. To detect quality deficits more quickly and within temporal context, we assess the use of log-likelihood CUSUM charts [2] in the EQA.

This exploration uses Markov chains and CUSUM run simulations in order to investigate optimal CUSUM designs for detecting quality deficits across hospital settings and assessing performances in different performance indicators.

It was possible: (1) to establish equivalence with alternative approaches (Bernoulli & Standard CUSUM); (2) to understand how the decision rule influences false-positive as well as false negative rates (missing relevant signals); (3) to get a better understanding how run lengths are distributed and the impact of the decision rule; (4) to check how the hospital size influences the detection rates; (5) to assess different patient risk mixes.

Our results add new aspects to the literature regarding the use and evaluation of SAHBs.

Reference 1: Bottle A, Aylin P. Intelligent Information: A National System for Monitoring Clinical Performance. Health Services Research. 2007;43(1p1):10-31.

Reference 2: Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. Biostatistics. 2000;1(4):441-52.

PC8 - W64: ChimerDB 3.0: an enhanced database for fusion genes from cancer transcriptome and literature data mining

Authors: Insu Jang[1], Byungwook Lee[1], Myunggyo Lee[2], Namhee Yu[2], Sanghyuk Lee[2], Kyubum Lee [3].
[1] *Korean Bioinformation Center, Republic of Korea,* [2] *Ewha Womans University, Republic of Korea,* [3] *Korea University, Republic of Korea.*

Fusion gene is an important class of therapeutic targets and prognostic markers in cancer. ChimerDB is a comprehensive database of fusion genes encompassing analysis of deep sequencing data and manual curations. In this update, the database cov- erage was enhanced considerably by adding two new modules of The Cancer Genome Atlas (TCGA) RNA-Seq analysis and PubMed abstract mining. ChimerDB 3.0 is composed of three modules of ChimerKB, ChimerPub and ChimerSeq.

ChimerKB represents a knowledgebase including 1066 fusion genes with manual curation that were compiled from public resources of fusion genes with experimental evidences. ChimerPub includes 2767 fusion genes obtained from text mining of PubMed abstracts. ChimerSeq module is designed to archive the fusion candidates from deep sequencing data. Importantly, we have analyzed RNA-Seq data of the TCGA project covering 4569 patients in 23 cancer types using two reliable programs of FusionScan and TopHat-Fusion. The new user interface supports diverse search options and graphic representation of fusion gene structure. ChimerDB 3.0 is available at http://ercsb.ewha.ac.kr/fusiongene/.

Reference 1: Mertens,F., Johansson,B., Fioretos,T. and Mitelman,F. (2015) The emerging complexity of gene fusions in cancer. Nat. Rev. Cancer, 15, 371–381.

Reference 2: Kim,P., Yoon,S., Kim,N., Lee,S., Ko,M., Lee,H., Kang,H., Kim,J. and Lee,S. (2010) ChimerDB 2.0–a knowledgebase for fusion genes updated. Nucleic Acids Res., 38, D81–D85.

PC8 - W66: 4D Nucleome viewer: a web-based dynamic browsing tool for chromatin interactions

Authors: Jinhyuk Choi[1], Minseo Kim[1], Insu Jang[1], Byungwook Lee[1], Dongchan Yang[2], Inkyung Jung[2]. [1]Korea Research Institute of Bioscience and Biotechnology, [2]Korea Advanced Institute of Science and Technology.

The majority of the human genome consists of non-coding sequences. Recent studies have suggested that many non-coding regions have significant functional implications on gene regulation through long-range enhancer-promoter interactions. Hi-C is the most widely used molecular technique to reveal such long-range chromatin interactions. There is a rapidly growing need for user-friendly tools capable of visualizing and extracting meaningful chromatin contact information from Hi-C results but it requires a lot of computational work due to the complicated nature of Hi-C and large data size. Here we present 4D (4-dimensional) Nucleome viewer to provide one-to-all chromatin interactions for user defined bait regions with selected Hi-C samples as input. The unique output features of 4D Nucleome viewer are dynamic browsing of chromatin interactions, visualizing normalized chromatin interactions after removing experimental biases, identifying significant interactions after removing distance-dependent background noises, and providing pairwise comparison between samples. All Hi-C data were processed as previously described with minor modifications. We believe that 4D Nucleome viewer will be widely used to many researchers to explore chromatin interactions such as enhancer-promoter relationships.

13th July Thursday

Mini-Symposia

SY1: Choosing Estimands in a Clinical Trial

Thursday 13th July - 09.30-16.00 h. - Room: Sala Mar 2
Promoters: Rosa Lamarca, *AstraZeneca R&D Centre, Barcelona, Spain* and Frank Bretz, *Novartis Pharma, AG, Switzerland*

Abstract: Defining the primary objective of a clinical trial in the presence of non-compliance or non-adherence to the assigned treatment is crucial for the choice of design, the statistical analysis and the interpretation of the results. This raises the need for a structured framework to specify the primary estimand (i.e. "what is to be estimated"). The missing data report released in 2010 by the National Academy of Science, "Prevention and Treatment of Missing Data in Clinical Trials", recommends explicit specification of a casual estimand in the protocol of a confirmatory trial. This is also reflected by the decision of the International Conference on Harmonization (ICH) to amend its E9 guidance in the coming years to discuss estimands and their role in clinical trials. The focus of this MiniSymposium is to discuss the estimand topic from different perspectives (statistical, clinical, epidemiological), covering regulatory, industry and academic viewpoints.

Session 1: New Scientific and Regulatory Environment ICH E9 Addendum

Thursday 13th July - 09.30-13.30 h.
Moderator: R. Lamarca

ICH E9 addendum on 'Estimands and Sensitivity Analysis'

Rob Hemmings, *Medicines and Healthcare products Regulatory Agency, United Kingdom*
Mouna Akacha, *Novartis Pharma AG, Switzerland*

Abstract: Randomized clinical trials are often considered the gold standard in drug development. Randomization is used to create comparable groups in terms of patient characteristics at baseline such that any difference observed after randomization must be due to the randomized treatments. However, randomization does not protect from confounding and bias due to intercurrent events that occur after randomization, e.g. discontinuation of treatment, treatment switching etc. Such events may themselves be impacted by the randomized treatments so that the choice of relevant treatment effects is complicated. In particular, the traditional treatment-policy effect targeted by the ITT approach may not always lead to clinically meaningful treatment effects. At present, these intercurrent events are dealt with implicitly by choices made about the data collection and statistical analysis. In order to improve transparency and ensure alignment between trial objectives and statistical approaches, it is necessary to clearly define the treatment effect (estimand) which is to be targeted in a clinical trial. Once there is agreement on the treatment effect of interest this will inform trial design, data capture and analysis. Through the publication of a concept paper the ICH has reinforced the need to more precisely define the measure of treatment benefit that is targeted in a trial and to clearly distinguish this from the method of estimation. Since then a working group has been tasked to develop a corresponding addendum to the main statistical guidance in drug development – the ICH E9. This addendum is expected to lead to a substantive evolution to traditional clinical trial design, conduct and analysis. In particular, this is a multi-disciplinary effort that requires a common understanding beyond the statistics community. In this presentation we

will provide a non-technical introduction and present an improved framework for trial planning. We will describe how estimands can be constructed and illustrate these aspects using a case study in patients suffering from pain.

Clinical perspective

Esther García Gil, *Aclidinium Franchise, AstraZeneca Barcelona, Spain*

Abstract: The treatment effect reflected by the treatment policy estimand (or de facto estimand) may be confounded by the patients inadequately controlled who discontinue study drug and switch to alternative therapy. Thus, the treatment policy estimand (randomized patients regardless of adherence) would tend to overstate the efficacy of an inferior treatment (such as placebo) by including data from patients who are rescued from inferior therapy and step-up their therapy. The clinical implications of using the treatment policy estimand would be discussed: How results should be interpreted?: Clinical interpretation of a treatment effect and its clinical relevance in clinical trials other than CV outcome studies, has been traditionally based on efficacy estimand, using data until discontinuation from study drug. Design of clinical trials to minimize missing data for patients on treatment Burden of data to be collected post study drug discontinuation Extend of analysis using different estimands to ensure clinicians/statisticians/regulators understand the magnitude of the treatment effect.

Session 2: Methodological Perspectives on the Estimand Framework

Thursday 13th July - 14.30-15.30 h.
Moderator: F. Bretz

Statistical perspective

James Roger, *London School of Hygiene & Tropical Medicine, United Kingdom*

Abstract: While a clear cut statement about the estimand should make life easier for the trial designer and the statistical analyst it reminds us of some long standing questions that may have been brushed under the carpet. This talk discusses some of the questions raised by this more formal approach.

The way that Intercurrent events are handled in the estimand has an important impact on how far the estimated value will change from population to population. We are used to baseline covariates such as severity of disease having an impact on size of outcome and these are regularly controlled in the trial design. The frequency of the intercurrent event will vary from population to population and also the potential impact of that event on the outcome will change in a similar way; say termination for lack of efficacy and differing health care systems. This is important when choosing a suitable estimand. Would one want to choose an estimand which is indifferent to such population characteristics?

There also remain problems of classic missing data due to patients abandoning the trial. For instance, with a treatment-policy estimand it will be difficult to "guess" those missed values as it requires a model for off-treatment experience for which, hopefully, we have little internal data. For hypothetical estimands it seems even less clear.

Causal inference perspective

Rhian Daniel, *London School of Hygiene & Tropical Medicine, United Kingdom*

Abstract: Due to the inherently more challenging and varied set of designs used in observational studies (compared with RCTs), it has been suggested since the 1980s that in observational studies, researchers should be unambiguous about the causal estimand being targeted by their study design and chosen data analysis. Although different suggestions have been made, the potential outcomes notation has been the favoured language in which to express these estimands in epidemiology. As well as facilitating clearer communication of conclusions, and a heightened appreciation of the strong assumptions under which the chosen causal conclusions are justified, this formal framework for drawing quantitative causal inferences from observational data has also led to novel statistical analysis methods that relax some of the assumptions required by traditional approaches. In this talk, I will briefly discuss some of the areas of the so-called "causal inference" approach – in particular instrumental variables and mediation analysis – that are relevant to the current discussion of moving beyond the ITT estimand in RCTs.

Session 3: The New Era in Clinical Research

Thursday 13th July - 15.30-16.00 h.
Moderator: R. Lamarca

SY2: Modelling Personalised Screening: a Step Forward on Risk Assessment Methods

Thursday 13th July - 09.30-16.00 h. - Room: Auditorio
Promoter: Montse Rué, *University of Lleida, Spain*

Abstract: In the era of precision medicine, a major challenge is to measure risk as precisely as possible and then to target appropriate interventions. For many years, in familial breast and ovarian cancers, risk prediction models have been developed and used for predicting the risk of developing cancer and the likelihood of carrying specific mutations. Cancer risk models have been developed for the general population but have been seldom used to personalise screening. On the one hand, there is a need for a thorough examination of models validity across multiple settings and, moreover newer approaches to risk prediction are needed. This symposium will review the state-of-the-art in risk assessment methods and their application.

The mini-symposium will focus on three areas of interest. The first morning session will be dedicated to modelling for high-risk individuals and applying risk prediction to target preventive interventions. The second morning session will cover modelling in population-based screening programs and applying risk prediction to customize screening. The afternoon session will review and debate modelling and validating models for clinical prediction. The three sessions will have panel discussions on methodological challenges and on the critical barriers for their application.

Session 1

Thursday 13th July - 09.30-11.00 h.

Modelling for high risk individuals

Antonis Antoniou, *University of Cambridge, United Kingdom*

Abstract: Several studies have demonstrated that a large fraction of all breast and ovarian cancers occur in a minority of the population who are susceptible. Currently women at high risk of breast or ovarian cancer are primarily identified on the basis of family history and through mutation screening in the high risk BRCA1 and BRCA2 genes. However, rare genetic variants conferring intermediate risks (e.g. PALB2, CHEK2, ATM for breast cancer; RAD51C, RAD51D, BRIP1 for ovarian cancer) and >150 common alleles (SNPs) conferring low risks have been identified. Hormonal/reproductive, lifestyle and imaging risk factors are also associated with cancer risks. This growing understanding of the multifactorial aetiology of breast and ovarian cancer suggests that more powerful and reliable cancer risk prediction can be achieved by combining data on all known genetic and other risk factors. The presentation will review some of the currently available breast and ovarian cancer risk prediction algorithms, the methods used in their development, their strengths and limitations. The presentation will also provide an overview of the latest developments and challenges in understanding the penetrance of mutations in genetically susceptible individuals (e.g. BRCA1, BRCA2, PALB2 mutation carriers) and will focus on the efforts to develop BOADICEA, a comprehensive risk prediction model for breast and ovarian cancer. These efforts are facilitated by capitalising on the data resources generated through large collaborative studies such as the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA), the International BRCA1/2 Carrier Cohort Study (IBCCS) and the Breast Cancer Association Consortium (BCAC).

Clinical practice perspective

Judith Balmaña, *Vall d'Hebron Institut of Oncology, Spain*

Abstract: Current advances in sequencing technology and its implementation in clinical practice are leading towards the concept of personalized and/or precision medicine. This also points out to achieving personalized cancer risk predictions to be able to provide precision cancer prevention. In this regard, some of the clinical questions are: where are we now, what do we need, which people may benefit from cancer risk predictions, how can we improve prediction, and how can we translate these estimates into prevention options, among other questions?

It is well know that cancer emerges from at least a combination of inherited genetic factors and exposure to environmental factors. Updated knowledge of new genetic factors (high, moderate, and low penetrant alleles) and environmental factors associated to cancer development, as well as their interaction and/or multiplicative effect is needed to get accurate predictions in the clinics.

All these factors are to be incorporated into prediction tools for specific types of cancer. Since risk factors have different associations with different types of cancer, and the latter require different types of prevention strategies, there is the need to achieve accurate estimations for each type of cancer in order to provide the adequate preventive recommendations for each cancer type. In the clinics, absolute short-term cancer risk predictions more than relative risk or lifetime risks might translate into more realistic prevention actions. These prevention and/or screening actions also need to demonstrate their magnitude of risk reduction and their clinical utility, in terms of demonstrating improvement in people's health status. Once average cancer risk in a specific population is known and medical interventions of screening or cancer prevention are established for that particular population risk, individuals and physicians want to know how to accurately predict individual cancer risk. For individuals with a known high penetrant genetic risk factor, such as a pathogenic germline variant in BRCA1 or BRCA2 genes, it would be helpful to predict their individual risk with the consideration of other low-risk alleles, family history, lifestyle and reproductive factors in order to be able to make age-specific preventive or screening recommendations. For individual decision-making, a short-term prediction might be clinically most useful than a lifetime risk prediction. If predictions are able to establish different levels of risk stratification, these will guide individual decision-making and population risk-reduction programs.

Finally, there is the need to identify the appropriate preventive intervention for each cancer risk estimate according to the type of cancer and the risk factor. Preventing a hormone receptor positive breast cancer is likely to need a different strategy than preventing a hormone receptor negative breast cancer. Therefore, research of the relationship between genotype-phenotype and environmental risk factors-phenotype are needed to guide the appropriate prevention options.

Session 2

Thursday 13th July - 11.30-13.00 h.

Collaborative modeling for cancer screening

Sandra J. Lee, *Harvard T.H. Chan School of Public Health and Dana Farber Cancer Institute, USA*

Abstract: Screening asymptomatic individuals for early detection of cancer is a public health initiative that is growing rapidly. The main motivation is to diagnose disease early before it progresses to advanced stages so that the benefit of treatment may be enhanced. The optimal screening programs for the general and high-risk populations are still debated. Models that characterize breast cancer progression are available to evaluate early detection programs. Modeling approaches provide insights into expected outcomes, positive and negative from early detection programs. The Cancer Intervention Surveillance Modeling Network (CISNET) is a consortium of investigators that includes modeling to improve our understanding of the impact of cancer control interventions (such as screening, treatment, prevention) on population trends in incidence and mortality. Collaborative modeling work of the CISNET-Breast Working Group on mammography screening strategies for the general and high-risk populations will be presented. These strategies will be compared with respect to the age of screening initiation, screening intervals, stopping age, overall benefits and harms of the each strategy. Current breast cancer screening guidelines recommend scheduling mammograms annually, biennially or triennially. A novel method for scheduling screening mammograms based on risk instead of time intervals will be introduced. The risk of undiagnosed disease at the age of screening initiation is used as a threshold for determining next screen. This method generates a risk-based screening schedule which reduces a total number of examinations while achieving a comparable screening benefit. These risk-based screening strategies will be presented. The second part of the presentation will focus on screening strategies for early detection of melanoma. A population based German skin cancer screening study in Schleswig-Holstein region has established a link between early detection of disease and a reduction in melanoma-related mortality. However the value of melanoma screening for the general population is debated. To identify sub-populations to target melanoma screening, factors associated with the risk of developing melanoma and of fatal melanoma are evaluated. The natural history model for melanoma progression is developed. A model-based result on the impact of early detection of melanoma by risk groups will be discussed.

Modelling for Clinical Prediction

Ewout W. Steyerberg, *Erasmus University Medical Center, Rotterdam, The Netherlands*

Abstract: There is increasing interest in the development and application of clinical prediction models in the current era of precision medicine. This is related to the exponential increase in biological knowledge on predictors of outcome, including biomarkers, imaging and omics technology. I aim to address some key issues in the development of prediction models and their clinical implementation.

Case study 1: A model was proposed to predict the likelihood of having a mutation in DNA mismatch-repair genes at the time of diagnosis of colorectal cancer (NEJM 2006;354:2751). I critically review the modeling strategy that relied on stepwise selection of categorized predictors from a large pool of candidate predictors in a cohort where only 38 mutations were found. Simulations in small subsets

of a large contemporary cohort with 2,051 mutations showed that this modeling strategy led to poor discrimination at validation (c = 0.75). Better performance was achieved by model pre-specification based on subject knowledge and using continuous predictors (c = 0.83).

Case study 2: The MINDACT trial enrolled 6693 women with early-stage breast cancer (NEJM 2016;375:717). A gene signature was used to determine genomic risk, and clinical risk was estimated by a categorized version of Adjuvant!Online. Women with discordant risk results were randomized to the use of chemotherapy. While the trial confirmed the independent prognostic effect of the signature (Hazard Ratio 2.4), the chemotherapy effect was uncertain. Further statistical and decision-analytic modeling showed wide variability in estimated risk and individual benefit within the high clinical risk group, and identified fewer women for genetic testing and chemotherapy treatment.

Conclusion 1: Many current prediction models are flawed if based on small numbers and developed with common but suboptimal statistical approaches. Alternative modeling strategies to best exploit available prognostic information merit wider implementation, with collaborative research to increase sample sizes.

Conclusion 2: High quality but pragmatic trials may be insufficient to directly inform clinical practice on the value of a biomarker or genomic test. More detailed prognostic modeling is required to support optimal clinical implementation and shared decision-making.

Session 3

Thursday 13th July - 14.30-16.00 h.

Validating Prediction Models

Inmaculada Arostegui, *University of the Basque Country and Basque Center for Applied Mathematics, Spain*

Abstract: Prediction models are currently relevant in medicine. They are often translated to prognostic severity scores for risk stratification, which are also known as clinical prediction rules (CPR). The development of a CPR has to follow strict methodological norms during the whole process in order to be useful in clinical practice. One of the most relevant methodological issues when developing CPRs is related to validation. The goal of this talk is to present and discuss the several steps that arise in the development of CPRs and to highlight the relevance of validation of the whole process.

First of all, when a CPR is designed for clinical practice, it should be based on routinely available clinical parameters. The second step is related to the statistical model, more precisely, the selection of the modelling technique, the predictors to be included in the model and the model specification. On the one hand, choosing between standard statistical models (such as logistic regression) or more complex machine learning algorithms (that have recently aroused increasing attention in medical research) is a hot topic of research. On the other hand, the precise selection of predictors and its coding are relevant issues too in the modelling phase. The third step consists on creating the CPR or prognostic severity scores that estimates the risk of event. For instance, when modelling is based on generalized linear models, the score is often created using a points system based on the estimated beta coefficients. The next step, forth, consists on categorizing the severity score into k levels that classify individuals into different risk

groups, with guaranties that it preserves the predictive accuracy of the original model. Moreover, the CPR has to be available as an easy to use tool for clinical decision-making. All the previously mentioned steps are closely related to the validation process. Variable selection techniques, categorization of continuous variables and correction for optimism of the predictive accuracy are related to internal validation of the prediction model. Bootstrapping techniques have been proposed in the literature as adequate at this modelling phase. Validation of the CPR and its discrimination ability compared to the original model and validation of the stratification criteria are also very important issues at the third and forth steps. Nevertheless, these two points have not generated so much attention in the field. External validation, or generalizability, would measure the performance of the CPR with independent data as compared with that used for the development of the rule. Finally, the first and the last steps are related to clinical validity, in the sense that the CPR should be easy to use for decision-making in clinical practice.

SY3: Statistical Methods for Medical Imaging and their Use in Clinical and Epidemiological Studies

Thursday 13th July - 09.30-16.00 h. - Room: Sala Mar 4
Promoter: María Durbán, *University Carlos III of Madrid, Spain*

Abstract: Medical imaging is an effective tool for clinical diagnosis, monitoring and development of new therapies. The statistical methods developed in recent years provide the necessary tools for processing and analyzing the information in the images in order to make more accurate decisions. This symposium will focus on some of these findings and different ways in which clinicians are using them. The talks will cover different statistical techniques that provide a more integrative and comprehensive understanding of different medical problems (regularization methods, statistical matching methods, machine learning techniques) and their application to areas such as: HIV-associated cognitive impairment, fingerprinting in the context of resting state fMRI, prediction of cardio-vascular diseases and replicability problems in neuroimaging.

Session 1

Thursday 13th July - 09.30-11.00 h.

Regression modeling with Laplacian-based regularization: theory and brain imaging applications

Jarek Harezlak, *Indiana University, Bloomington, USA*

Abstract: Majority of multimodal neuroimaging studies are analyzed separately for each modality. Importantly, statistical methods that simultaneously assess multimodal data provide a more integrative and comprehensive understanding of the brain. We propose an extension to the statistical regularization methods in the linear model setting with a Laplacian-based penalty operator. Model parameters are estimated by a unified approach directly incorporating structural connectivity information into the estimation by exploiting the joint eigenproperties of the predictors and the

penalty operator. We present the closed-form solution for the estimators, test its properties via a simulation study and apply it to find the best predictive imaging markers of HIV-associated cognitive impairment. Introducing a priori information minimized spurious findings by assigning penalty weights in such a way that highly connected regions associated with the outcome were less penalized than other regions that had no association with the outcome. Future work will incorporate functional connectivity and finer cortical parcellation.

Am I My Connectome? Fingerprinting With Repeated Resting State Functional MRI Data

Brian Caffo, *Bloomberg School of Public Health, John Hopkins University, Baltimore, USA*

Abstract: In the context of resting state functional MRI (rs-fMRI), fingerprinting is the practice of matching a set of subjects to themselves using only rs-fMRI correlations. The quality of the matching is then validated using the subjects' IDs. A statistical inference on this matching is often performed using permutation tests. We discuss many aspects of this process in this talk. First, we discuss desired invariances in the matching process and distance metric. Secondly, we discuss matching statistics and strategies and the resulting null distributions they induce. Thirdly, we discuss variations on the null hypothesis, which is typically left unspecified despite the calculation of a permutation based null distribution. We discuss these topics in the context of the rich history of this problem, spanning over two centuries from Montmort's matching problem.

Session 2

Thursday 13th July - 11.30-13.00 h.

Carotid Artery Characterization in Ultrasound Imaging using Machine Learning Techniques

Maria del Mar Vila, *Hospital de Mar Medical Research Institute, Barcelona, Spain*

Abstract: Atherosclerosis is the main pathogenic process causing most Cardio Vascular Diseases (CVD). The presence of subclinical atherosclerosis has been pointed as a potential biomarker to identify individuals with high risk of developing a clinical CVD event. Carotid artery Ultra Sound (US) image is currently used to identify and quantify the presence of subclinical atherosclerotic vascular disease and for evaluating CVD risk. In particular, this high-resolution and non-invasive technique provides one of the best methods for the manual detection of atherosclerotic disease. On the other hand, it is well known that disruption of an atherosclerotic plaque plays a crucial role in the pathogenesis of CVD event and that plaque disruption is characterized by the content of different histologic components. Therefore, B-mode US characterization of plaque morphology is useful in assessment of the vulnerability of the atherosclerotic lesion. Our purpose is to study a big dataset of carotid artery ultrasound images in order to build an automatic method able to automatically extract image features and analyze their potential clinical value. To do this, we

will focus on the possible mechanisms involved in plaque formation and its vulnerability. Then, we will use these characteristics together with common biomarkers and we will analyze their predictive power in cardiovascular diseases. Our proposal uses image processing and novel machine learning techniques to achieve 3 objectives. First, create a fully automatic method for the Intima-Media segmentation region and to detect the presence of plaque. Second, create a method to characterize the histologic plaque components. Third, explore new image features to improve the predictive method for cardiovascular events (REGICOR risk function [1]). [1] Validity of an adaptation of the Framingham cardiovascular risk function: the VERIFICA study Jaume Marrugat, Isaac Subirana, Eva Comín, Carmen Cabezas, Joan Vila, Roberto Elosua, Byung-Ho Nam, Rafel Ramos, Joan Sala, Pascual Solanas, Ferran Cordón, Joan Gené-Badia, Ralph B D'Agostino J Epidemiol Community Health 2007.

Statistical Extraction of the Arterial Input Function in Dynamic PET Imaging Studies: A Review and Some Recent Developments

Finbarr O'Sullivan, *University College Cork, Cork, Ireland*

Abstract: Positron emission tomography (PET) imaging is widely used in the clinical management of several types of cancer. In dynamic mode, PET scanning produces a time-sequence of 3-D volumetric images of the distribution of the injected PET radiotracer in the scanner field of view. The interpretation of such dynamic time-course data using kinetic models can provide insight into detailed physiological and metabolic characteristics of the tissue being imaged. Kinetic modeling requires the time-course of the radiotracer in the arterial blood as an input function (AIF). In a clinical setting the direct measurement of the AIF by catheterized arterial sampling is impractical. In pre-clinical small animal PET imaging studies, blood sampling may also be problematic because the total volumes of blood present are often insufficient. Thus in either pre-clinical or clinical settings direct extraction of the AIF from the imaging data is practically important. The work reviews a number of statistical approaches to this problem. Image segmentation is useful as an initial step, however it is important that the interpretation of segment time-courses properly account for background spillover and also the non-arterial signals the may be present. The theoretical limitation of fully non-parametric extraction is described. We discuss how historical directly sampled AIF data can be used construct empirical Bayesian type approaches. This includes consideration of a Markov Chain whole body circulation model [1] for representation of arterial and non-arterial sources that are often present in the scanner field of view. The methods discussed are illustrated with examples from cancer imaging studies with (18F) fluoro-deoxyglucose (FDG) for assessment of glucose metabolism, (18F) fluoro-thymidine (FLT) as a marker of DNA synthetic rate and (15O) H20 for assessment of local blood flow characteristics in the context of breast and brain-tumor imaging. [1] Huang J and O'Sullivan F. An Analysis of Whole Body Tracer Kinetics in Dynamic PET Studies With Application to Image-Based Blood Input Function Extraction. IEEE Trans. Med. Imaging 2014, 33, 1093-1108. Acknowledgement Work supported in part by Science Function Ireland (SFI) under the award SFI-PI-11/1027.

Session 3

Thursday 13th July - 14.30-16.00 h.

Validity of psychological and neuroimaging findings

Manuel Desco, *University General Hospital Gregorio Marañón and University Carlos III of Madrid, Spain*

Abstract: We will discuss about the interpretability and replicability problems that affect many scientific fields with a special focus on the fields psychology and neuroimaging. We will disseminate some of the factors that contribute to the interpretability and replicability crises and will present some of the potential solutions proposed by the scientific community to minimize them.

Students' Day

Thursday 13th July - 09.25-18.00 h. - Room: Sala Terra 2
Organizer: Nadine Binder, *University of Freiburg, Germany; and the Students' Day organising group*

Abstract: The aim of the day is to stimulate discussion on how to be a good researcher, how to come up with biostatistical research projects of ongoing or future interest and to discuss difficulties that students face during their degree. It also provides an opportunity to practice your presentation skills in a semi-formal environment. Although research presentations are welcome from first time attendees, we strongly encourage students to instead use this opportunity to share their personal experiences about doing biostatistical research and how they have dealt with pitfalls that have come up in the research process. It doesn't matter if you are studying for a masters, just about to start with your PhD or will hand in your thesis soon, the Students' Day provides a perfect opportunity to meet your peers, exchange thoughts and ideas, and to get to know more about how to shape a career in biostatistics.

Welcome

09.25-09.30 h.

Invited Speaker: How Not to Plan A Career in University, Big Pharma, Small Biotech, Medium Pharma, Consulting and back to University

09.30-10.00 h.
David W. Warne. *Consultant Biostatistician and Academic Fellow at University of Geneva, Switzerland*

Abstract: About 30 years ago, in the late 1980's I too was an applied statistics student... Due to various mostly unplanned career changes, I've had the chance to apply statistics in various places and this talk will give examples of the challenges which may be faced in each, not only statistical but also personal. Moving abroad, learning new languages, coping with difficult colleagues, finding and losing jobs, being closed down, getting older but not necessarily wiser, work-life balance, especially finding time for yourself. I'll also focus on the highlights, such as making a difference to patients' lives, career progression and taking on more responsibilities, moving into related areas of biostatistics such as people, process and quality management. Finally, I will emphasise the wonderful opportunities presented by getting involved in a statistical society such as ISCB.

Session 1: Good research process: From programming discipline to adherence to guidelines

10.00-11.00 h.
Chair: Maja von Cube

From scripts to packages: how to be disciplined about your R code

Theodor Adrian Balan. *Leiden University Medical Center, The Netherlands*

Abstract: During a PhD programme, it is common to write a lot of code. Quite quickly, this can become difficult to manage. Going back to code that was written one year ago might seem like a horror experience. Small changes in data sets might make methods programmed a while ago unusable. From some old projects, an impossible to navigate labyrinth of folders and files is left behind.

There is a lot of talk about producing reproducible research in biostatistics. As this will get more important in the future, structuring code in one way or another becomes a necessity. Some journals already require authors to provide the code used for simulation or data analysis. And those that do not do that, probably will at some point in the future.

Soon after I started my PhD, I realized how unsustainable the (lack of) organization of my R scripts was. However, discipline did not get built over night. It was a gradual process, adopting one useful tool at a time. Years later, as I saw my first R package on CRAN, I realized how much time and energy I could have saved. All I had to do was to invest a bit of time in the beginning.

In this talk, I will present lessons I learned in transitioning from writing scripts to writing packages. This includes writing tidy functions and tidy code, properly documenting scripts and functions with rmarkdown, version control with git, developing, checking and documenting R packages with RStudio. I will discuss advantages and caveats of the different tools. Looking back, I wish someone would have told me all these things earlier. Finally, I hope that this will underline the importance of discipline in coding for a young researcher in biostatistics.

Exploring results from simulation studies interactively

Alessandro Gasparini. *University of Leicester, United Kingdom*

Abstract: Simulation studies represent a powerful tool with a multiplicity of aims: among others, evaluating new or existing statistical methods, comparing them, assessing the impact of modelling assumption violations, and helping with the understanding of statistical concepts. The increased availability of powerful computational tools (both personal and high-performance cluster computers) to the average researcher surely contributed to the rise of simulation studies in current literature, with the search for simulation study in Scopus returning less than 1,000 papers in 1994 and almost

7,000 in 2016. Additionally, the increased computational capabilities allow researchers to simulate an ever-growing number of scenarios, making reporting results a non-trivial task. Dissemination of results plays a focal role in simulation studies: (1) it can drive practitioners and applied statisticians to methods that have been shown to perform well in their practical settings (e.g.: small sample size, high proportion of missing values), (2) it can guide researchers to develop new methods in a promising direction, and (3) it can provide insights into less established methods. As a consequence, several design and reporting guidelines emerged, often tailor-made to a specific research area (e.g. health technology assessment, medical statistics, social sciences).

To bridge the gap between evaluating a plethora of scenarios in simulation studies and dissemination of results, we developed an online tool for exploring results interactively. Our tool is developed using R and the shiny framework. It requires the researcher to upload a dataset in a standardised, tidy format (observations are in rows, variables are in columns) containing results from a simulation study. Then, it computes performance measures such as bias, coverage probability, Monte Carlo errors, and empirical standard errors automatically. Finally, it presents results and performance summaries both in tabular and graphical fashion (via bar plots and lolly plots) and allows the reader to vary simulation parameters and choose estimands of interest for further investigations. We believe this tool could supplement reporting of simulation studies to a great extent, allowing researchers to share all the results from their simulations and readers to explore them freely.

Increasing research value, reducing waste: interventions to improve adherence to reporting guidelines in the context of MiRoR project

David Blanco de Tena-Dávila. *Universitat Politècnica de Catalunya, Spain*

Abstract: Tens of billions of Euros are wasted each year on studies that are redundant, flawed in their design, never published or poorly reported. Turning around this situation was one of the main reasons why the Methods in Research on Research (MiRoR) project was created. The MiRoR project is a joint doctoral training programme in the field of clinical research funded by Marie Skłodowska-Curie Actions. I and other fourteen students, as well as a group of high-level senior researchers and partner institutions (like The BMJ, Cochrane, EQUATOR Network, or BioMed Central, among others), are joining forces to try to meet the ambitious goal of the project: to increase research value and to reduce waste.

In my talk, I will first introduce the MiRoR project, the topics we are working on, the way we are collaborating with prominent researchers and our partner institutions, as well as the challenges we are facing. Second, I will give an overview of my PhD project, which is focused on investigating what actions can be taken to improve adherence to reporting guidelines.

Reporting guidelines are sets of recommendations aiming to provide advice on how to report research methods and findings. They have been developed since early 1990s to help improving the completeness and transparency of published articles, which helps decision makers to judge the applicability of the research, and enhances reproducibility. There is evidence that the use of some RGs, such as CONSORT, is associated with improved standards of reporting. However, the current

levels of adherence to reporting guidelines are poor, far from desired. For this reason, different actions aiming to enhance compliance with reporting guidelines have been taken over the last years.

The ultimate goal of my PhD project is to explore what strategies to improve adherence to reporting guidelines could be implemented and formally assessed. To achieve this objective, we are first performing a scoping review in order to explore what interventions to enhance adherence to reporting guidelines have been assessed and what further ideas have been suggested. This review could send a message to funders, authors, and editors about how the problem of adhering better to RGs has been tackled from different perspectives. In addition, it could be a major first step towards developing future strategies to improve compliance with reporting guidelines.

Discussion

Invited Speaker: Portrait of a Consultant Biostatistician

11.30-12.00 h.
Irene Schmidtmann. *University Medical Center, Johannes Gutenberg University Mainz, Germany*

Abstract: In the last decades, biostatistics have arisen as a key discipline in biomedical research. The use of rigorous statistical procedures contributes to the validity of the results and increases the possibility of publication in high-impact journals. Complexity of statistical methods commonly used in medical research projects has increased steadily over recent years, making essential the contribution from a biostatistician. However, incorporation of statisticians to biomedical research teams and research centers is not widespread.

In this talk, the role of statistics in biomedical research is analyzed from the point of view of a biostatistician working in a clinical research unit at a Spanish hospital. Areas in which biostatistician is routinely involved will be analyzed, from the protocol design to the analysis and manuscript writing. Desired skills and academic background to work in a medical context will also be covered, focusing on the complex relationship between biostatisticians and biomedical researchers. Finally, the current status of statisticians enrollment to hospitals will be analyzed.

Session 2: On comparing and improving biostatistical approaches

12.00-13.00 h.
Chair: David Blanco de Tena-Dávila

Tuned penalized regression – a solution to the problem of separation in logistic regression?

Hana Šinkovec. *Medical University of Vienna, Austria*

Abstract: A frequent problem in logistic regression analyses is non-existence of finite maximum likelihood parameter estimates. This typically occurs in small or sparse data sets with rare outcomes, rare exposures or covariates with strong correlations or effects. It has been termed »separation« as the two outcome groups are perfectly separated by the values of a covariate or a linear combination of covariates. Unfortunately, it can rarely be assumed that the outcome is perfectly predictable in the underlying population. For this reason we should consider separation as a consequence of random data variation and evaluate possible solutions to achieve a stable fit. Penalized regression models seem to be a natural choice as they are intended to provide shrinkage of the parameter estimates. The penalty parameter is often found by optimization of the AIC or cross-validated deviance. However, as we will demonstrate and theoretically explain for a 2x2 table these methods have limitations in presence of separation. In particular, the optimal penalty parameter is often zero leading to the standard maximum likelihood solution. Paradoxically adding noise predictors will often remove the problem but it is unclear whether this really improves estimation. Thus, the general performance of both tuning approaches is evaluated by means of real and simulated data, also including multivariable settings. Ridge, LASSO, Firth and log-F penalties are considered. We conclude that tuned penalized regression is a questionable solution to separation. In contrast, pre-specifying the value of the penalty parameter could yield finite parameter estimates with better properties for all the considered penalization methods.

Bayesian Inference for skew-normal linear mixed models with covariates measurements errors

Oludare Ariyo. *KU Leuven, Belgium*

Abstract: The extensively used linear mixed model assumes normality for the random error and random effects. However, in practices, this assumption may not be realistic, especially when the data present skewness. Also, the time-varying covariates are often mistimed especially in longitudinal data. Hence, we consider some extensions of the linear mixed model to allow for a skew normal distribution for the random error and random effects, taking into account possible measurement error in the covariates. These models are estimated under a Bayesian approach. The problem of

model selection is also considered here, where the most common selection criterion, namely the deviance information criterion (DIC), has different definitions for models with latent variables as the ones hereby discussed. We consider two versions of the DIC: the conditional deviance information criterion (cDIC) which is usually selected by the researchers given its easy computation, and the observed deviance information criterion (oDIC) which involves less straightforward computation but has been found to be more appropriate in literature. In this paper, we compute the oDIC using importance sampling with an approximation to the effective number of parameters. This method was evaluated via simulation studies are real data sets.

Generalizations of the receiver operating characteristic curve

Sonia Pérez-Fernández. *University of Oviedo, Spain*

Abstract: The receiver operating characteristic (ROC) curve is a popular graphical method used to study and compare the diagnostic accuracy of a considered marker. However, the diagnostic process is closely related to the classification subsets considered, which are some particular ones in the usual ROC curve. There exist some generalizations of the curve extending these classification subsets and supporting different techniques to study how good a considered marker is to distinguish between two groups in population.

Nevertheless, the major drawback of these flexible approaches is the possible loss of interpretability of the study under consideration. In order to control it, what restrictions should be applied? A logical one is that the classification subsets may be self-contained, but this kind of considerations brings us to some open issues such as how to build the optimal ROC curve under restrictions and if it is possible to design some specific algorithms to achieve it under different assumptions. It is also important to place value on computational implementations of these techniques in statistical software (R) and problems associated such as computational times.

Discussion

Moderated round on open questions

14.30-15.15 h.
Chair: Maral Saadati and Michael Grayling

Session 3: Time-to-event data: From simple to complex solutions

15.15-16.30 h.
Chair: Theodor A. Balan

Simple parametric analysis for a multi-state model in hospital epidemiology

Maja von Cube. *University of Freiburg, Germany*

Abstract: In modern clinical research, multi-state models have become an established and adequate tool to quantify determinants and consequences of hospital-acquired infections (HAIs). These models avoid typical analytical pitfalls such as time-dependent or competing risks bias which are serious and very common in clinical research of HAIs. However, multi-state models are mathematically complex and the interpretation of the results is often challenging. When evaluation data with a multi-state model the statistical quantities of interest are the transition-specific hazard rates and the time-dependent transition probabilities as well as attributable mortality (AM) and the population-attributable fraction (PAF). In the most general case a direct calculation of these expressions is infeasible.

When assuming time-constant hazards calculation of the mortality risk and excess mortality due to infections is facilitated. In this situation the transition probabilities can be expressed in closed mathematical forms. Estimators that quantify the excess mortality can be easily obtained from these forms. We derive the closed mathematical forms of an extended illness-death model to perform a complete but simplified analysis of a real and publicly available data set from a cohort study of intensive care unit patients.

Only a few simply accessible values (event counts and patient-days) and a pocket calculator are needed to get basic insights into cumulative risk and clinical outcomes of HAI. Specifically in our data example the derived statistical quantities closely approximate their non-parametrically estimated counterparts. The approach avoids typical pitfalls such as time-dependent or competing risks bias.

No sophisticated statistical software is required for applying a complex however unadjusted multi-state model to obtain a basic insight to the data structure. The proposed method offers an easy access into the field of multi-state modelling to clinicians and professionals new to the field. The approach can be applied to any data situation with competing risks and a binary time-dependent covariate. A typical example would be the wish to investigate the effect of bone-marrow transplant on the risk of death due to cancer. In this data setting transplant is a time-dependent covariate and death due to other cause the competing risk.

Joint model for longitudinal marker and time-to-event in presence of competing risks: application on cystic fibrosis

Lionelle Nkam. *Conservatoire National des Arts et Métiers, France*

Abstract: The management of various diseases requires an ongoing monitoring of patients, in which many markers are collected. The analysis of such markers allows us to describe the progression of the disease and anticipate the occurrence of clinical events. In the case of Cystic Fibrosis (CF), one of the most important markers is the lung function (FEV1). Joint models are helpful in the case of CF, in modelling the dependency between FEV1 and the occurrence of clinical events, especially lung transplantation (LT) and death.

The aim of this study was to develop a model which identifies different profiles of the evolution of CF and provides dynamic predictions of cause-specific hazard (CSH) of LT and CSH of death taking into account the evolution of FEV1.

We developed a joint latent class model which assumes that the population can be divided into G latent homogeneous classes. This model was made up of three submodels. A multinomial logistic regression model which described the probability of being in a specific class. A mixed model with a quadratic function of age from 18 years old, which described the FEV1 trajectory over time. Lastly, proportional hazard models which provided class-specific risks of the events.

We selected the model with 3 latent classes based on BIC criteria and computed subject-specific predictions at various landmark ages with a prediction window of 3 years.

A total of 1625 adults from the French CF Registry were included in the study, over the period 2007-2013. A total of 86 (5.3%) patients died and 277 (17%) patients received a lung transplant during this period. The developed joint model identified three evolution profiles. Class 1 (55% of subjects), characterized by a slight decrease of FEV1 with age and a risk of death and LT close to zero before 25 years old. Class 2 (27% of subjects) with a deep decline of FEV1 and high risk of death and LT. Class 3 (18% of subjects) with high values of FEV1 and low risk of death and LT. The estimated AUCs corresponding to the prediction model were high (ranged from 0.7 to 0.9 for both LT and death). The estimated Brier scores were low (values less than 0.1) for predictions of CSH of LT and death.

Bayesian approach to modelling unobserved household and cluster effect on under five mortality rate in Zambia. An approach to inform strategic program implementation and decision marking in the health sector

Isaac Fwemba. *University of Ghana/Leuven Inter Biostatistics and Statistical Bioinformatics Center, Zambia/Belgium/Ghana*

Abstract: Under-five mortality has been used as a proxy indicator to measure the health coverage and access of the general population. In Zambia the child mortality rate stands at 75 per 1000 deaths. Numbers of children dying should be reduced if Zambia is to attain the Sustainable Development Goals of reducing the child deaths to 25 deaths per 1000 deaths by 2030. Implementation of interventions to improve child health outcome should look at the individual and contextual determinants of child survival. Evidence from this study will guide policy decision in the implementation of child health programs so as to maximise child survival. Similarly, we assume that implementation of child programs should be influenced by factors that are shared by the individuals in regions or clusters such as biological, parental competence, genetic, customs and other unobserved factors that are not accounted for by the observable factors at household level.

In this study, we propose the application of Clayton-type counting process formulations for clustered survival data using gamma frailties, which have routinely been applied in analyses of clustered survival data. Frailty models have been successfully used to model dependence in clustered survival models. The frailty components in the proposed model will be captured in terms of the family size and the clusters. The area-specific unobserved frailty, will be captured as the risk of the unobserved variables- to be modeled non-parametrically. The above approach is being proposed, as a more suitable approach compared to most methods known in epidemiological studies involving mortality and morbidity mapping which assumes that disease risks are independent across geographical areas.

We propose analysis of the effects of unobserved family heterogeneity in children survival times using a full Bayesian approach. We shall rely on survey data from the ZDHS conducted in 2013/14 and use a proportional hazard model with multiplicative random effects based on the assumption that independence of observations is unrealistic.

Several studies have been conducted on child mortality in Zambia using variety of statistical methods including geoadditive modelling. As far as we are aware, this is the first time that the unobserved effect using Bayesian survival frailty model is being proposed for such studies.

The study is being supported by the University of Ghana through TDR/WHO scholarship. The authors acknowledge the support received for this work.

Data Science in HIV studies

Yovaninna Alarcón Soto. *Universitat Politècnica de Catalunya, Spain*

Abstract: The goal of this talk is two-fold. First, we present an overview of my thesis proposal and the main results found so far. Second, we would like to discuss some difficulties and challenges faced during the beginning of the project.

The thesis project aims to cover the following two large blocks: one block is rather applied and consists of developing new techniques to work with data related to HIV (like data from trials with therapeutic vaccines and enzyme-linked immunospot (ELISPOT) assays), and the other block is more theoretical, consisting of joint models of survival and multi-omics data. The development of immunologic interventions to control viral rebound in HIV infection is a major goal of the HIV-1 cure field. Therapeutic vaccination with "kick and kill" strategies has been proposed to control viral replication after discontinuation of Antiretroviral Therapy. Kick and kill strategies mean to wake up or activate latent reservoir cells (kick) and to teach the cells of the immune system to recognize these activated cells and kill them.

Besides, concerning HIV-related data, there is little information so far, on data obtained from ELISPOT assays. The ELISPOT assay is an immunoassay that measures the frequency of cytokine-secreting cells at the single-cell level. A single cell forms a colored "footprint" (spot) on the bottom of the well representing its secretory activity.

The main idea of the analyses of the data from therapeutic vaccination and ELISPOT assays is to identify which kind of variables are correlated with clinical parameters and if some of them can be important for a prediction model.

On the more theoretical arm, to identify biomarkers it is essential to study different omics layers and associate them with the survival of patients. Different approaches have been developed here1,2, but nothing in the context of HIV. Because of this, one aim of my thesis is to study existing and develop new joint models for survival and omics data.

Collecting these ideas and establishing a doctoral thesis project in the field of biostatistics has its difficulties, from which also the great challenges arise. As an example, one difficulty is to manage a common vocabulary between the statistician and the clinician, and then to be able to find mutual profit. The clinician wants to answer a question and the statistician is interested in developing a new methodology to give the response, this is how a great network of multidisciplinary cooperation is born.

- [1] Dokyoon Kim et al. Predicting censored survival data based on the interactions between meta-dimensional omics data in breast cancer. Journal of Biomedical Informatics 56 (2015) 220-228.
- [2] Dimitrieva et al. Prognostic value of cross-omics screening for kidney clear cell renal cancer survival. Biology Direct (2016) 11:68

Discussion

Invited Speaker: Using Statistical Theory, Epidemiology and Clinical Trials for Public Health Impact

16.30-17.00 h.
Deborah Ashby, *Imperial College London, United Kingdom*

Abstract: Statistical theory, making epidemiological observations and carrying out incisive analyses to understand those observations, and designing and carrying out elegant clinical trials are all absorbing, challenging academic activities in their own right. However, sometimes we can be left wondering 'So what?'. Using those underpinning sciences to make a tangible difference to people's health sometimes means going the extra mile, and venturing into worlds beyond the ivory tower.

Drawing on the speaker's experiences from the worlds of academia, drug regulation and research funding, we look at some examples that have made a difference, reflecting on both the scientific underpinnings, and the pathways to making an impact.

Author Index

Author Index

C

D

E

F

K

L

Q

R

S

T

U

V

W

X

Y

Z

Sponsors

GOLDEN

Universidade Vigo

SiDOR
Statistical Inference
Decision & Operations Research Group
Universidade Vigo

CONCELLO DE VIGO

VITRASA

XUNTA DE GALICIA

SILVER

Cytel

BRONZE

bcam
basque center for applied mathematics

EXHIBITORS

WILEY

CRC Press
Taylor & Francis Group

Springer

Universidade Vigo

Since its **foundation in 1990,** the University of Vigo has consolidated itself as a reference of **modernity and innovation in Galicia.** The University of Vigo offers more than **40 undergraduate degrees** (including three double programs), **50 postgraduate degrees** and **40 PhD programs in the science, humanities, technology and legal-social fields.** Across its three campuses in Ourense, Pontevedra and Vigo, the University has more than **20 faculties** and a network of own research centres (Ecimat, Cacti, Cinbio, CITI…) that completes its research infrastructure map. The University of Vigo pursues academic excellence through the specialisation of its campuses: Campus da Auga in Ourense, Campus CREA S2i in Pontevedra and Vigo Tecnolóxico.

The University of Vigo is committed to knowledge transfer with society. It owns more than **200 patents** and has **active collaboration with the surrounding companies.** This academic institution encourages student entrepreneurship in every campuses through business incubation programs called Incuvi-Emprende and Incuvi-Avanza that provide students with the opportunity to develop their ideas.

In response to society's demands, the University of Vigo fosters the intellectual, professional and personal development of its students by providing them with leisure facilities in its three campuses including swimming pools, sports halls, theatres and exhibition halls.

The University **promotes student mobility** and is the number one university in Galicia for the number of incoming and outgoing students.

Gender equality and inclusion of students with disabilities are two of the University of Vigo's priorities.

CÍES A JEWEL OF DIVERSITY

Three islands that play at being two: Monteagudo, Faro and San Martiño. A protected paradise inhabited by a huge variety of flora and fauna. These are the Cíes Islands, the heart of the Galician Atlantic Islands Maritime-Terrestrial National Park, an authentic jewel of diversity that emerges as a magnificent gateway to the city of Vigo and its surroundings.

The Cíes are conserved as a natural paradise that is only open in summer and at Easter to a maximum of 2,200 people a day. The most valuable asset of this unique space is its biodiversity, from the microscopic beauty in the fields of seaweed to the grandeur of its aquatic mammals.

The scenic beauty of this setting, a National Park since 2002, captivates you from the moment you arrival with a magnificent view of what The Guardian newspaper proclaimed internationally as "the Best Beach in the World", connecting the islands of Monteagudo and Faro.

The marine environment of this natural paradise possesses great wildlife diversity. Along the rocky coasts are the habitats of barnacles, mussels and limpets; as we penetrate further into the depths anemones, sea urchins, crabs and octopus appear; and seeking refuge in the sand are razor shells, cockles and clams. This abundance of marine organisms sustains populations of marine birds such as the yellow-legged gull and the European shag cormorant, which form one of the largest breeding colonies in Europe.

And if the Cíes are a treasure by day – the night opens up a whole new perspective. Last year, the Islands became one of fourteen destinations in the world with "Starlight" accreditation.

galicia
o bo camiño

XUNTA DE GALICIA
galicia.es

Cytel is shaping the future of drug development. As the world's largest Biometrics CRO, we improve our customers' chances of success through expert trial design, efficient operational implementation and reliable clinical data management. The Cytel Strategic Consulting Team has designed and successfully implemented more adaptive trials than any other service provider. Learn more about our expertise and experience, together with our statistical and operations-support software, at **www.cytel.com**.

SAVE THE DATE

Joint International Society for Clinical Biostatistics and Australian Statistical Conference 26-30 August 2018

ISCB ASC18

26-30 AUGUST 2018

MELBOURNE, AUSTRALIA

Confirmed Keynote Speakers

Chris Holmes University of Oxford

Louise Ryan University of Technology, Sydney

Susan Murphy University of Michigan

Thomas Lumley University of Auckland

Key Dates

Registration opens **August 2017**

Abstract submission opens **October 2017**

Abstract submission closes **March 2018**

Early bird registration deadline **May 2018**

To receive updates, please register your interest:

Web: www.iscbasc2018.com

Email: iscbasc2018@arinex.com.au

Joint International Society for Clinical Biostatistics and Australian Statistical Conference 2018

HOSTED BY: International Society for Clinical Biostatistics; Statistical Society of Australia

MANAGED BY: ISCB ASC 2018 Conference Managers: Arinex Pty Ltd
91–97 Islington St, Collingwood, VIC 3066, Australia
Ph: +61 3 8888 9500

Technical Secretariat: Orzán Congres
Phone: +34 981 900 700
Email: iscb2017@orzancongres.com

ISCB