ISCB 42
LYON 2021

# 42nd Annual Conference of the International Society for Biostatistics

18-22 July 2021

## Final Programme & Book of Abstracts

ISCB 42
LYON 2021

The Local Organising Committee of the 42nd Conference of the International Society for Clinical Biostatistics (ISCB) would like to thank its sponsors for their support.

BIOMÉRIEUX

CLARA
CANCÉROPÔLE LYON AUVERGNE RHÔNE - ALPES

INSTITUT NATIONAL DU CANCER

izyCardio

LYRICAN
LYon Recherche Innovation contre le CANcer

Santé publique France

# Welcome to the ISCB 2021 Lyon Virtual Conference

On behalf of the Scientific Programme Committee (SPC), it is my great pleasure to welcome you to the **42nd Annual Conference of the International Society for Clinical Biostatistics (ISCB)**, The SPC members have been working to propose an exciting scientific program that devotes a large place to the main current challenges in Biostatistics.

The three short-courses focus on very recent tools and methods in Biostatistics: Joint Models under the Bayesian approach by **Dimitris Rizopoulo**s (Erasmus Univ, NL), Causal Mediation Analysis by **Linda Valeri** (Columbia Univ, US), Flexible penalized hazard model for time-to-event data by **Mathieu Fauvernier** (Univ of Lyon, FR).

For the plenary conferences, we welcome two distinguished speakers. The President's Invited Speaker **Per Kragh Andersen** (Copenhagen Univ, DK), and Keynote Speaker **Geert Molenberghs** (KU Leuven Univ, BE) will discuss modern methods for survival data and hierarchical data analysis.

The 8 invited sessions will focus on various hot topics in biostatistics such as epidemic modelling for COVID19, personalized medicine, cluster randomized trials, challenges and opportunities from long-term registries, causal inference for dense longitudinal data from wearable devices, combining deep learning and modelling, selective inference after variable selection and variance modelling for multilevel data.

For this virtual conference, according to the founding objectives of the ISCB, and in close collaboration with the LOC, the program was organized to maximize the participation from most time zones and the interactions between participants. Recorded sessions will be available on the platform for 2 months after the conference. Speakers have the opportunity to pursue discussion with attendees in dedicated virtual room. The posters are available at any time and for 2 months on the platform and each poster presenter will have 30 minutes virtual room during the conference for discussion with attendees. The long break time will enable networking thanks to the platform functionalities.

The last day, the **Early Career Biostatisticians' (ECB)** Day will give young researcher the opportunity to share experience while the mini-symposium will gather eminent biostatisticians from the **STRATOS** (STRengthening Analytical Thinking for Observational Studies) initiative.

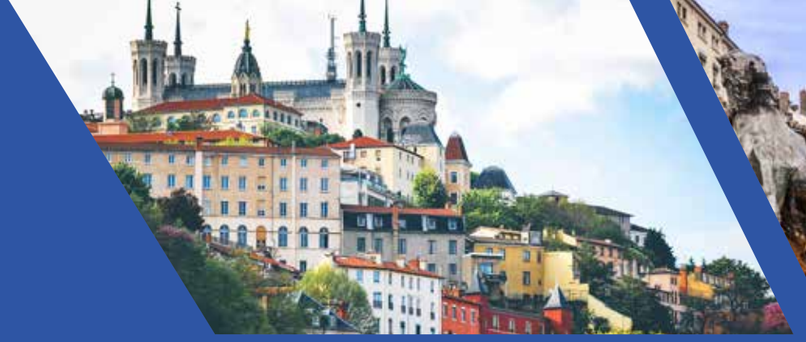
In this period when the COVID crisis has highlighted the essential role that biostatisticians hold in both clinical research and public health decision, this conference will be a great opportunity to present innovative methods and applications in biostatistics and to initiate new collaborations.

***Hélène Jacqmin-Gadda***
*Chair*
*Scientific Programme Committee*

ISCB 42
LYON 2021

# Welcome to the ISCB 2021 Virtual Conference

## Lyon, France, 18-22 July



On behalf of the Local Organizing Committee (LOC), I am very happy to welcome you to the **42nd Annual Conference of the International Society for Clinical Biostatistics (ISCB)**. To prepare this Conference, the members of the LOC worked in close collaboration with the Scientific Programme Committee (SPC) and wished, besides the strictly scientific content of the Conference, to offer the participants various aspects of the city of Lyon and its surrounding region (history, architecture, culture, etc.) with the hope that this glimpse will spark a desire to learn more about that charming city and a wish to visit it as soon as possible; hopefully, next summer.

Lyon is located in the Southeast of France (~470 km from Paris), at the confluence of two rivers: Rhône and Saône. It is currently the heart of the second-largest urban area in France and its old centre (year 476 to Renaissance) is registered on the UNESCO list of World Heritage Sites.

The city was founded by the Romans in 43 BC on a hill named Colline de Fourvière. It became soon an important site on the Roman road. Parts of this road, the Roman aqueduct, and the Roman amphitheatre are still visible close to a modern museum (Musée Gallo-romain) that displays interesting excavation remains. Lyon was the birthplace of two Emperors: Claudius and Caracalla. In this city, early Christians were martyred under Emperors of whom Marcus Aurelius and Septimius Severus. To recall this period, you may visit the Espace Culturel du Christianisme à Lyon, which is close to the ruins, the museum, and the famous mosaic-rich Basilique de Fourvière. From the Esplanade of Fourvière, you can admire nearly the whole city and, on clear days, the Alps and Mont Blanc on the horizon! You may next come down the hill to visit the medieval Vieux Lyon, its famous Traboules, and its impressive gothic Cathédrale Saint Jean.

In the 15th century, the introduction of fairs by Italian merchants made Lyon an important economic centre. During the Renaissance, under an Italian influence, Lyon was the first French city engaged in producing, waving, and trading silk. A few silk workshops, such as Maison des Canuts, are still active and open for visits. The Musée des Tissus et des Arts Décoratifs is dedicated to the history and astonishing produces of silk-making. Several shops in 'Vieux Lyon' and 'Presqu'île' propose genuine handicraft clothing articles or accessories.

The Italian Renaissance influence on the city architecture is still visible through many buildings from this period. By that time, Lyon was also active in book-publishing whose history is recalled in the Musée de l'Imprimerie et de la Communication graphique.

Lyon was the place where Lumière brothers (Auguste and Louis) invented the cinematograph in 1895 and produced the first French black and white films. Later, in 1903, they invented the Autochrome plates. Their Villa Lumière is now a museum dedicated to their memory and works.

During World War II, Lyon was a Nazi command centre and a Resistance stronghold. This period is also recalled in the Centre d'Histoire de la Résistance et de la Déportation whose buildings were once a Kommandantur and a detention centre.

Lyon hosts currently a number of international agencies (such as Interpol and the International Agency for Research on Cancer) and will soon host the WHO Academy. It is also a major centre for chemical industry (Solvay, Arkema, Total, Bayer, Air Liquide…), pharmaceutical and biotech industries (bioMérieux, Sanofi Pasteur, Genzyme, Merck Serono…), and software industries (Amaris, Tech Soft 3D, Inova Software, Digital Virgo…). Besides, the city hosts a great number of famous research and training centres (Universities, École Normale Supérieure, École Centrale, Institut National des Sciences Appliquées, etc.). Université Claude Bernard and Hospices Civils de Lyon (the main institutions that support the Conference) are putting many efforts to develop Biostatistics, Bioinformatics, and, more recently, Artificial Intelligence for medical research and applications. Nevertheless, they are giving more priority to humanities through Science/Medicine + Art/Culture association, an interface that is involving many researchers and artists to remind everyone that medicine deals with human life and well-being.

The heart of the city is Place Bellecour, the third-largest public square in France. The place displays nice small playgrounds, rest areas, and the famous statues of Louis XIV and Le Petit Prince. From there you can dwell between Renaissance or Haussman-style buildings, visit the Hôtel-Dieu, the Clocher de la Charité, the Célestins, the Opera, the Hôtel de Ville, the Fontaine de Bartholdy, the Musée des Beaux-Arts and its gardens… and do some shopping (rue de la République and roundabouts) or go antique hunting (rue Auguste Comte).

The Parc de la Tête d'Or is a real "heaven of peace" at the end of a busy day. This large grassy park (1.2 km²) includes a lake (160.000 m², boating possible in summer) and l'Île du Souvenir (war memorial). The Park is famous for its Porte des Enfants du Rhône (wrought iron majestic gate), its large botanical garden and greenhouse, its zoo including the Plaine africaine (nearly 130 animals in appropriate environment: giraffes, elephants, felines, deer, reptiles, primates, etc.) and, mainly, its worldwide-known Roseraie. This magnificent rose garden, especially in June, covers 50.000 m² and shows nearly 30,000 plants from hundreds of cultivars.

If it rains, Lyon will invite you to various other cultural sites: museums (Confluences, Gadagne, Beaux-Arts, Art Contemporain,……) or theaters (Célestins, TNP, Maison de la Danse, Orchestre National de Lyon, Opéra National de Lyon…).

Finally, the city is known for its cuisine and gastronomy. In the centre of the city, you may visit a Bouchon (traditional restaurant and local cuisine) or the Ancient Grand Hôtel-Dieu (historical hospital) that hosts now the Cité Internationale de la Gastronomie where you can find a wide variety of fancy cafés and restaurants. Not far from this site, you can find at La Halle Paul Bocuse a selection of the finest local products.

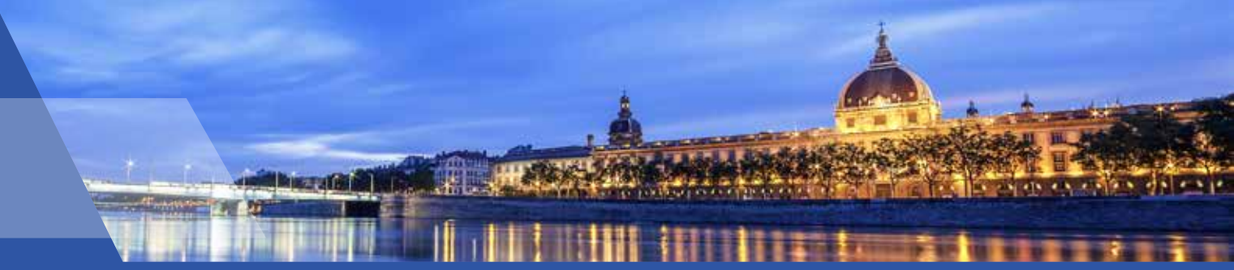
***For still more information you may visit websites like "This is Lyon".***

Welcome once more to the Conference. Its Social Programme will remind you of the all those good reasons to visit Lyon as soon as the travel and stay conditions will improve.

***Pascal Roy***
*Head of the Local Organizing Committee*

ISCB 42
LYON 2021

# Table of Contents





## Book of Abstracts

# Organisation – Committees

## Scientific Programme Committee

**Chair:**
Hélène Jacqmin-Gadda (FR)

**Members:**
Ismael Ahmed (FR)
Harald Binder (DE)
Andrew Copas (UK)
Stacia Desantis (US)
Tim Friede (DE)
Roch Giorgi (FR)
Els Goetghebeur (BE)
Donald Hedeker (US)
Georg Heinze (AT)
Boris Hejblum (FR)
Karen Leffondre (FR)
Catherine Legrand (BE)
Delphine Maucort-Boulch (FR)
Havi Murad (IL)
Muriel Rabilloud (FR)
Dimitris Rizopoulos (NL)
Pascal Roy (FR, LOC Chair)
Linda Sharples (UK)
Brian Tom (UK)
Linda Valeri (US)
Sarah Zohar (FR)

## Local Organising Committee

**Chair:**
Pascal Roy

**Vice-Chair:**
Delphine Maucort-Boulch

**Treasurers:**
Muriel Rabilloud
Stéphanie Dupinay

**Secretary:**
Fabien Subtil

**Relationships with Geneva / Lausanne Mini-symposia Coordinator:**
David Warne

**Webmaster:**
Mathieu Fauvernier

**Scientific Programme Committee Coordinator:**
Hélène Jacqmin-Gadda

**Sponsor acquisition:**
Jean Iwaz
Nadine Bossard

**Conference facilities (technical equipment, social programme):**
Paola Damaso
Stéphanie Dupinay

## Organisation

**University Lyon 1**
Cellule Congrès - ISCB 2021
Batiment Géode, 43 bd du 11 Novembre 1918, 69622 Villeurbanne Cedex, France
Phone: +33 (0) 4 72 43 35 90 • Email: iscb2021@univ-lyon1.fr • Website: www.iscb2021.org

# About ISCB

The International Society for Clinical Biostatistics (ISCB) was founded in 1978 to stimulate research into the principles and methodology used in the design and analysis of clinical research and to increase the relevance of statistical theory to the real world of clinical medicine.
Membership is open to all interested individuals who share the Aims of the Society. ISCB's membership include clinicians, statisticians and members of other disciplines, such as epidemiologists, clinical chemists and clinical pharmacologists, working or interested in the field of clinical biostatistics.
ISCB has an Executive Committee and 6 Subcommittees: Conference Organising, Early Career Biostatisticians, Education, National Groups, Statistics in Regulatory Affairs (SiRA), Student Conference Awards.

## Executive Committee

**President:**
Zdenek Valenta (CZ)

**Vice-President:**
Tomasz Burzykowski (BE)

**Secretary:**
Elaine Pascoe (AU)

**Treasurer:**
Chris Metcalfe (UK)

**Members:**
Vana Sypsa (GR)
Dimitris Rizopoulos (NL)
Gerta Ruecker (DE)
Jonathan Bartlett (UK)
Krystyna Szafraniec (PL)
Laure Wynants (NL)
Lehana Thabane (CA)
Nan van Geloven (NL)
Nicole Close (US)
David W. Warne (CH)

## Honorary Members

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

## ISCB Permanent Office

convin

29 K. Varnali Street, 152 33 Chalandri,
Athens, Greece
**Call center:** +30 210 6833600
**Mobile contact:** +30 6956 665669
**Website:** www.iscb.info
**Email:** secretariat@iscb.info

# 2021 Award Recipients

### Student Conference Awards (StCA)

| Name | Country | University | Title | Session |
|---|---|---|---|---|
| Solon Karapanagiotis | UK | MRC Biostatistics Unit | Tailored Bayesian variable selection for risk prediction modelling under unequal misclassification costs | **OC4D** 20 Jul |
| Annabel Davies | UK | University of Manchester | Network meta-analysis and random walks | **OC1C** 19 Jul |
| Subodh Selukar | US | University of Washington | RECeUS: Ratio Estimation of Censored Uncured Subjects for Studying Sufficient Follow-Up in Studies of Long-Term Survivors | **OC5E** 20 Jul |

### Conference Awards for Scientists (CAS)

The National Groups Subcommittee has decided not to grant any award for ISCB Conference Awards for Scientists.

### Conference Fund for Developing Countries (CFDC)

| Name | Country | Title | Session |
|---|---|---|---|
| Hao Yuan | CN | Impact of competing event in COVID-19 clinical data analysis | **OC4E** 20 Jul |
| Hongji Wu | CN | Analysis of competing risks data using restricted mean time lost | **OC4E** 20 Jul |
| Moumita Chatterjee | IN | A martingale based approach for modelling the alternating recurrences in Cystic Fibrosis patients | **OC1F** 19 Jul |
| Noor Jahan Akter | BD | Mixed Modeling of Regional Infant Mortality Data over Twenty Years in North Rhine-Westphalia. | **OC1G** 19 Jul |
| Rubaiya Rubaiya | BD | Relation between women empowerment and birth interval: A survival analysis approach | **P07** 20 Jul |

# Keynote Speakers

## President's Invited Speaker

**Per Kragh Andersen**

Professor, Section of Biostatistics, Department of Public Health University of Copenhagen, DK

https://biostat.ku.dk/

**MONDAY, 19 JULY 2021 | 13:00-14:00**

**Lecture Title:** ***The Joy of Pseudo-Values***

## Keynote Invited Speaker

**Geert Molenberghs**

Professor, Faculty of Medicine, Leuven Biostatistics and Statistical Bioinformatics Centre (L-BioStat), KU Leuven, BE

https://www.kuleuven.be/wieiswie/en/person/00056633

**WEDNESDAY, 21 JULY 2021 | 17:00-18:00**

**Lecture Title:** ***Handling Negative Correlation and/or Over/Underdisperson in Gaussian and Non-Gaussian Hierarchical Data***

ISCB 42 LYON 2021

# 42nd Annual Conference of the International Society for Biostatistics

18-22 July 2021

## Programme Overview

| Time in CEST | SUN 18 JUL 21 | | | MON 19 JUL 21 | | TUE 20 JUL 21 | | | WED 21 JUL 21 | | THU 22 JUL 21 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | HALL 1 | HALL 2 | HALL 3 | HALL 1 | HALLS 2-8 | HALL 1 | HALL 2 | HALLS 3-8 | HALL 1 | HALLS 2-8 | HALL 1 | HALL 2 |
| 12:00-12:15 | PCC1 (A) | PCC2 (A) | PCC3 (A) | OPENING CEREMONY | | | | | | | | |
| 12:15-12:30 | | | | | | | | | | | | |
| 12:30-12:45 | | | | | | | | | | | | |
| 12:45-13:00 | | | | | | | | | | | | |
| 13:00-13:15 | | | | PRESIDENT'S INVITED SPEAKER | | IS3 | OC3A OC3B OC3C OC3D OC3E OC3F OC3G | | IS7 | OC6A OC6B OC6C OC6D OC6E OC6F | MS (A) | ECB (A) |
| 13:15-13:30 | | | | | | | | | | | | |
| 13:30-13:45 | | | | | | | | | | | | |
| 13:45-14:00 | | | | | | | | | | | | |
| 14:00-14:15 | Break – Exchange time | | | **Break – Exchange time**<br>14:00-14:30 \| Virtual Rooms P1 & P2<br>14:30-14:35 \| Time for Art 1 | | **Break – Exchange time**<br>14:30-15:00 \| Virtual Rooms IS3<br>14:30-15:00 \| Virtual Rooms OC3A to OC3G<br>14:30-15:00 \| Virtual Rooms P6<br>15:00-15:05 \| Time for Art 3 | | | **Break – Exchange time**<br>14:30-15:00 \| Virtual Rooms IS7<br>14:30-15:00 \| Virtual Rooms OC6A to OC6F<br>14:30-15:00 \| Virtual Rooms P9<br>15:00-15:05 \| Time for Art 5 | | Break – Exchange time | |
| 14:15-14:30 | | | | | | | | | | | | |
| 14:30-14:45 | PCC1 (B) | PCC2 (B) | PCC3 (B) | IS1 | OC1A OC1B OC1C OC1D OC1E OC1F OC1G | IS4 | IS5 | OC4A OC4B OC4C OC4D OC4E | IS8 | OC7A OC7B OC7C OC7D OC7E OC7F OC7G | MS (B) | ECB (B) |
| 14:45-15:00 | | | | | | | | | | | | |
| 15:00-15:15 | | | | | | | | | | | | |
| 15:15-15:30 | | | | | | | | | | | | |
| 15:30-15:45 | | | | | | | | | | | | |
| 15:45-16:00 | | | | | | | | | | | | |
| 16:00-16:15 | | | | | | | | | | | | |
| 16:15-16:30 | | | | **Break – Exchange time**<br>16:15-16:45 \| Virtual Rooms IS1<br>16:15-16:45 \| Virtual Rooms OC1A to OC1G<br>16:15-16:45 \| Virtual Rooms P3<br>16:45-16:50 \| Time for Art 2 | | **Break – Exchange time**<br>16:45-17:15 \| Virtual Rooms IS4 & IS5<br>16:45-17:15 \| Virtual Rooms OC4A to OC4E<br>16:45-17:15 \| Virtual Rooms P7<br>17:15-17:20 \| Time for Art 4 | | | Break – Exchange time | | | |
| 16:30-16:45 | Break – Exchange time | | | | | | | | | | | |
| 16:45-17:00 | | | | | | | | | | | | |
| 17:00-17:15 | PCC1 (C) | PCC2 (C) | PCC3 (C) | IS2 | OC2A OC2B OC2C OC2D OC2E OC2F OC2G | IS6 | OC5A OC5B OC5C OC5D OC5E OC5F OC5G | | KEYNOTE INVITED SPEAKER | | CLOSING & ISCB 2022 | |
| 17:15-17:30 | | | | | | | | | | | | |
| 17:30-17:45 | | | | | | | | | | | | |
| 17:45-18:00 | | | | | | | | | | | | |
| 18:00-18:15 | | | | | | | | | Break – Exchange time | | **Break – Exchange time**<br>18:00-18:10 \| Time for Art 7 | |
| 18:15-18:30 | | | | | | | | | | | | |
| 18:30-18:45 | | | | 18:30-19:00 \| Virtual Rooms IS2<br>18:30-19:00 \| Virtual Rooms OC2A to OC2G<br>18:30-19:00 \| Virtual Rooms P4 & P5 | | | | | ISCB AGM<br>1815-1930 | | | |
| 18:45-19:00 | | | | | | | | | | | | |
| 19:00-19:15 | Break – Exchange time | | | **Virtual Get Together**<br>19 Jul 1900-2030 | | 19:00-19:30 \| Virtual Rooms IS6<br>19:00-19:30 \| Virtual Rooms OC5A to OC5G<br>19:00-19:30 \| Virtual Rooms P8 | | | **Break – Exchange time**<br>19:30-20:00 \| Virtual Rooms IS8<br>19:30-20:00 \| Virtual Rooms OC7A to OC7G<br>19:30-20:00 \| Virtual Rooms P10<br>20:00-20:05 \| Time for Art 6 | | | |
| 19:15-19:30 | **Welcome to the Conference** | | | | | **Virtual Gala** | | | | | | |
| 19:30-19:45 | **Escape Game** | | | | | | | | | | | |
| 19:45-20:00 | | | | | | | | | | | | |
| 20:00-20:15 | | | | | | **Gala Concerts 2000-2200** | | | | | | |
| 20:15-20:30 | | | | | | | | | | | | |
| 20:30-20:45 | | | | | | | | | | | | |
| 20:45-21:00 | | | | | | | | | | | | |

## Overview Sunday, 18 July 2021

| Time in CEST | SUNDAY \| 18 JULY 2021 | | |
|---|---|---|---|
| | PRE-CONFERENCE COURSES | | |
| | HALL 1 | HALL 2 | HALL 3 |
| 12:00-12:15 | **PCC1** Joint Models under the Bayesian approach (Part A) | **PCC2** Causal Mediation Analysis (Part A) | **PCC3** Flexible penalized hazard model for time-to-event data (Part A) |
| 12:15-12:30 | | | |
| 12:30-12:45 | | | |
| 12:45-13:00 | | | |
| 13:00-13:15 | | | |
| 13:15-13:30 | | | |
| 13:30-13:45 | | | |
| 13:45-14:00 | | | |
| 14:00-14:15 | **Break – Exchange time** | | |
| 14:15-14:30 | | | |
| 14:30-14:45 | | | |
| 14:45-15:00 | **PCC1** Joint Models under the Bayesian approach (Part B) | **PCC2** Causal Mediation Analysis (Part B) | **PCC3** Flexible penalized hazard model for time-to-event data (Part B) |
| 15:00-15:15 | | | |
| 15:15-15:30 | | | |
| 15:30-15:45 | | | |
| 15:45-16:00 | | | |
| 16:00-16:15 | | | |
| 16:15-16:30 | | | |
| 16:30-16:45 | **Break – Exchange time** | | |
| 16:45-17:00 | | | |
| 17:00-17:15 | **PCC1** Joint Models under the Bayesian approach (Part C) | **PCC2** Causal Mediation Analysis (Part C) | **PCC3** Flexible penalized hazard model for time-to-event data (Part B) |
| 17:15-17:30 | | | |
| 17:30-17:45 | | | |
| 17:45-18:00 | | | |
| 18:00-18:15 | | | |
| 18:15-18:30 | | | |
| 18:30-18:45 | | | |
| 18:45-19:00 | | | |
| 19:00-19:15 | **Break – Exchange time** | | |
| 19:15-19:30 | **Welcome to the Conference** | | |
| 19:30-19:45 | **Escape Game** | | |
| 19:45-20:00 | | | |
| 20:00-20:15 | | | |
| 20:15-20:30 | | | |
| 20:30-20:45 | | | |
| 20:45-21:00 | | | |

## Overview Monday, 19 July 2021

| Time in CEST | MONDAY \| 19 JULY 2021 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | HALL 1 | HALL 2 | HALL 3 | HALL 4 | HALL 5 | HALL 6 | HALL 7 | HALL 8 |
| 12:00-12:15 | **BIENVENUE!** **ISCB 2021 OPENING CEREMONY** | | | | | | | |
| 12:15-12:30 | | | | | | | | |
| 12:30-12:45 | | | | | | | | |
| 12:45-13:00 | | | | | | | | |
| 13:00-13:15 | **PRESIDENT'S INVITED SPEAKER** | | | | | | | |
| 13:15-13:30 | | | | | | | | |
| 13:30-13:45 | | | | | | | | |
| 13:45-14:00 | | | | | | | | |
| 14:00-14:15 | **Break – Exchange time** 14:00-14:30 \| Virtual Rooms P1: Clinical trial design 14:00-14:30 \| Virtual Rooms P2: Longitudinal data analysis 14:30-14:35 \| Time for Art 1 | | | | | | | |
| 14:15-14:30 | | | | | | | | |
| 14:30-14:45 | | | | | | | | |
| 14:45-15:00 | **IS1** Modelling the global spread of Covid19 and impact of interventions | **OC1A** Mendelian randomi-sation & causal inference | **OC1B** Bayesian clinical trial design -1 | **OC1C** Network meta analysis | **OC1D** Omics and genetic studies | **OC1E** Multiple testing and randomiza-tion tests | **OC1F** Methods for survival analysis | **OC1G** Mixed effects model |
| 15:00-15:15 | | | | | | | | |
| 15:15-15:30 | | | | | | | | |
| 15:30-15:45 | | | | | | | | |
| 15:45-16:00 | | | | | | | | |
| 16:00-16:15 | | | | | | | | |
| 16:15-16:30 | **Break – Exchange time** 16:15-16:45 \| Virtual Rooms IS1 16:15-16:45 \| Virtual Rooms OC1A to OC1G 16:15-16:45 \| Virtual Rooms P3: Causal inference 16:45-16:50 \| Time for Art 2 | | | | | | | |
| 16:30-16:45 | | | | | | | | |
| 16:45-17:00 | | | | | | | | |
| 17:00-17:15 | **IS2** Personalized Medicine with Dynamic Predictions | **OC2A** Mediation analysis | **OC2B** Bayesian clinical trial design -2 | **OC2C** Epidemic model-ling of COVID19 | **OC2D** Meta-analysis: network and other | **OC2E** Some thought on research in biostatistics | **OC2F** Methods for anal-ysis of electronic health records | **OC2G** Analysis of gene expression and omics data |
| 17:15-17:30 | | | | | | | | |
| 17:30-17:45 | | | | | | | | |
| 17:45-18:00 | | | | | | | | |
| 18:00-18:15 | | | | | | | | |
| 18:15-18:30 | | | | | | | | |
| 18:30-18:45 | **18:30-19:00 \| Virtual Rooms IS2 18:30-19:00 \| Virtual Rooms OC2A to OC2G** 18:30-19:00 \| Virtual Rooms P4: High Dimensional Data 18:30-19:00 \| Virtual Rooms P5: Machine Learning | | | | | | | |
| 18:45-19:00 | | | | | | | | |
| 19:00-19:15 | **Virtual Get Together 1900-2030** | | | | | | | |
| 19:15-19:30 | | | | | | | | |

## Overview Tuesday, 20 July 2021

| Time in CEST | TUESDAY \| 20 JULY 2021 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | HALL 1 | HALL 2 | HALL 3 | HALL 4 | HALL 5 | HALL 6 | HALL 7 | HALL 8 |
| 13:00-13:15<br>13:15-13:30<br>13:30-13:45<br>13:45-14:00<br>14:00-14:15<br>14:15-14:30 | **IS3**<br>Optimal design of longitudinal cluster randomised trials | **OC3A**<br>Propensity score in causal studies | **OC3B**<br>COVID19 modelling | **OC3C**<br>Dynamic prediction | **OC3D**<br>Meta-analysis for prediction models | **OC3E**<br>Multi-state model | **OC3F**<br>Neural network and machine learning | **OC3G**<br>Study designs |
| 14:30-14:45<br>14:45-15:00<br>15:00-15:15 | **Break – Exchange time**<br>14:30-15:00 \| Virtual Rooms IS3 14:30-15:00 \| Virtual Rooms OC3A to OC3G<br>14:30-15:00 \| Virtual Rooms P6: Epidemiology<br>15:00-15:05 \| Time for Art 3 | | | | | | | |
| 15:15-15:30<br>15:30-15:45<br>15:45-16:00<br>16:00-16:15<br>16:15-16:30<br>16:30-16:45 | **IS4**<br>Challenges and opportunities for learning from long term disease registers | **IS5**<br>The best of both worlds: combining deep learning and modeling | **OC4A**<br>Missing data in causal studies | **OC4B**<br>Cluster randomized trials | **OC4C**<br>Meta-analysis | **OC4D**<br>Prediction model for omics data | **OC4E**<br>Competing risks and multi-state models | |
| 16:45-17:00<br>17:00-17:15<br>17:15-17:30 | **Break – Exchange time**<br>16:45-17:15 \| Virtual Rooms IS4 & IS5 16:45-17:15 \| Virtual Rooms OC4A to OC4E<br>16:45-17:15 \| Virtual Rooms P7: Survival analysis<br>17:15-17:20 \| Time for Art 4 | | | | | | | |
| 17:30-17:45<br>17:45-18:00<br>18:00-18:15<br>18:15-18:30<br>18:30-18:45<br>18:45-19:00 | **IS6**<br>Causal inference in continuous time for dense longitudinal data from wearable devices | **OC5A**<br>Bayesian clinical trial analysis | **OC5B**<br>Prediction by Machine learning | **OC5C**<br>Screening and Diagnostic studies | **OC5D**<br>Bayesian Joint models for longitudinal data and time-to-event | **OC5E**<br>Cure and mixture models | **OC5F**<br>Longitudinal data analysis | **OC5G**<br>Missing data & measurement error |
| 19:00-19:15<br>19:15-19:30 | 19:00-19:30 \| Virtual Rooms IS6 19:00-19:30 \| Virtual Rooms OC5A to OC5G<br>19:00-19:30 \| Virtual Rooms P8: Epidemic modeling | | | | | | | |
| 19:30-19:45<br>19:45-20:00 | **Virtual Gala** | | | | | | | |
| 20:00-20:15<br>20:15-20:30 | Gala Concerts 2000-2200 | | | | | | | |

## Overview Wednesday, 21 July 2021

| Time in CEST | WEDNESDAY \| 21 JULY 2021 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | HALL 1 | HALL 2 | HALL 3 | HALL 4 | HALL 5 | HALL 6 | HALL 7 | HALL 8 |
| 13:00-13:15<br>13:15-13:30<br>13:30-13:45<br>13:45-14:00<br>14:00-14:15<br>14:15-14:30 | **IS7**<br>Selective inference after variable selection | **OC6A**<br>Causal inference | **OC6B**<br>Adaptive clinical trial design | **OC6C**<br>Joint models & 2-stage approach | **OC6D**<br>Time-to-event methods for non-proportional hazards | **OC6E**<br>Relative survival and net benefit estimation | **OC6F**<br>Deep learning | |
| 14:30-14:45<br>14:45-15:00<br>15:00-15:15 | **Break – Exchange time**<br>14:30-15:00 \| Virtual Rooms IS7 14:30-15:00 \| Virtual Rooms OC6A to OC6F<br>14:30-15:00 \| Virtual Rooms P9: Individual prediction and precision medicine<br>15:00-15:05 \| Time for Art 5 | | | | | | | |
| 15:15-15:30<br>15:30-15:45<br>15:45-16:00<br>16:00-16:15<br>16:15-16:30<br>16:30-16:45 | **IS8**<br>Variance modelling for multi-level data and joint models | **OC7A**<br>Causal inference in survival analysis | **OC7B**<br>Clinical trial design and sample size calculation | **OC7C**<br>Frailty model and recurrent events | **OC7D**<br>Functional data analysis | **OC7E**<br>Selection and validation of prediction models | **OC7F**<br>Flexible modelling and spatial data analysis | **OC7G**<br>Methods for clinical research |
| 16:45-17:00 | **Break – Exchange time** | | | | | | | |
| 17:00-17:15<br>17:15-17:30<br>17:30-17:45<br>17:45-18:00 | **KEYNOTE INVITED SPEAKER** | | | | | | | |
| 18:00-18:15 | **Break – Exchange time** | | | | | | | |
| 18:15-18:30<br>18:30-18:45<br>18:45-19:00<br>19:00-19:15 | **ISCB ANNUAL GENERAL MEETING**<br>**1815-1930** | | | | | | | |
| 19:15-19:30<br>19:30-19:45<br>19:45-20:00 | 19:30-20:00 \| Virtual Rooms IS8 19:30-20:00 \| Virtual Rooms OC7A to OC7G<br>19:30-20:00 \| Virtual Rooms P10: Meta analysis<br>20:00-20:05 \| Time for Art 6 | | | | | | | |

ISCB 42
LYON 2021

## Overview Thursday, 22 July 2021

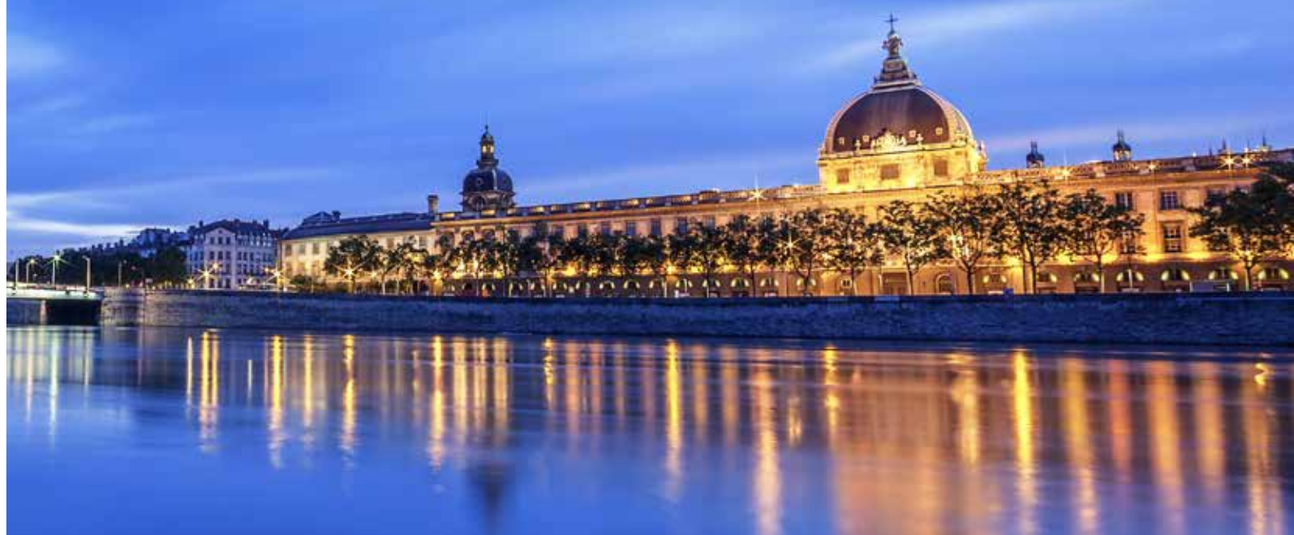
| Time in CEST | THURSDAY \| 22 JULY 2021 | |
|---|---|---|
| | **HALL 1** | **HALL 2** |
| 13:00-13:15<br>13:15-13:30<br>13:30-13:45<br>13:45-14:00<br>14:00-14:15<br>14:15-14:30 | STRATOS<br>Mini-Symposium<br>(Part A) | Early Career<br>Biostatisticians' Day<br>(Part B) |
| 14:30-14:45<br>14:45-15:00<br>15:00-15:15<br>15:15-15:30 | **Break – Exchange time** | |
| 15:30-15:45<br>15:45-16:00<br>16:00-16:15<br>16:15-16:30<br>16:30-16:45<br>16:45-17:00 | STRATOS<br>Mini-Symposium<br>(Part B) | Early Career<br>Biostatisticians' Day<br>(Part B) |
| 17:00-17:15<br>17:15-17:30<br>17:30-17:45<br>17:45-18:00 | **CLOSING CEREMONY & INVITATION TO ISCB 43 NEWCASTLE** | |
| 18:00-18:15<br>18:15-18:30 | **Break – Exchange time**<br>18:00-18:10 \| Time for Art 7 | |



ISCB 42
LYON 2021

# Detailed Programme





# Sunday, 18 July 2021

HALL 1

## 12:00-19:00 PRE-CONFERENCE COURSE 1: Joint Models under the Bayesian approach

**ORGANISED BY: Dimitris Rizopoulos,** Rotterdam, NL

**TIMELINE:**

12:00-14:00 → **Part A**
14:00-14:30 → BREAK
14:30-16:30 → **Part B**
16:30-17:00 → BREAK
17:00-19:00 → **Part C**

**OUTLINE:**

1. Introduction
2. Linear Mixed-Effects Models
3. Relative Risk Models
4. The Basic Joint Model
5. Extensions of Joint Models
6. Dynamic Predictions
7. Closing
8. Practicals

**Abstract:** In follow-up studies, different types of outcomes are typically collected for each subject. These include longitudinally measured responses (e.g., biomarkers), and the time until an event of interest occurs (e.g., death, dropout). Often these outcomes are separately analyzed, but on many occasions, it is of scientific interest to study their association. This type of research question has given rise in the class of joint models for longitudinal and timeto-event data. These models constitute an attractive paradigm for the analysis of follow-up data that is mainly applicable in two settings: First, when the focus is on a survival outcome, and we wish to account for the effect of endogenous time-dependent covariates measured with error. Second, when the focus is on the longitudinal outcome, and we want to correct for nonrandom dropout.This full-day course is aimed at applied researchers and graduate students and will provide a comprehensive introduction to this modeling framework. We will explain when these models should be used in practice, which are the key assumptions behind them, and how they can be utilized to extract relevant information from the data. Emphasis is given on applications, and after the end of the course, participants will be able to define appropriate joint models to answer their questions of interest.

**Necessary background:** This course assumes knowledge of basic statistical concepts, such as standard statistical inference using maximum likelihood and regression models. Also, a basic knowledge of R would be beneficial but is not required. Participants are required to bring their laptop with the battery fully charged. Before the course instructions will be sent for installing the required software.

**Target audience:** Professional statisticians working in applied environments where hierarchical modeling and survival analysis are key issues; this would include biostatisticians working in the pharmaceutical industry, regulatory agencies, or academic centers.

**Learning objectives:** After this course, participants should be able to identify settings in which a joint modeling approach is required. From the course, it will become clear which joint models can be used depending on the actual research questions to be answered, and which model-building strategies are currently available. Further, participants should be able to construct and fit an appropriate joint model, correctly interpret the obtained results, and extract additional useful information (e.g., plots) that can help communicate the results better.The course will be explanatory rather than mathematically rigorous. Therefore emphasis is given in sufficient detail for participants to obtain a clear view of the different joint modeling approaches and how they should be used in practice. To this end, we first motivate joint modeling using real datasets and then illustrate in detail the virtues and drawbacks of each of the presented joint modeling approaches. For completeness and throughout the course, references are provided to material with more technical information.

**Presenter background:** Dimitris Rizopoulos is a Professor in Biostatistics at the Erasmus University Medical Center. He received an M.Sc. in statistics (2003) from the Athens University of Economics and Business, and a Ph.D. in Biostatistics (2008) from the Katholieke Universiteit Leuven. Dr. Rizopoulos wrote his dissertation, as well as a number of methodological and applied articles on various aspects of models for survival and longitudinal data analysis, and he is the author of a recent book on the topic of joint models for longitudinal and time-to-event data. He has also written two freely available packages to fit this type of models in R under maximum likelihood (i.e., package JM) and the Bayesian approach (i.e., package JMbayes). He currently serves as co-Editor for Biostatistics.

# Sunday, 18 July 2021

HALL 2

## 12:00-19:00 PRE-CONFERENCE COURSE 2: Causal Mediation Analysis

**ORGANISED BY: Linda Valeri,** New York NY, US

**TIMELINE:**

12:00-14:00 → **Part A**
14:00-14:30 → BREAK
14:30-16:30 → **Part B**
16:30-17:00 → BREAK
17:00-19:00 → **Part C**

**OUTLINE:**

1a. Traditional statistical approaches for mediation analysis and their limitations
1b. Theory of counterfactual based mediation analysis
2a. Regression based approaches for mediation analysis theory and application with R package CMAverse
2b. Decomposition of the total effect in mediating and interactive mechanisms theory and application with R package CMAverse
3. Sensitivity analyses for unmeasured confounding theory and application with R package CMAverse

**Abstract:** Mediation analysis concerns assessing the mechanisms and pathways by which causal effects operate. The course will cover the relationship between traditional methods for mediation in epidemiology and the social sciences and new methods in causal inference. All theoretical concepts will be set into the context of research questions in the biomedical social sciences. The course will enable the participants to conduct their own mediation analyses in settings with either single or multiple mediators. R packages, SAS and Stata commands to implement these techniques will be covered and distributed to course participants. The use and implementation of sensitivity analysis techniques to assess the how sensitive conclusions are to violations of assumptions will be covered.

**Target audience:** All participants are expected to have ample experience with the application of regression-based modeling (i.e. multiple linear regressions and logistic regressions as a minimum) and familiarity with R. Participants are not expected to have had prior exposure to the potential outcome framework for causal inference.

**Learning outcomes:**
At the end of the course, participants should:

1. Understand the limitations of traditional mediation analyses based on structural equation models.
2. Understand the meaning of natural direct and indirect effects.
3. Be able to assess the plausibility of the assumptions required for the identification of natural direct and indirect effects.
4. Be able to interpret causal mediation analysis models.
5. Be able to conduct mediation analyses based on CMAverse in R.
6. Be able to communicate the results and assumptions of a mediation analysis.

**Presenter background:** Linda Valeri is an assistant professor in Biostatistics at the Columbia University Mailman School of Public Health. Dr. Valeri joined the Department of Biostatistics at Columbia University in 2018 after 3 years as faculty at the Laboratory of Psychiatric Biostatistics of McLean Hospital and the Department of Psychiatry at Harvard Medical School. Dr. Valeri received her PhD in Biostatistics from Harvard University in 2013, where her dissertation focused on statistical methods for causal mediation analysis. Dr. Valeri is an expert in causal inference with a focus on statistical methods for causal mediation analysis, measurement error, and missing data. She is interested in translating statistical methods in public health to improve our understanding of mental health, environmental determinants of health, and health disparities.

# Sunday, 18 July 2021

HALL 3

12:00-19:00 **PRE-CONFERENCE COURSE 3**
Flexible penalized hazard model for time-to-event data

**ORGANISED BY:**
**Mathieu Fauvernier**, Lyon, FR

**CO-INSTRUCTORS:**
**Laurent Remontet**, Lyon, FR / **Laurent Roche**, Lyon, FR / **Zoé Uhry**, Lyon, FR / **Emmanuelle Dantony**, Lyon, FR

**TIMELINE:**

| | | |
|---|---|---|
| 12:00-14:00 | → | **Part A** |
| 12:00-12:05 | **Mathieu Fauvernier** | Introduction |
| 12:05-13:00 | **Emmanuelle Dantony** | Basic concepts in hazard regression |
| 13:00-14:00 | **Mathieu Fauvernier** | Penalized hazard model – Part 1 |
| 14:00-14:30 | → | BREAK |
| 14:30-16:30 | → | **Part B** |
| 14:30-16:30 | **Mathieu Fauvernier** | / Penalized hazard model – Part 2 |
| 16:30-17:00 | → | BREAK |
| 17:00-19:00 | → | **Part C** |
| 17:00-18:00 | **Zoé Uhry** | / Practical examples in R – Part 1 |
| 18:00-19:00 | **Mathieu Fauvernier** | / Practical examples in R – Part 2 |

**OUTLINE:** The course will include a theoretical and a practical session.

1. In the **theoretical session**, we will present key aspects of flexible parametric hazard models focusing on the notion of dynamics of the hazard. Then we will present penalized hazard models and their extension to excess hazard models.
2. In the **practical session**, participants will learn how to fit suitable models on real data using R package survPen, how to properly predict various outcomes from their fitted models and how to graphically assess the validity of their models.

**Target audience:** The course is presented at a level that can be handled by all statisticians/biostatisticians, or epidemiologists and medical researchers who are familiar with regression analyses, including survival models. Experience with R software is also required.

**Presenters' background:**

**Mathieu Fauvernier:** I graduated in Actuarial Science (University of Lyon, France) in 2013 and worked for two years as an actuary in a French consulting firm near Lyon. In June 2016, I graduated in Biostatistics (Lyon) and started a PhD in Biostatistics which I completed in November 2019. I am currently working at the Department of Biostatistics (Hospices Civils de Lyon, France). My research project is about modelling the effects of predictive factors, along with their interactions, on various epidemiological indicators (incidence, mortality, and net survival) using smooth functions. In that regard, I have developed the survPen R package that allows fitting penalized (excess) hazard models. Number of publications: 8.

**Emmanuelle Dantony:** I graduated in Bioinformatics and Modeling (INSA engineering school, Lyon, France) in 2005. I joined the Department of Biostatistics (Hospices Civils de Lyon, France) in 2006. My main research topics were analyses of high-dimensional data (especially transcriptomic data) and multi-state modeling with application to end-stage renal disease. I also participated in various clinical research projects. In 2015, I joined the Department's team specialized in cancer epidemiology and participated since then in issuing country-wide statistics on cancer incidence, mortality, and survival. Number of publications: 30.

**Zoé Uhry:** I graduated in Stochastic Modeling (University of Grenoble, France) and in Applied Mathematics and Informatics (ENSIMAG engineering school) in 1997. I first worked as a biostatistician in genetic epidemiology in an INSERM research unit. In 2000, I integrated Santé Publique France and worked in the field of cancer epidemiology on two main topics: cancer screening evaluation and development of a calibration model to estimate cancer incidence at a local level in areas without registry. In 2014, I joined the Department of Biostatistics (Hospices Civils de Lyon, France) in an outposted SPF position, where my main duties are developing/implementing statistical methods to provide national epidemiological indicators on cancer in France (incidence, mortality and survival). I was also involved in a study on cancers net survival trends in six Latin European countries. Number of publications: 37.

**Laurent Remontet:** I graduated in Applied Mathematics and Sociology (University XI, Paris, France) in 1992 then worked for four years on non-linear mixed models for growth curves in an INSERM research unit in epidemiology. In 1998, I integrated the Department of Biostatistics (Hospices Civils de Lyon, France) in which I worked for 15 years with Professor Jacques Esteve in the field of statistical methods for descriptive cancer epidemiology. The Department being in charge of producing the French cancer epidemiological statistics (at the national and sub-national level), my current main responsibility is to coordinate this activity which involves the development and implementation of statistical methods for estimating incidence, mortality, survival, and net survival using, in particular GAM and penalized survivals models. Number of publications: 95.

**Laurent Roche:** L. Roche first graduated in Mathematics in 2004 (Université Claude Bernard Lyon 1, France) then taught mathematics in high schools. In 2008, he obtained a Master's degree in applied mathematics (statistics, informatics, and numerical methods) and, in 2009, integrated the Department of Biostatistics at the Hospices Civils de Lyon. In this Department, he is involved in various research projects on various diseases (observational studies, clinical trials, etc.). However, his main and favorite research topic is the development of statistical methodologies in cancer epidemiology, mostly flexible modeling of incidence/survival data, especially that the Department is regularly issuing French cancer epidemiological statistics (incidence, mortality, survival, and net survival).



# Sunday, 18 July 2021

NETWORKING ROOM

19:15-19:30 **WELCOME TO THE CONFERENCE**

19:30-21:00 **ESCAPE GAME**

# Monday, 19 July 2021

PLENARY HALL

12:00-13:00 **BIENVENUE - ISCB 2021 OPENING CEREMONY**

| | | |
|---|---|---|
| 12:00-12:20 | **Virtual Presentation of the City of Lyon** | |
| 12:20-12:45 | **Welcome Messages** | |
| | **Hélène Jacqmin-Gadda**, SPC Chair<br>**Pascal Roy**, LOC Chair<br>**Frédéric Fleury**, Université Claude Bernard Lyon 1 President | **Zdenek Valenta**, ISCB President<br>**Gilles Rode**, Faculty of Medicine Lyon Est Dean<br>**Gregory Doucet**, Mayor of Lyon |
| 12:45-13:00 | **Awards Presentation** | |
| Student Conference Awards presented by: | **Jonathan Bartlett**, StCA Subcommittee Chair | |
| Poster Awards presented by: | **Hélène Jacqmin-Gadda**, SPC Chair | |

PLENARY HALL

13:00-14:00 **PRESIDENT'S INVITED SPEAKER**

**CHAIR: Zdenek Valenta**, ISCB President, Prague, CZ

The Joy of Pseudo-Values
**PRESIDENT'S INVITED SPEAKER: Per Kragh Andersen**, University of Copenhagen, DK

14:00-14:45 **BREAK – EXCHANGE TIME**

PRESENTERS' VIRTUAL ROOMS

14:00-14:30 **POSTER SESSION 01:** Clinical trial design

| | | |
|---|---|---|
| P01-01 | **Ryuji Uozumi,** Kyoto, JP | An early-phase clinical trial design in oncology with generalization ability |
| P01-02 | **Rainer Muche,** Ulm, DE | A flipped classroom approach for teaching medical statistics and statistical software training |
| P01-03 | **Ben Lanza,** Coventry, UK | A comparison of dual biomarker threshold identification procedures within a confirmatory clinical trial |
| P01-04 | **Vittorio Simeon,** Naples, IT | Multifactor intervention efficacy on MACE and mortality in diabetic kidney disease: a cluster-randomized controlled trial |
| P01-05 | **Ioana R. Marian,** Oxford, UK | Challenges in Factorial Design Randomized Control Trials |
| P01-06 | **Eisuke Hida,** Osaka, JP | Non-inferiority trials with indirect evidence of assay sensitivity using network meta-analysis |
| P01-07 | **Akiyoshi Nakakura,** Kyoto, JP | Biomarker-based Bayesian randomized clinical trial for population finding with efficient reduction of sample size |

# Monday, 19 July 2021

PRESENTERS' VIRTUAL ROOMS

## 14:00-14:30 POSTER SESSION 01: Clinical trial design

| | | |
|---|---|---|
| P01-08 | **Zhulin Yin,** London, UK | A methodological review of phase I designs with late-onset toxicities and incomplete follow-up |
| P01-09 | **M. Iftakhar Alam,** Dhaka, BD | Patient-specific dose finding in seamless phase I/II clinical trials |
| P01-10 | **Audrey Mauguen,** New York, US | The true power of clinical trials in pediatric cancers and other rare diseases |
| P01-11 | **Myanca Rodrigues,** Hamilton ON, CA | Outcomes reported in randomized clinical trials of depression in geriatric patients: a methodological review |
| P01-12 | **Akimitsu Miyake,** Osaka, JP | One small clinical trial design to provide additional evidence of treatment effects than single-arm trials |
| P01-13 | **Philipp Mildenberger,** Mainz, DE | SteppedPower, an R Package for Power Calculation in Stepped Wedge Cluster Randomised Designs |
| P01-14 | **Annette Kopp-Schneider,** Heidelberg, DE | Investigating the operating characteristics of clinical trials with borrowing from external data |
| P01-15 | **Lena Jiricka,** Vienna, AT | Introducing GINGER – A General simulation-INterpolation tool for designing multiGroup ExpeRiments |
| P01-16 | **Marius Sieverding,** Berlin, DE | Substitution of study control group by historic controls: Effect on study results using the example pain therapy for endometriosis |
| P01-17 | **Andrej Schwabe,** Hennigsdorf, DE | Design optimization and intermediate safety reporting for a randomized controlled biomarker trial |
| P01-18 | **Soheila Aghlmandi,** Basel, CH | Using routinely collected data to conduct a pragmatic randomized controlled trial: an example addressing antibiotic prescription and resistance monitoring in Swiss primary care |
| P01-19 | **Bart K.M. Jacobs,** Antwerp, BE | Statistical considerations in using a novel consensus building technique to estimate action thresholds in clinical decision making |
| P01-20 | **Stephan Wojciekowski,** Biberach Riss, DE | Flexible software framework to compare Bayesian hierarchical models across basket trial designs |
| P01-21 | **Nicole Close,** Kitty Hawk NC, US | Assessment of Clinical Trial Missing Data During a Pandemic: A Tipping Point Analysis Case Study |

PRESENTERS' VIRTUAL ROOMS

## 14:00-14:30 POSTER SESSION 02: Longitudinal data analysis

| | | |
|---|---|---|
| P02-01 | **Rheanna Mainzer,** Parkville VIC, AU | A comparison of multiple imputation strategies for handling missing data in repeatedly measured multi-item scales |
| P02-02 | **Alphanie Midelet,** Grenoble, FR | A Hidden Markov model segmentation to identify trajectories in sleep apnoea patients |
| P02-03 | **Courtney McDermott,** Dublin, IE | A comparison of statistical methods to compensate for missing data in longitudinal cluster-randomised trials |
| P02-04 | **Lisanne A. Horvat-Gitsels,** London, UK | Effect of impaired vision on physical activity from childhood to adolescence |
| P02-05 | **Mina Yakoub,** Alexandria, EG | Predictors of Multidrug Resistance in Nosocomial Pneumonia among Intensive Care Units' Patients of a Tertiary Hospital, Egypt |
| P02-06 | **David Stevens,** Liverpool, UK | Modelling of longitudinal data to predict cardiovascular disease risk: a methodological review |
| P02-07 | **Mélanie Guhl,** Paris, FR | Impact of model misspecification on model-based bioequivalence |
| P02-08 | **Kazue Yamaoka,** Tokyo, JP | Cluster randomised controlled trial of lifestyle intervention for adolescents' health using 'SPRAT' programme |



# Monday, 19 July 2021

PRESENTERS' VIRTUAL ROOMS

## 14:00-14:30 POSTER SESSION 02: Longitudinal data analysis

| | | |
|---|---|---|
| P02-09 | **Matteo Petrosino,** Milan, IT | Assessing the role of hyperventilation in patients with traumatic brain injury: longitudinal data analysis from the CENTER-TBI |
| P02-10 | **Francesca Graziano,** Monza, IT | A new approach to measure frailty in the context of COVID-19 population |
| P02-11 | **Carla Henriques,** Viseu, PT | Pain Management in Immediate Life Support Ambulances |
| P02-12 | **Kostas Tryposkiadis,** Birmingham, UK | Statistical methods for estimating sources of variability in count-based biomarkers |
| P02-13 | **Leena Elhussein,** Oxford, UK | Longitudinal progression of frailty in older population and risk of adverse events: An application of joint models |
| P02-14 | **Simon Baldwin,** Birmingham, UK | Semi-variogram approach to estimate within-subject variability in repeated measurements |
| P02-15 | **Mireya Diaz,** Kalamazoo MI, USA | Non-linear dynamic indices summarize densely sampled longitudinal data |
| P02-16 | **Stanislav Katina,** Brno, CZ | Kernel density estimation for circular data about COVID-19 in the Czech Republic |
| P02-17 | **Stanislav Katina,** Brno, CZ | Functional analysis of temporal data about patient's health condition after total knee replacement |
| P02-18 | **Anikó Lovik,** Solna, SE | Multivariate Analysis of Blood Biomarkers in Amyotrophic Lateral Sclerosis |
| P02-19 | **Milada Cvancarova Smastuen,** Oslo, NO | Possibilities and challenges when analysing large longitudinal data from a population based Norwegian registry (MoBa study) |
| P02-20 | **Thomas Debray,** Utrecht, NL | Estimating time to confirmed disease progression in observational data sources with irregular visit schedules |

## 14:30-14:35 TIME FOR ART 1

HALL 1

## 14:45-16:15 INVITED SESSION 1
### Modelling the global spread of Covid19 and impact of interventions

**CHAIR: Roch Giorgi,** Marseille, FR

| | | |
|---|---|---|
| 14:45-15:15 | IS1-1 | **Joseph T. Wu,** Hong Kong, HK-CN<br>Nowcasting the spread of COVID-19 to inform control policies in Hong Kong |
| 15:15-15:45 | IS1-2 | **Daniela De Angelis,** Cambridge, UK<br>Statistical challenges in the monitoring of the SARS-CoV-2 pandemic |
| 15:45-16:15 | IS1-3 | **Pierre-Yves Boëlle,** Paris, FR<br>Modelling the COVID-19 pandemic: initial introduction, lockdown assessment, phasing out |

HALL 2

## 14:45-16:15 SESSION OC1A: Mendelian randomisation & causal inference

**CHAIR: Vanessa Didelez,** Bremen, DE

| | | |
|---|---|---|
| 14:45-15:03 | OC1A-1 | **Claudia Coscia,** Madrid, ES<br>How to deal with collider bias in Mendelian randomization analysis? |

# Monday, 19 July 2021

## HALL 2

### 14:45-16:15 SESSION OC1A: Mendelian randomisation & causal inference

| | | |
|---|---|---|
| 15:03-15:21 | OC1A-2 | **Daniele Bottigliengo,** Bolzano, IT<br>The impact of instruments' selection on Mendelian randomization results: a case study |
| 15:21-15:39 | OC1A-3 | **Sharon Lutz,** Boston MA, US<br>Caution When Inferring the Effect Direction in Mendelian Randomization |
| 15:39-15:57 | OC1A-4 | **Jungyeon Choi,** Leiden, NL<br>Tying research question and analytical strategy when variables are affected by medication use |
| 15:57-16:15 | OC1A-5 | **Bas Penning de Vries,** Leiden, NL<br>Identification of causal effects in case-control studies |

## HALL 3

### 14:45-16:15 SESSION OC1B: Bayesian clinical trial design (Part 1)

**CHAIR: Stephen Senn**, Edinburgh, UK

| | | |
|---|---|---|
| 14:45-15:03 | OC1B-1 | **Moritz Pohl,** Heidelberg, DE<br>Modular components in basket trials and connections among the applied tools |
| 15:03-15:21 | OC1B-2 | **Benjamin Duputel,** Paris, FR<br>Seamless Master Protocol with account for correlations within subgroups |
| 15:21-15:39 | OC1B-3 | **Lukas Baumann,** Heidelberg, DE<br>Monotonicity Rules for Inference in Basket Trials |
| 15:39-15:57 | OC1B-4 | **David Robertson,** Cambridge, UK<br>Response-adaptive randomization in clinical trials: from myths to practical considerations |
| 15:57-16:15 | OC1B-5 | **Luke Ouma,** Newcastle Upon Tyne, UK<br>Treatment allocation strategies for umbrella trials in the presence of multiple biomarkers: A comparison of methods |

## HALL 4

### 14:45-16:15 SESSION OC1C: Network meta analysis

**CHAIR: Tim Friede**, Göttingen, DE

| | | |
|---|---|---|
| 14:45-15:03 | OC1C-1 | **Becky Turner,** London, UK<br>Loop-splitting in network meta-analysis: a new approach to evaluating loop inconsistency |
| 15:03-15:21 | OC1C-2 | **Konstantina Chalkou,** Bern, CH<br>Decision curve analysis for treatment benefit in a network meta-analysis framework |
| 15:21-15:39 | OC1C-3 | **Virginia Chiocchia,** Bern, CH<br>The Risk Of Bias due to Missing Evidence in Network meta-analysis (ROB-MEN) tool: web application and implementation in a network of antidepressant drugs |
| 15:39-15:57 | OC1C-4 | **Annabel Davies,** Manchester, UK - *StCA WINNER*<br>Network meta-analysis and random walks |
| 15:57-16:15 | OC1C-5 | **Janharpreet Singh,** Leicester, UK<br>Bayesian network meta-analysis methods for combining IPD and aggregate data from single-arm studies and RCTs |



# Monday, 19 July 2021

## HALL 5

### 14:45-16:15 SESSION OC1D: Omics and genetic studies

**CHAIR: Martin Treppner**, Freiburg, DE

| | | |
|---|---|---|
| 14:45-15:03 | OC1D-1 | **Hélène Ruffieux,** Cambridge, UK<br>ATLASQTL and EPISPOT: two joint hierarchical approaches for detecting and interpreting hotspots in molecular QTL studies |
| 15:03-15:21 | OC1D-2 | **Nicholas Schreck,** Heidelberg, DE<br>Explained Variation in the Linear Mixed Model |
| 15:21-15:39 | OC1D-3 | **Lars van der Burg**, Leiden, NL<br>Reconstructing KIR haplotypes taking ambiguous and missing data into account |
| 15:39-15:57 | OC1D-4 | **Andreas Ziegler,** Davos, CH; Hamburg, DE<br>COMET: An R package to identify sample cross-contamination in whole genome sequencing studies |
| 15:57-16:15 | OC1D-5 | **Yue Zhai,** Lyon, FR<br>Evaluating DNA sequencing performance: concordance-discordance model and latent class model |

## HALL 6

### 14:45-16:15 SESSION OC1E: Multiple testing and randomization tests

**CHAIR: Robin Ristl**, Vienna, AT

| | | |
|---|---|---|
| 14:45-15:03 | OC1E-1 | **Jinyu Zhu,** Lancaster, UK<br>Confidence intervals for the treatment effect in the Magnusson-Turnbull adaptive enrichment design |
| 15:03-15:21 | OC1E-2 | **Frank Bretz,** Basel, CH<br>Graphical approaches for the control of generalized error rates |
| 15:21-15:39 | OC1E-3 | **Stephanie Wied,** Aachen, DE<br>Evaluation of the Fill-it-up design to combine data from observational trials and RCTs |
| 15:39-15:57 | OC1E-4 | **Julia Niewczas,** Gothenburg, SE<br>Improved group sequential Holm procedures for testing multiple correlated hypotheses over time |
| 15:57-16:15 | OC1E-5 | **Diane Uschner,** Washington DC, US<br>Randomization tests to address disruptions in clinical trials |

## HALL 7

### 14:45-16:15 SESSION OC1F: Methods for survival analysis

**CHAIR: Michal Abrahamowicz**, Montreal QC, CA

| | | |
|---|---|---|
| 14:45-15:03 | OC1F-1 | **Jacobo de Uña-Álvarez**, Vigo, SP<br>Testing for ignorable sampling bias under random double truncation |
| 15:03-15:21 | OC1F-2 | **Moumita Chatterjee**, Kolkata, IN - *CFCD WINNER*<br>A martingale based approach for modelling the alternating recurrences in Cystic Fibrosis patients |
| 15:21-15:39 | OC1F-3 | **Steve Ferreira Guerra**, Montreal QC, CA<br>Impact of measurement error in time-varying prescription-based drug exposures in time-to-event analysis |
| 15:39-15:57 | OC1F-4 | **Marina Roshini Sooriyarachchi**, Colombo, LK<br>Developing a Goodness of fit test for a joint model: The case of clustered survival and count data |
| 15:57-16:15 | OC1F-5 | **Ali Shariati,** Sydney, AU<br>Confidence bands of the MRL function in some right censored prevalent cohort studies via the empirical likelihood |

# Monday, 19 July 2021

**HALL 8**

## 14:45-16:15 SESSION OC1G: Mixed effects model

**CHAIR: Geert Molenberghs**, Hasselt, BE

| | | |
|---|---|---|
| 14:45-15:03 | OC1G-1 | **Charlotte Castel,** Saint-Maurice, FR<br>Bayesian multi-response nonlinear mixed-effect model: application of two recent HIV infection biomarkers |
| 15:03-15:21 | OC1G-2 | **Bernard Francq,** Rixensart, BE<br>Confidence, Prediction, Tolerance Intervals in Linear Mixed Models: Applications in (Non)-Clinical Trials |
| 15:21-15:39 | OC1G-3 | **Aglina Lika,** Rotterdam, NL<br>Bayesian multivariate longitudinal data analysis assuming different association structures |
| 15:39-15:57 | OC1G-4 | **Hannah Morgan,** Sydney, AU<br>The use of mixed logistic modelling in the analysis of HIV latency study data in the context of low sample size and low outcome rates |
| 15:57-16:15 | OC1G-5 | **Noor Jahan Akter,** Dhaka, BD - *CFCD WINNER*<br>Mixed Modeling of Regional Infant Mortality Data over Twenty Years in North Rhine-Westphalia |

## 16:15-17:00 BREAK – EXCHANGE TIME

**VIRTUAL ROOMS**

## 16:15-16:45 INTERACTION WITH SPEAKERS
INVITED SESSION 1

## 16:15-16:45 INTERACTION WITH PRESENTERS
SESSIONS OC1A-OC1B-OC1C-OC1D-OC1E-OC1F-OC1G

**PRESENTERS' VIRTUAL ROOMS**

## 16:15-16:45 POSTER SESSION 03: Causal inference

| | | |
|---|---|---|
| P03-01 | **Nazneen Shariff,** Edinburgh, UK | Statistical methods to analyse ordinal categorical data arising from the clinical trial of drugs from the pharmaceutical industry |
| P03-02 | **Daisy Shepherd,** Melbourne, AU | Causal inference with skewed outcome data: Moving beyond the "ignore or transform" approach |
| P03-03 | **Awa Diop,** Quebec QC, CA | Marginal Structural Models with Latent Class Growth Modeling of Time-varying Treatment |
| P03-04 | **Nadia Dardenne,** Liege, BE | Multidimensional mediators: Comparison between statistical methods using Data simulation |
| P03-05 | **Priska Heinz,** Zurich, CH | The role of the matching algorithm in an analysis of the effect of hemoadsorption in patients with sepsis |
| P03-06 | **Adrienne O'Donnel** Boston MA, US | Estimating the causal effect of direct-acting antiviral agents on kidney function in a clinical cohort of chronic Hepatitis C patients |
| P03-07 | **Jiaxin Zhang,** Melbourne, AU | Should multiple imputation be stratified by exposure when estimating causal effects via outcome regression? |
| P03-08 | **Daniela Schlüter** Liverpool, UK | Pathways to inequalities in child mortality: a population level study in Wales |
| P03-09 | **Daniel Tompsett,** London, UK | Target Trial Emulation and Missing Eligibility Data: A study of Palivizumab for child respiratory illness |



# Monday, 19 July 2021

**PRESENTERS' VIRTUAL ROOMS**

## 16:15-16:45 POSTER SESSION 03: Causal inference

| | | |
|---|---|---|
| P03-10 | **Chris Kennedy,** Boston MA, US | Targeted causal quantile estimation for measurement of postdischarge opioid use in a text-to-web survey |
| P03-11 | **Khalil El Asmar,** Paris, FR | Does early weight gain mediate the causal effects of severity and duration of depression on the onset of metabolic syndrome? |
| P03-12 | **Maurice O'Connell,** Galway, IE | Pathway specific population attributable fractions |

## 16:45-16:50 TIME FOR ART 2

**HALL 1**

## 17:00-18:30 INVITED SESSION 2: Personalized Medicine with Dynamic Predictions

**CHAIR: Dimitris Rizopoulos**, Rotterdam, NL

| | | |
|---|---|---|
| 17:00-17:30 | IS2-1 | **Anirudh Tomer,** Rotterdam, NL<br>Personalized Schedules for Invasive Diagnostic Tests: With Applications in Surveillance of Chronic Diseases |
| 17:30-18:00 | IS2-2 | **Liang Li,** Houston TX, US<br>Comparing risk prediction models in longitudinal context |
| 18:00-18:30 | IS2-3 | **Layla Parast,** Santa Monica CA, US<br>Testing for Heterogeneity in the Utility of a Surrogate Marker |

**HALL 2**

## 17:00-18:30 SESSION OC2A: Mediation analysis

**CHAIR: Mark van der Laan**, Berkeley CA, US

| | | |
|---|---|---|
| 17:00-17:18 | OC2A-1 | **Wen Wei Loh,** Ghent, BE<br>Nonlinear mediation analysis with high-dimensional mediators whose causal structure is unknown |
| 17:18-17:36 | OC2A-2 | **Vanessa Didelez,** Bremen, DE<br>Separable Causal Effects as alternative Estimands in Epidemiology and Biostatistics |
| 17:36-17:54 | OC2A-3 | **Margarita Moreno-Betancur,** Melbourne, AU<br>Simulating hypothetical interventions on multiple mediators: Extending methods and practical guidance |
| 17:54-18:12 | OC2A-4 | **Kamaryn Tanner,** London, UK<br>Mediation with a survival outcome and time-varying mediator: empirical comparison and use in observational data |
| 18:12-18:30 | OC2A-5 | **David Cheng,** Boston MA, US<br>Mediation with Irregular Longitudinal Biomarkers: An Application to Examining Obesity and Severe Disease in COVID-19 Patients |

**HALL 3**

## 17:00-18:30 SESSION OC2B: Bayesian clinical trial design (Part B)

**CHAIR: Sarah Zohar**, Paris, FR

| | | |
|---|---|---|
| 17:00-17:18 | OC2B-1 | **Masahiro Kojima,** Tokyo, JP<br>Early completion of phase I cancer clinical trials with Bayesian optimal interval design |
| 17:18-17:36 | OC2B-2 | **Elias Laurin Meyer,** Vienna, AT<br>Decision rules for identifying combination therapies in open-entry, randomized controlled platform trials |

# Monday, 19 July 2021

## HALL 3

### 17:00-18:30 SESSION OC2B: Bayesian clinical trial design (Part B)

| | | |
|---|---|---|
| 17:36-17:54 | OC2B-3 | **Robert Mahar,** Melbourne, AU<br>A Bayesian decision-theoretic approach to outcome-adaptive sequential multiple assignment randomised trials (SMARTs) with distinct intermediate binary endpoints |
| 17:54-18:12 | OC2B-4 | **Aris Perperoglou,** Cambridge, UK<br>Modelling time varying recruitment rates and site activation prediction in multicentre clinical trials: A comparison study |
| 18:12-18:30 | OC2B-5 | **Silvia Calderazzo,** Heidelberg, DE<br>Compromise Bayesian test decisions under type I error rate constraint |

## HALL 4

### 17:00-18:30 SESSION OC2C: Epidemic modelling of COVID19

**CHAIR: Boris Hejblum**, Bordeaux, FR

**Ornella Wafo Noubissie,** Montreal QC, CA

| | | |
|---|---|---|
| 17:00-17:18 | OC2C-1 | Dynamic predictive modelling of the first wave of the COVID-19 pandemic in Canada using a deterministic density-dependent susceptible-exposed-infected-recovered (SEIR) model that accounts for age-stratified ageing, reporting delays and mortality risks |
| 17:18-17:36 | OC2C-2 | **Shaun Seaman,** Cambridge, UK<br>Nowcasting CoVID-19 Deaths in England by Age and Region |
| 17:36-17:54 | OC2C-3 | **Charl Janse van Rensburg,** Cape Town, ZA<br>An evaluation of the stability, precision and performance of phenomenological models applied to COVID-19 cases and deaths in South Africa |
| 17:54-18:12 | OC2C-4 | **Vera Arntzen,** Leiden, NL<br>Estimation of incubation time and latency time distribution of SARS-CoV-2: the impact of distributional assumptions |
| 18:12-18:30 | OC2C-5 | **Stefan Heyder,** Ilmenau, DE<br>Regional estimates of reproduction numbers with application to COVID-19 |

## HALL 5

### 17:00-18:30 SESSION OC2D: Meta-analysis: network and other

**CHAIR: Ian White**, London, UK

| | | |
|---|---|---|
| 17:00-17:18 | OC2D-1 | **Loukia Spineli,** Hannover, DE<br>Quantifying the robustness of primary analysis results: a case study on missing outcome data in pairwise and network meta-analysis |
| 17:18-17:36 | OC2D-2 | **Katerina Papadimitropoulou,** Leiden, NL<br>Meta-analysis of randomised trials with continuous outcomes: methods that adjust for baseline should be used |
| 17:36-17:54 | OC2D-3 | **Tasnim Hamza,** Bern, CH<br>Flexible generic framework for evidence synthesis in health technology assessment |
| 17:54-18:12 | OC2D-4 | **Ellesha Smith,** Leicester, UK<br>Component Network Meta-Analysis Including Individual Participant Data and Summary Aggregate Data |
| 18:12-18:30 | OC2D-5 | **Ludovic Trinquart,** Boston MA, US<br>Bayesian multivariate network meta-analysis model for the difference in restricted mean survival times |



# Monday, 19 July 2021

## HALL 6

### 17:00-18:30 SESSION OC2E: Some thought on research in biostatistics

**CHAIR: David W. Warne**, Thonex, CH

| | | |
|---|---|---|
| 17:00-17:18 | OC2E-1 | **Georg Heinze,** Vienna, AT<br>Phases of methodological research in biostatistics – A proposal to increase transparency & reproducibility |
| 17:18-17:36 | OC2E-2 | **Karla Hemming,** Birmingham, UK<br>The p-value conundrum: how can a Bayesian analysis help? A case study in reproductive and maternal-fetal medicine |
| 17:36-17:54 | OC2E-3 | **Tim Morris,** London, UK<br>On the marginality principle, with a focus on ratios and interactions |
| 17:54-18:12 | OC2E-4 | **Kym Snell,** Newcastle, UK<br>An extension to reporting guidelines for systematic reviews of prediction model studies (TRIPOD-SRMA) |
| 18:12-18:30 | OC2E-5 | **Jonathan Bartlett,** Bath, UK<br>Reference based multiple imputation for trials – what's the right variance and how to estimate it? |

## HALL 7

### 17:00-18:30 SESSION OC2F: Methods for analysis of electronic health records

**CHAIR: Ellen Hamaker**, Utrecht, NL

| | | |
|---|---|---|
| 17:00-17:18 | OC2F-1 | **Freya Tyrer,** Leicester, UK<br>Immortal time bias for life-long conditions in retrospective observational studies using electronic heath records |
| 17:18-17:36 | OC2F-2 | **Michail Katsoulis,** London, UK<br>Identifying high-risk groups for BMI change using electronic health records from 2.3 million adults |
| 17:36-17:54 | OC2F-3 | **Colin Everett,** Leeds, UK<br>Multiple imputation of sporadically-missing continuous time data by Brownian bridge stochastic interpolation |
| 17:54-18:12 | OC2F-4 | **Mia Tackney,** London, UK<br>Handling missing data from wearable devices in clinical trials |
| 18:12-18:30 | OC2F-5 | **Agus Salim,** Melbourne, AU<br>Using wristwear device to assess impact of COVID19 lockdown on physical activity |

## HALL 8

### 17:00-18:30 SESSION OC2G: Analysis of gene expression and omics data

**CHAIR: Jörg Rahnenführer**, Dortmund, DE

| | | |
|---|---|---|
| 17:00-17:18 | OC2G-1 | **Lara Cavinato,** Milan, IT<br>Prostate cancer intratumor heterogeneity assessment by depth measures analysis on imaging texture features |
| 17:18-17:36 | OC2G-2 | **Zhujie Gu,** Utrecht, NL<br>Investigating Down syndrome by integrating methylation and glycomics using supervised PO2PLS |
| 17:36-17:54 | OC2G-3 | **Julia Duda,** Dortmund, DE<br>Model selection characteristics when using MCP-Mod for dose-response gene expression data |
| 17:54-18:12 | OC2G-4 | **Franziska Kappenberg,** Dortmund, DE<br>Information sharing across genes for improved parameter estimation in dose-response curves |
| 18:12-18:30 | OC2G-5 | **Sara Venkatraman,** Ithaca NY, US<br>A Bayesian approach to estimating dynamic models of co-regulated gene expression |

# Monday, 19 July 2021

**VIRTUAL ROOMS**

**18:30-19:00** — **INTERACTION WITH SPEAKERS**
INVITED SESSION 2

**18:30-19:00** — **INTERACTION WITH PRESENTERS**
SESSIONS OC2A-OC2B-OC2C-OC2D-OC2E-OC2F-OC2G

**PRESENTERS' VIRTUAL ROOMS**

## 18:30-19:00 POSTER SESSION 04: High Dimensional Data

| | | |
|---|---|---|
| P04-01 | **Pradeep Virdee,** Oxford, UK | A methodological approach to assess data quality from the Clinical Practice Research Datalink |
| P04-02 | **Beryl Ang'Iro,** Hasselt, BE | Modeling Child Mortality in the presence of Clustering |
| P04-03 | **Jean-Michel Nguyen** Saint-Martin-d'Hères, FR | Nguyen's Information Criteria (NIC) |
| P04-04 | **Sonia Gran,** Nottingham, UK | Long-term oral prednisolone exposure for bullous pemphigoid: a population-based study using 'big data' and missing data algorithms |
| P04-05 | **Ilaria Gandin,** Trieste, IT | Exploring risk stratification in cardiomyopathies using a deep learning approach for survival prediction |
| P04-06 | **Stanislav Katina,** Brno, CZ | Can animal studies on rodents help better understand Alzheimer's disease in humans? |
| P04-07 | **Elena Colicino,** New York NY, US | Non-linear and non-additive associations between the pregnancy exposome and birthweight |

**PRESENTERS' VIRTUAL ROOMS**

## 18:30-19:00 POSTER SESSION 05: Machine Learning

| | | |
|---|---|---|
| P05-01 | **Mari Carmen Robustillo Carmona** Badajoz, ES | Machine learning to support Reinke's edema diagnosis from voice recordings |
| P05-02 | **Kelly Reeve,** Zurich, CH | The Need for Expanded Standardized Reporting for Machine Learning Methods in Clinical Prediction |
| P05-03 | **Francesca Ieva,** Milan, IT | Virtual biopsy in action: a radiomic-based model for CALI prediction |
| P05-04 | **Roma Puronaitė,** Vilnius, LT | Creation of adverse drug reactions assessment tool |
| P05-05 | **Romane Péan,** Paris, FR | Use of machine learning models combined with innovative interpretation methods to identify prognostic factors |

**NETWORKING ROOM**

## 19:00-20:30 VIRTUAL GET TOGETHER

Choose between

- ✔ ***Live cooking lesson*** *with chef Gregory Cuilleron*
- ✔ ***"Lumière ! l'aventure commence"***
  A fascinating French documentary film directed by Thierry Frémaux (2016) on Auguste and Louis Lumière who invented the cinematograph in Lyon. Accessible to ISCB 2021 delegates only until 30 July 2021.



# Tuesday, 20 July 2021

**HALL 1**

## 13:00-14:30 INVITED SESSION 3
Optimal design of longitudinal cluster randomised trials

**CHAIR: Stacia DeSantis**, Houston TX, US
**DISCUSSANT: Karla Hemming**, Birmingham, UK

| | | |
|---|---|---|
| 13:00-13:23 | IS3-1 | **Jessica Kasza,** Melbourne VIC, AU<br>Faster and more agile designs: speeding up the stepped wedge with batched designs |
| 13:23-13:46 | IS3-2 | **Richard Hooper,** London, UK<br>Longitudinal cluster randomised trials with continuous recruitment |
| 13:46-14:09 | IS3-3 | **Andrew Copas,** London, UK<br>Optimal design of cluster randomized trials with baseline data comparing routine care to a new intervention |
| 14:09-14:30 | | **DISCUSSANT** |

**HALL 2**

## 13:00-14:30 SESSION OC3A: Propensity score in causal studies

**CHAIR: Clemence Leyrat**, London, UK

| | | |
|---|---|---|
| 13:00-13:18 | OC3A-1 | **Corentin Ségalas,** London, UK<br>Multiple imputation in propensity score matching: obtaining correct confidence intervals |
| 13:18-13:36 | OC3A-2 | **Malte Braitmaier,** Bremen, DE<br>Effectiveness of screening colonoscopy in reducing colorectal cancer incidence: emulated target trials from German claims data |
| 13:36-13:54 | OC3A-3 | **Valérie Gares,** Rennes, FR<br>Variance estimators for weighted and stratified linear dose-response function estimators using generalized propensity score |
| 13:54-14:12 | OC3A-4 | **Wen Wei Loh**, Ghent, BE<br>Confounder selection strategies targeting stable treatment effect estimators |
| 14:12-14:30 | OC3A-5 | **Bénédicte Colnet,** Paris, FR<br>Causal inference for combining RCTs and observational studies: methods comparison and medical application |

**HALL 3**

## 13:00-14:30 SESSION OC3B: COVID19 modelling

**CHAIR: Daniela De Angelis**, Cambridge, UK

| | | |
|---|---|---|
| 13:00-13:18 | OC3B-1 | **Vittorio Simeon,** Naples, IT<br>Factors involved in COVID-19 prognosis of patients hospitalized in Campania Region. Findings from COVOCA Study |
| 13:18-13:36 | OC3B-2 | **Oswaldo Gressani,** Hasselt, BE<br>Laplace approximations for fast Bayesian inference of the time-varying reproduction number under misreported epidemic data |
| 13:36-13:54 | OC3B-3 | **Pantelis Samartsidis,** Cambridge, UK<br>Evaluating the effectiveness of local tracing partnerships on NHS test and trace for Covid19 |
| 13:54-14:12 | OC3B-4 | **Marie Alexandre,** Bordeaux, FR<br>Accounting for time-dependant confounding variables in mechanistic ODE model: simulations and application to a vaccine trial |
| 14:12-14:30 | OC3B-5 | **Mui Pham,** Utrecht, NL<br>Interventions to control nosocomial transmission of SARS-CoV-2: a modelling study |

# Tuesday, 20 July 2021

**HALL 4**

## 13:00-14:30 SESSION OC3C: Dynamic prediction

**CHAIR: Brian Tom**, Cambridge, UK

| | | |
|---|---|---|
| 13:00-13:18 | OC3C-1 | **Anthony Devaux,** Bordeaux, FR<br>Individual dynamic prediction of clinical endpoint from large dimensional longitudinal biomarker history: a landmark approach |
| 13:18-13:36 | OC3C-2 | **Clémence Moreau,** Angers, FR<br>Comparison of multiple dynamic predictive accuracies |
| 13:36-13:54 | OC3C-3 | **Francesca Gasperoni,** Cambridge, UK<br>Spatio-temporal score driven modeling of resting state fMRI data |
| 13:54-14:12 | OC3C-4 | **Rickard Strandberg,** Stockholm, SE<br>Breast cancer risk prediction in mammography screening cohorts: an approach based on modeling tumor onset and growth |
| 14:12-14:30 | OC3C-5 | **Sarah Booth,** Leicester, UK<br>Accounting for improvements in survival when developing risk prediction models in a competing risks setting |

**HALL 5**

## 13:00-14:30 SESSION OC3D: Meta-analysis for prediction models

**CHAIR: Ronald Geskus**, Oxford, UK

| | | |
|---|---|---|
| 13:00-13:18 | OC3D-1 | **Laure Wynants,** Maastricht, NL & Leuven, BE<br>COVID-PRECISE: A living methodological review of prediction models for diagnosis and prognosis of covid-19 |
| 13:18-13:36 | OC3D-2 | **Doranne Thomassen,** Leiden, NL<br>A Bayesian model for heterogeneous treatment effects on the additive risk scale in meta-analysis |
| 13:36-13:54 | OC3D-3 | **Brooke Levis,** Newcastle, UK<br>Assessing risk of bias in individual participant data meta-analyses for prediction model research |
| 13:54-14:12 | OC3D-4 | **Lucinda Archer,** Newcastle, UK<br>Using meta-analysis for external validation of prediction models in big data, accounting for competing risks |
| 14:12-14:30 | | Discussion |

**HALL 6**

## 13:00-14:30 SESSION OC3E: Multi-state model

**CHAIR: Linda Sharples**, London, UK

| | | |
|---|---|---|
| 13:00-13:18 | OC3E-1 | **Nikolaos Skourlis,** Solna, SE<br>MSMplus: A dynamic interactive web tool for presentation of multi-state model analysis results |
| 13:18-13:36 | OC3E-2 | **Laura Bondi,** Milan, IT<br>Statistical models for the natural history of breast cancer, with application to data from a Milan cohort study |
| 13:36-13:54 | OC3E-3 | **Nadine Binder,** Freiburg, DE<br>Reevaluating dementia incidence trends: The critical role of adequate design and methodology |
| 13:54-14:12 | OC3E-4 | **Alexandra Niessl,** Ulm, DE<br>Statistical inference for transition probabilities in non-Markov multi-state models subject to both random left-truncation and right-censoring |
| 14:12-14:30 | OC3E-5 | **Jerome Lambert,** Paris, FR<br>Harmonization of endpoints in ICU trials using multi-state modelling |



# Tuesday, 20 July 2021

**HALL 7**

## 13:00-14:30 SESSION OC3F: Neural network and machine learning

**CHAIR: Anne-Laure Boulesteix**, Munich, DE

| | | |
|---|---|---|
| 13:00-13:18 | OC3F-1 | **Georgios Kantidakis,** Brussels, BE<br>Neural networks for survival prediction in medicine: a review and critical appraisal |
| 13:18-13:36 | OC3F-2 | **Alexia Sampri,** Manchester, UK<br>Comparison of imputation methods that solve granularity problem resulting from healthcare structured data integration |
| 13:36-13:54 | OC3F-3 | **Elvire Roblin,** Paris, FR<br>Survival Predictions and Uncertainty Measures with Censored Data |
| 13:54-14:12 | OC3F-4 | **Johannes Vey,** Heidelberg, DE<br>Parametric and non-parametric variable selection methods for predictive modeling with binary response |
| 14:12-14:30 | OC3F-5 | **Ian White,** London, UK<br>Performance measures for assessing machine learning algorithms in clinical trials |

**HALL 8**

## 13:00-14:30 SESSION OC3G: Study designs

**CHAIR: Katie Harron**, London, UK

| | | |
|---|---|---|
| 13:00-13:18 | OC3G-1 | **Stella Erdmann,** Heidelberg, DE<br>Using Historical Data to Predict Health Outcomes – The Prediction Design |
| 13:18-13:36 | OC3G-2 | **Maximilian Linde,** Groningen, NL<br>Bayes Factors for Equivalence, Non-inferiority, and Superiority Designs Using baymedr |
| 13:36-13:54 | OC3G-3 | **Juliette Ortholand,** Paris, FR<br>Forecast Alzheimer's disease progression to better select patients for clinical trials |
| 13:54-14:12 | OC3G-4 | **Caroline Kristunas,** Birmingham, UK<br>Design effects and analysis considerations for the split-mouth design with an unequal numbers of sites per patient |
| 14:12-14:30 | OC3G-5 | **Roma Puronaitė,** Vilnius, LT<br>Challenges of using big health data to identify patterns of anxiety and depression in multimorbid population |

**VIRTUAL ROOMS**

## 14:30-15:00 INTERACTION WITH SPEAKERS
INVITED SESSION 3

## 14:30-15:00 INTERACTION WITH PRESENTERS
SESSIONS OC3A-OC3B-OC3C-OC3D-OC3E-OC3F-OC3G

# Tuesday, 20 July 2021

PRESENTERS' VIRTUAL ROOMS

## 14:30-15:00 POSTER SESSION 06: Epidemiology

| | | |
|---|---|---|
| P06-01 | **Yue Zhai,** Lyon, FR | The impact of left truncation of exposure in environmental case-control studies: evidence from breast cancer risk associated with airborne dioxin |
| P06-02 | **Tanja Bülow,** Aachen, DE | Lower Limit of Quantification in various distributed data: examining confidence interval variations |
| P06-03 | **Asuka Nemoto,** Tokyo, JP | Diversity indices and statistical methods used in studies addressing dysbiosis applied to composition data of the gut microbiota |
| P06-04 | **Lucie Pehalova,** Prague, CZ | Subsequent primary neoplasms in bladder cancer patients |
| P06-05 | **Jan Klaschka,** Prague, CZ | On heuristic detection of maternal-age-related increase of birth defect risk: Experience, issues, alternatives |
| P06-06 | **Helen Blake,** Liverpool, UK | Linkage of national clinical datasets without patient identifiers using probabilistic methods |
| P06-07 | **Bibi F. Ngaleu**, Lyon, FR | Chronic exposure to multiple air pollutant and risk of breast cancer: A nested case - control within the E3N cohort |
| P06-08 | **Adrien Darbier,** Paris, FR | Use of innovative methods to estimate a reliable French pathological complete response rate on real world data |
| P06-09 | **Linda Nab,** Leiden, NL | Sensitivity analyses for measurement error using regression calibration or simulation-extrapolation |

## 15:05-15:10 TIME FOR ART 3

HALL 1

## 15:15-16:45 INVITED SESSION 4
### Challenges and opportunities for learning from long term disease registers

**CHAIR: Els Goetghebeur**, Ghent, BE
**DISCUSSANT: Arvid Sjölander**, Solna, SE

| | | |
|---|---|---|
| 15:15-15:38 | IS4-1 | **Katie Harron,** London, UK<br>Data linkage for creating electronic birth cohorts: handling bias due to linkage error |
| 15:38-16:01 | IS4-2 | **Ingeborg Waernbaum,** Uppsala, SE<br>Challenges and opportunities for learning from long term disease registers: Causal inference |
| 16:01-16:24 | IS4-3 | **Elizabeth Stuart,** Baltimore MD, US<br>Methods for combining experimental and population data to estimate population average treatment effects |
| 16:24-16:45 | | **DISCUSSANT** |



# Tuesday, 20 July 2021

HALL 2

## 15:15-16:45 INVITED SESSION 5
### The best of both worlds: combining deep learning and modeling

**CHAIR: Harald Binder**, Freiburg, DE

| | | |
|---|---|---|
| 15:15-15:45 | IS5-1 | **Christopher Rackauckas,** Cambridge MA, US<br>Pharmacometrics-Informed Deep Learning with DeepNLME |
| 15:45-16:15 | IS5-2 | **Austin R. Benson,** Ithaca NY, US<br>Temporal and relational machine learning for biostatistical and other scientific applications |
| 16:15-16:45 | IS5-3 | **Göran Köber,** Freiburg, DE<br>Individualizing deep dynamic models for psychological resilience data |

HALL 3

## 15:15-16:45 SESSION OC4A: Missing data in causal studies

**CHAIR: Saskia Le Cessie**, Leiden, NL

| | | |
|---|---|---|
| 15:15-15:33 | OC4A-1 | **Audinga-Dea Hazewinkel,** Bristol, UK<br>Sensitivity to MNAR dropout in clinical trials: use and interpretation of the Trimmed Means Estimator |
| 15:33-15:51 | OC4A-2 | **François Bettega,** Grenoble, FR<br>Comparison of two causal inference methods for multiple treatments in clinical research on observational data |
| 15:51-16:09 | OC4A-3 | **S. Ghazaleh Dashti,** Melbourne, AU<br>Handling missing data for causal effect estimation in longitudinal cohort studies using Targeted Maximum Likelihood Estimation: a simulation study |
| 16:09-16:27 | OC4A-4 | **Camila Olarte Parra**, Bath, UK<br>Estimands in clinical trials: making the hypothetical strategy concrete |
| 16:27-16:45 | OC4A-5 | **Emily Roberts,** Ann Arbor MI, US<br>Incorporating baseline covariates to validate surrogate endpoints with a constant biomarker under control arm |

HALL 4

## 15:15-16:45 SESSION OC4B: Cluster randomized trials

**CHAIR: Andrew Copas**, London, UK

| | | |
|---|---|---|
| 15:15-15:33 | OC4B-1 | **Kendra Davis-Plourde,** New Haven CT, US<br>Sample size calculation for stepped wedge cluster randomized trials with multiple levels of clustering |
| 15:33-15:51 | OC4B-2 | **Christina Easter,** Birmingham, UK<br>Stepped-wedge cluster randomised trials with binary outcomes and small numbers of clusters: a case study |
| 15:51-16:09 | OC4B-3 | **Rhys Bowden,** Melbourne, AU<br>Inference for the treatment effect in longitudinal cluster randomized trials when treatment effect heterogeneity is ignored |
| 16:09-16:27 | OC4B-4 | **Jennifer Thompson,** London, UK<br>Cluster randomised trials and a small number of clusters: Analysis method for a binary outcome |
| 16:27-16:45 | OC4B-5 | **Laura Marsden,** Leeds, UK<br>Under what conditions do open-cohort cluster RCTs provide improvements over conventional designs? A simulation study |

# Tuesday, 20 July 2021

## HALL 5

### 15:15-16:45 SESSION OC4C: Meta-analysis

**CHAIR: Tomasz Burzykowski**, Hasselt, BE

| | | |
|---|---|---|
| 15:15-15:33 | OC4C-1 | **Jinran Zhan,** Coventry, UK<br>Comparison of frequentist and Bayesian methods for two-arm borrowing of historical data |
| 15:33-15:51 | OC4C-2 | **Theodosia Salika,** London, UK<br>Implications of Analysing Time-to-Event Outcomes as Binary in Meta-analysis |
| 15:51-16:09 | OC4C-3 | **Miriam Hattle,** Newcastle, UK<br>Exploring non-linear treatment-covariate interactions at multiple time points using multivariate IPD meta-analysis |
| 16:09-16:27 | OC4C-4 | **Peter Godolphin,** London, UK<br>A comprehensive framework for 'deft' (within-trials) interactions in meta-analysis |
| 16:27-16:45 | OC4C-5 | **Anastasios Papanikos,** Leicester, UK<br>Inclusion of real world data in surrogate endpoint evaluation: a Bayesian meta-analytic approach |

## HALL 6

### 15:15-16:45 SESSION OC4D: Prediction model for omics data

**CHAIR: Michael Kammer,** Vienna, AT

| | | |
|---|---|---|
| 15:15-15:33 | OC4D-1 | **Solon Karapanagiotis,** Cambridge, UK - *STCA WINNER*<br>Tailored Bayesian variable selection for risk prediction modelling under unequal misclassification costs |
| 15:33-15:51 | OC4D-2 | **Zhi Zhao,** Oslo, NO<br>Feature selection in multivariate varying-coefficient mixed models for drug response prediction |
| 15:51-16:09 | OC4D-3 | **Darren Scott,** London, UK<br>Variational Bayes for Model Averaging for Multivariate models using Compositional Microbiome predictors |
| 16:09-16:27 | OC4D-4 | **Mirrelijn van Nee,** Amsterdam, NL<br>Fast marginal likelihood estimation of penalties for group-adaptive elastic net |
| 16:27-16:45 | OC4D-5 | **Jeroen Goedhart,** Amsterdam, NL<br>Improving model performance estimation in high-dimensional data settings by using learning curves |

## HALL 7

### 15:15-16:45 SESSION OC4E: Competing risks and multi-state models

**CHAIR: Nadine Binder**, Freiburg, DE

| | | |
|---|---|---|
| 15:15-15:33 | OC4E-1 | **Serge M.A. Somda**, Bobo-Dioulasso, BF<br>Bayesian inference for the direct approach for competing risk modeling with Gompertz distribution |
| 15:33-15:51 | OC4E-2 | **Hongji Wu,** Guangzhou, CN - *CFCD AWARD*<br>Analysis of competing risks data using restricted mean time lost |
| 15:51-16:09 | OC4E-3 | **Hao Yuan,** Guangzhou, CN - *CFCD AWARD*<br>Impact of competing event in COVID-19 clinical data analysis |
| 16:09-16:27 | OC4E-4 | **Gustavo Soutinho,** Porto, PT<br>Parametric Landmark estimation of the transition probabilities in survival data with multiple events |
| 16:27-16:45 | OC4E-5 | **Anais Andrillon,** Paris, FR<br>Phase I/II dose-finding design for right censored toxicity endpoints with competing disease progression |



# Tuesday, 20 July 2021

## VIRTUAL ROOMS

### 16:45-17:15 INTERACTION WITH SPEAKERS
INVITED SESSIONS 4-5

### 16:45-17:15 INTERACTION WITH PRESENTERS
SESSIONS OC4A-OC4B-OC4C-OC4D-OC4E

## PRESENTERS' VIRTUAL ROOMS

### 16:45-17:15 POSTER SESSION 07: Survival analysis

| | | |
|---|---|---|
| P07-01 | **Shoichi Irie,** Tokyo, JP | Modeling Non-Proportional Hazards for Overall Survival Time for Cancer Treatments |
| P07-02 | **Isabelle Smith,** Leeds, UK | Sensitivity of results to missing data for clinical trials with discrete, longitudinal outcome measurements |
| P07-03 | **Masako Nishikawa,** Tokyo, JP | New Application of competing risks model in IgA nephropathy to explore the severity-dependent urinary remission |
| P07-04 | **Robabeh Ghodssi-Ghassemabadi,** Tehran, IR | Frailty Multi-state Model with Time-dependent Covariate for Prediction of Colorectal Cancer Progression |
| P07-05 | **Carly Milliren,** Boston MA, US | Time to readmission among newborns: time for a reappraisal? |
| P07-06 | **Caterina Gregorio,** Trieste, IT | Studying the longitudinal trajectory of potassium in heart failure patients through dynamic survival models |
| P07-07 | **Ilse Cuevas Andrade**, Leicester, UK | Use of electronic health records to enhance data from a single clinical trial evaluating maintenance therapy in non-small cell lung cancer patients |
| P07-08 | **Michael Lauseker,** Munich, DE | Modelling the length of stay of COVID-19 patients using a multistate approach |
| P07-09 | **Isobel Barrott,** Cambridge, UK | Landmarking: An R package for analysis using landmark models |
| P07-10 | **Stefania Chounta,** Paris, FR | The risk of valvular heart disease after childhood cancer: contribution of dose-volume histogram parameters |
| P07-11 | **James Schmidt,** Leicester, UK | The estimation of adjustment factors for expected mortality rates with application in comorbidity-adjusted lifetables |
| P07-12 | **Ana Matos,** Viseu, PT | Multiple Cox regression analysis to investigate a biomarker in IgA nephropathy disease: different approaches |
| P07-13 | **Stanislav Katina,** Brno, CZ | Analysis of Results of Total Knee Replacement Failure Using Cox Proportional Hazard Model with Time-Dependent Covariates |
| P07-14 | **Yuqin Cao,** Lyon, FR | Joint contribution of positive and total lymph nodes number in predicting overall survival of esophageal cancer |
| P07-15 | **Mamadou Saliou Kalifa Diallo,** Montpellier, FR | Modelling the duration of recurrent events from interval-censored data |
| P07-16 | **Rubaiya Rubaiya,** Dhaka, BD *CFCD AWARD* | Relation between women empowerment and birth interval: A survival analysis approach |
| P07-17 | **Hugues de Courson,** Bordeaux, FR | Impact of model choice when studying the relationship between blood pressure variability and risk of stroke |
| P07-18 | **Lubomír Štěpánek,** Prague, CZ | A jack-knifed version of the log-rank test in small samples: when bias meets variance to increase test power |

### 17:15-17:20 TIME FOR ART 4

# Tuesday, 20 July 2021

## HALL 1

### 17:30-19:00 INVITED SESSION 6: Causal inference in continuous time for dense longitudinal data from wearable devices

**CHAIR: Linda Valeri,** New York NY, US

| | | |
|---|---|---|
| 17:30-18:00 | IS6-1 | **Mark van der Laan**, Berkeley CA, US<br>TMLE for Causal Effects based on continuous time longitudinal data structures |
| 18:00-18:30 | IS6-2 | **Walter Dempsey,** Ann Arbor MI, US<br>Assessing Time-Varying Causal Effect Moderation in the Presence of Cluster-Level Treatment Effect Heterogeneity |
| 18:30-19:00 | IS6-3 | **Xiaoxuan Cai,** New York NY, US<br>Missing data imputation for non-stationary time series in mHealth data |

## HALL 2

### 17:30-19:00 SESSION OC5A: Bayesian clinical trial analysis

**CHAIR: Moreno Ursino**, Paris, FR

| | | |
|---|---|---|
| 17:30-17:48 | OC5A-1 | **Emma Gerard,** Paris, FR<br>Bayesian joint modeling of a bivariate toxicity for dose-regimens in early phase oncology |
| 17:48-18:06 | OC5A-2 | **Ronan Fougeray,** Suresnes, FR<br>Interim analysis of a clinical trial using the predictive probability of success based on a surrogate endpoint |
| 18:06-18:24 | OC5A-3 | **Stephen Senn,** Edinburgh, UK<br>Trials of Vaccine Efficacy for COVID-19: Inferential and Practical Issues |
| 18:24-18:42 | OC5A-4 | **Hongchao Qi,** Rotterdam, NL<br>Incorporating multiple parameters from historical controls using the meta-analytic-predictive (MAP) prior |
| 18:42-19:00 | OC5A-5 | **Alma Revers,** Amsterdam, NL<br>Bayesian hierarchical modeling for MedDRA coded adverse events in RCTs |

## HALL 3

### 17:30-19:00 SESSION OC5B: Prediction by Machine learning

**CHAIR: Göran Köber**, Freiburg, DE

| | | |
|---|---|---|
| 17:30-17:48 | OC5B-1 | **Valia Baralou,** Athens, GR<br>Individual risk prediction: comparing Random Forests with Cox proportional-hazards model by a simulation study |
| 17:48-18:06 | OC5B-2 | **Thomas Cowling,** London, UK<br>Predicting patient mortality from large sets of diagnosis codes using logistic regression and machine learning |
| 18:06-18:24 | OC5B-3 | **Dimitra-Kleio Kipourou,** London, UK<br>Predicting individual life years lost due to cancer using pseudo-observations with random forest, in the absence of cause of death information |
| 18:24-18:42 | OC5B-4 | **Francisco Ojeda,** Hamburg, DE<br>A comprehensive comparison of approaches for the calibration of probability machines |
| 18:42-19:00 | OC5B-5 | **Paula Dhiman,** Oxford, UK<br>Methodological conduct of clinical prediction models using machine learning methods in oncology needs to be improved |



# Tuesday, 20 July 2021

## HALL 4

### 17:30-19:00 SESSION OC5C: Screening and Diagnostic studies

**CHAIR: Muriel Rabilloud**, Lyon, FR

| | | |
|---|---|---|
| 17:30-17:48 | OC5C-1 | **Per-Henrik Zahl,** Oslo, NO<br>The Natural History of Invasive Breast Cancers Detected in the Scandinavian Mammography Screening Programs: A Cohort Study |
| 17:48-18:06 | OC5C-2 | **Enzo Cerullo,** Leicester, UK<br>Meta-analysis of dichotomous and polytomous diagnostic tests without a gold standard |
| 18:06-18:24 | OC5C-3 | **Rocío Aznar-Gimeno,** Zaragoza, ES<br>Comparison of methods for the linear combination of biomarkers under Youden Index optimisation criterion |
| 18:24-18:42 | OC5C-4 | **Rose Sisk,** Manchester, UK<br>Single and multiple imputation combined with missing indicators in clinical prediction models: a simulation study |
| 18:42-19:00 | OC5C-5 | **Alfred Kipyegon Keter**, Antwerp, BE<br>Diagnosing Latent Class Analysis for Analyzing Diagnostic Tests in the Absence of a Gold Standard |

## HALL 5

### 17:30-19:00 SESSION OC5D: Bayesian Joint models for longitudinal data and time-to-event

**CHAIR: Cécile Proust-Lima**, Bordeaux, FR

| | | |
|---|---|---|
| 17:30-17:48 | OC5D-1 | **Marion Kerioui,** Paris, FR<br>Bayesian multilevel nonlinear joint model to characterize the variability in the response to immunotherapy |
| 17:48-18:06 | OC5D-2 | **Christos Thomadakis,** Athens, GR<br>Multistate inference based on longitudinal/competing risks joint modeling under misclassified cause of failure |
| 18:06-18:24 | OC5D-3 | **Zixuan Yao,** Kyoto, JP<br>Bayesian Predictive Model Averaging for Joint Model of Survival and Longitudinal Data: Application to an Immunotherapy Trial |
| 18:24-18:42 | OC5D-4 | **Pedro Miranda Afonso,** Rotterdam, NL<br>A Joint Model for Multiple Longitudinal Outcomes, Recurrent and Terminal Events using CF Patient Registry Data |
| 18:42-19:00 | OC5D-5 | **Yuriko Takeda,** Yokohama, JP<br>Joint Modeling of Incomplete Longitudinal Data and Time-to-Event Data |

## HALL 6

### 17:30-19:00 SESSION OC5E: Cure and mixture models

**CHAIR: Philip Hougaard**, Valby, DK

| | | |
|---|---|---|
| 17:30-17:48 | OC5E-1 | **Subodh Selukar,** Seattle WA, US - *StCA WINNER*<br>RECeUS: Ratio Estimation of Censored Uncured Subjects for Studying Sufficient Follow-Up in Studies of Long-Term Survivors |
| 17:48-18:06 | OC5E-2 | **Maria Quelhas,** Leiden, NL<br>Multiple imputation for survival analysis with missing data and a cure fraction: a study of osteosarcoma |
| 18:06-18:24 | OC5E-3 | **Laura Botta,** Milan, IT<br>A simulation analysis of reliability and robustness of a cancer cure model accounting for extra non-cancer mortality |
| 18:24-18:42 | OC5E-4 | **Thanh-Huan Vo,** Rennes, FR<br>An extension of Fellegi-Sunter record linkage model for mixed-type data with application to SNDS |
| 18:42-19:00 | | DISCUSSION |

# Tuesday, 20 July 2021

**HALL 7**

## 17:30-19:00 SESSION OC5F: Longitudinal data analysis

**CHAIR: Michael Crowther**, Solna, SE

| | | |
|---|---|---|
| 17:30-17:48 | OC5F-1 | **Rushani Wijesuriya,** Melbourne, AU<br>Multiple imputation approaches for handling incomplete three-level data with time varying cluster memberships |
| 17:48-18:06 | OC5F-2 | **Angelika Geroldinger,** Vienna, AT<br>Separation in Marginal Logistic Regression Models |
| 18:06-18:24 | OC5F-3 | **Robin Ristl,** Vienna, AT<br>A geometric Brownian motion model with non-normal random effect for the prediction of the growth of abdominal aortic aneurysms |
| 18:24-18:42 | OC5F-4 | **Mariam R. Rizkallah**, Bremen, DE<br>Predicting Patient Risk for Adverse Drug Events in Health Care Claims Data using Functional Targets |
| 18:42-19:00 | | DISCUSSION |

**HALL 8**

## 17:30-19:00 SESSION OC5G: Missing data & measurement error

**CHAIR: Havi Murad**, Tel Hashomer, IL

| | | |
|---|---|---|
| 17:30-17:48 | OC5G-1 | **Melissa Middleton,** Melbourne, AU<br>Multiple imputation for missing data in case-cohort studies: simulation and case study |
| 17:48-18:06 | OC5G-2 | **Nidhiali Menon,** Canberra, AU<br>Application of three level multiple imputation to national surveys |
| 18:06-18:24 | OC5G-3 | **Lilith Faucheux,** Paris, FR<br>Profiles of COVID 19-Hematological patients: Franco-Brazilian observational cohort study |
| 18:24-18:42 | OC5G-4 | **Edouard Chatignoux,** Saint-Maurice, FR<br>Prediction of cancer incidence in areas without registries using proxy and registry data |
| 18:42-19:00 | OC5G-5 | **Edouard Chatignoux,** Saint-Maurice, FR<br>What is the real prevalence of hypertension in France? |

**VIRTUAL ROOMS**

## 19:00-19:30 INTERACTION WITH SPEAKERS

INVITED SESSION 6

## 19:00-19:30 INTERACTION WITH PRESENTERS

SESSIONS OC5A-OC5B-OC5C-OC5D-OC5E-OC5F-OC5G



# Tuesday, 20 July 2021

**PRESENTERS' VIRTUAL ROOMS**

## 19:00-19:30 POSTER SESSION 08: Epidemic modeling

| | | |
|---|---|---|
| P08-01 | **Jiayu He,** Grinnell IA, US | Exploring the Sensitivity of Extended SIR Models Through Randomized Simulations and Multiple Factor Analysis |
| P08-02 | **Nikolai Saperkin,** Nizhny Novgorod, RU | Monte Carlo simulation of the COVID-19 spread using an agent-based modelling in Russian regions |
| P08-03 | **Cristina Gena Dascalu,** Iasi, RO | Similarities between the COVID-19 spread in Romanian counties identified through data clustering |
| P08-04 | **Ettore Rocchi** Urbino, IT | A new epidemic model for the Covid-19 pandemic |
| P08-05 | **Ornella Wafo Noubissie,** Montreal QC, CA | Spatial analyses of the first wave of COVID-19 cases in Hong-Kong using Poisson intrinsic and Besag York Mollié conditional autoregressive models under a Bayesian paradigm |
| P08-06 | **Kirsi Manz,** Munich, DE | Indices of inequality to monitor temporal and geographic trends in COVID-19 incidence and death data |
| P08-07 | **Kirsi Manz,** Munich, DE | Bayesian disease mapping of standardized infection fatality rate using the example of COVID-19 in Bavaria |
| P08-08 | **Kazumi Omata,** Tokyo, JP | Mathematical Modelling of COVID-19 Epidemics in Tokyo Metropolitan and New York City |
| P08-09 | **Jacob Doody,** Baltimore MD, US | Application of a spatio-temporal SVEIRD model to COVID-19 epidemic in the Czech Republic |
| P08-10 | **Achilleas Tsoumanis,** Antwerp, BE | Impact of STI screening intensity on antibiotic exposure: A modelling study among men who have sex with men in Belgium |
| P08-11 | **Marco Bonetti,** Milan, IT | Epilocal: A real-time tool for local epidemic monitoring |

**NETWORKING ROOM**

## 19:30-20:00 VIRTUAL GALA

**BREAK-OUT ROOMS**

## 20:00-22:00 GALA CONCERTS

# Wednesday, 21 July 2021

**HALL 1**

## 13:00-14:30 INVITED SESSION 7
Selective inference after variable selection

**CHAIR: Georg Heinze**, Vienna, AT

| | | |
|---|---|---|
| 13:00-13:30 | IS7-1 | **Oliver Dukes,** Ghent, BE<br>Valid post-selection inference for cox regression parameters, with and without the proportional hazards assumption |
| 13:30-14:00 | IS7-2 | **Michael Kammer,** Vienna, AU<br>Selective inference for the Lasso in statistical practice |
| 14:00-14:30 | IS7-3 | **Lisa McShane,** Bethesda MD, US<br>Biologically-informed development of treatment selection scores from high-dimensional omics data |

**HALL 2**

## 13:00-14:30 SESSION OC6A: Causal inference

**CHAIR: Jonathan Bartlett**, Bath, UK

| | | |
|---|---|---|
| 13:00-13:18 | OC6A-1 | **Kim Luijken,** Leiden, NL<br>Incident and prevalent-user designs and the definition of study time origin in pharmacoepidemiology: a systematic review |
| 13:18-13:36 | OC6A-2 | **Joris Hautekiet,** Ghent, BE<br>A framework for meta-analysis of studies with baseline exposure through standardized survival curves |
| 13:36-13:54 | OC6A-3 | **Susanne Strohmaier,** Vienna, AT<br>Causal inference in practice: two case studies in nephrology |
| 13:54-14:12 | OC6A-4 | **Arthur Chatton,** Nantes, FR<br>Machine learning, G-computation and small sample sizes: a simulation study |
| 14:12-14:30 | OC6A-5 | **Niloofar Moosavi,** Umea, SE<br>Valid Uncertainty Interval for the Average Causal Effect in a High-dimensional Setting |

**HALL 3**

## 13:00-14:30 SESSION OC6B: Adaptive clinical trial design

**CHAIR: Babak Choodari-Oskooei**, London, UK

| | | |
|---|---|---|
| 13:00-13:18 | OC6B-1 | **Jan Meis,** Heidelberg, DE<br>Adaptive designs for three-arm gold-standard non-inferiority trials |
| 13:18-13:36 | OC6B-2 | **Fulvio Di Stefano**, Turin, IT<br>A comparison of estimation methods adjusting for selection bias in adaptive enrichment designs with time-to-event endpoints |
| 13:36-13:54 | OC6B-3 | **Martin Law,** Cambridge, UK<br>A stochastically curtailed two-arm randomised phase II trial design for binary outcomes |
| 13:54-14:12 | OC6B-4 | **Alyssa Vanderbeek,** London, UK<br>Fast-tracking clinical trial innovations – A COVID-19 silver lining |
| 14:12-14:30 | OC6B-5 | **Marta Bofill Roig,** Vienna, AT<br>Adaptive clinical trials with selection of composite endpoints and sample size reassessment |



# Wednesday, 21 July 2021

**HALL 4**

## 13:00-14:30 SESSION OC6C: Joint models & 2-stage approach

**CHAIR: Ruth Keogh**, London, UK

| | | |
|---|---|---|
| 13:00-13:18 | OC6C-1 | **Mohadeseh Shojaei Shahrokhabadi**, Pretoria, ZA<br>Marginalized two-part joint modeling of longitudinal semi-continuous responses and survival data: with application to medical expenses |
| 13:18-13:36 | OC6C-2 | **Cécile Proust-Lima,** Bordeaux, FR<br>Healthy life expectancy computation using the Item Response Theory combined with a joint modeling approach |
| 13:36-13:54 | OC6C-3 | **Marta Spreafico,** Milan, IT<br>Modelling the effect of longitudinal latent toxicity profiles on survival: an application to osteosarcoma |
| 13:54-14:12 | OC6C-4 | **Célia Touraine,** Montpellier, FR<br>Joint model versus linear mixed model to analyze longitudinal data of health-related quality of life in cancer clinical trials |
| 14:12-14:30 | OC6C-5 | **Anaïs Rouanet,** Bordeaux, FR<br>Joint modelling of the temporal relationships between multivariate longitudinal markers of Alzheimer's disease progression and clinical endpoints |

**HALL 5**

## 13:00-14:30 SESSION OC6D: Time-to-event methods for non-proportional hazards

**CHAIR: Mathieu Fauvernier**, Lyon, FR

| | | |
|---|---|---|
| 13:00-13:18 | OC6D-1 | **Rouven Behnisch,** Heidelberg, DE<br>Methods for analyzing time-to-event endpoints in immuno-oncological trials with delayed treatment effects |
| 13:18-13:36 | OC6D-2 | **Dominic Magirr,** Riehen, CH<br>Non-proportional hazards in immuno-oncology: Is an old perspective needed? |
| 13:36-13:54 | OC6D-3 | **Bharati Kumar,** Bath, UK<br>Weighted hazard ratio for time to event endpoints under non proportional hazards |
| 13:54-14:12 | OC6D-4 | **Liliane Manitchoko,** Paris, FR<br>Comparison of statistical methods for estimating time-varying treatment effect on adverse event risk |
| 14:12-14:30 | OC6D-5 | **Suzanne Freeman,** Leicester, UK<br>Evidence Synthesis of Time-To-Event Outcomes in the Presence of Non-Proportional Hazards |

**HALL 6**

## 13:00-14:30 SESSION OC6E: Relative survival and net benefit estimation

**CHAIR: Virginie Rondeau**, Bordeaux, FR

| | | |
|---|---|---|
| 13:00-13:18 | OC6E-1 | **Christopher Büsch,** Heidelberg, DE<br>Additional benefit method assessment for time-to-event endpoints – A comparison of ESMOs and IQWiGs approach |
| 13:18-13:36 | OC6E-2 | **Rachael Stannard,** Leicester, UK<br>Obtaining long-term stage-specific relative survival estimates in the presence of incomplete historical stage information |
| 13:36-13:54 | OC6E-3 | **Robert Darlin Mba**, Marseille, FR<br>A latent class model for the estimation of the excess mortality hazard for correcting inaccurate background mortality |
| 13:54-14:12 | OC6E-4 | **Roland Matsouaka,** Durham NC, US<br>Robust statistical inference for the matched net benefit and win ratio |
| 14:12-14:30 | OC6E-5 | **Alessia Eletti,** London, UK<br>A unifying framework for flexible excess hazard modeling with applications in cancer epidemiology |

# Wednesday, 21 July 2021

HALL 7

## 13:00-14:30 SESSION OC6F: Deep learning

**CHAIR: Christopher Rackauckas**, Boston MA, US

| | | |
|---|---|---|
| 13:00-13:18 | OC6F-1 | **Halehsadat Nekoee Zahraei**, Liege, BE<br>Cluster analysis on emergency COVID-19 data: A result-based multiple imputation for missing data |
| 13:18-13:36 | OC6F-2 | **Martin Treppner,** Freiburg, DE<br>Statistical Power for Single Cell Representations |
| 13:36-13:54 | OC6F-3 | **Alexandre Bailly,** Lyon, FR<br>Effects of Interactions and Dataset Size in Neural Networks and Other Machine Learning Approaches |
| 13:54-14:12 | OC6F-4 | **Corentin Blanc,** Lyon, FR<br>CamemBERT word-embedding for Information Extraction in a Biomedical Context |
| 14:12-14:30 | OC6F-5 | **Lucas Kook,** Zurich, CH<br>Interpretable effect estimates of semi-structured predictors in deep distributional regression models |

## 14:30-15:15 BREAK – EXCHANGE TIME

VIRTUAL ROOMS

## 14:30-15:00 INTERACTION WITH SPEAKERS
INVITED SESSION 7

## 14:30-15:00 INTERACTION WITH PRESENTERS
SESSIONS OC6A-OC6B-OC6C-OC6D-OC6E-OC6F



# Wednesday, 21 July 2021

PRESENTERS' VIRTUAL ROOMS

## 14:30-15:00 POSTER SESSION 09: Individual prediction and precision medicine

| | | |
|---|---|---|
| P09-01 | **Samer Kharroubi,** Beirut, LB | A Bayesian nonparametric approach for modeling SF-6D health state utility scores |
| P09-02 | **Math Candel,** Maastricht, NL | Small-sample accuracy of approximations of individual polynomial growth curves' prediction error in linear mixed regression |
| P09-03 | **Harald Heinzl,** Vienna, AT | Assessment of performance measures for external validation of multivariable prediction models: Simulations |
| P09-04 | **Ettore Rocchi,** Urbino, IT | Mathematical proof of the equivalence between Post-test Predictive Probabilities and Predictive Values |
| P09-05 | **Laura Savaré,** Milan, IT | Development and validation of a clinical risk score to predict the risk of SARS-CoV-2 infection from administrative data |
| P09-06 | **Mariella Gregorich,** Vienna, AT | Subject-specific networks as features for predictive modelling - A scoping review of methods |
| P09-07 | **Alexandra Lavalley-Morelle,** Paris, FR | Individual dynamic predictions in joint analysis of non-linear longitudinal model and parametric competing risk model: application to sepsis patients |
| P09-08 | **Yoann Blangero,** Vienna, AT | Estimating the treatment selection ability of a marker: review, comparison, and improvement of existing approaches |
| P09-09 | **Laura Antolini,** Milan, IT | Double cut-point identification of continuous diagnostic test variables for clinical use |
| P09-10 | **Diane Duroux,** Liege, BE | Patient heterogeneity assessments via network-based ANOVA |
| P09-11 | **Ryan Chung,** Cambridge, UK | A comparison of prioritisation tools for CVD risk assessment in UK Biobank: Primary care records versus polygenic risk scores |
| P09-12 | **Vaiva Deltuvaite-Thomas,** Louvain-la-Neuve, BE | Operational characteristics of generalized pairwise comparisons for hierarchically ordered endpoints |
| P09-13 | **Behnaz Alafchi,** Hamadan, IR | Landmark Prediction of Survival for HIV-infected Patients by Considering AIDS as an Intermediate Event |

## 15:00-15:05 TIME FOR ART 5

HALL 1

## 15:15-16:45 INVITED SESSION 8
Variance modelling for multilevel data and joint models

**CHAIRS: Donald Hedeker**, Chicago, IL & **Karen Leffondré**, Bordeaux, FR

| | | |
|---|---|---|
| 15:15-15:45 | IS8-1 | **Michael Crowther,** Stockholm, SE<br>A general framework and implementation for variance modelling in joint model settings |
| 15:45-16:15 | IS8-2 | **Ellen Hamaker,** Utrecht, NL<br>Dynamic structural equation modeling for intensive longitudinal data |
| 16:15-16:45 | IS8-3 | **Jessica Barrett,** Cambridge, UK<br>Jointly modelling longitudinal heteroscedasticity and a time-to-event outcome |

# Wednesday, 21 July 2021

## HALL 2

### 15:15-16:45 SESSION OC7A: Causal inference in survival analysis

**CHAIR: Margarita Moreno-Betancur**, Melbourne, AU

| | | |
|---|---|---|
| 15:15-15:33 | OC7A-1 | **Yoshinori Takeuchi,** Tokyo, JP<br>Marginal structural Cox models in nested case-control studies with time-varying treatments |
| 15:33-15:51 | OC7A-2 | **Ilaria Prosepe,** Leiden, NL<br>Benefit-based organ allocation in liver transplantation |
| 15:51-16:09 | OC7A-3 | **Nan van Geloven**, Leiden, NL<br>Extending multistate models with g-computation to evaluate the effect of treatment timing |
| 16:09-16:27 | OC7A-4 | **Quentin Le Coënt**, Bordeaux, FR<br>Causal assessment of surrogacy for time-to-event endpoints using meta-analytic data |
| 16:27-16:45 | OC7A-5 | **Ruth Keogh,** London, UK<br>Estimating treatment effects on survival in an entirely treated cohort: negative controls in longitudinal data |

## HALL 3

### 15:15-16:45 SESSION OC7B: Clinical trial design and sample size calculation

**CHAIR: Richard Hooper**, London, UK

| | | |
|---|---|---|
| 15:15-15:33 | OC7B-1 | **Samuel Kilian,** Heidelberg, DE<br>Analysis and sample size calculation for a survival model conditional on a binary surrogate endpoint |
| 15:33-15:51 | OC7B-2 | **Katy E. Morgan**, London, UK<br>Dangers of wrongly assuming linearity in a trial sample size calculation when treatment affects rate of change |
| 15:51-16:09 | OC7B-3 | **Derek Dinart,** Bordeaux, FR<br>Estimating sample size for biomarker-strategy designs with survival endpoints |
| 16:09-16:27 | OC7B-4 | **Cornelia Ursula Kunz**, Biberach, DE<br>Multidimensional Go/No-Go decision rules |
| 16:27-16:45 | OC7B-5 | **Sonja Zehetmayer,** Vienna, AT<br>Online control of the False Discovery Rate in platform trials |

## HALL 4

### 15:15-16:45 SESSION OC7C: Frailty model and recurrent events

**CHAIR: Catherine Legrand**, Louvain, BE

| | | |
|---|---|---|
| 15:15-15:33 | OC7C-1 | **Juste Goungounga,** Dijon, FR<br>A non-mixture cure model with frailty correction for inaccurate background mortality to estimate time-to-cure |
| 15:33-15:51 | OC7C-2 | **Maximilian Bardo,** Göttingen, DE<br>A family of discrete random effect distributions for modelling bivariate time-to-event data |
| 15:51-16:09 | OC7C-3 | **Hadrien Charvat,** Tokyo, JP<br>A general approach to fit flexible hazard regression models with multiple random effects |
| 16:09-16:27 | OC7C-4 | **Florence Gillaizeau,** Saint-Herblain, FR<br>Handling recurrent events in the context of clinical trials |
| 16:27-16:45 | OC7C-5 | **Gaëlle Munsch,** Bordeaux, FR<br>Retro-prospective modelling of recurrent events |



# Wednesday, 21 July 2021

## HALL 5

### 15:15-16:45 SESSION OC7D: Functional data analysis

**CHAIR: Fabien Subtil**, Lyon, FR

| | | |
|---|---|---|
| 15:15-15:33 | OC7D-1 | **Amna Klich,** Lyon, FR<br>Trajectory clustering using mixed classification models |
| 15:33-15:51 | OC7D-2 | **Beatrice Charamba,** Galway, IE<br>Bayesian concurrent functional regression for sensor data |
| 15:51-16:09 | OC7D-3 | **Michela Carlotta Massi,** Milan, IT<br>ERNEST: A Representation Learning-based Cross-Subject EEG Channel Selection Algorithm |
| 16:09-16:27 | OC7D-4 | **Vojtech Sindlar,** Brno, CZ<br>Linear statistical models and ridge regression used in shape index calculation on human face |
| 16:27-16:45 | OC7D-5 | **Annamaria Porreca,** Chieti, IT<br>Unsupervised classification of ECG signals via FDA to look for different patterns of variation among patients |

## HALL 6

### 15:15-16:45 SESSION OC7E: Selection and validation of prediction models

**CHAIR: Delphine Maucort-Boulch**, Lyon, FR

| | | |
|---|---|---|
| 15:15-15:33 | OC7E-1 | **Hana Sinkovec,** Vienna, AT<br>To tune or not to tune, a case study of ridge logistic regression in small or sparse datasets |
| 15:33-15:51 | OC7E-2 | **Valentijn de Jong**, Utrecht, NL<br>Propensity-based standardization to enhance the interpretation of prediction model discrimination |
| 15:51-16:09 | OC7E-3 | **Paola M.V. Rancoita**, Milan, IT<br>A novel score for prognostic index assessment with event-free survival outcome |
| 16:09-16:27 | OC7E-4 | **Richard Riley,** Newcastle, UK<br>Minimum sample size for external validation of a clinical prediction model with a binary or time-to-event outcome |
| 16:27-16:45 | OC7E-5 | **Edwin Kipruto,** Freiburg, DE<br>The nonnegative garrote as a flexible approach for model selection in low and high dimensional data |

## HALL 7

### 15:15-16:45 SESSION OC7F: Flexible modelling and spatial data analysis

**CHAIR: Christine Wallisch**, Vienna, AT

| | | |
|---|---|---|
| 15:15-15:33 | OC7F-1 | **Yin Bun Cheung**, Singapore, SG<br>Conditional and unconditional logistic regression analysis of matched case-control studies of recurrent events |
| 15:33-15:51 | OC7F-2 | **Willi Sauerbrei,** Freiburg, DE<br>On the implications of influential points for the selection and reproducibility of MFP models |
| 15:51-16:09 | OC7F-3 | **Maxime Bonjour,** Lyon, FR<br>Worldwide age specific Human Papillomavirus prevalence patterns: a mixed effects binomial method to cluster trajectories |
| 16:09-16:27 | OC7F-4 | **Chuen Seng Tan**, Singapore, SG<br>Treating ordinal outcomes as continuous quantities: when, why and how |
| 16:27-16:45 | OC7F-5 | **Karen Lamb,** Melbourne, AU<br>Health map for HealthGap: Defining a geographical catchment to examine cardiovascular risk in Australia |

# Wednesday, 21 July 2021

HALL 8

## 15:15-16:45 SESSION OC7G: Methods for clinical research

**CHAIR: Stephen Senn**, Edinburgh, UK

| | | |
|---|---|---|
| 15:15-15:33 | OC7G-1 | **Sandrine Boulet,** Paris, FR<br>Bayesian extrapolation from pre-clinical data to human |
| 15:33-15:51 | OC7G-2 | **Anne-Laure Boulesteix,** Munich, DE<br>Analysing ordinal endpoints in two-arm randomized clinical trials |
| 15:51-16:09 | OC7G-3 | **Cattram Nguyen,** Melbourne, AU<br>Simulations to assess the impact of non-adherance due to COVID-19 in a cluster-randomized non-inferiority trial |
| 16:09-16:27 | OC7G-4 | **Kim May Lee**, London, UK<br>Analysis methods for personalized randomized controlled trial (PRACTical) |
| 16:27-16:45 | OC7G-5 | **Sara Prada Alonso**, A Coruña, ES<br>MultiNet: a computational algorithm to analyze high-dimensional epigenetic correlation structures |

## 16:45-17:00 BREAK – EXCHANGE TIME

PLENARY HALL

## 17:00-18:00 KEYNOTE INVITED SPEAKER

**CHAIR: Hélène Jacqmin-Gadda,** SPC Chair

Handling Negative Correlation and/or Over/Underdisperson in Gaussian and Non-Gaussian Hierarchical Data

**KEYNOTE SPEAKER: Geert Molenberghs**, Hasselt University & KU Leuven, Belgium

## 18:00-18:15 BREAK – EXCHANGE TIME

PLENARY HALL

## 18:15-19:30 ISCB ANNUAL GENERAL ASSEMBLY



# Wednesday, 21 July 2021

VIRTUAL ROOMS

## 19:30-20:00 INTERACTION WITH SPEAKERS

INVITED SESSION 8

## 19:30-20:00 INTERACTION WITH PRESENTERS

SESSIONS OC7A-OC7B-OC7C-OC7D-OC7E-OC7F-OC7G

PRESENTERS' VIRTUAL ROOMS

## 19:30-20:00 POSTER SESSION 10: Meta analysis

| | | |
|---|---|---|
| P10-01 | **Luca Bertolaccini,** Milan, IT | Video-Assisted Non-Intubated Lobectomies for Lung Cancer: A Systematic Review and Meta-Analysis |
| P10-02 | **Hannah Ensor,** Edinburgh, UK | Bias in evaluation of discrete surrogate outcomes, due to separation: a penalized likelihood solution |
| P10-03 | **Clareece Nevill,** Leicester, UK | Developing a Novel Interactive Multifaceted Graphic for Treatment Ranking within Network Meta-Analysis |
| P10-04 | **Carissa Reid,** Heidelberg, DE | Propensity Score-Integrated Meta-Analytic-Predictive Priors |
| P10-05 | **Molly Wells,** Leicester, UK | Using threshold analysis to assess the robustness of public health intervention recommendations from network meta-analyses |
| P10-06 | **Elizabeth Korevaar,** Clayton VIC, AU | Meta-analysis methods used in systematic reviews of interrupted time series studies |
| P10-07 | **Lehana Thabane,** Hamilton ON, CA | Effects of probiotics on mortality and morbidity in preterm infants: a Bayesian network meta-analysis of randomized and non-randomized studies |

## 20:00-20:05 TIME FOR ART 6

# Thursday, 22 July 2021

HALL 1

## 13:00-17:00 STRATOS MINI-SYMPOSIUM

**CHAIRS: Willi Sauerbrei**, Freiburg, DE & **Els Goetghebeur**, Ghent, BE

### PART A

| | | |
|---|---|---|
| 13:00-13:05 | | INTRODUCTION |
| 13:05-13:30 | MS-1 | **Paul Gustafson,** Vancouver BC, CA<br>Understanding and accounting for measurement error when prediction equations are used in observational studies |
| 13:30-14:00 | MS-2 | **David J McLernon**, Aberdeen, UK / **Terry Therneau,** Rochester MN, US<br>Guidance for performance assessment in prediction models for survival outcomes |
| 14:00-14:30 | MS-3 | **Saskia le Cessie,** Leiden, NL; Ghent, BE / **Els Goetghebeur,** Ghent, BE<br>Guiding the path from Patient Reported Outcomes to treatment registration based on randomised and single arm studies: STRATOS engaged in the European IMI-SISAQOL project. *(https://qol.eortc.org/projectqol/sisaqol/)* |

## 14:30-15:30 BREAK – EXCHANGE TIME

### PART B

| | | |
|---|---|---|
| 15:30-15:55 | MS-4 | **Jörg Rahnenführer,** Dortmund, DE<br>Statistical analysis of high-dimensional biomedical data: Analytical goals, common approaches and challenges |
| 15:55-16:25 | MS-5 | **Christine Wallisch,** Vienna, AT / **Geraldine Rauch,** Berlin, DE<br>Current education and practical guidance in statistical (non-linear) modeling for researchers with limited statistical background (level-1) |
| 16:25-16:55 | MS-6 | **Michal Abrahamowicz,** Montreal QC, CA / **Victor Kipnis,** Bethesda MD, US<br>In defense of correct use of statistical significance |
| 16:55-17:00 | | CLOSING |

# Thursday, 22 July 2021

HALL 2

## 13:00-17:00 EARLY CAREER BIOSTATISTICIANS' DAY

### PART A

**CHAIR: Alexandre Bailly**, Lyon, FR

| | | |
|---|---|---|
| 13:00-13:45 | ECB-1 *Invited* | **Anne-Laure Boulesteix,** Munich, DE<br>A replication crisis in methodological statistical research? |
| 13:45-14:00 | ECB-2 *OC* | **Max Behrens,** Freiburg, DE<br>From Consulting to Research |
| 14:00-14:15 | ECB-3 *OC* | **Jinran Zhan,** Coventry, UK<br>It's all about collaboration: PhD as an academia industry partnership in an interdisciplinary environment |
| 14:15-14:30 | ECB-4 *OC* | **Constant Josse,** Malakoff, FR<br>Two-piecewise model to assess the impact of a treatment on a quantitative repeated endpoint |

## 14:30-15:30 BREAK – EXCHANGE TIME

### PART B

**CHAIR: Daniella Zöller**, Freiburg, DE

| | | |
|---|---|---|
| 15:30-16:15 | ECB-5 *Invited* | **Mavuto Mukaka,** Bangkok, TH<br>Studying and working in developing nations |
| 16:15-16:30 | ECB-6 *OC* | **Charlotte Bolch,** Glendale AZ, US<br>A biostatistician's journey navigating ethical issues |
| 16:30-16:45 | ECB-7 *OC* | **Zheng Jing Hu**, Hamilton ON, CA<br>Developing a prediction model: Challenges and Lessons Learned |
| 16:45-17:00 | | DISCUSSION |

PLENARY HALL

## 17:00-18:00 CLOSING CEREMONY & INVITATION TO ISCB43 NEWCASTLE

## 18:00-18:30 EXCHANGE TIME

## 18:00-18:05 TIME FOR ART 7

# Keynote Lectures

PRESIDENT'S INVITED SPEAKER

**Per Kragh Andersen**
Section of Biostatistics, University of Copenhagen, Denmark

## PS | The Joy of Pseudo-Values

Survival analysis is characterized by the need to deal with incomplete observation of the outcome variable, most frequently caused by right-censoring, and several – now standard – inference procedures have been developed to deal with this. Examples include the Kaplan-Meier estimator for the survival function and partial likelihood for estimating regression coefficients in the proportional hazards (Cox) model. During the last decades, methods based on pseudo-values have been studied. Here, the idea is to apply a transformation of the incompletely observed survival data and, thereby, to create a more simple data set for which 'standard' techniques (i.e., for complete data) may be applied, e.g., methods using generalized estimating equations.
An advantage of this approach is that it applies quite generally to(marginal)parameters for which no or few other regression methods are directly available (including average time spent in a state of a multi-state model). Another advantage is that it allows the use of a number of graphical techniques, otherwise unavailable in survival analysis. Disadvantages include that the method is not fully efficient and that it, in its simplest form, assumes covariate-independent censoring (though generalizations to deal with this have been developed).
We will review the development in the field since the idea was put forward by Andersen, Klein and Rosthøj in a 2003 Biometrika paper. Focus will be on graphical methods but the theoretical properties of the approach will also be touched upon.

KEYNOTE INVITED SPEAKER

**Geert Molenberghs**[1,2]
Interuniversity Institute for Biostatistics and Statistical Bioinformatics
[1] I-BioStat, Hasselt University, Belgium
[2] I-BioStat, KU Leuven, Belgium

## KS | Handling Negative Correlation and/or Over/Underdisperson in Gaussian and Non-Gaussian Hierarchical Data

The occurrence and interpretation of negative variance components in the context of linear mixed models is well understood at this point, even though the issue is surrounded by subtle issues for estimation and testing (Verbeke and Molenberghs 2003, Molenberghs and Verbeke 2007). Broadly, negative variance components often point to negative within-cluster correlation. It is even possible to give such linear mixed models a meaningful hierarchical interpretation (Molenberghs and Verbeke 2011). Matters are more complicated when the outcomes are non-Gaussian, either in the context of the generalized linear mixed model, or extensions thereof that allow for flexible modeling of both within-unit correlation as well as overdispersion (Molenberghs et al. 2010). An additional complication is that, in practice, not only negative variance components due to negative correlation, but also underdispersion instead of overdispersion can occur, sometimes even jointly. With focus on both continuous and count data, we describe how models can be made sufficiently flexible and, in a number of cases, interpreted hierarchically (Luyts et al. 2019).

**References:** *Luyts, M., Molenberghs, G., Verbeke, G., Matthijs, K., Demétrio, C.G.B., and Hinde, J. (2019). A Weibull-count approach for handling under- and/or over-dispersed clustered data structures. Statistical Modeling, 19, 569-589.*
*Molenberghs, G. and Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter space. The American Statistician, 61, 1-6. Molenberghs, G. and Verbeke, G. (2011). A note on a hierarchical interpretation for negative variance components. Statistical Modeling, 11, 389-408. Molenberghs, G., Verbeke, G., Demétrio, C.G.B., and Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. Statistical Science, 25, 325-347.*
*Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. Biometrics, 59, 254-262.*

# Invited Lectures

INVITED SESSION 1 — Modelling the global spread of Covid19 and impact of interventions

## IS1-1 Nowcasting the spread of COVID-19 to inform control policies in Hong Kong

**Joseph T. Wu**
WHO Collaborating Center for Infectious Disease Epidemiology and Control, School of Public Health, The University of Hong Kong

In this talk, I will discuss two of the COVID-19 nowcasting projects in my group that help informed the Hong Kong government on its pandemic control policies.
In the first project, we developed a new framework that parameterizes disease transmission models with age-specific digital mobility data. By fitting the model to case data in Hong Kong, we were able to accurately track the local effective reproduction number of COVID-19 in near real-time (i.e. no longer constrained by the delay of around 9 days between infection and reporting of cases) which is essential for quick assessment of the effectiveness of interventions on reducing transmissibility. Our findings showed that accurate nowcast and forecast of COVID-19 epidemics can be obtained by integrating valid digital proxies of physical mixing into conventional epidemic models.
In the second project, we assessed the relative transmissibility of two new SARS-CoV-2 lineages with the N501Y mutation in the receptor-binding domain of the spike protein have spread rapidly in the United Kingdom in late-2020. We estimated that the earlier 501Y lineage without amino acid deletion Δ69/Δ70 circulating mainly between early September to mid-November was 10% (6-13%) more transmissible than the 501N lineage, and the 501Y lineage with amino acid deletion Δ69/Δ70 circulating since late September was 75% (70-80%) more transmissible than the 501N lineage.

## IS1-2 Statistical challenges in the monitoring of the SARS-CoV-2 pandemic

**Daniela De Angelis**
Medical Research Council Biostatistics Unit, University of Cambridge, United Kingdom

In England policy decisions during the SARS-COV-2 pandemic have relied on prompt scientific evidence on the state of the pandemic. Through participation in advisory groups to the government we have contributed to the real time pandemic assessment for the last fifteen months, from the early phase, to the current potential third wave, providing estimates of key epidemic quantities and short-to-medium-term projections of severe disease. In this talk I describe the statistical transmission modelling framework we have used throughout and the model developments introduced at various stages of the pandemic to tackle the challenges posed by the changing epidemiology, including: the changing relationship between the available data streams; the introduction of vaccination; and the ecological effects of new variants. In addition, I discuss the computational efforts made to ensure that model developments would not jeopardise the timely provision of relevant outputs to policy makers.

## IS1-3 Modelling the COVID-19 pandemic: initial introduction, lockdown assessment, phasing out

**Pierre-Yves Boëlle**
Sorbonne Université, Paris, France

Since its initial description in 2020, COVID-19 totaled millions of cases worldwide. Modelling has been used at all stages of the pandemic and will be illustrated here with examples taken with the French situation. Initial spread was analyzed by looking at the volume of exportation jointly with the pattern of flights out of the first recognized source in China and shed some light on features of the disease. After initial introduction, the disease spread in many countries, leading to unprecedented decisions for control including lockdown and curfews. Estimating the extent of disease spread after the first wave and to what extent the trace & test policies covered the population presented several challenges. Computing real time trends improved small term projections of disease spread. As vaccines were developed, optimal allocation in a context of uncertainty and limited availability could be studied as a means to phase out of social distancing. More recently though, the emergence of variants of increased transmissibility led to ask questions regarding overall spread and control anew.

# Invited Lectures

## INVITED SESSION 2 — Personalized Medicine with Dynamic Predictions

### IS2-1 Personalized Schedules for Invasive Diagnostic Tests: With Applications in Surveillance of Chronic Diseases

**Anirudh Tomer[1], Daan Nieboer[2,3], Monique J. Roobol[3], Ewout W. Steyerberg[2,4], Dimitris Rizopoulos[1]**
1 Department of Biostatistics, Erasmus University Medical Center, Rotterdam, Netherlands
2 Department of Public Health, Erasmus University Medical Center, Rotterdam, Netherlands
3 Department of Urology, Erasmus University Medical Center, Rotterdam, Netherlands
4 Department of Biomedical Data Sciences, Leiden University Medical Center, Netherlands

Benchmark surveillance tests for diagnosing disease progression (biopsies, endoscopies, etc.) in early-stage chronic non-communicable disease patients (e.g., cancer, lung diseases) are usually invasive. For detecting progression timely, over their lifetime, patients undergo numerous invasive tests planned in a fixed one-size-fits-all manner (e.g., biannually). We present personalized test schedules based on the risk of disease progression, that aim to optimize the number of tests (burden) and time delay in detecting progression (shorter is beneficial) better than fixed schedules. Our motivation comes from the problem of scheduling biopsies in prostate cancer surveillance studies.
Using joint models for time-to-event and longitudinal data, we consolidate patients' longitudinal data (e.g., biomarkers) and results of previous tests, into an individualized future cumulative-risk of progression. To use this predicted risk profile for scheduling invasive tests we propose three different approaches. First, by minimizing loss functions motivated by Bayesian decision theory, under the predicted risk profile, to obtain the optimal time of conducting a future invasive test. Second, planning tests on those future visits where the predicted cumulative-risk of the patient is above a particular threshold (e.g., 5% risk). Third, by estimating the expected number of tests (burden) and expected time delay in detecting progression (shorter is beneficial) for all possible test schedules, and then optimizing a utility function of the expected number of tests and delay to find a mathematically optimal schedule. Since we estimate the expected number of tests and delay in a personalized manner, they can be used directly by patients and doctors to compare various test schedules for their benefit and burden and make a shared decision for a suitable schedule. In all three approaches, a common specialty of our schedules is that they automatically update on each follow-up with newly gathered data. We compare our methodologies with currently used heuristic schedules, and schedules based on partially observable Markov decision processes. We also evaluate criteria for the selection of risk-threshold (e.g., Youden index, F1 score, net benefit) in schedules based on risk-threshold. Last, we implement our methodology in a web-application for prostate cancer patients.

### IS2-2 Comparing risk prediction models in longitudinal context

**Liang Li**
Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston TX, United States

In this presentation, I will review our recent research on comparing risk prediction models in longitudinal cohort studies. The research was conducted under the commonly encountered clinical context where patient characteristics are measured at baseline, biomarkers and clinical history are collected over time as repeated measures until censoring or the occurrence of a terminal clinical event of predictive interest. First, in a comparison between static vs. dynamic prediction models, the latter always showed equivalent or improved prediction accuracy, and the magnitude of improvement depended critically on the concept of baseline and how the at-risk population vary over time. This result suggests that when longitudinal data are available, dynamic prediction is a useful tool that can replace the widely used static prediction approaches. Second, in a comparison between two dynamic prediction model approaches, landmark modeling vs. joint modeling, neither approach dominated the other in terms of prediction accuracy, and their performance under the simulation context depended on model misspecification. This comparison was made feasible by using a novel algorithm to simulate longitudinal data that satisfied infinitely many landmark models simultaneously. The result suggests that further research on both types of models is warranted. Third, in a comparison among various joint models for dynamic prediction, it was found that misspecification of longitudinal trajectories could decrease the prediction accuracy, which highlights the importance of future research on joint models with multivariate longitudinal data and flexible trajectories. A computational approach to facilitate the estimation of this kind of joint models will be presented.



# Invited Lectures

## INVITED SESSION 2 — Personalized Medicine with Dynamic Predictions

### IS2-3 Testing for Heterogeneity in the Utility of a Surrogate Marker

**Layla Parast[1], Lu Tian[2], Tianxi Cai[3]**
1 Statistics Group, RAND, Santa Monica CA, United States
2 Biomedical Data Science, Stanford University, Stanford CA, United States
3 Biostatistics, Harvard University, Cambridge MA, United States

For many clinical outcomes, randomized clinical trials to evaluate the effectiveness of a treatment or intervention require long-term follow-up of participants. In such settings, there is substantial interest in identifying and using a surrogate marker that can measured earlier and be used to evaluate the treatment effect. Several statistical methods have been proposed to evaluate potential surrogate markers; however, these methods generally do not account for or address the potential for a surrogate to vary in utility by certain patient characteristics. Such heterogeneity is important to understand, particularly if the surrogate is to be used in a future trial to potentially predict or replace the primary outcome. Here, we develop a nonparametric approach to examine and test for heterogeneity in the strength of a surrogate marker of interest. Specifically, we propose an approach and estimation procedure to measure the surrogate strength as a function of a baseline covariate, W. We then propose a testing procedure to test for evidence of heterogeneity and, if there is heterogeneity, an approach to identify a region of interest i.e., a subgroup of W where the surrogate is useful. Lastly, we extend these methods to a setting where there are multiple baseline covariates. We examine the performance of these methods using a simulation study and illustrate the methods using data from the Diabetes Prevention Program clinical trial.

## INVITED SESSION 3 — Optimal design of longitudinal cluster randomised trials

### IS3-1 Faster and more agile designs: speeding up the stepped wedge with batched designs

**Jessica Kasza, Rhys Bowden, Andrew Forbes**
School of Public Health and Preventive Medicine, Monash University, Clayton VIC, Australia

Stepped wedge designs are an increasingly popular variant of longitudinal cluster randomised trial designs. Stepped wedge designs roll interventions out across clusters in a randomised, but step-wise fashion, and gain power over standard cluster randomised trials through within-cluster comparisons. However, the standard stepped wedge design is typically neither fast nor agile. All clusters must start and end trial participation at the same time, implying that ethics approvals and data collection procedures must be in place in all clusters before a stepped wedge trial can start in any cluster. Hence, although stepped wedge designs are useful for testing the impacts of many cluster-based interventions on outcomes, this requirement means that there can be lengthy delays before a trial can commence.
In this talk we will discuss the "batched" stepped wedge design. Batched variants of stepped wedge designs allow for clusters to come on-line to the study in batches, instead of all at once, and thus can be deployed more quickly. However, like the stepped wedge, the batched step wedge rolls the intervention out to all clusters in a randomised and step-wise fashion. Provided that the effect of time is appropriately included in the regression model for the outcome, sample size calculations are straightforward and the power of the study will be robust to delays with the start-up of batches. Researchers can also modify sample size calculations to accommodate adaptations such as early stopping for futility or success, or for sample size re-calculation.

# Invited Lectures

## INVITED SESSION 3 — Optimal design of longitudinal cluster randomised trials

### IS3-2 Longitudinal cluster randomised trials with continuous recruitment

**Richard Hooper**
Institute of Population Health Sciences, Queen Mary University of London, United Kingdom

When a stepped wedge or other longitudinal cluster randomised trial recruits/identifies a consecutive sample of participants from a continuous stream presenting at clusters over a given calendar period, it is quite a different prospect to sampling in a series of discrete, cross-sectional slices. For one thing, introducing an intervention mid-stream to a cluster could contaminate participants recently recruited under the routine care condition. For another, it is inadequate to speak of distinct time "periods": two individuals recruited at either end of the "same" period may have less in common than two individuals recruited just on either side of a "division" between periods. A continuous timescale also offers a continuously adaptable framework for designing a longitudinal trial: timing when to intervene, and when to start or stop recruitment.

This talk focuses particularly on maximising statistical efficiency in two very different design problems. In the simple case of a trial randomising clusters to two groups, intervention and routine care, with an initial, prospective baseline period during which all clusters receive routine care, I show how close-to-optimal efficiency is generally obtained either with no baseline period at all, or with a baseline period that divides the available time in half. This finding is robust to the form of the underlying, fixed effect of time, assuming this is correctly specified in the analysis model (I hope to have simulation results looking at how well different approaches to analysis fare under mis-specification.)

At the other end of the spectrum of design complexity is the case of a longitudinal cluster randomised trial where we choose when each cluster crosses from routine care to the intervention along a continuous timescale, and try to achieve the required statistical power by recruiting the smallest number of participants out of the total presenting at all clusters over the calendar period – an incomplete stepped wedge design. Search algorithms identify surprising solutions – in some instances resembling a series of before-and-after studies rather than concurrent comparisons of intervention and control – though for a design robust to the form of the underlying time effect a smooth "staircase" design may be preferable.

### IS3-3 Optimal design of cluster randomized trials with baseline data comparing routine care to a new intervention

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

**Background:** In cluster randomised trials (CRTs) it is sometimes possible to choose a different cluster size (number of individuals measured per cluster) between trial arms, or between baseline and endline e.g. in the SNEHA-TARA trial where clusters are large communities and only a sample of individuals are surveyed. In most trials clusters can be allocated unequally to arms if desired. An optimal design minimises the total number of measurements required for a given number of trial clusters. For CRTs with cross sectional data and a continuous outcome, it is known how to (i) optimally allocate measurements between baseline and endline when the cluster autocorrelation (CAC) is the same across trial arms [1], and (ii) optimally allocate clusters and measurements when the variance or intra-cluster correlation coefficient (ICC) are affected by the intervention [2].

**Objective:** To extend previous work to trials comparing routine care to a new intervention, assuming a similar ICC and variance for both trial arms at baseline and in the routine care arm at endline, that the intervention is likely to reduce the CAC, and may affect the ICC and variance.

**Results:** We present algebraic results, and graphical methods, to help identify optimal desgns for this setting. The reduction in number of measurements required compared to the standard design, where clusters are allocated equally to arms and the cluster size is equal over time and between trial arms, can be substantial where cluster sizes or ICC values are large. If the intervention reduces the CAC, but does not affect the variance or ICC, then the optimal design will typically involve (i) smaller cluster size in the intervention arm compared to rountine care at both baseline and endline, and (ii) more clusters allocated to the intervention arm.

**Conclusions:** Optimal designs can save resources but designs must be chosen to maintain power across plausible ranges for the correlation and variance parameters which will often be wide. We recommend trialists report these parameters separately by arm to inform the design of future trials.

**References:** *[1] Copas AJ, Hooper R Cluster randomised trials with different numbers of measurements at baseline and endline: sample size and optimal allocation. Clin Trials 2020 17:69-76; [2] Copas AJ, Hooper R Optimal design of cluster randomised trials allowing unequal allocation of clusters and measurements Stat Med Under revision*



# Invited Lectures

## INVITED SESSION 4 — Challenges and opportunities for learning from long term disease registers

### IS4-1 Data linkage for creating electronic birth cohorts: handling bias due to linkage error

**Katie Harron**
UCL Great Ormond Street Institute of Child Health, London, United Kingdom

Data linkage is a valuable tool for creating datasets with which to understand long term trajectories of health and disease. Linkage can provide a low-cost, efficient means of collecting extensive and detailed data on interactions with health and other services. These data can be used to create population-level electronic cohorts that offer the ability to answer questions that require large sample sizes or detailed data on hard to reach populations, and to generate evidence with a high level of external validity and applicability for policy-making.

Lack of access to unique or accurate identifiers means that linkage of the same individual across different data sources or time can be challenging. Errors occurring during linkage (false-matches and missed-matches) disproportionally affect particular subgroups of individuals and can lead to substantial bias in results based on linked data.

This talk will first describe methods for creating electronic birth cohorts using data linkage. We will then explore the impact of linkage error, drawing on examples from the literature. We will demonstrate a range of methods for evaluating linkage quality and discuss how bias due to linkage error can be handled within analyses.

### IS4-2 Challenges and opportunities for learning from long term disease registers: Causal inference

**Ingeborg Waernbaum**
Department of Statistics, Uppsala University, Sweden

Disease incidence registers are longitudinal data bases giving opportunities to study the development of chronic diseases in response to various treatments. For time-to-event emerging long term disease registers have at least two important advantages. They let many events build up over time, thus generating a rich information basis. They uniquely allow to study long term effects of treatment choices. When populations and available treatments change over calendar time, however, unavoidable administrative censoring patterns bring new estimation challenges as well as new questions on transportability of results.

Here, we describe two Swedish examples, the Swedish Renal Registry, recording incident cases of End Stage Renal Disease since 1991 and the Swedish Childhood Diabetes Register, a population-based incidence register, recording incident cases of childhood onset diabetes mellitus (T1DM) since 1977. We describe challenges and opportunities when controlling for baseline confounding for a point-treatment (immediate kidney transplant vs. start with dialysis) when the causal estimand is the difference in average potential survival curves. For the Swedish Childhood Diabetes Register, we describe selection bias and bounds when applying multiple inclusion criteria using extensions of previous results of Smith and Vanderweele (2019) and Sjölander (2020).

# Invited Lectures

## INVITED SESSION 4 Challenges and opportunities for learning from long term disease registers

### IS4-3 Methods for combining experimental and population data to estimate population average treatment effects

**Elizabeth A. Stuart**
Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore MD, United States

With increasing attention being paid to the relevance of studies for real-world practice (such as in education, international development, and comparative effectiveness research), there is also growing interest in external validity and assessing whether the results seen in randomized trials would hold in target populations. While randomized trials yield unbiased estimates of the effects of interventions in the sample of individuals (or physician practices or hospitals) in the trial, they do not necessarily inform about what the effects would be in some other, potentially somewhat different, population. While there has been increasing discussion of this limitation of traditional trials, relatively little statistical work has been done developing methods to assess or enhance the external validity of randomized trial results. This talk will discuss design and analysis methods for combining experimental and population data to assess and increase external validity, as well as general issues that need to be considered when thinking about external validity. Implications for how future studies should be designed in order to enhance the ability to estimate population effects will also be discussed.

## INVITED SESSION 5 The best of both worlds: combining deep learning and modeling

### IS5-1 Pharmacometrics-Informed Deep Learning with DeepNLME

**Chris Rackauckas[1,2], Vijay Ivaturi[2,3]**
1 Department of Mathematics, Massachusetts Institute of Technology, Cambridge MA, United States
2 Pumas-AI, United States
3 School of Pharmacy, University of Maryland, College Park MD, United States

Nonlinear mixed effects modeling (NLME) is commonly employed throughout the clinical pharmacometrics community in order to uncover covariate relationships and understand the personalized effects involved in drug kinetics. However, in many cases a full model of drug dynamics is unknown. Even further, common models used throughout clinical trials ignore many potentially predictive covariates as their connection to drug effects is unknown. Given the rise of machine learning, there have been calls to utilize deep learning techniques to potentially uncover these unknown relationships, but common deep learning techniques are unable to incorporate the prior information captured in known predictive models and thus are not predictive with the minimal data available. Thus, the question: is it possible to bridge the gap between deep learning and nonlinear mixed effects modeling?

In this talk we will describe the DeepNLME method for performing automatic discovery of dynamical models in NLME along with discovery of covariate relationships. We will showcase how this extension of the universal differential equation framework is able to generate suggested models in a way that hypothesizes testable mechanisms, predicts the covariates of interest, and allows incorporating data in the form of images and sequences into the personalized precision dosing framework. This framework and the automated model discovery process will be showcased in the Pumas pharmaceutical modeling and simulation environment. We will end by describing how this is being combined with recent techniques from Bayesian Neural Ordinary Differential Equations in order to give probabilistic estimates to the discovered models and allow for direct uncertainty quantification. Together this demonstrates a viable path for incorporating all of the knowledge of pharmacometricians into the data-driven future.



# Invited Lectures

## INVITED SESSION 5 The best of both worlds: combining deep learning and modeling

### IS5-2 Temporal and relational machine learning for biostatistical and other scientific applications

**Austin R. Benson**
Cornell University, Ithaca NY, United States

Machine learning models have made it easier to reason about not only big, but also complex, data that is pervasive throughout science and engineering. In this talk, I will discuss some of our recent research on designing machine learning models for making predictions about data with complex structure, specifically temporal and relational data, for various applications, including biostatistics. A theme of the talk will be the advantages and disadvantages of deep neural network approaches to modeling such data. As part of this, I will highlight applications in time series forecasting where deep models have been particularly useful, as well as applications where deep models are unnecessary and computationally expensive. In the latter case, domain knowledge has enabled us to design algorithms that are faster and scale to larger datasets, without sacrificing accuracy. The talk will also cover some reasons why biostatistics has unique opportunities compared to other common application domains, such as social and information network analysis.

### IS5-3 Individualizing deep dynamic models for psychological resilience data

**Göran Köber[1], Shakoor Pooseh[2], Haakon Engen[3], Andrea Chmitorz[4], Miriam Kampa[4], Anita Schick[3], Alexandra Sebastian[5], Oliver Tüscher[4], Michèle Wessa[5], Kenneth S.L. Yuen[3], Raffael Kalisch[3], Jens Timmer[2], Harald Binder[1]**
1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Germany
2 Institute of Physics, University of Freiburg, Germany
3 Neuroimaging Center, Johannes Gutenberg University Medical Center, Mainz, Germany
4 Leibniz Institute for Resilience Research (LIR), Mainz, Germany
5 Department of Psychiatry and Psychotherapy, Johannes Gutenberg University Medical Center, Mainz, Germany

Deep learning approaches can uncover complex patterns in data. In particular, variational autoencoders (VAEs) achieve this by a non-linear mapping of data into a low-dimensional latent space. Motivated by an application to psychological resilience in the Mainz Resilience Project (MARP), which features intermittent longitudinal measurements of stressors and mental health, we propose an approach for individualized, dynamic modeling in this latent space. Specifically, we utilize ordinary differential equations (ODEs) and develop a novel technique for obtaining person-specific ODE parameters even in settings with a rather small number of individuals and observations, incomplete data, and a differing number of observations per individual. This technique allows us to subsequently investigate individual reactions to stimuli, such as the mental health impact of stressors. A potentially large number of baseline characteristics can then be linked to this individual response by regularized regression, e.g., for identifying resilience factors. Thus, our new method provides a way of connecting different kinds of complex longitudinal and baseline measures via individualized, dynamic models. The promising results obtained in the exemplary resilience application indicate that our proposal for dynamic deep learning might also be more generally useful for other application domains.

# Invited Lectures

## INVITED SESSION 6 — Causal inference in continuous time for dense longitudinal data from wearable devices

### IS6-1 TMLE for Causal Effects based on continuous time longitudinal data structures

**Mark van der Laan[1], Helene Rytgaard[2]**
1 UC Berkeley, Berkeley CA, USA
2 University of Copenhagen, Denmark

In many applications one is concerned with estimation of the causal impact of a multiple time point intervention on a final outcome based on observing an i.i.d. sample of longitudinal data structures. We consider the case that subjects are monitored at a random finite set of time-points on a continuous time-scale, and at these monitoring times treatment actions and or time-dependent covariates and outcomes are collected. Current methods based on sequential regression break down under this setting. We develop a new targeted maximum likelihood estimator that still avoids estimation of the conditional densities for outcome and covariates of likelihood, but instead estimates a conditional mean function. The TMLE uses maximum likelihood-based estimation of the monitoring, treatment, censoring, and survival process intensities. We also consider a TMLE that involves estimation of all the conditional densities, including the time-dependent outcome and covariate mechanisms. We develop highly adaptive lasso estimators of the nuisance functions and establish asymptotic efficiency of the TMLE under minimal conditions. In particular, we demonstrate these new TMLEs for estimation of treatment specific survival functions for single time-point interventions on competing survival times. Advantages relative to first discretizing the time scale and using currently available corresponding TMLE are discussed. Various applications are presented and simulation results are used to demonstrate the theoretical properties.

### IS6-3 Missing data imputation for non-stationary time series in mHealth data

**Xiaoxuan Cai**
Columbia University Mailman School of Public Health, New York NY, United States

Missing data is a ubiquitous problem in studies of psychiatry, epidemiology, social, political science and many other biomedical and social science disciplines, when large number of variables are collected (especially with repeated measurements over time), and thus complete data are rarely available. As mobile devices (e.g., cell phones and fitness activity tracker bracelets) are being more widely adopted, a new way of collecting personal health data densely or even in realtime using mobile devices has realized, and revolutionized data collection methods for personalized health outcomes. Multivariate time series of outcomes, exposures, and covariates data evoke new challenges in handling missing data in order to get unbiased estimate of causal quantities of interest and call for more efficient data imputation approaches. We conducted a comprehensive comparison of the performance of complete-case analysis with most commonly used imputation methods, including mean imputation, last-observation-carried forward imputation, multiple imputation, multiple imputation with significantly longer history information, as well as state-space model in estimating causal quantities in mHealth data. Validity of most imputation methods rely on the stationarity of the time series, in the sense that both variance and treatment effect do not change over time, failing to reflect that the intervention effects in psychiatry may change with severity of disease as well as the fluctuation of the patient's mood. We propose a new imputation method derived from state space model, that accommodates the non-stationarity of time series when treatment effect and variance may change over time, to learn causal effects of interventions. We consider possible missing data in the outcome variable, exposure variable, or both under different missing rates and missing data mechanisms.



Invited Lectures

## INVITED SESSION 7 — Selective inference after variable selection

### IS7-1 Valid post-selection inference for cox regression parameters, with and without the proportional hazards assumption

**Oliver Dukes[1], Kelly Van Lancker[1], Stijn Vansteelandt[1,2]**
1 Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium
2 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom

The problem of how to best select variables for confounding adjustment forms one of the key challenges in the evaluation of exposure or treatment effects in observational studies. A major drawback of common selection procedures is that there is no finite sample size at which they are guaranteed to deliver tests and confidence intervals with adequate performance. In this talk, I will consider this problem in the context of estimating a conditional causal hazard ratio in an observational study where all key confounders are measured. An added complication with time-to-event outcomes is that one must adjust not only for confounders, but also variables that render censoring non-informative. Assuming first that the hazard ratio of interest is constant, I will describe a simple procedure for obtaining valid inference for the conditional hazard ratio after variable selection. This procedure involves three different selection steps, in order to best capture variables that account for confounding and informative censoring, and can be implemented using standard software for the Lasso. Our proposed estimator (along with common alternatives) has the disadvantage that it may not converge to something easily interpretable when the proportional hazards assumption fails. The resulting tests and confidence intervals also typically lose their validity under misspecification. I will therefore outline an alternative proposal based on a nonparametric estimand that reduces to a Cox model parameter under the proportional hazards assumption, but which continues to capture the association of interest under arbitrary types of misspecification. I will then outline how to obtain valid inference for this novel estimand whilst incorporating variable selection. The different proposals will be illustrated in the analysis of an observational study on the predictive relationship between the initial concentration of serum monoclonal protein and the progression to multiple myeloma or another plasma-cell cancer.

# Invited Lectures

## INVITED SESSION 7 — Selective inference after variable selection

### IS7-2 Selective inference for the Lasso in statistical practice

**Michael Kammer[1,2], Daniela Dunkler[1], Stefan Michiels[3], Georg Heinze[1]**
1 Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria
2 Department for Internal Medicine III, Division of Nephrology and Dialysis, Medical University of Vienna, Austria
3 Service de Biostatistique et d'Epidémiologie, Gustave Roussy; INSERM, CESP U1018, University Paris-Saclay, France

Nowadays multivariable clinical models are ubiquitous, facilitating personalized medicine and guiding therapy decisions. Such models are often developed by the use of variable selection, but the uncertainty introduced by selection is often ignored in the dissemination. Part of the reason is that classical methods for statistical inference to estimate e.g. confidence intervals are not applicable after selection. One way to facilitate proper post-selection inference is by means of the selective inference framework, which addresses inference for statistical hypotheses that are formulated and analysed using the same set of data. In recent years the methodology was developed for the widely used Lasso method, i.e. L1-penalized regression, but there are also approaches agnostic of the model selection procedure. We will present some practical considerations when working with the selective inference framework for regression problems. We will discuss a systematic simulation study in linear and logistic Lasso regression. Our focus lies on the properties of selective confidence intervals derived from different approaches, in particular selective coverage, power to exclude zero and stability. We elaborate on the practical use and interpretation of selective inference using two real-data case-studies of typical applications. First, selective inference after the selection of anthropometric features to estimate body fat in men. Second, selective inference for the main exposure after confounder selection in a cardiology dataset. We found selective inference to be challenging in terms of interpretation and computation. Development of corresponding user-friendly software is still in its infancy. Lasso confidence intervals tended to be very wide and quite variable, but could potentially improve model selection properties, in particular false positive findings. Simple selection agnostic methods showed unsatisfactory trade-offs between selection and inference accuracy, while modern approaches were much more conservative and computationally expensive, limiting their practical usability. In conclusion, selective inference using the Lasso is a promising tool for statistical modelling, but remains difficult to use in practice.

### IS7-3 Biologically-informed development of treatment selection scores from high-dimensional omics data

**Lisa M. McShane[1], Ming-Chung Li[1], George Wright[1], Ting Chen[2], Lori Long[2], Qian Xie[3], Jyothi Subramanian[2], Zhiwei Zhang[1], Yingdong Zhao[1]**
1 Biometric Research Program, National Cancer Institute, National Institutes of Health, Bethesda MD, USA
2 Emmes Corporation, United States (under contract to the U.S. National Cancer Institute)
3 General Dynamics Information Technology, United States (under contract to the U.S. National Cancer Institute)

Precision medicine therapeutic approaches rely on matching mechanism of action of a therapy to biological and molecular characteristics of a patient or the patient's disease. Therapies that can correct for aberrant or missing gene products or compensate for a disrupted biological pathway hold promise for the treatment of the corresponding disease. In oncology, many predictors based on multivariable scores generated from high dimensional omics data have been developed for purposes of prognosis; some of those have been secondarily assessed for their value in informing choice between different therapy options. Repurposing a prognostic predictor may be suboptimal because there are fundamental differences between the goals of prognostication and therapy selection. The modified covariates method of Tian and colleagues (JASA 2014;109:2350-2358) is one approach that has been proposed specifically for development of a therapy selection predictor. Biologically informed enhancements of the modified covariates approach that use information about biological pathways significantly associated with outcome or that use pre-specified variable groupings are proposed in this talk. These biologically informed approaches are found to yield treatment selection predictors with improved performance relative to that of the original modified covariates method on some real omics data from patients with cancer. The discussion additionally highlights challenges in identification of data that are suitable for treatment selection predictor development as well as issues to consider in selection of metrics to evaluate performance of these predictors.



# Invited Lectures

## INVITED SESSION 8 — Variance modelling for multilevel data and joint models

### IS8-1 A general framework and implementation for variance modelling in joint model settings

**Michael J. Crowther**
Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

The rise in availability of electronic health record data enables us to answer more detailed clinical questions; however, the associated increased complexity raises substantial statistical and computational challenges. Recent work in the area of joint models has introduced an extended mixed effects framework, encompassing multiple outcomes of any type, each of which could be repeatedly measured (longitudinal), with any number of levels, and with any number of random effects at each level (Crowther, 2020). This allows for sharing and linking between outcome models in an extremely flexibly way, either by linking random effects directly, or the expected value of one outcome (or function of it) within the linear predictor of another. Non-linear and time-dependent effects are also seamlessly incorporated to the linear predictor through the use of splines or fractional polynomials. In this talk, I'll present an extension to the framework to further allow modelling of variance components directly, allowing each random effect or residual variance to have its own complex linear predictor, such as allowing for heteroskedasticity, which in turn provides new tools for joint modelling. Throughout my talk I will illustrate an accompanying user-friendly implementation in Stata, showing how to build and estimate a joint longitudinal-survival model with complex variance components, quantifying how between-subject variation in the level 1 variance structure of a continuous biomarker (e.g., blood pressure), can be associated with survival. Dynamic predictions from such a joint model will also be derived and presented. Due to the generality of the implementation, multiple outcomes, such as multiple biomarkers or competing risks, are also immediately available.

**Reference:**

*Crowther MJ. merlin - a unified framework for data analysis and methods development in Stata. Stata Journal 2020;20(4):763-784.*

### IS8-2 Dynamic structural equation modeling for intensive longitudinal data

**Ellen L. Hamaker**
Methodology and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University, Netherlands

**Objective:** Recent methodological innovations have opened up an exciting new horizon of research opportunities. Technological developments such as smartphones and other wearable devices have created new data collection methods such as ambulatory assessments, experience sampling, and ecological momentary assessments. Characteristic of the intensive longitudinal data that are obtained with these methods is that they consist of large numbers of repeated measurements of what individuals are doing, thinking and feeling while living their daily life. As such, these data provide us the opportunity to study individual processes as they unfold over time, and to investigate individual differences therein. But to fully realize this unique potential of intensive longitudinal data, we need new statistical techniques that adequately deal with the specifics of these data, and that can uncover the meaningful patterns hidden in them. This has led to diverse innovations, including the development of dynamic structural equation modeling (DSEM), a new toolbox in the software package Mplus.
**Statistical Methods:** DSEM forms a combination of: a) time series analysis to model the lagged relations between variables, thereby accounting for the autocorrelation structure within individuals; b) multilevel modeling to ensure the proper decomposition of variance into within-person and between-person components, and to allow for individual differences in means, slopes, and variances; and c) structural equation modeling such that multiple observed variables can be analyzed simultaneously, and can be combined using factor analysis and/or mediation analysis. Additionally, DSEM can account for unequal time-intervals between observations, and for trends and cycles over time.
**Application:** In this talk, I will briefly sketch the general DSEM framework, and then discuss several specific ways in which DSEM can be used to analyze particular data structures and tackle specific research questions. These include data from a design in which a randomized controlled trial is combined with intensive longitudinal measurements before and after the intervention, and data for which there may be a need to account for patters due to cycles such as day-of-the-week patterns or circadian rhythms. These empirical examples illustrate the wide variety of modeling opportunities that are offered by DSEM.

# Invited Lectures

## INVITED SESSION 8 — Variance modelling for multilevel data and joint models

### IS8-3 Jointly modelling longitudinal heteroscedasticity and a time-to-event outcome

**Jessica Barrett[1], Richard Parker[2], Raphael Huille[3], Yuichiro Yano[4], Michael Griswold[5]**
1 MRC Biostatistics Unit, University of Cambridge, United Kingdom
2 MRC Integrative Epidemiology Unit, University of Bristol, United Kingdom
3 ENSAE, France
4 Department of Preventive Medicine, University of Mississippi Medical Center, Jackson MS, United States
5 Center of Biostatistics and Bioinformatics, University of Mississippi Medical Center, Jackson MS, United States

In the clinical literature it has been shown that individuals with higher variability in their blood pressure measurements have a greater risk of cardiovascular disease. This is typically explored by calculating a variability measure, e.g. the standard deviation, from a set of blood pressure measurements per individual, and including this as an explanatory variable in a regression model for the time to the first cardiovascular event. However, this leads to regression dilution bias in the estimated association parameter because the variability measure is subject to measurement error.

We will explore statistical models which allow within-individual variability, as well as the mean, to depend on covariates and/or random effects, e.g. mixed effects location scale models (Hedeker et al, 2008). We propose a joint model with mixed effects location scale and time-to-event sub-models for the longitudinal blood pressure measurements and time to first cardiovascular event respectively (Barrett et al, 2019). The time-to-event sub-model incorporates the random effect associated with the longitudinal within-individual variability, which allows direct estimation of the association between blood pressure variability and the risk of CVD events.

We use simulation studies and data from the Atherosclerosis Risk in Communities (ARIC) study to compare the joint model with the usual method used in the literature and a two-stage method. We demonstrate that substantial bias may be incurred by the usual method and slight to moderate bias with the two-stage method, especially when blood pressure measurements are taken concurrently with the time-to-event follow-up. From the analysis of ARIC study data, the estimated hazard ratio for the association between visit-to-visit systolic blood pressure variability and cardiovascular disease from a joint model with random intercept, slope and variability effects is 1.08 (95% CI 1.04, 1.09) per unit increase in systolic blood pressure standard deviation.

**References:** *[1] Hedeker et al (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. Biometrics 64: 627-634.*
*[2] Barrett et al (2019). Estimating the association between blood pressure variability and cardiovascular disease: An application using the ARIC Study. Statistics in Medicine 38:1855-1868.*

# SESSION OC1A

## SESSION OC1A — Mendelian randomisation & causal inference

### OC1A-1 How to deal with collider bias in Mendelian randomization analysis?

**Claudia Coscia[1,2], Teresa Perez[2], Núria Malats[1], Stephen Burgess[3]**
1 Spanish National Cancer Research Centre, Madrid, Spain
2 Department of Statistics and Data Science, Complutense University of Madrid, Spain
3 MRC Biostatistics Unit, University of Cambridge, United Kingdom

**Background:** Mendelian randomization (MR) is a method used to estimate the causal effect of a risk factor on an outcome by using genetic variants as instrumental variables. An instrumental variable must be associated with the exposure, and with the outcome only through the exposure and cannot be associated with any confounders. When the aim of the study is to obtain a causal effect for a particular subgroup of the population rather than a population causal effect (for example by stratifying on a specific variable), collider bias could be generated. Collider bias appears when we control for one variable that it is influenced by other two variables This bias could lead to an association between the instrument and the outcome, violating the instrument assumptions, and it could lead to invalid results.
**Objectives:** To identify potential collider bias in MR studies, to assess its impact on the causal effect estimates and to evaluate a new technique in MR as a solution for controlling the collider bias.
**Methods:** We proposed the stratification approach in MR analysis to study the causality between the risk factor and the outcome controlling the collider bias. This method creates a new variable based on both the collider and the instrument, which is afterwards categorized into quartiles and then the stratum-specific causal effects can be estimated. With this solution, we control the collider bias and therefore the estimates effect are unbiased. We simulate several datasets considering different levels of collider bias and strongness of the instrument. We apply this method to a real dataset.
**Results:** This is an on-going study, and our aim is to be able to identify those scenarios where the collider could generate biased causal estimates and to find if the stratification approach is an appropriate solution. This method will be applied to estimate the causal effect of diabetes mellitus on pancreatic cancer among different levels of body mass index.
**Conclusions:** We detail a methodological process to analyse the impact of the collider bias and a potential solution to estimate the local causal effect.

**References:** *Burgess S, Davies NM, Thompson SG; EPIC-InterAct Consortium. Instrumental variable analysis with a nonlinear exposure-outcome relationship. Epidemiology. 2014;25(6):877-885. doi:10.1097/EDE.0000000000000161*
*Gkatzionis A, Burgess S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? Int J Epidemiol. 2019 Jun 1;48(3):691-701. doi: 10.1093/ije/dyy202. PMID: 30325422; PMCID: PMC6659463.*



# SESSION OC1A

## OC1A-2 The impact of instruments' selection on Mendelian randomization results: a case study

**Daniele Bottigliengo[1], Amke Caliebe[2], Inke R. König[3], Fabiola Del Greco M.[1]**
1 Institute for Biomedicine, EURAC Research, Bolzano, Italy
2 Institute of Medical Informatics and Statistics, Christian–Albrechts–Universität zu Kiel, Germany
3 Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany

Mendelian randomization (MR) investigates the causal effect of modifiable exposures on health outcomes within the Instrumental variable (IV) framework, using genetic variants as instruments. Therefore, a genetic variant should satisfy the IV assumptions to be considered as a valid instrument. MR analyses performed on summary data from large genetic association studies are increasingly being used and characterized by an increasing number of potential candidate instruments and a greater power. But, when multiple instruments are available, IV assumptions can be questionable and findings might highly depend on how those instruments are selected [1]. The present study aims at evaluating how MR estimates are sensitive to sets of instruments selected using different criteria. Applying different genetic data sources, the causal role of inflammation on Parkinson's disease (PD) is investigated as a motivating example.
Considering two-sample MR methods available for summary genetic data, two main practical issues related to the instruments' selection were explored: the Linkage Disequilibrium (LD) threshold and the pleiotropy strategy. Regarding LD, three sets of instruments were identified according to different thresholds of correlation among IVs. To rule out the presence of pleiotropy, two sets of instruments were derived excluding: (i) IVs with extreme Q-statistic; (ii) IVs identified as outliers by the Radial plot method. Causal effects were estimated using both fixed and random effects Inverse Variance Weighted methods and nine robust MR estimators [2]. The inflammation marker C-reactive Protein was used as exposure and PD as the outcome.
When different LD thresholds were set, no important differences in MR estimates were observed within each method. And there was no consensus between results obtained with different methods. Moreover, a different exclusion strategy for reducing pleiotropic IVs led to opposite effect estimates.
In MR studies, the findings can be sensitive to the choice of the instruments. Identification of appropriate strategies for selecting genetic variants is necessary and potential study-specific issues that may arise when different sets of instruments are considered should be accounted for.

**References:** *[1] Swerdlow DI, Kuchenbaecker KB, Shah S, Sofat R, Holmes MV, White J, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. Int J Epidemiol. 2016;45:1600–16.*
*[2] Slob EAW, Burgess S. A comparison of robust Mendelian randomization methods using summary data. Genetic Epidemiology. 2020;44:313–29.*

## OC1A-3 Caution When Inferring the Effect Direction in Mendelian Randomization

**Sharon M. Lutz[1,2], Stijn Vansteelandt[3,4], Christoph Lange[2]**
1 Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston MA, United States
2 Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston MA, United States
3 Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium
4 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom

In genetic association studies, Mendelian Randomization (MR) has gained in popularity as a concept to assess the causal relationship between two phenotypes. Some methods, the MR Steiger and bidirectional MR approaches, have been proposed as tools that can infer the causal direction between two phenotypes. Through simulation studies, we extend previous work to examine the ability of the MR Steiger and bidirectional MR approaches to correctly determine the effect direction in the presence of pleiotropy, measurement error, and unmeasured confounding. In addition, we examined the performance of these approaches when there is a longitudinal causal relationship between the two phenotypes, under weak instrument variables, and differing distributions for the phenotypes (binary, Poisson, etc). We also applied the Steiger method and bidirectional MR approach to the COPDGene study, a case-control study of Chronic Obstructive Pulmonary Disease (COPD) in current and former smokers, to examine the role of smoking on lung function in the presence of pleiotropy and measurement error.



# SESSION OC1A

## OC1A-4 Tying research question and analytical strategy when variables are affected by medication use

**Jungyeon Choi[1], Olaf M. Dekkers[2], Saskia le Cessie[3]**
1 Department of Clinical Epidemiology, Leiden University Medical Center, Netherlands
2 Department of Clinical Epidemiology & Department of Endocrinology and Metabolism, Leiden University Medical Center, Netherlands
3 Department of Clinical Epidemiology & Department of Biomedical Data Sciences, Leiden University Medical Center, Netherlands

In an epidemiological setting where measurements fluctuate over time due to intercurrent events, ill-defined research questions could be particularly problematic. Medication use is one important cause for this change, as it is prescribed to target specific measures. When a research question fails to specify how the medication use should be handled methodologically, arbitrary decisions may be made during the analysis phase, which likely leads to a mismatch between the intended research question and the performed analysis. The mismatch can result in vastly different interpretations of the estimated effects depending on how one handled medication use. In some cases, the estimated effect may not even provide any clinical relevance. Thus, a research question such as 'what is the effect of X on Y?' requires further elaboration, and it should take into account whether and how medication use has affected the measurements of interest.
The importance of well defining a research question when intercurrent events occur has been stressed by causal inference experts [1]. In the field of randomized trials, several different estimands for intercurrent events such as post-randomization medication use have been proposed [2]. Despite this, a recently conducted literature review by us on the handling of medication use in medical studies (in preparation) showed that the research question was formulated vaguely in a majority of studies and their intended aims were unclear.
Therefore, in our study, we will discuss how well-defined research questions can be formulated when medication use is involved in observational studies. We will distinguish between a situation where a (possible time-varying) exposure is affected by medication use and where the outcome of interest is affected by medication use. For each setting, we will give examples of different research questions that could be asked depending on how medication use is taken into account in the estimand and discuss methodological considerations under each research question.

**References:** *[1] Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, et al. A causal framework for classical statistical estimands in failure-time settings with competing events. Statistics in Medicine 2020;39(8):1199-236.*
*[2] ICH E9 working group, ICH E9 (R1): addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. European Medicines Agency, 2020.*

## OC1A-5 Identification of causal effects in case-control studies

**Bas B.L. Penning de Vries[1], Rolf H.H. Groenwold[2]**
1 Department of Clinical Epidemiology, Leiden University Medical Center, Netherlands
2 Departments of Clinical Epidemiology and Biomedical Data Sciences, Leiden University Medical Center, Netherlands

**Introduction:** Case-control designs are an important yet commonly misunderstood tool in the epidemiologist's arsenal for causal inference. We reconsider classical concepts, assumptions and principles and explore when the results of case-control studies can be endowed a causal interpretation.
**Methods:** We present a framework that is rich enough to articulate and study various target causal quantities (i.e., estimands) relating to intention-to-treat or per-protocol effects. We then establish how, and under which conditions, these estimands can be identified based on the data that are collected under popular sampling schemes (case-base, survivor, and risk-set sampling, with or without matching).
**Results:** We present a concise summary of our identification results that link the estimands to the (distribution of the) available data and articulate under which conditions these links hold.
**Conclusion:** The modern epidemiologist's arsenal for causal inference is well-suited to make transparent for case-control designs what assumptions are necessary or sufficient to endow the respective study results with a causal interpretation and, in turn, help resolve or prevent misunderstanding. Our approach may inform future research on different estimands, other variations of the case-control design or settings with additional complexities.

# SESSION OC1B

## SESSION OC1B — Bayesian clinical trial design (Part 1)

### OC1B-1 Modular components in basket trials and connections among the applied tools

**Moritz Pohl, Johannes Krisam, Meinhard Kieser**
Institute of Medical Biometry and Informatics, Heidelberg University, Germany

Basket trials have been a hot-topic in medical and statistical research throughout the last ten years. The goal of basket trials is to analyse one treatment among several diseases in one trial based on the medical justification of a common (genetic) pathway of the treatment. For this setting many different designs with different statistical techniques were proposed in the literature which resulted in an unclear situation of methodological opportunities.
Therefore, a modular approach is presented to introduce a structure to basket trial designs and to the available techniques. Currently used designs employ either only frequentist techniques [1], only Bayesian methods [2] or a combination of them throughout the course of the trial. Within the three genereal techniques a variety of different tools are applied. To clarify the situation, the available tools are classified and simplified, also connections are presented which show differences, similarities or even equivalence of tools that at first appear different. Knowing about connections of the tools facilitates the planning of basket trials and also their conduct in hands-on medical research because researchers have the opportunity to design the basket trial in the sense of 'as simple as possible and as complex as necessary' which makes a basket trial more tangible for all stakeholders of new clinical trials. We consider the presented work as a required basis to promote knowledge about basket trials, to motivate further research and to ultimately set the path for individualized treatments that showed their benefit in basket trials.

**References:** *[1] Cunanan K M, Iasonos A, Shen R, Begg C B, Gönen M. An efficient basket trial design. Statistics in Medicine. 2017; 36(10): 1568-1579. [2] Fujikawa, K, Teramukai, S, Yokota, I, Daimon, T. A Bayesian basket trial design that borrows information across strata based on the similarity between the posterior distributions of the response probability. Biometrical Journal. 2020; 62: 330– 338*

### OC1B-2 Seamless Master Protocol with account for correlations within subgroups

**Benjamin Duputel[1,2], Sarah Zohar[1], François Montestruc[2], Moreno Ursino[1,3]**
1 INSERM, Centre de Recherche des Cordeliers, Université de Paris, Sorbonne Université, France
2 eXYSTAT, Malakoff, France
3 F-CRIN PARTNERS platform, AP-HP, Université de Paris, France

**Introduction:** Master Protocol designs (Basket, Umbrella, Platform) allow simultaneous comparison of multiple treatments or disease subgroups. Master protocols can be also designed as seamless studies, in which two or more clinical phases are considered within the same trial. They can be divided in two categories: operational seamless, in which the two phases are separated as two independent studies or inferential seamless, in which the interim analysis is considered as an adaptation of the study. Bayesian designs taking into account correlation between treatments and doses are scarcely studied.
**Aim:** To propose and compare two Bayesian seamless Phase II/III designs (operational and inferential) using a binary endpoint for the first stage and a time to event endpoint for the second step.
**Methods:** For the first stage, a Bayesian hierarchical model accounting for multiple doses of multiple treatments while taking partial ordering the correlation structure into account was developed. After treatment and dose selection, based on posterior and predictive probabilities, the results of the first phase were incorporated into prior distributions of a time-to-event model. Extensive simulations were performed in order to compare the robustness and operating characteristics of the two seamless designs depending on several prior variabilities or effective sample sizes.
**Results:** The inferential seamless has in average better operating characteristics in terms of sample size required and precision. If one which to obtain the same operating characteristics under the operational seamless design, a bigger sample size is needed.
**Conclusion:** When using stage one data to build the prior distributions for the time to event model, it should be done carefully in order to not overpower the posterior distributions and influence trial results. Our proposition allows to avoid this kind of issue.



SESSION OC1B

### OC1B-3 Monotonicity Rules for Inference in Basket Trials

**Lukas Baumann, Johannes Krisam, Meinhard Kieser**
Institute of Medical Biometry and Informatics, University of Heidelberg, Germany

A basket trial is a new type of clinical trial where several subgroups are exposed to a new treatment. They are especially popular in oncology, where the subgroups, denoted as so-called baskets, usually comprise patients with different primary tumor sites but a common biomarker or mutation. Most basket trials are single-arm phase II trials that investigate a binary endpoint such as tumor response.
Several designs for such trials have been proposed. In simple approaches, the results of subgroups are pooled before the analysis if the results are similar. Advanced designs allow a more nuanced combination of results, many of which use Bayesian tools to share information between baskets. An exciting method was recently proposed by Fujikawa et al. (2020). In their design, the individual baskets are initially analyzed using a beta-binomial model. Information is then shared by computing the posterior distribution parameters as a weighted sum of the parameters of the individual posterior distributions. The weights are derived from a similarity measure between the individual posterior distributions and further tuning parameters. Compared to other basket trial designs with Bayesian components, this design is computationally very inexpensive. Operating characteristics can be calculated directly without having to rely on simulations.
While information sharing increases the power, it can also lead to undesired results. Using the example of Fujikawa et al.'s design, we show that the number of null hypotheses that can be rejected is not always monotonically increasing in the number of observed responses. As a consequence, there are scenarios where the treatment is declared futile in all baskets even when there are at least as many responses in all baskets as in other scenarios where the treatment is declared efficacious in at least one basket. Furthermore, there are scenarios where the treatment is declared efficacious in some baskets but is declared futile in other baskets with more responses. We define monotonicity rules for the inference in basket trials and assess how the choice of the tuning parameter values that determine the amount of information sharing has an influence on whether these rules hold.

**References:** *Fujikawa, K., Teramukai, S., Yokota, I., & Daimon, T. (2020). A Bayesian basket trial design that borrows information across strata based on the similarity between the posterior distributions of the response probability. Biometrical Journal, 62(2), 330-338.*

### OC1B-4 Response-adaptive randomization in clinical trials: from myths to practical considerations

**David Robertson, Kim Lee, Boryana Lopez-Kolkovska, Sofia Villar**
MRC Biostatistics Unit, University of Cambridge, United Kingdom

Response-adaptive randomization (RAR) is part of a wider class of data-dependent sampling algorithms, for which clinical trials have commonly been used as a motivating application. In that context, patient allocation to treatments is defined using the accrued data on responses to alter randomization probabilities, in order to achieve different experimental goals. RAR has received abundant theoretical attention from the biostatistical literature since the 1930's and has been the subject of numerous debates. Recently it has received renewed consideration from the applied community due to successful practical examples and its widespread use in machine learning. Many position papers on the subject each give a specific or partial view on its usefulness and these views may be difficult for the non-expert to navigate through. This work aims to address this gap by providing a unified and updated review of methodological and practical issues to consider when debating the use of RAR in clinical trials.

# SESSION OC1B

## OC1B-5 Treatment allocation strategies for umbrella trials in the presence of multiple biomarkers: A comparison of methods

**Luke Ouma[1], Michael Grayling[1], Haiyan Zheng[2], James Wason[1,2]**
1 Population Health Sciences Institute, Newcastle University, United Kingdom
2 MRC Biostatistics Unit, University of Cambridge, United Kingdom

**Background:** Umbrella trials are an innovative trial design where different treatments are matched with subtypes of a disease, with the matching typically based on a set of predictive biomarkers. In practice, patients can test positive for multiple targeted biomarkers, hence eligible for multiple treatment arms. Consequently, different approaches could be applied to allocate multiple biomarker patients to a specific treatment arm. However, it is currently unclear how the method used to account for multiple biomarkers will affect the statistical properties of the trial.
**Methods:** We conduct a simulation study to compare five approaches that have been or could be implemented to guide treatment allocation in the presence of multiple biomarkers – equal randomisation; randomisation with fixed probability of allocation to control; Bayesian adaptive randomisation (BAR); constrained randomisation (CR); and hierarchy of biomarkers. We evaluate these approaches on six operating characteristics under different scenarios in the context of a hypothetical phase II biomarker-guided umbrella trial. We assume a binary endpoint and restrict our focus to a setting of four targeted biomarkers and their linked treatments plus a single control. We define the pairings representing the pre-trial expectations on efficacy as linked pairs.
**Results:** The hierarchy and BAR approaches have the highest power to detect a treatment-biomarker linked interaction. However, the BAR method is more robust to the biomarker ordering being invalid, a scenario when the hierarchy procedure performs poorly. When a treatment delivers an unanticipated detrimental effect, the BAR method allocates a higher proportion of multiple biomarker patients to the most promising treatments. On the other hand, The CR procedure allocates on average a higher proportion of patients to experimental treatments and thus best balances allocation to all treatment arms. Notably, all methods have reasonable bias in all scenarios.
**Conclusion:** The choice of an approach to deal with treatment allocation in the presence of multiple biomarkers may considerably influence the trial operational characteristics. However, no method is optimal in all settings. Thus, pre-specification of an approach is important, and should be considered carefully in the context of the trial sample size, prevalence of the biomarkers, and prevalence of individual overlaps.

# SESSION OC1C

SESSION OC1C — Network meta analysis

## OC1C-1 Loop-splitting in network meta-analysis: a new approach to evaluating loop inconsistency

**Becky Turner[1], Tim Band[2], Tim Morris[1], David Fisher[1], James Carpenter[1], Julian Higgins[3], Ian White[1]**
1 MRC Clinical Trials Unit, UCL, United Kingdom
2 Department of Earth Sciences, UCL, United Kingdom
3 Population Health Sciences, University of Bristol, United Kingdom

In a network meta-analysis, clinical trials evaluating multiple different treatment comparisons are modelled simultaneously, and estimation is informed by a combination of direct and indirect evidence. Network meta-analysis relies on an assumption of consistency, meaning that direct and indirect evidence should agree for each treatment comparison. Tests for inconsistency may be local or global. Existing tests do not handle treatments symmetrically and global tests based on a design-by-treatment interaction approach lack power. Here we propose new local and global tests for inconsistency and demonstrate their application to two example networks.
We apply the local test to a loop of treatments in the network meta-analysis. We define a model with one inconsistency parameter that can be interpreted as loop inconsistency. The model builds on the existing ideas of node-splitting or side-splitting in network meta-analysis. To provide a global test for inconsistency, the model can be extended across multiple independent loops with one degree of freedom per loop. We describe an algorithm for identifying independent loops within a network meta-analysis.
The models are applied first to a small network meta-analysis comparing 4 treatments for promoting smoking cessation. Local tests for inconsistency are applied to each loop and show no evidence of local loop inconsistency. Global tests for loop inconsistency are applied to every combination of independent loops. We demonstrate the invariance of the global model to choice of loops and find no global evidence of inconsistency (p=0.67). Next, the models are applied to a large network meta-analysis comparing 12 antidepressant drugs in adults with major depressive disorder. We describe how to identify a set of 31 independent loops in this network. The global model is applied and shows no global evidence of inconsistency (p=0.51).
Our proposed models handle treatments symmetrically and are invariant to choice of reference treatment, which makes interpretation easier. The global model is invariant to choice of independent loops and we have shown how to identify a set of independent loops. In comparison with the existing approach to testing for global inconsistency in network meta-analysis, our model uses fewer degrees of freedom and is expected to improve power.

# SESSION OC1C

## OC1C-2 Decision curve analysis for treatment benefit in a network meta-analysis framework

Konstantina Chalkou, Georgia Salanti
Institution of Social and Preventive Medicine, University of Bern, Switzerland

**Background:** Predicting individualized treatment effects is of great importance, so that a treatment might be targeted to individuals who benefit and be avoided by those who don't. Traditional measures of predictive accuracy, such as discrimination, do not evaluate the clinical usefulness of a model. Decision curve analysis can be used to determine whether a model should be applied in clinical practice or not, and is well described in the literature for models that compare two treatments in a randomized clinical trial [1].
**Objectives:** Our main objective is to extend the decision curve analysis methodology into a network meta-analysis framework, where several treatment options are compared in several trials. We also exemplify our methodology on a prediction model for treating patients diagnosed with relapsing-remitting multiple sclerosis (RRMS).
**Methods:** A threshold probability, based on considerations about the harms associated with each treatment and with an event, need to be determined for each treatment. Our methodology includes 12 steps to estimate the net benefit, i.e. the difference between the absolute benefits and the harms under possible strategies: treat all patients with the most effective treatment on average, treat patients based on a prediction model about multiple treatments, and treat none. The net benefit per strategy can then be plotted for a range of threshold probabilities, to reveal the most clinically useful strategy.
**Results:** We applied our methodology on a network meta-analysis individualized prediction model, which we have previously developed for RRMS [2]. We illustrated our decision curve analysis methodology using two different options: a common threshold probability for all treatments, and different threshold probabilities for each treatment. Our model appears to be clinical useful for a small range of threshold probabilities under both options.
**Conclusions:** As models to compare different treatments for individualized prediction models are becoming widely used, our methodology could be an important tool for assessing the impact of such models in the clinical practice.
**Acknowledgements:** This project is funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 825162.
**References:** *[1] Vickers AJ, Kattan MW, Sargent DJ. Method for evaluating prediction models that apply the results of randomized trials to individual patients. Trials. 2007;8(1):14. doi:10.1186/1745-6215-8-14*
*[2] Chalkou K, Steyerberg E, Egger M, Manca A, Pellegrini F, Salanti G. A two-stage prediction model for heterogeneous effects of many treatment options: application to drugs for Multiple Sclerosis. ArXiv200413464 Stat. Published online April 28, 2020. Accessed October 11, 2020. http://arxiv.org/abs/2004.13464*



# SESSION OC1C

## OC1C-3 The Risk Of Bias due to Missing Evidence in Network meta-analysis (ROB-MEN) tool: web application and implementation in a network of antidepressant drugs

Virginia Chiocchia[1], Adriani Nikolakopoulou[2], Theodoros Papakonstantinou[1], Andrea Cipriani[3], Toshi A. Furukawa[4], Julian P.T. Higgins[5], Matthew J. Page[6], Matthias Egger[1,5], Georgia Salanti[1]
1 Institute of Social and Preventive Medicine, University of Bern, Switzerland
2 Institute of Medical Biometry and Statistics, University of Freiburg, Germany
3 Department of Psychiatry, University of Oxford, United Kingdom
4 Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Japan
5 Population Health Sciences, Bristol Medical School, University of Bristol, United Kingdom
6 School of Public Health and Preventive Medicine, Monash University, Clayton VIC, Australia

The risk of bias due to missing evidence, also called reporting bias, threatens the validity of systematic reviews and meta-analysis and, therefore, potentially affects clinical decision-making. Various methods are available to assess selective outcome reporting and publication bias separately, but a rigorous framework to evaluate the impact of both sources of bias on the meta-analysis results of a network of interventions is still lacking. We developed a framework and tool to assess the Risk Of Bias due to Missing Evidence in Network meta-analysis (ROB-MEN).
We built on the tool for assessing Risk Of Bias due to Missing Evidence (ROB-ME, https://www.riskofbias.info) developed by Page et al. We expanded the ROB-ME framework to network meta-analysis (NMA). We used qualitative and quantitative methods to combine the risk of bias due to missing evidence in pairwise comparisons with their impact on the network estimates.
Our framework first evaluates the risk of bias due to missing evidence for each direct comparison separately. We consider possible bias due to the presence of studies with unavailable results (known unknowns) and the potential for unpublished studies (unknown unknowns) before reaching an overall judgement about the risk of bias due to missing evidence in each comparison. Then, we evaluate the risk of bias due to missing evidence in each NMA estimate, which is our tool's final output. The bias and contributions from direct comparisons to the NMA are thus combined with the likelihood of small study-effects as evaluated by network meta-regression and the bias from unobserved comparisons.
We present the ROB-MEN tool and illustrate its application in an NMA of 18 antidepressants from head-to-head studies [1] using the R Shiny app that we developed to facilitate the assessment process. The ROB-MEN tool will also be implemented in the open-source CINeMA web application (https://cinema.ispm.unibe.ch/) to supplement the reporting bias domain.
**Reference:** *[1] Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. The Lancet 2018;391:1357–66. doi:10.1016/S0140-6736(17)32802-7*

# SESSION OC1C

## OC1C-4 Network meta-analysis and random walks

➜ *STCA AWARD WINNER*

**Annabel L. Davies[1], Theodoros Papakonstantinou[2,3], Adriani Nikolakopoulou[2], Gerta Rücker[2], Tobias Galla[1,4]**

1 Theoretical Physics, Department of Physics and Astronomy, School of Natural Sciences, University of Manchester, United Kingdom
2 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Centre-University of Freiburg, Germany
3 Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland
4 Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), Campus Universitat Illes Balears, Palma de Mallorca, Spain

Network meta-analysis (NMA) has been established as a central tool of evidence synthesis in clinical research. The results of an NMA depend critically on the quality of evidence being pooled. In assessing the validity of an NMA, it is therefore important to know the proportion contributions of each direct treatment comparison to each network treatment effect. To construct the matrix of contributions, Papakonstantinou et al (2018) presented an algorithm that identifies the flow of evidence in each evidence 'path' and decomposes it into direct comparisons. This method is based on the observation by König et al (2013) that each row of the hat matrix represents an evidence flow network for each treatment comparison. However, in certain cases, the algorithm presented by Papakonstantinou et al. [1] is associated with ambiguity according to the selection of paths.

The aim of our work is to demonstrate the analogy between NMA and random walks. We also aim to illustrate the clinical application of this analogy in deriving the proportion contribution matrix.

A random walk on a graph is a stochastic process that describes a succession of random 'hops' between vertices which are connected by an edge. The weight of an edge relates to the probability that the walker moves along that edge. In statistical mechanics, there exists a well-established connection between electrical networks and random walks. We use the existing analogy between meta-analytic and electrical networks to construct the transition matrix for a random walk on the network of evidence. We show that the net number of times a walker crosses each edge of the network is directly related to the evidence flow graph in König et al [2]. By then defining a random walk on the directed evidence flow network, we derive analytically the matrix of proportion contributions.

The interdisciplinary analogy between NMA and random walks provides new insight into NMA methodology. In particular, the analogy leads to a derivation of the proportion contribution matrix without the ambiguity of previous algorithms. Our approach can therefore be used to reliably quantify the contribution of individual study limitations to the resulting network treatment effects.

**References:** *[1] T. Papakonstantinou, A. Nikolakopoulou, G. Rücker, A. Chaimani, G. Schwarzer, M. Egger, and G. Salanti. Estimating the contribution of studies in network meta-analysis: paths, flows and streams. F1000Research, 7:610, 2018.*
*[2] J. König, U. Krahn, and H. Binder. Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. Stat. Med., 32(30):5414–5429, 2013.*

# SESSION OC1C

## OC1C-5 Bayesian network meta-analysis methods for combining IPD and aggregate data from single-arm studies and RCTs

**Janharpreet Singh[1], Sandro Gsteiger[2], Clare L. Gillies[3], Keith R. Abrams[4], Sylwia Bujkiewicz[1]**

1 Department of Health Sciences, University of Leicester, United Kingdom
2 Global Access, F. Hoffmann-La Roche Ltd, Switzerland
3 Diabetes Research Centre, University of Leicester, United Kingdom
4 Centre for Health Economics, University of York, United Kingdom

**Background:** In health technology assessment, evidence from a single-arm study assessing effectiveness of a new treatment, where individual participant data (IPD) are available, may need to be synthesised with aggregate data in a network of randomised controlled trials (RCTs) assessing existing treatments.
**Objective:** We aim to develop methods to perform a network meta-analysis (NMA) combining IPD and aggregate data from single-arm studies and RCTs, under either contrast- or arm-based parametrisations, and to compare the methods using an applied example in rheumatoid arthritis (RA).
**Methods:** We extend the contrast- and arm-based NMA methods by Hong et al [1] for IPD only, using the approach by Saramago et al [2] to combine IPD and aggregate data via shared model parameters. We apply the methods to a network of RCTs assessing biologic therapies as first-line treatments for RA, with American College of Rheumatology (ACR20) response criteria as the outcome measure. The network consists of aggregate-level data from 10 RCTs and IPD from one single-arm study. We also apply a method with independent baseline effects [2], to understand the impact of assuming exchangeable baseline effects.
**Results:** For tocilizumab compared to placebo, there was an increase in uncertainty when incorporating IPD from an additional single-arm study vs. an additional RCT; for the contrast-based approach, pooled log odds ratio (OR) = 1.55 (0.16, 2.85) vs. 1.54 (0.38, 2.74), and for the arm-based approach, log OR = 1.50 (0.32, 2.76) vs. 1.51 (0.38, 2.62). However, the difference between the estimates was not sufficiently large enough to change conclusions as neither of the 95% credible intervals contained zero. The estimates were also similar to those obtained when assuming independent baseline effects; log OR = 1.56 (0.35, 2.81).
**Conclusions:** Incorporating IPD from a single-arm study into a network of RCTs, to estimate the relative effect of new vs. existing treatments, can be achieved via an assumption of exchangeable baseline effects. This assumption may introduce some bias (discrepancy). However, in the example presented here it did not result in significant bias. Further research is required to understand when bias can arise and to what extent it can be adjusted for.

**References:** *[1] Hong H, Fu H, Price KL, Carlin BP. Incorporation of individual‐patient data in network meta‐analysis for multiple continuous endpoints, with application to diabetes treatment. Statistics in Medicine. 2015 Sep 10;34(20):2794-819.*
*[2] Saramago P, Sutton AJ, Cooper NJ, Manca A. Mixed treatment comparisons using aggregate and individual participant level data. Statistics in medicine. 2012 Dec 10;31(28):3516-36.*

# SESSION OC1D

## SESSION OC1D — Omics and genetic studies

### OC1D-1 ATLASQTL & EPISPOT: two joint hierarchical approaches for detecting and interpreting hotspots in molecular QTL studies

**Hélène Ruffieux[1], Leonardo Bottolo[2], Benjamin P. Fairfax[3], Sylvia Richardson[1]**
1 MRC Biostatistics Unit, University of Cambridge, United Kingdom
2 Department of Medical Genetics, University of Cambridge, United Kingdom
3 Department of Oncology, Weatherall Institute for Molecular Medicine, University of Oxford, United Kingdom

We present ATLASQTL and EPISPOT, two complementary Bayesian sparse regression approaches for joint analysis of molecular quantitative trait locus (QTL) data. ATLASQTL performs QTL mapping on a genome-wide scale, while EPISPOT refines this mapping and generates mechanistic hypotheses by exploiting large panels of epigenetic annotations as predictor-level information. Both approaches consider a series of parallel regressions combined in a hierarchical manner to flexibly accommodate high-dimensional responses (molecular levels) and predictors (genetic variants). This novel framework allows information-sharing across outcomes and variants, and directly models the propensity of variants to be trans hotspots, i.e., to remotely control the levels of many gene products, via a dedicated top-level representation. EPISPOT also couples QTL mapping with a hypothesis-free selection of annotations which contribute to the primary QTL effects. Both methods implement efficient annealed variational inference procedures that improve exploration of multimodal spaces and allow simultaneous analysis of data comprising hundreds of thousands of predictors, and thousands of responses and samples. This unified learning boosts statistical power and sheds light on the mechanistic basis of the uncovered hits. ATLASQTL and EPISPOT therefore mark a step forward to improving the thus far challenging detection and functional interpretation of trans-acting genetic variants, including hotspots. We illustrate the advantages of our framework in simulations emulating real-data conditions and in a monocyte expression QTL study, which confirms known hotspots and reveals new ones, as well as plausible mechanisms of action. Software for our methods is publicly available as packages implemented in R and C++: https://github.com/hruffieux/atlasqtl and https://github.com/hruffieux/epispot.

**References:** *H. Ruffieux, A. C. Davison, J. Hager, J. Inshaw, B. Fairfax, S. Richardson, and L. Bottolo. A global-local approach for detecting hotspots in multiple response regression. The Annals of Applied Statistics, 14:905-928, 2020, 10.1214/20-AOAS1332.*
*H. Ruffieux, B. P. Fairfax, I. Nassiri, E. Vigorito, C. Wallace, S. Richardson, L. Bottolo. EPISPOT: An epigenome-driven approach for detecting and interpreting hotspots in molecular QTL studies, biorxiv 10.1101/2020.09.21.305789.*

### OC1D-2 Explained Variation in the Linear Mixed Model

**Nicholas Schreck**
Biostatistics, DKFZ Heidelberg, Germany

The coefficient of determination is a standard characteristic in linear models with quantitative response variables. It is widely used to assess the proportion of variation explained, to determine the goodness-of-fit and to compare models with different covariates. For models with categorical covariables only, the coefficient of determination reduces to the explained sum of squares known from ANOVA. However, there has not been an agreement on a similar quantity for the class of linear mixed models yet. We introduce a natural extension of the well-known adjusted coefficient of determination in linear models to the variance components form of the linear mixed model. This extension is dimensionless, has an intuitive and simple definition in terms of variance explained, is additive for several random effects and reduces to the adjusted coefficient of determination in the linear model without random effects.
To this end, we prove a full decomposition of the sum of squares of the independent variable into the explained and residual variance. Based on the restricted maximum likelihood equations, we propose a novel measure for the explained variation which we allocate specifically to the contribution of the fixed and the random covariates of the model. In particular, hierarchical data (clustered data, repeated measurements, longitudinal data) can be investigated by this approach. We illustrate the usefulness of our approach on a typical real dataset with repeated measures, where we are able to partition the variability in the clinical endpoint to fixed effects (such as age, sex, health status), as well as random effects (patients).
Another important application is the estimation of the single nucleotide polymorphism heritability in genome-wide association studies with complex traits. We compare our approach with existing approaches (GCTA-GREML, LD score regression) on real datasets of model organisms such as Arabidopsis thaliana.



SESSION OC1D

### OC1D-3 Reconstructing KIR haplotypes taking ambiguous and missing data into account

**Lars van der Burg[1], Liesbeth C. de Wreede[1,2], Henning Baldauf[2], Jürgen Sauter[2], Johannes Schetelig[2,3], Hein Putter[1], Stefan Böhringer[1]**
1 Biomedical Data Sciences, LUMC, Leiden, Netherlands
2 DKMS, Dresden/Tübingen, Germany
3 Department of Internal Medicine I, University Hospital Carl Gustav Carus, Dresden, Germany

**Background:** The role of KIR (killer-cell immunoglobulin-like receptors) genes in improving outcomes after allogeneic hematopoietic stem cell transplantation is under debate. It is hard to assess, since the KIR region is genetically complex; exhibiting copy number variations, a huge allelic diversity and alleles are often measured with ambiguities. For better modelling the biological impact of KIR genes, haplotypes have to be reconstructed based on these complex data [1].
**Objectives:** We present a method to reconstruct haplotype gene content based on genotype calls from independently genotyped KIR genes with a high percentage of ambiguities, i.e., partially missing data. We developed an Expectation-Maximization (EM)-algorithm to deal with the missing data components, taking linkage disequilibrium between different KIR genes into account, to estimate haplotype frequencies. A simulation study was performed to evaluate the effect of combining the information of multiple genes into one analysis and to compare the EM-algorithm with a naïve imputation method.
**Methods:** The complete data is obtained by grouping the donor's possible diplotypes of all KIR genes. Due to the high data dimensionality, the EM-algorithm has to be refined via heuristic grouping strategies, e.g., summarizing rare information. To further improve efficiency, haplotype frequencies are estimated iteratively, by adding genes one at a time. Our strategy ensures accurate estimates combined with a user controlled dimensionality.
**Results:** Estimated haplotype frequencies with this new method are approximately unbiased in simulations. Real data analysis estimates are close to those obtained in an independent study [1]. Our simulation study shows that the EM-algorithm combines information from multiple genes when linkage disequilibrium is high, whereas increased ambiguity levels increase standard errors. In comparison with a naïve imputation method, the EM-algorithm is more efficient.
**Conclusions:** Our new EM-algorithm based method is the first to account for the full genetic architecture of the KIR loci. This algorithm can handle the numerous observed ambiguities, and allows for the grouping of haplotypes to perform implicit dimension reduction. Combining information from multiple genes improves the accuracy of estimates and allows for better haplotype reconstruction. This method can also be applied to other sets of genes with missing data.

**Reference:** *[1] Solloch et al., "Estimation of German KIR Allele Group Haplotype Frequencies," Frontiers in immunology, vol. 11, pp. 429-441, 2020.*

### OC1D-4 COMET: An R package to identify sample cross-contamination in whole genome sequencing studies

**Alexandre Thiéry[1], Tanja Zeller[2], Stefan Blankenberg[2], Andreas Ziegler[1,2,3]**
1 Cardio-CARE, Medizincampus Davos, Switzerland
2 University Heart Center Hamburg, University Medical Center Hamburg-Eppendorf, Germany
3 School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

Identification of sample cross-contamination is crucial in next generation sequencing (NGS) studies because undetected contamination may lead to bias in association studies. In PCR-free germline multiplexed whole genome sequencing (WGS) studies, sample cross-contamination may be investigated by studying the excess of non-matching reads at homozygous sites compared to the expected sequencing error fraction. In this presentation, we propose a probabilistic method to infer contaminated samples and their contaminant for low levels of contamination. The distance on the well plate between the contaminant and the contaminated sample may be penalized. The method is implemented in a free of charge R package. We compare it with the three alternative methods ART-DeCo, VerifyBamID2 and the built-in function in Illumina's DRAGEN platform and demonstrate its accuracy on simulated data. We illustrate the method using real data from the pilot phase of a large-scale NGS experiment with 9000 whole genome sequences. In the real data, our method was able to successfully identify cross-contamination. Sample cross-contamination in NGS studies can be identified using a simple-to-use R package.

# SESSION OC1D

## OC1D-5 Evaluating DNA sequencing performance: concordance-discordance model and latent class model

**Yue Zhai[1,2], Anthony Herzig[3], Claire Bardel[4], Damien Sanlaville[5,6], Emanuelle Genin[3,7], Pascal Roy[1,2,8]**
1 Université Claude Bernard Lyon 1, Lyon, France
2 Laboratoire de Biométrie et Biologie Évolutive, CNRS UMR 5558, Villeurbanne, France
3 Inserm, Univ Brest, EFS, UMR1078, GGB, Brest, France
4 Cellule Bioinformatique de la Plateforme NGS, Hospices Civils de Lyon, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR5558, Lyon 1 University, Bron, France
5 Service de Génétique, Hospices Civils de Lyon, Bron, France
6 INSERM U1028, CNRS UMR5292, UCBL1, GENDEV Team, Neurosciences Research Center of Lyon, Bron, France
7 CHU Brest, France
8 Service de Biostatistique Bioinformatique, Pôle Santé Publique, Hospices Civils de Lyon, France

Although the use of next generation sequencing (NGS) has been rapidly growing, evaluation of the performance of NGS often encounters difficulties, such as lacking gold standard, biased gold standard, especially when concerning real individual's sequencing data. To evaluate the accuracy of NGS sequencing and reduce error, researchers often use technical replicates, biological replicates or results from multiple pipelines. The concordance rates between these replicates are also used as an indicator of sequencing accuracy. However, the appropriateness of these substitute criteria has rarely been questioned. This study aimed to analyze whether the concordance rate is an adequate criterion for evaluating DNA sequencing performance, as well as the modelization of association between covariates and the error rates.
Two approaches were compared to estimate error rate given DNA characteristics and other covariates. The appropriateness of using concordance/discordance criteria as an alternative in the absence of gold standard was firstly analyzed. Situations with different values of sensitivity, specificity of NGS as well as different prevalences of variants were studied. The contribution of latent class models was then investigated, the true status of base-pairs being the latent variable.
Finally, the clinical contribution of these two approaches for medical practice was analyzed according to the clinical context of DNA sequencing prescription.

# SESSION OC1E

**SESSION OC1E** Multiple testing and randomization tests

## OC1E-1 Confidence intervals for the treatment effect in the Magnusson-Turnbull adaptive enrichment design

**Jinyu Zhu, Andrew Titman, Fang Wan**
Mathematics and Statistics Department, Lancaster University, United Kingdom

Establishing treatment efficacy in patient subgroups presents statistical challenges due to multiplicity and the potential for small samples which could lead to misleading conclusions. Thus, adaptive enrichment group sequential designs have been proposed for phase II/III trials. Such designs focus on ensuring maximum power to detect a treatment effect in either the whole population or a selected subgroup. However, the adaptive nature of the procedure makes quantification of uncertainty in treatment effects difficult. We therefore consider the problem of constructing individual and simultaneous confidence intervals for the treatment effects for subgroups at the termination of an adaptive enrichment trial. Focusing on the two-stage version of the enrichment designs function proposed by Magnusson and Turnbull (2013), our approach involves devising a suitable p-value function for the combined statistics based on space ordering methods. By inverting the relevant p-value function, one-sided individual confidence intervals with exact coverage for either the selected group or an individual subgroup. The construction of simultaneous confidence intervals for every group in the trial, using either a simple Bonferroni approach or the weighted Bonferroni method of Brannath and Schmidt (2014), is also explored. Future research should focus on the extension to other adaptive enrichment designs.

**References:** *Brannath, W., & Schmidt, S. (2014). A new class of powerful and informative simultaneous confidence intervals. Statistics in medicine, 33(19), 3365-3386.*
*Magnusson, B. P., & Turnbull, B. W. (2013). Group sequential enrichment design incorporating subgroup selection. Statistics in medicine, 32(16), 2695-2714.*

## OC1E-2 Graphical approaches for the control of generalized error rates

**Frank Bretz[1], David Robertson[2], James Wason[3]**
1 Novartis, Basel Switzerland
2 University of Cambridge, United Kingdom
3 Newcastle University, United Kingdom

When simultaneously testing multiple hypotheses, the usual approach in the context of confirmatory clinical trials is to control the familywise error rate (FWER), which bounds the probability of making at least one false rejection. In many trial settings, these hypotheses will additionally have a hierarchical structure that reflects the relative importance and links between different clinical objectives. The graphical approach of Bretz et al (2009) is a flexible and easily communicable way of controlling the FWER while respecting complex trial objectives and multiple structured hypotheses. However, the FWER can be a very stringent criterion that leads to procedures with low power, and may not be appropriate in exploratory trial settings. This motivates controlling generalized error rates, particularly when the number of hypotheses tested is no longer small. We consider the generalized familywise error rate (k-FWER), which is the probability of making k or more false rejections, as well as the tail probability of the false discovery proportion (FDP), which is the probability that the proportion of false rejections is greater than some threshold. We also consider asymptotic control of the false discovery rate, which is the expectation of the FDP. In this presentation, we show how to control these generalized error rates when using the graphical approach and its extensions. We demonstrate the utility of the resulting graphical procedures on clinical trial case studies.

# SESSION OC1E

## OC1E-3 Evaluation of the Fill-it-up design to combine data from observational trials and RCTs

**Stephanie Wied[1], Nicole Heussen[1,2], Ralf-Dieter Hilgers[1]**
1 Department of Medical Statistics - RWTH Aachen University, Germany
2 Center for Biostatistics and Epidemiology, Medical School – Sigmund Freud Private University, Vienna, Austria

**Context:** The most appropriate method to investigate the effects of interventions in clinical research is to conduct a randomised controlled clinical trial (RCT). However, conducting an RCT is often challenging. In particular, multicentre and multinational trials face some constraints for clinical researchers. The EPISTOP trial was conducted to compare conventional and preventive therapy for epilepsy in children with tuberous sclerosis complex [1]. In some centres, however, permission to conduct an RCT was not granted by the ethics board due to different guidelines for conducting clinical trials in children, resulting in observational data on the one hand and randomised data on the other. The statistical analysis and combination of the randomised and observational data poses a great challenge.
**Objectives:** The aim is to extent the Fill-it-up design for the combination of observational data and data from RCTs. To avoid biased estimates of the treatment effect, observational data should be similar to randomised data to a reasonable extent and with a certain confidence.
**Method:** The combination of observational and randomised data should only be considered if their equivalence is confirmed in an equivalence pretest. We therefore propose to pause the originally planned trial when a certain sample size of all study arms is reached in order to conduct the equivalence pretest. If equivalence is confirmed, the observational and randomised data will be pooled and no further recruitment is carried out. If equivalence cannot be confirmed, the observational data will be described separately and the final statistical analysis will be performed based on the randomised data only. For this purpose, the recruitment of the original study is continued. We investigate the performance of this study design in terms of adherence to the familywise error rate and overall power.
**Results:** We show how the significance levels of the separate tests need to be adjusted to maintain the overall type-I-error probability and overall power of our design within acceptable limits while reducing the total randomised sample size in case of equivalence.
**References:** *[1] Kotulska, K. et al.(2020), Prevention of Epilepsy in Infants with Tuberous Sclerosis Complex in the EPISTOP Trial. Ann Neurol, 89: 304-314. https://doi.org/10.1002/ana.25956*

## OC1E-4 Improved group sequential Holm procedures for testing multiple correlated hypotheses over time

**Fredrik Öhrn, Julia Niewczas, Carl-Fredrik Burman**
Early Biometrics and Statistical Innovation, Data Science and Artificial Intelligence, R&D, AstraZeneca, Gothenburg, Sweden

Clinical trials can typically feature two different types of multiple inference: testing of more than one null hypothesis and testing at multiple time points. These modes of multiplicity are closely related mathematically but distinct statistically and philosophically. Regulatory agencies require strong control of the family-wise error rate (FWER), the risk of falsely rejecting any null hypothesis at any analysis. The correlations between test statistics at interim analyses and the final analysis are therefore routinely used in group sequential designs to achieve less conservative critical values. However, the same type of correlations between different comparisons, endpoints or sub-populations are less commonly used. As a result, FWER is in practice often controlled conservatively for commonly applied procedures.
Repeated testing of the same null hypothesis may give changing results, when an efficacy boundary is crossed at an interim but not at the final analysis. The mathematically correct overall rejection is at odds with an inference theoretic approach and with common sense. We discuss these two issues, of incorporating correlations and how to interpret time-changing conclusions, and provide case studies where power can be increased while adhering to sound statistical principles.



# SESSION OC1E

## OC1E-5 Randomization tests to address disruptions in clinical trials

**Diane Uschner**
The Biostatistics Center & Department of Biostatistics and Bioinformatics, George Washington University, Washington DC, United States

**Background:** In early 2020, the World Health Organization declared the novel corona virus disease (COVID-19) a pandemic. On top of prompting various trials to study treatments and vaccines for COVID-19, COVID-19 also had numerous consequences for ongoing clinical trials. People around the globe restricted their daily activities to minimize contagion, which led to missed visits and cancelling or postponing of elective medical treatments. For some clinical indications, COVID-19 may lead to a change in the patient population or treatment effect heterogeneity.
**Methods:** We present three models for clinical trial disruptions. The first model will account for the change in patient population based on chronological bias. The other will model the disruption based on the assessment of an early biomarker that is correlated with the primary outcome. The third model will account for missed visits. We will measure the effect of the disruption on randomization tests. Randomization tests are a non-parametric, design-based method of inference. We derive a methodological framework for randomization tests that allows for the assessment of clinical trial disruptions, and we will conduct a simulation study to assess the impact of disruptions on type I error probability and power in practice. Finally, we will illustrate the results with a simulation study and a case study based on a clinical trial that was interrupted by COVID-19.
**Results:** We show that randomization tests are robust against clinical trial disruptions in certain scenarios, namely if the disruption can be considered an ancillary statistic to the treatment effect. As a consequence, randomization tests maintain type I error probability and power at their nominal levels.
**Conclusions:** Randomization tests can provide a useful sensitivity analysis in clinical trials that are affected by clinical trial disruptions.

# SESSION OC1F

## SESSION OC1F — Methods for survival analysis

### OC1F-1 Testing for ignorable sampling bias under random double truncation

**Jacobo de Uña-Álvarez**
Department of Statistics and OR & Biomedical Research Center (CINBIO), University of Vigo, Spain

The issue of random double truncation is ubiquitous in clinical and epidemiological research. It occurs, for instance, in Survival Analysis with interval sampling, when the observation of the target lifetime is restricted to events within two given calendar dates (Zhu and Wang, 2014). Unlike one-sided truncation, which always shifts the target distribution in the observable world, double truncation may result in an ignorable sampling bias. This is because the left and right truncation variables may balance each other, in the sense of giving the lifetimes the same chances of observation regardless their size. In such a case, the ordinary empirical distribution function is consistent, and no correction for double truncation is needed. This is interesting, since the recovery of a distribution function from doubly truncated data typically suffers from a large variance. See de Uña-Álvarez and Van Keilegom (2021) for nonparametric maximum likelihood estimation (NPMLE) with doubly truncated data.
Even when testing for ignorable sampling bias from doubly truncated data is highly relevant, no methods for this goal have been proposed so far. In this work I introduce a formal, omnibus test for the null hypothesis that the double truncation induces no observational bias on the target. The test is based on a distance between the NPMLEs of the sampling probability function under the null and under the alternative. The asymptotic null distribution of the test is derived and the consistency of the test is proven. A bootstrap resampling procedure to approximate the null distribution of the test in practice is introduced. This is not immediate, however, since the null hypothesis does not characterize the distribution of the truncation couple and, thus, Wang's obvious bootstrap for truncated data is not applicable. The finite sample performance of the test is investigated through simulations. The method is applied to two different studies with interval sampling: (a) age at diagnosis for childhood cancer, and (b) age at diagnosis for Parkinson's disease. While no relevant sampling bias is found in the first study, the hypothesis of ignorable sampling bias is largely rejected for the Parkinson's disease data. Practical recommendations are given.
**References:** *[1] de Uña-Álvarez J, Van Keilegom I (2021). Efron-Petrosian integrals for doubly truncated data with covariates: an asymptotic analysis. Bernoulli 27, 249-273. [2] Zhu H and Wang MC (2014) Nonparametric inference on bivariate survival data with interval sampling: Association estimation and testing. Biometrika 101, 519-533.*

### OC1F-2 A martingale based approach for modelling the alternating recurrences in Cystic Fibrosis patients ➜ *CFCD AWARD WINNER*

**Moumita Chatterjee,** Department of Mathematics and Statistics, Aliah University, Kolkata, India

**Background:** Alternating recurrent events are often evidenced in varied clinical experiments (Kalbfleisch and Prentice, 2002). The history of the analysis of alternating recurrent events has evidenced several techniques. However, in most of the cases the association between the two alternating recurrences are considered using marginal methods (Yan and Fine, 2008; Sen Roy and Chatterjee, 2015 and many more). Some sparse attempts were made to address the issue of incorporating the association (Li et al., 2010; Chatterjee and Sen Roy, 2018).
**Objective:** In the present article, we will use martingale based methods to take care of the dependence between the two recurrences for the above types of data. Using this method, the present article aims to analyze the Cumulative Mean Functions (CMF) for the alternating recurrent events.
**Method:** Nelson and Doganaksoy, 1989 illustrates a method of estimating the CMF's for recurrent events from the processes which are identically distributed. This was later generalized by Lawless and Nadeau, 1995 by incorporating regressors. Nonparametric estimation methods for CMFs were taken care of by them. However, only a few attempts were made where the CMF's were modeled for alternating recurrent episodes. In the present article, we will consider the counting process approach and the idea of combining the two martingales (Fleming and Harrington, 1991) will be used to bring in the association between the two recurrent episodes. Also, this will help us to extend the existing log rank test for the alternating recurrent events, using the combination of the martingales.
**Potential results & Conclusion:** For illustration purpose, a data (as reported by Fuchs et al., 1994) on the disease cystic fibrosis will be considered. Patients suffering from this disease often alternatively face recurrent episodes of illness and cure. The study reports time to relapse and time to cure of 647 patients being treated either to placebo (325) or to rhDNase (322). The data had previously been analyzed by many (Yan and Fine, 200; Chatterjee and Sen Roy, 2018, 2020 and many more). Our newly proposed distribution free method involving the mean functions can potentially give a new direction towards finding out the nature and dependence between the cure and ill events.



# SESSION OC1F

### OC1F-3 Impact of measurement error in time-varying prescription-based drug exposures in time-to-event analysis

**Steve Ferreira Guerra, Robert Platt, Michal Abrahamowicz**
McGill University, Montreal QC, Canada

Administrative databases have grown in popularity as sources of data for pharmaco-epidemiological studies since these often include drug prescription registries used to determine individual drug exposures. Yet, an inherent problem is that a filled prescription does not guarantee actual medication intake. A consequence of this non-adherence to the prescribed treatment is that true episodes of drug exposure potentially diverge from those reconstructed based on the history of filled prescriptions recorded in the study database. It is well known from the statistical measurement error (ME) literature that analyses using error-prone exposures may result in biased estimates of their associations with the outcome, and incorrect inference. Despite this, ME in prescription-based exposures is often overlooked and, even when researchers are aware of the presence of ME, very few studies aim to correct it. We aim to illustrate the impact of ME due to prescription non-adherence on estimated drug associations, with specific focus on the estimation of complex treatment effects in time-to-event analyses with time-varying exposures.
Specifically, we will investigate through simulation studies the effects of exposure ME on estimated hazard ratios (HR) from the Cox model. The first simulation study will analyze Berkson ME in (i) time-fixed and (ii) time-varying exposures considering both linear and non-linear effects of exposure on the log hazard. Then, to evaluate the effects of prescription non-adherence, a plasmode simulation study will be conducted in which the observed prescription histories will be resampled from a real pharmaco-epidemiological dataset and then 'true' unobserved time-varying exposure patterns will be simulated under various assumed patterns of non-adherence. Finally, biases caused by existing ad hoc methods to reconstruct drug exposure episodes (e.g. filling gaps) will also be investigated. Based on preliminary results, we expect to show non-negligeable bias of the estimated HR and distortion of the estimated non-linear effect of exposure on the log-hazard.
In conclusion, we aim to show that exposure ME due to assumed exposures based on prescription registries is a non-negligeable source of bias that should not be overlooked in time-to-event analyses using time-varying exposures and thus bridge the gap to a more warranted implementation of ME correction methods.

### OC1F-4 Developing a Goodness of fit test for a joint model: The case of clustered survival and count data

**Marina Roshini Sooriyarachchi, Nayani Adikari**
Department of Statistics, University of Colombo, Sri Lanka

In Epidemiology, infectious viral diseases such as dengue, encephalitis, hepatitis to name a few often result in responses on the survival and count of patients. Many studies have found these two responses to be correlated and thus joint models are recommended over univariate models. Often this scenario is more complex with data clustered within geographical units. A method that models the responses using a GLMM with a single random effect for the geographical units is considered. This model assumes exchangeability structure for the covariance matrix. The objective of this research was to develop a goodness of fit test, examine its Type one error and power and to apply this GOF test to infectious disease data clustered within geographical units. For the survival data a Discrete Time Hazard Model (DTHM) is used and a Negative Binomial (NB) model is used for the count data. These two responses were jointly modeled using a GLMM model. Once the joint model was fitted, the Hosmer-Lemeshow test was generalized to this scenario in order to group the expected values and get indicator variables. Under the null hypothesis of a well-fitting model the coefficients of the indicator variables have to be simultaneously equal to zero. A detailed simulation study has been used to examine the type one error and power of this test. A variety of sample sizes have been examined together with several intra-cluster correlation (ICC) values. A real data set on the infectious disease Dengue from Sri Lanka is used to illustrate the theory. The important conclusions from the study are that that the test maintains type one error for almost all the scenarios considered. The power of the test is good for at least moderate sample sizes (around 500 patients). The cluster effect was taken to be district and the model is shown to fit well. Theoretically, the importance of this research is that it fills a gap in the literature as there is no known GOF test for this situation. Practically there are many applications that can make use of this development particularly in Epidemiology and other areas of Medicine.

# SESSION OC1F

## OC1F-5 Confidence bands of the MRL function in some right censored prevalent cohort studies via the empirical likelihood

**Ali Shariati, Hassan Doosti, Justin Wishart**
Department of Mathematics and Statistics, Macquarie University, Sydney, Australia

Survival data collected in a cohort of prevalent cases may be used to draw statistical inference on the natural progress of a disease. Since non-random sampling of subjects is involved, the data collected in this sampling scheme are biased. The most common case for modelling this bias is known as length-bias. It assumes the initiating event follows a stationary Poisson process. It is also often necessary to take into account the inability to follow-up some subjects, i.e. the presence of censored data. Length-bias is modelled in the context of survival data described above but can be seen in other sampling schemes.
Life expectancy is a key concept in survival analysis. In this talk, an empirical likelihood-based procedure is proposed to obtain a simultaneous confidence band for the mean residual life (MRL) function of the unbiased survival time of interest where the available sample includes right censored length-biased data. The empirical log-likelihood is revealed to converge weakly to a mean zero Gaussian process. As a direct result, it is also shown that the weak convergence implies that the limiting distribution of the log-likelihood ratio statistic is chi-square. These asymptotic results are employed to obtain confidence bands and intervals for the MRL function. The finite sample performance of the proposed method is inspected through a simulation study. The proposed method is illustrated by modelling the MRL and the respective confidence bands/intervals for elderly residents of a retirement centre.

# SESSION OC1G

SESSION OC1G — Mixed effects model

## OC1E-1 Bayesian multi-response nonlinear mixed-effect model: application of two recent HIV infection biomarkers

**Charlotte Castel[1,2], Cécile Sommen[1], Edouard Chatignoux[1], Yann Le Strat[1], Ahmadou Alioum[3,4]**
1 French Institute for Public Health Surveillance, Data science division, Saint-Maurice, France
2 University of Paris-Est, Champs-sur-Marne, France
3 Inserm Center U1219- Bordeaux Population Health, Biostatistical Team, Bordeaux, France
4 University of Bordeaux, ISPED, Inserm Center U1219- Bordeaux Population Health, Bordeaux, France

Since the discovery of the human immunodeficiency virus (HIV) 35 years ago, the epidemic is still ongoing in France. To monitor the dynamics of HIV transmission and assess the impact of prevention campaigns, the main indicator is the incidence. One method to estimate HIV incidence is based on the knowledge of the dynamics of two recent HIV infection biomarkers. Estimating HIV incidence from biomarkers first requires modeling their dynamics since infection using external longitudinal data.
The main objective was to jointly model the dynamics of two recent HIV infection biomarkers using a nonlinear mixed-effect model. We considered one random effect for each biomarker and a correlation of random effects to take into account the correlation of the biomarkers. Parameters were estimated using the Hamiltonian Monte Carlo method.
This procedure was first applied to the real data of the PRIMO cohort. This cohort recruited 298 volunteers infected with primary HIV infection. The patients were examined biologically at inclusion, at 1, 3 and 6 months, then every 6 months. At each visit, the concentration of the two antibodies was measured. We also simulated 200 datasets closed to the PRIMO cohort data.
The goodness of fit of our model was assessed by comparing the observed individual trajectories and those predicted by the model using real data. The Bayesian estimation procedure was validated through a simulation study using conventional indicators (bias, coverage rate and RMSE).
The results on the real data and on the simulation study indicate that the proposed approach gives satisfactory results for the estimation of the dynamics of the two biomarkers. For the real data, the predicted trajectories are closed of observed trajectories for all individuals. For the simulation study the absolute relative bias is between 0% and 8%, RMSEs are between 0.11 and 2.1, and the coverage rate is between 90% and 98%.
To our knowledge, this work is the first attempt to jointly study the dynamics of two biomarkers in a Bayesian nonlinear mixed-effect model. This modeling can potentially be used to estimate HIV incidence from HIV diagnoses surveillance data with values of markers at diagnosis.

# SESSION OC1G

## OC1G-2 Confidence, Prediction, Tolerance Intervals in Linear Mixed Models: Applications in (Non)-Clinical Trials

**Bernard G. Francq[1], Dan Lin[2]**
1 Technical R&D - CMC Statistical Sciences, GSK, Rixensart, Belgium
2 Pre-Clinical & Research - Biostatistics and Statistical Programming, GSK, Rixensart, Belgium

Confidence intervals (CI) have been widely accepted and used in the medical literature. However, such intervals mainly focus on uncertainty of "average" effects. In practice, it is often useful to predict the primary outcome for future patients, and to predict the ranges where most of patients will lie. This is where prediction intervals (PI) and tolerance intervals (TI) can provide much more information to medical researchers and to the patients. These intervals focus on future patients with an exchangeable interpretation under both frequentist or Bayesian paradigms. The literature about PI and TI in linear mixed models is usually developed for continuous response variables with some specific designs, which is a main limitation to their use.
We propose a formula of the two-sided PI which is generalizable under a wide variety of designs in mixed models (one random factor; nested and crossed designs with multiple random factors; balanced or unbalanced designs). Construction of two-sided TI are also detailed by using the expected mean squares for random effects. Computation of prediction and tolerance intervals can also be performed under the clinical trial continuous endpoint with repeated measures.
A simulation study is carried out to compare the widths and coverage probabilities of CI, PI and TI, to their nominal levels. Finally, these intervals are applied to real datasets from orthopedic surgery study (intralesional resection risk). While marginal prediction and tolerance intervals are not implemented in most of the statistical software, it will be shown how to calculate and interpret these intervals.

**References:** *[1] Francq B, Lin D, Hoyer W. Confidence, Prediction and Tolerance in Linear Mixed Models. Statistics in Medicine 2019; 38: 5603-5622. [2] Francq B, Berger M, Boachie C. To Tolerate or To Agree: A Tutorial on Tolerance Intervals in Method Comparison Studies With BivRegBLS R Package. Statistics in Medicine 2020; 39:4334–4349.*

## OC1G-3 Bayesian multivariate longitudinal data analysis assuming different association structures

**Aglina Lika, Dimitris Rizopoulos, Michelle Kruijshaar, Ans Van der Ploeg, Elrozy Andrinopoulou**
Erasmus Medical Center, Rotterdam, Netherlands

Studies in life course epidemiology often involve different types of outcomes being collected on individuals, who are followed over time. These outcomes are mainly analysed separately, although it may be of scientific interest to study their associations. To model the correlation of multiple longitudinal outcomes, it is common to assume a multivariate normal distribution for the corresponding random effects. This approach, however, has its limitations in terms of interpreting the strength of association between the outcomes. To overcome this, we can include several longitudinal outcomes, as time-dependent covariates, in the model of the main longitudinal outcome. Another advantage is that several features of these longitudinal predictors could be used, namely the current value, the slope, and the area under the curve. We propose a multivariate mixed model that incorporates different functional forms to link multiple outcomes assuming the Bayesian framework. This approach was motivated by a dataset of Dutch adult patients with Late-onset Pompe disease. Late-onset Pompe disease is an autosomal recessive metabolic disorder and is characterized by progressive muscle weakness and loss of respiratory function. This disease is caused by an accumulation of glycogen in the lysosome due to deficiency of the lysosomal acid alpha-glucosidase enzyme. For these patients we have physical and patient-report outcomes. Clinicians are interested in investigating how the physical outcomes (e.g., Force vital capacity (FVC)) are associated with the patient-reported outcomes over time.



SESSION OC1G

## OC1G-4 The use of mixed logistic modelling in the analysis of HIV latency study data in the context of low sample size and low outcome rates

**Hannah J. Morgan[1], Bethany A. Horsburgh[2,3], Bonnie Hiener[2,3], Sarah Palmer[2,3], Timothy E. Schlub[1]**
1 Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Australia
2 Centre for Virus Research, The Westmead Institute of Medical Research, The University of Sydney, Australia
3 Sydney Medical School, Westmead Clinical School, Faculty of Medicine and Health, The University of Sydney, Australia

**Background:** The goal of Human Immunodeficiency Virus (HIV) research is to find a cure. Although treatment greatly reduces viral levels to clinically undetectable levels, a latent reservoir of virus lies dormant during treatment and is reactivated if treatment is stopped. The low numbers of latent virus make studying and measuring this reservoir difficult and intrusive. Studies on HIV latency are therefore characterised by few participants with as much data extracted from biological samples as possible. This data is frequently analysed with mixed logistic regressions, however the low participant numbers, low rates of virus, and high across participant variability presents a number of challenges in using a mixed logistic model.
**Method:** This project investigates the use of mixed logistic models in this context, when the sample size at participant level and rate of outcome are very low, and through simulation outlines the effects that this has on the accuracy of parameter and variance estimates, and their associated standard error. The impact of sample size on power at low outcome rates is explored, in combination with varying random effect sizes, reflecting the nature of the data seen in HIV latency studies.
**Results:** We find that model performance is adversely affected when infection frequencies are below 1 cell per million in combination with low sample sizes (10 participants or less) and large variance components. Simulation demonstrates failure in model convergence, increased variability in simulated parameter estimates and inadequate power achieved.
**Conclusions:** The low infection frequencies used in simulations in this study are seen in latency studies when analysing intact HIV virus in participants on anti-retroviral therapy. When applying the results of this study to latency data, issues with model performance indicate that when participant numbers are less than 10 in combination with such low outcome rates, other methods should be explored for analysing such data. Considering this, we also use the developed simulations to estimate sample size requirements for future HIV latency studies to guide future study design.

## OC1G-5 Mixed Modeling of Regional Infant Mortality Data over Twenty Years in North Rhine-Westphalia

**Noor Jahan Akter[1], Dankmar Böhning[2]**
1 Institute of Statistical Research and Training, University of Dhaka, Dhaka, Bangladesh
2 University of Southampton, Highfield Campus, Southampton, United Kingdom

Infant mortality is one of the major indicators of quality of health and health provision. In North Rhine-Westphalia (NRW), as in other parts of Germany, infant mortality is declining. However, regions vary in the way they decline with respect to infant mortality. Hence, it is important to choose a model that accounts for the variability of infant mortality in each region to draw a valid conclusion. In this paper, we have built an appropriate statistical model using random coefficients approach and exemplified how the proposed model can be beneficially used for exploring trends in administratively regionally aggregated data of infant mortality in NRW from the year 1988 to 2010. We have measured the region-specific character by calculating the area under the infant mortality curves to identify which regions of NRW had higher infant mortality and which had lower in the specified duration. Our analysis reveals that the regions Gelsenkirchen, Mönchengladbach, Duisburg, Oberhausen, Krefeld, Oberbergischer Kreis, Hagen, Siegen-Wittgenstein had higher infant mortality, whereas Höxter, Solingen, Rheinisch-Bergischer Kreis, Rhein-Sieg-Kreis, Münster, Warendorf, Gütersloh, Herford had lower infant mortality. Moreover, we have found that the rural areas of NRW had less infant mortality compared to the urban areas.

# SESSION OC2A

## SESSION OC2A — Mediation analysis

### OC2A-1 Nonlinear mediation analysis with high-dimensional mediators whose causal structure is unknown

**Wen Wei Loh[1], Beatrijs Moerkerke[1], Tom Loeys[1], Stijn Vansteelandt[2,3]**
1 Department of Data Analysis, Ghent University, Ghent, Belgium
2 Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium
3 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom

**Clinical research problem and statistical challenges:** In settings that involve multiple possible mediators, fine-grained decompositions of path-specific effects are only valid under stringent assumptions. The assumptions are violated when – as often – the causal structure among the mediators is unknown, or there is unobserved confounding among the mediators. For example, the effect of a microRNA expression on the three-month mortality of brain cancer patients may be potentially mediated by a set of gene expression values whose internal causal structure is unknown. In contrast, interventional indirect effects for multiple mediators can be identified under much weaker conditions, while providing scientifically relevant causal interpretations. Nonetheless, current estimation approaches require (correctly) specifying a model for the joint mediator distribution, which can be difficult in practice when there is a high-dimensional set of possibly continuous and noncontinuous mediators, such as in gene regulation networks or protein signalling networks.
**Objective:** In this article, we develop a definition of interventional effects that avoids the need to model the mediators' joint distribution, by building on a previous suggestion for longitudinal mediation [1]. We exploit the definitions to propose a novel estimation strategy that uses nonparametric estimates of the (counterfactual) mediator distributions. Noncontinuous outcomes can be accommodated using nonlinear outcome models.
**Statistical Methods:** Estimation proceeds via Monte Carlo integration and requires an outcome model. When treatment is non-randomly assigned, a model for the treatment given the observed baseline covariates is required. No models for the mediators' (joint) distribution are required.
**Results:** The validity of the methods are examined via simulation studies. The method is illustrated using a reanalysis of a high-dimensional mediation analysis by Huang and Pan [2]. Huang and Pan determined that the effect of a microRNA expression on mortality within three months among 490 patients suffering from a malignant brain tumor was mediated jointly via each of nine subsets of 1220 gene expression values in the tumor genome.
**Conclusion:** Analysis of the brain cancer study using the proposed interventional indirect effects detected the individual genes (within each of the nine subsets) whose expression value mediated the effect of the microRNA expression on three-month mortality.
**References:** *[1] VanderWeele, T. J. and Tchetgen Tchetgen, E. J. (2017). Mediation analysis with time varying exposures and mediators. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79, 917-938.*
*[2] Huang, Y.-T. and Pan, W.-C. (2016). Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. Biometrics 72, 402-413.*



SESSION OC2A

### OC2A-2 Separable Causal Effects as alternative Estimands in Epidemiology and Biostatistics

**Vanessa Didelez[1,2]**
1 Department of Biometry and Data Management, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany
2 Department of Mathematics and Computer Science, University of Bremen, Germany

Separable effects have been proposed as an alternative approach to causal mediation analyses especially with view to time-dependent settings. The basic idea refers to a way of elaborating our causal model in order to better motivate a mediational research question. Specifically, separable effects are concerned with situations where the treatment (or exposure) can be decomposed into two or more components that could (possibly hypothetically) be intervened upon independently and thus separately 'activate' different causal pathways. Formulating time-dependent mediational research questions in this way is appealing as it sidesteps 'cross-world' notions and assumptions involved in the analysis of natural (in)direct effects based on nested counterfactuals. Crucially, separable effects turn out to be especially useful in time-dependent setting, e.g. when the outcome is survival and mediation is via a process (Didelez, 2019). Moreover, they lead to a novel estimand in competing events settings, where the separable direct effect on the event of interest is the effect of the treatment component that only activates causal paths avoiding the competing event (Stensrud et al., 2020).
The presentation will discuss the practical use of separable effects, and compared it in detail to more traditional approaches. Their interpretation, structural assumptions and estimation will be illustrated in the context of three examples typical for epidemiological and bio-statistical applications: (i) In an application of causal mediation analysis we aim at investigating how an exposure might affect cognitive decline in the elderly – these are processes taking place over time which will be addressed by the statistical analysis. (ii) Typically, survival analyses in cancer research are faced with competing risks, e.g. death due to other reasons, which no approach can ignore; we discuss the separable effect estimand and methods of analysis for such settings. (iii) Finally we consider the analysis of an RCTs that is affected by an undesirable intercurrent event, such as an adverse drug reaction; here, we show that separable effects specifically address research questions relevant to future drug development as they explicitly represent effects of modified treatments. This is based on joint work with Mats Stensrud.
**References:** *[1] Didelez (2019). Defining causal mediation with a longitudinal mediator and a survival outcome. Lifetime Data Analysis 25, 593 - 610. [2] Stensrud, Young, Didelez, Robins & Hernán (2020) Separable effects for causal inference in the presence of competing events, JASA (online).*

### OC2A-3 Simulating hypothetical interventions on multiple mediators: Extending methods and practical guidance

**Margarita Moreno-Betancur[1,2], John B Carlin[1,2]**
1 Clinical Epidemiology and Biostatistics Unit, Department of Paediatrics, University of Melbourne, Australia
2 Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Melbourne, Australia

Many epidemiological questions concern the pathways presumed to mediate an association, particularly in life course and social epidemiology. Invariably, the translational intent of such research questions is to inform potential intervention targets, but until recently causal mediation analysis methods did not define mediation effects in a way that acknowledged this interventional intent. In recent work [1,2] we have proposed a novel framework conceptualising mediation effects by mapping to a target randomised trial evaluating mediator interventions. This approach is particularly relevant in the context of mediators that do not correspond to well-defined interventions, a scenario where it is plausible, and indeed perhaps necessary, to consider hypothetical interventions that would shift the mediators' distributions. The approach consists in specifying a target trial relevant to the research question, regarding the impact of shifting joint mediator distributions to user-specified distributions. These estimand assumptions are distinguished from identifiability assumptions, which are needed to emulate the effects of those shifts with the observed data. Estimation is done by simulation via a g-computation approach. By its very nature, the approach is context-specific: the target effects must be tailored to each specific question and context. Drawing on learnings from its application to several longitudinal cohort studies, some of which are already published, this paper presents further developments of the method to tackle practical issues encountered in its application. This includes alternative approaches to effect definition according to the question, consideration of different types of exposures and mediators, and guidance on handling issues such missing data, clustering and reporting. Findings will assist researchers in applying the method while tackling practical issues as soundly and easily as possible.
**References:** *[1] Moreno-Betancur M, Moran P, Becker D, Patton G, Carlin J. Mediation Effects That Emulate a Target Randomised Trial: Simulation-Based Evaluation of Ill-Defined Interventions on Multiple Mediators. Statistical Methods in Medical Research (in press)*
*[2] Moreno-Betancur M. The target trial: A powerful device beyond well-defined interventions. Epidemiology. 2020;32(1):291-294.*

# SESSION OC2A

## OC2A-4 Mediation with a survival outcome and time-varying mediator: Empirical comparison and use in observational data

**Kamaryn T. Tanner[1], Linda D. Sharples[1], Rhian M. Daniel[2], Ruth H. Keogh[1]**

1 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom

2 Division of Population Medicine, Cardiff University, United Kingdom

Mediation analyses seek an understanding of the mechanisms through which an exposure affects an outcome but there are statistical challenges for survival outcomes with longitudinal covariates from observational data. We are motivated by the study of cystic fibrosis-related diabetes (CFRD), a common comorbidity of cystic fibrosis (CF). CFRD negatively affects survival but the mechanisms are not well understood. We aim to illustrate and compare two recently proposed methods for mediation analysis in this setting using observational data from the UK CF Registry. Further, the sensitivity of each method to model misspecification and data availability is examined via a simulation study. In causal mediation analyses, analytical and identification challenges arise when working with a time-dependent mediator and covariates. Survival outcomes pose another difficulty: because the exposure affects the mediator both directly and indirectly through its effect on survival time, survival is a post-treatment confounder. In the observational data, individuals have different entry times and, when the exposure is the diagnosis of a disease, there is no natural time zero. We propose a stacked analysis dataset, constructed similarly to a landmark dataset, to maximally exploit the longitudinal data. Two mediation methods are employed. Aalen et al. (Biometrical Journal 2018) described a method based on exposure splitting and an additive hazards model for quantifying indirect effects. Vansteelandt et al. (Statistics in Medicine 2019) proposed a method based on path-specific effects that accommodates time-varying confounders and flexible outcome models.
In the study population of 5,553 individuals there were 1,180 incident cases of CFRD. The indirect effect of CFRD on survival via lung function (FEV1%) increased slightly over time with the proportion mediated by FEV1% estimated at 9% [95% CI: 4%,18%] at 5 years post CFRD diagnosis by the method of Aalen and at 7% [95% CI: -14%, 18%] for the method of Vansteelandt. The simulation study suggests that both methods may suffer from bias when mediator measurements are infrequent; analogous to measurement error, the indirect effect estimates are attenuated. Also, if the assumption of no time-varying confounders is not valid, Aalen's methods may produce significant bias in effect estimates.

## OC2A-5 Mediation with Irregular Longitudinal Biomarkers: An Application to Examining Obesity and Severe Disease in COVID-19 Patients

**David Cheng, Andrea Foulkes**

Biostatistics Center, Massachusetts General Hospital, Harvard Medical School, Boston MA, United States

Obesity is an established risk factor for severe disease in patients with COVID-19. As inflammatory biomarkers, such as C-reactive protein (CRP), have been associated with both obesity and severe disease, effects of obesity on severe disease may be mediated in part through CRP. Data on laboratory values from electronic health records, which are repeatedly measured at irregular times during hospitalization, can help evaluate this hypothesis. However, common approaches to mediation analysis in causal inference generally assume that mediators are measured either at a single time point or longitudinally over regularly-spaced intervals. Though summary measures, such as the maximum over the course of follow-up, could be used, loss of information is expected from neglecting the full longitudinal trajectory. We consider an approach that accounts for an irregular longitudinal mediator through functional regression modeling when estimating natural direct and indirect effects (NDE and NIE). The longitudinal mediator is regarded to be sampled points from an underlying mediator stochastic process over time that has time-varying effects on the outcome and exposure. A functional logistic regression (Müller 2005) is used to estimate the probability of exposure given the mediator process and covariates, modeling the cumulative effects of the mediator process over time. These propensity score estimates are then applied in estimators for the NDE and NIE based on inverse-probability weighting representations of the mediation formula (Pearl 2001), assuming nonparametric identification conditions. We analyzed data from 983 patients hospitalized with COVID-19 at Massachusetts General Hospital from March through May 2020. In preliminary results, we find that there are significant population-average natural direct and indirect effects of obesity (BMI>30 kg/m2) through CRP on a binary outcome for intensive care unit admission within 28 days (NDE=0.051, 95% CI -0.002-0.103; NIE=0.043, 95% CI 0.021-0.069), with 46% of the effect mediated through longitudinal CRP measurements. In contrast, using only the maximum value of CRP over follow-up resulted in lower degree of mediated effects (NDE=0.078, 95% CI 0.023-0.136; NIE=0.017, 95% CI 0.004-0.034), with 18% of the effect mediated. These results suggest that functional modeling can be a promising approach to incorporating irregular longitudinal biomarkers in mediation analysis.

**References:** *[1] Müller, H. G. (2005). Functional modelling and classification of longitudinal data. Scandinavian Journal of Statistics, 32(2), 223-240. [2] Pearl, J. (2001). Direct and indirect effects. In Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence (UAI'01), 411-420. Morgan Kaufmann Publishers Inc.*



# SESSION OC2B

SESSION OC2B — Bayesian clinical trial design (Part B)

## OC2B-1 Early completion of phase I cancer clinical trials with Bayesian optimal interval design

**Masahiro Kojima**

Kyowa Hakko Kirin Co., Ltd., Graduate University for Advanced Studies, Tokyo, Japan

Phase I cancer clinical trials have been proposed novel designs such as algorithm-based, model-based, and model-assisted designs. Model-based and model-assisted designs have a higher identification rate of maximum tolerated dose (MTD) than algorithm-based designs, but are limited by the fact that the sample size is fixed. Hence, it would be very attractive to estimate the MTD with sufficient accuracy and complete the trial early. O'Quigley proposed the early completion of a trial with the continual reassessment method (CRM) among model-based designs when the MTD is estimated with sufficient accuracy. However, the proposed early completion method based on the binary outcome trees has a problem that the calculation cost is high when the number of remaining patients is large. Among model-assisted designs, the Bayesian optimal interval (BOIN) design provides the simplest approach for dose adjustment.
We propose the novel early completion method for the clinical trials with the BOIN design when the MTD is estimated with sufficient accuracy. This completion method can be easily calculated. In addition, the method does not require many more patients treated for the determination of early completion. We confirm that the BOIN design applying the early completion method has almost the same MTD identification rate compared to the BOIN design through simulations conducted based on over 30,000 scenarios.

## OC2B-2 Decision rules for identifying combination therapies in open-entry, randomized controlled platform trials

**Elias Laurin Meyer[1], Peter Mesenbrink[2], Cornelia Dunger-Baldauf[3], Ekkehard Glimm[3], Yuhan Li[2], Franz König[1]**

1 Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Austria

2 Novartis Pharmaceuticals Corporation, One Health Plaza, East Hanover NJ, United States

3 Novartis Pharma AG, Basel, Switzerland

The design and conduct of platform trials have become increasingly popular for drug development programs, attracting interest from statisticians, clinicians and regulatory agencies. Many statistical questions related to designing platform trials - such as what is the impact of decision rules, sharing of information across cohorts, and allocation ratios on operating characteristics and error rates – remain unanswered. In many platform trials, the definition of error rates is not straightforward as classical error rate concepts are not applicable. In particular, the strict control of the family-wise Type I error rate may not be applicable in certain settings. For an open-entry, exploratory platform trial design comparing combination therapies to the respective monotherapies and standard-of-care, we define a set of error rates and operating characteristics and then use these as a measure to compare a set of design parameters under a range of simulation assumptions. When setting up the simulations, we aimed for realistic trial trajectories, e.g. in case one compound is found to be superior to standard-of-care, it could become the new standard-of-care in future cohorts. Our results indicate that the method of data sharing, exact specification of decision rules and quality of the biomarker used to make interim decisions all strongly contribute to the operating characteristics of the platform trial. Together with the potential flexibility and complexity of a platform trial, which also impact the achieved operating characteristics, this implies that utmost care needs to be given to evaluation of different assumptions and design parameters at the design stage.

# SESSION OC2B

## OC2B-3 A Bayesian decision-theoretic approach to outcome-adaptive sequential multiple assignment randomised trials (SMARTs) with distinct intermediate binary endpoints

**Robert K. Mahar[1,2], Julie A. Simpson[1], Bibhas Chakraborty[3,4,5], Agus Salim[1,6,7]**
1 Centre for Epidemiology and Biostatistics, University of Melbourne, Australia
2 Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Melbourne, Australia
3 Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore
4 Department of Statistics and Applied Probability, National University of Singapore, Singapore
5 Department of Biostatistics and Bioinformatics, Duke University, Durham NC, United States
6 School of Mathematics and Statistics, University of Melbourne, Australia
7 Baker Department of Cardiometabolic Health, University of Melbourne, Australia

The "COVID-19 prevention and treatment in cancer: a sequential multiple assignment randomised trial (SMART)" is an innovative multi-stage design that randomises high-risk cancer patients to prophylaxis and, if they develop COVID-19, re-randomises them to an experimental treatment conditional on their disease severity (NCT04534725). Although SMARTs are excellent designs to identify personalised treatment sequences that are tailored to a patient's evolving clinical status, also known as "dynamic treatment regimens", they are typically analysed at completion so the dynamic treatment regimens are optimised only for future patients. But identifying and implementing an efficacious COVID-19 prophylaxis and treatment regimen for cancer patients is an immediate priority. Outcome-adaptive randomisation is one approach that could increase the chance that patients are randomised to the most promising treatment and also completely stop randomisation to a clearly inferior treatment, enabling rapid clinical implementation.
Routinely used outcome-adaptive randomisation algorithms do not account for potential treatment effects in the later stages of SMART designs. Such approaches may randomise patients to suboptimal regimens of prophylaxis and treatment. Methods that are typically used to optimise dynamic treatment regimens from SMART data could be used to inform the adaptive randomisation, but they are complicated, so very few guiding examples exist.
Q-learning, a statistical dynamic programming method that can be used to analyse SMART data, is one of the few algorithms that has been used to perform outcome-adaptive randomisation for a SMART [1,2]. In its simplest form, Q-learning uses stage-wise statistical models and backward induction to incorporate later-stage 'payoffs' (i.e., clinical outcomes) into early-stage 'actions' (i.e., treatments). We propose a Bayesian decision-theoretic Q-learning method to perform outcome-adaptive randomisation. This approach allows dynamic treatment regimens with distinct binary endpoints at each stage to be evaluated, a known limitation of the Q-learning method. Our simulation study, motivated by the cancer trial, aims to examine whether the Bayesian decision-theoretic Q-learning method can expedite treatment optimisation and, compared to routinely used adaptive randomisation approaches that do not consider later stages of the SMART, assign more trial participants to optimal dynamic treatment regimens.

**References:** *[1] Cheung YK, Chakraborty B, Davidson KW. Sequential multiple assignment randomized trial (SMART) with adaptive randomization for quality improvement in depression treatment program. Biometrics. 2015;71(2):450–9.*
*[2] Lee J, Thall PF, Ji Y, Müller P. A decision-theoretic phase I–II design for ordinal outcomes in two cycles. Biostatistics. 2016;17(2):304–19.*



# SESSION OC2B

## OC2B-4 Modelling time varying recruitment rates and site activation prediction in multicentre clinical trials: A comparison study

**Aris Perperoglou[1,2], Szymon Urbas[1,2], Youyi Zhang[1], Kevin Quach[1]**
1 Advanced Analytics, Data Science & Artificial Intelligence, R&D, AstraZeneca, Cambridge, United Kingdom
2 School of Mathematics, Statistics and Astrophysics, University of Newcastle, United Kingdom
3 School of Mathematics, Lancaster University, United Kingdom

Multicentre Phase II/III clinical trials are large scale operations that often include hundreds of recruiting sites (centres) in several countries. Planning of operational aspects of a clinical trial requires selection of sites and countries to adhere to study protocol and recruitment requirements. It is thus critical to accurately predict site activation and recruitment timelines, to optimize success of a trial. Such predictions occurring prior to trial initiation assist study teams with trial monitoring progress, and also assist them to take proper actions during the trial when recruitment data indicate deviations from the study plan.
In this work we showcase our experience from modelling recruitment in clinical trials sponsored by AstraZeneca between 2010-2020. We show that recruitment rates tend to vary during a trial, depending on therapeutic area and country. However, industry standard has often employed a homogeneous Poisson model (Anisimov et al 2007) which models patient recruitment rates as a time-constant function. Instead, we show how a non-homogenous Poisson modelling approach (Urbas et al 2020) is used to account for time-varying recruitment rates, and we demonstrate improved accuracy in trial prediction timelines. The latter approach utilises and ensemble of five models, four of which explicitly model time-varying recruitment rate and one assuming a homogenous process. Bayesian modelling averaging is used to combine estimations from models.
We will present a thorough descriptive analysis of our data per therapeutic area and investigate the impact of model misspecifications under the following scenarios: (a) when recruitment rates are modelled as constant but change during the trial, (b) where recruitment rates are modelled as time-varying on homogeneous data, and (c) when time-varying effect is not correctly specified by the model. Additionally, we will investigate how failure of predicted site activation and number of sites can significantly change predictions against existing study plans.

**References:** *[1] Anisimov, V. V., & Fedorov, V. V. (2007). Design of multicentre clinical trials with random enrolment. In Advances in statistical methods for the health sciences (pp. 387-400). Birkhäuser Boston.*
*[2] Urbas, Szymon, Chris Sherlock, and Paul Metcalfe. "Interim recruitment prediction for multi-centre clinical trials."Biostatistics (2020). https://doi.org/10.1093/biostatistics/kxaa036*

## OC2B-5 Compromise Bayesian test decisions under type I error rate constraint

**Silvia Calderazzo, Annette Kopp-Schneider**
Division of Biostatistics, German Cancer Research Center, Heidelberg, Germany

Bayesian clinical trials can benefit of historical information through the elicitation of informative prior distributions. Concerns are however often raised about the potential for prior-data conflict and the impact of Bayes test decisions on frequentist operating characteristics, in particular type I error rates. Indeed, power gains through incorporation of historical information are typically not possible when requiring strict conditional type I error rate control, even when historical information is dynamically discounted based on the observed degree of prior-data conflict (Kopp-Schneider et al., 2020). This observation motivates the development of principled borrowing mechanisms, which strike a balance between frequentist and Bayesian decisions. Ideally, the trust assigned to historical information defines the degree of robustness to prior-data conflict one is willing to sacrifice. However, such relationship is often not directly available when explicitly considering inflation of type I error rates.
We thus investigate a rationale for inflation of conditional type I error rate in a one-arm one-sided test situation which explicitly and linearly relates the amount of borrowing and the amount of frequentist type I error rate inflation, while satisfying Bayesian optimality criteria under the (inflated) type I error rate constraint. To this aim, we exploit the known duality between test error costs and prior probabilities (e.g. Berger, 1985). The solution is equivalent to a slight modification of the restricted Bayes solution of Hodges and Lehmann (1952), and a characterization is made in the spirit of Efron and Morris (1971), who addressed a closely related problem in the context of estimation. Connections with the robust mixture prior approach, particularly in relation to the choice of the mixture weight and robust component, are made. Simulations are performed to show the properties of the approach for normal and binomial outcomes, and extensions to two-arm and two-sided situations are discussed.

# SESSION OC2C

## SESSION OC2C — Epidemic modelling of COVID19

### OC2C-1 Dynamic predictive modelling of the first wave of the COVID-19 pandemic in Canada using a deterministic density-dependent susceptible-exposed-infected-recovered (SEIR) model that accounts for age-stratified ageing, reporting delays and mortality risks

**Ornella Wafo Noubissie**
McGill University, Montreal QC, Canada

**Background:** As the COVID-19 pandemic continues to evolve, examining its differential dynamics by demographic and behaviour patterns is crucial. However, intersectional models are limited.
**Objectives:** Develop a dynamic epidemic model for COVID-19 in Canada that accounts for the impacts of age and health capacity, to uncover related associations and trends.
**Methodology:** Using age-specific COVID-19 mortality risk estimates and sociodemographic data, I modelled COVID-19 cases with a deterministic age-stratified density-dependent Susceptible-Exposed-Infectious-Recovered (SEIR) model, that accounts for ageing, reporting and testing delays, and mortality by age group. Cases were grouped into three age strata, namely 0-39, 40-59 and 60+, according to disease-related mortality risks. Stratum-specific reporting delays were assigned based on assumptions about socio-behavioural patterns and timelines of symptom onset. The model was calibrated using line-list data from Statistics Canada and Worldometre.
**Results:** The model modestly overestimates cumulative cases but underestimates two-thirds of all deaths, and the 60+ age group is most affected. After calibration, the fit between model-predicted time-series data and surveillance data improves to near-perfection. Age-stratum-specific risks of death from COVID-19 in Canada differ from global statistics. People aged 60+ are overrepresented in COVID-19 deaths despite a relatively even distribution of cases by age stratum. The reported burden of COVID-19 is affected by reporting delays, which vary significantly by age category.
**Conclusion:** Modelling incidence and prevalence data reveals that the course of the epidemic and disease outcome vary significantly by age, and the reported burden of COVID-19 is affected by reporting delays, which also vary by age category.
**Advocacy** message: The near-perfect fit between the calibrated model and line-list surveillance data validates the pertinence of a comprehensive age-stratified SEIR model to contextualize the pandemic. The model can be applied to other contexts, and can facilitate identifying high-risk populations, super-spreaders and limitations in health capacity, which would serve to mitigate the spread of COVID-19. Nonetheless, as the pandemic and public health response continue to evolve, epidemic trends may change.



SESSION OC2C

### OC2C-2 Nowcasting CoVID-19 Deaths in England by Age and Region

**Shaun Seaman[1], Pantelis Samartsidis[1], Meaghan Kall[2], Daniela De Angelis[1,2]**
1 MRC Biostatistics Unit, University of Cambridge, United Kingdom
2 National Infection Service, Public Health England, London, United Kingdom

Understanding the trajectory of the daily numbers of deaths in people with CoVID-19 is essential to decisions on the response to the CoVID-19 pandemic. Estimating this trajectory from data on numbers of deaths is complicated by the delay between deaths occurring and their being reported to the authorities. In England, Public Health England receives death reports from a number of sources and the reporting delay is typically several days, but can be several weeks. Delayed reporting results in considerable uncertainty about the number of deaths that occurred on the most recent days. Adjusting for reporting delays is complicated by day-of-the-week reporting effects, changes over calendar time in the delay distribution, and excess variability (overdispersion) in the daily numbers of reports.
Our aim is to estimate the number of deaths on each day in each of five age strata within seven English regions. Such estimates are known as 'nowcasts'. We use a Bayesian hierarchical model that involves a submodel for the number of deaths per day and a submodel for the reporting delay distribution. This model accounts for reporting-day effects and longer-term changes over time in the delay distribution. There is also a computationally efficient way to fit the model when the delay distribution is the same in multiple strata, e.g. over a wide range of ages. In this presentation, we shall describe this model and show an example of the resulting nowcasts for England.
**Reference:** *Nowcasting CoVID-19 Deaths in England by Age and Region. Seaman SR, Samartsidis P, Kall M, De Angelis D. MedRXiv. doi: https://doi.org/10.1101/2020.09.15.20194209*

### OC2C-3 An evaluation of the stability, precision and performance of phenomenological models applied to COVID-19 cases and deaths in South Africa

**Charl Janse van Rensburg[1], Tarylee Reddy[1], Ziv Skhedy[2], Samuel Manda[1]**
1 Biostatistics Unit, South African Medical Research Council, Cape Town, South Africa
2 Censtat, Hasselt University, Belgium

**Background:** Phenomenological models present users with a purely data driven approach for modelling infectious disease data. Although these models are well established and have been successfully applied to other outbreaks, the use of these models for estimating epidemiological parameters and doing short- and longterm prediction for COVID-19 cases and deaths has been limited.
**Objectives:** We present an evaluation of models fitted to South African data by investigating fit to data, convergence, stability of model parameters over time, use of piecewise models, change in predictions as the estimation period changes and sensitivity of model to perturbations in data. Models are fitted using least squares estimation with normal as well as Poisson assumptions of underlying conditional distribution. A simple SEIR model may be used to compare important model estimates.
**Methods:** Publicly available data on COVID-19 cases and deaths in South Africa were used. Well known nonlinear phenomenological models, including the logistic, Gompertz and Richards models, were fitted using nonlinear least squares. Robust confidence intervals for model parameters were estimated using parametric bootstrap. Piecewise models were fitted where the data dictated its use. Models are fitted to daily as well as weekly data in order to improve convergence. Fit criteria are used to compare models. Provincial models may be investigated.
**Results:** The model that fitted the data most consistently was the three-parameter logistic. Greater stability in parameters were observed for models fitted to case data whereas death data was perturbed by noise affecting model fitting. The Gompertz model tended to overestimate the final size of the pandemic, whereas the Richards model underestimated it.
**Conclusion:** Phenomenological models may provide a robust way to support findings from mathematical models, as well as be a tool to introduce complex disease modelling concepts to the general public. Phenomenological models are a useful tool for short term prediction, but may provide unreliable long term predictions early in the outbreak. Further results will be presented on the use of the piecewise models, the Poisson vs Normal distributions as well as the modelling of the second wave.

# SESSION OC2C

## OC2C-4 Estimation of incubation time and latency time distribution of SARS-CoV-2: The impact of distributional assumptions

**Vera H. Arntzen[1], Marta Fiocco[1,2,3], Ronald B. Geskus[4,5]**
1 Mathematical Institute, Leiden University, Netherlands
2 Statistics, Leiden University Medical Center, Netherlands
3 Princess Maxima Center for Child Oncology, Netherlands
4 Oxford University Clinical Research Unit, Ho Chi Minh, Vietnam
5 Centre for Tropical Medicine and Global Health, University of Oxford, United Kingdom

**Background:** The distribution of incubation time (from infection to symptom onset) and latency time (from infection to start of infectiousness) are key quantities in the analysis of infectious diseases. Both guide decisions on contact tracing and quarantine policies. Typically, the event time of symptom onset is observed exactly, whereas for the time origin (infection) only an exposure interval is known. Assuming that the risk of infection is constant within that interval, data can be made single interval-censored by transforming the time scale. Although the role of pre- and asymptomatic transmission is evident, estimates of the latency time distribution are lacking. Since start of infectiousness is determined by sequential test results, data on latency time are doubly interval censored. To simplify estimation of incubation and latency time, it is common practice to use parametric distributions. The appropriateness of such distributions remains unclear, especially for the right tail of the distribution which informs quarantine policy. As in the tail there are less observations, its estimate will strongly depend on the assumed parametric distribution. Hence, we hypothesize that a nonparametric approach is more appropriate.
**Methods:** Incubation and latency time distributions are estimated from Vietnamese data. Intensive contact tracing and large-scale testing has created a unique data set. We use smoothed nonparametric maximum likelihood methods and compare them with their parametric counterparts. We also consider an alternative approach based on renewal processes (Deng et al., 2020). We compare all approaches and their sensitivity to the imposed assumptions in a simulation study.
**Impact:** Accurate estimates of the incubation time and latency time distribution are important to optimize quarantine policy in Vietnam and elsewhere. We contribute to methodological literature, by comparing (non)parametric approaches, as well as to clinical literature: estimation of latency time for SARS-CoV-2 is a novelty.
**Reference:** *Deng Y, You C, Liu Y et al. Estimation of incubation period and generation time based on observed length-biased epidemic cohort with censoring for COVID-19 outbreak in China. Biometrics 2020; DOI:10.1111/biom.13325.*

## OC2C-5 Regional estimates of reproduction numbers with application to COVID-19

**Stefan Heyder[1], Thomas Hotz[2], Tyll Krüger[3], Jan Pablo Burgard[4]**
1 Institute for Mathematics, Technische Unversität Ilmenau, Germany
2 Institute for Mathematics, Technische Unversität Ilmenau, Germany
3 Wroclaw University of Science and Technology, Poland
4 Economic and Social Statistics Department, University of Trier, Germany

**Context:** In an ongoing epidemic, public health decisions are based on real-time monitoring of the spread of the disease. To this end one often considers reproduction numbers which measure the amount of secondary cases produced by a single infectious individual. While non-pharmaceutical interventions are applied on a subnational level, estimation of reproduction numbers on this level may be difficult due to low incidences.
**Objectives:** The study aims to provide reasonable estimates of reproduction numbers on the county level during periods of low incidence.
**Methods:** We start with the well-known renewal equation and, using techniques from small-area estimation, assume county level reproduction numbers to be random with a common distribution. Under this model we use maximum-likelihood estimation to obtain estimates of reproduction numbers on both the county and national level. Both estimators are analysed by a simulation study and applied to German case data from Robert-Koch Institute with focus on a local outbreak in summer 2020.
**Results:** The simulation study shows that the estimator yields sensible estimates of both the national and county reproduction numbers. It can handle low case counts, and may be used to distinguish local outbreaks from more widespread ones. For scenarios where incidences are low it handles local outbreaks, such as the one considered in the German case data, better than previous methods.
**Conclusions:** The new estimator provides insight on the spread of an epidemic on the subnational level despite low case counts.



# SESSION OC2D

SESSION OC2D — Meta-analysis: network and other

## OC2D-1 Quantifying the robustness of primary analysis results: a case study on missing outcome data in pairwise and network meta-analysis

**Loukia M. Spineli[1], Chrysostomos Kalyvas[2], Katerina Papadimitropoulou[3,4]**
1 Midwifery Research and Education Unit, Hannover Medical School, Germany
2 Biostatistics and Research Decision Sciences, MSD Europe Inc., Brussels, Belgium
3 Clinical Epidemiology, Leiden University Medical Center, Netherlands
4 Data Science and Biometrics, Danone Nutricia Research, Utrecht, Netherlands

Conducting sensitivity analyses is an integral part of the systematic review process to explore the robustness of results derived from the primary analysis. When the primary analysis results can be sensitive to assumptions concerning a model's parameters (e.g. missingness mechanism to be missing at random), sensitivity analyses become necessary. However, what can be concluded from sensitivity analyses is not always clear. For instance, in pairwise and network meta-analysis, conducting sensitivity analyses usually boils down to examining how 'similar' the estimated treatment effects are from different re-analyses to the primary analysis or placing undue emphasis on the statistical significance. To establish objective decision rules regarding the robustness of the primary analysis results, we propose an intuitive index, which uses the whole distribution of the estimated treatment effects under the primary and alternative re-analyses. This novel index summarises the robustness of primary analysis results in a single number per treatment comparison, and it is compared to an objective threshold to infer the presence or lack of the robustness. In the case of missing outcome data, we additionally propose a graph that contrasts the primary analysis results to those of alternative scenarios about the missingness mechanism in the compared arms. When robustness is questioned according to the proposed index, the suggested graph can demystify the scenarios responsible for producing results inconsistent with the primary analysis. The proposed decision framework is immediately applicable to a broad set of sensitivity analyses in pairwise and network meta-analysis. We illustrate our framework in the context of missing outcome data in pairwise and network meta-analysis using published systematic reviews.

# SESSION OC2D

## OC2D-2 Meta-analysis of randomised trials with continuous outcomes: methods that adjust for baseline should be used

**Katerina Papadimitropoulou[1,2], Richard D. Riley[3], Olaf M. Dekkers[1], Theo Stijnen[4], Saskia le Cessie[3,4]**

1 Clinical Epidemiology, Leiden University Medical Center, Netherlands
2 Data Science, Danone Nutricia Research, Utrecht, Netherlands
3 Centre for Prognosis Research, Research Institute for Primary Care – Keele University, United Kingdom
4 Biomedical Data Sciences, Leiden University Medical Center, Netherlands

We revisit a methodological case study on methods for meta-analysis of trials reporting aggregate/summary results for continuous outcomes measured at baseline and follow-up. The Trowman method, proposed in the original methodological case study to adjust for baseline imbalance, is compared with three aggregate data (AD) meta-analytic approaches that synthesize: 1) follow-up scores; 2) change scores; or 3) ANCOVA estimates recovered from reported data; and also, to a novel individual participant data (IPD) meta-analysis approach, where we generate pseudo IPD (based on the reported AD) followed by ANCOVA. The pseudo IPD analysis provides identical estimates to the true IPD analysis given that it is a likelihood-based approach making use of the appropriate sufficient statistics for ANCOVA. The methods are demonstrated on two real datasets, the original Trowman example, where considerable imbalance at baseline occurred and on a second example of meta-analysis of obstructive sleep apnea studies with moderate baseline imbalance and substantial effect modification by the baseline measurements.

The Trowman method makes strong, unrealistic assumptions and may provide erroneous conclusions regarding within- and across-trials effects. ANCOVA methods yield more precise treatment effect estimates than standard AD approaches and are recommended for the meta-analysis of continuous outcomes, with or without baseline imbalance. The pseudo IPD approach, generating pseudo IPD and fitting standard IPD models under one- or two-stage methods, is further advocated as it can additionally investigate potential differential responses to treatment. The one-stage approach comes with a plethora of modelling options e.g., study-stratified or random study intercepts, adjustment terms and allows to adopt more clinically plausible scenarios for the within-trial residual variances in a straightforward manner. The two-stage pseudo ANCOVA approach will often give very similar results to the one-stage approach (under the same modelling assumptions) and it is easier to perform as it requires less statistical expertise. Recovering ANCOVA estimates using AD could serve as good alternative, if no effect modification is expected. To initiate a paradigm shift in undertaking more ANCOVA meta-analysis in the future we developed an interactive tool using R Shiny, where the user is guided in algebraic data calculations followed by appropriate analysis.

**References:** *Trowman, R, J. C Dumville, D. J Torgeson, and G. & Cranny. 2007. "The impact of trial baseline imbalances should be considered in systematic reviews: a methodological case study." Journal of Clinical Epidemiology 60 (12): 1229-1233.*



# SESSION OC2D

## OC2D-3 Flexible generic framework for evidence synthesis in health technology assessment

**Tasnim Hamza[1], Fabio Pellegrini[2], Jens Kuhle[3,4], Pascal Benkert[4], Suvitha Subramaniam[4], Sabine Schaedelin[5], Cynthia Iglesias[6], Andrea Manca[7], Konstantina Chalkou[1], Georgia Salanti[1]**

1 Institute of Social and Preventive Medicine, University of Bern, Switzerland
2 Biogen International GmbH, Baar, Switzerland
3 Departments of Medicine, Biomedicine, and Clinical Research, University of Basel, Switzerland
4 Department of Clinical Research, University of Basel, Switzerland
5 Department of Clinical Research, Clinical Trial Unit, University Hospital Basel, Switzerland
6 Department of Health Sciences, University of York, United Kingdom
7 Centre for Health Economics, University of York, United Kingdom

**Background:** Network meta-analysis (NMA) is commonly used to simultaneously assess multiple competing interventions. It facilitates the quantitative synthesis of the evidence base, which is reported using different data formats, individual participant data (IPD) or aggregate data (AD). Moreover, the evidence can come from non-randomized studies (NRS) or randomized controlled trials (RCT). RCTs are considered the highest quality of evidence due to its low risk to be associated with selection bias, however, the restricted settings of RCTs limit the generalizability of their results. Combining RCT and NRS evidence in NMA may help overcome some of RCTs limitations. Our aim is to develop a generic NMA framework to synthesise IPD and AD evidence from RCTs and NRS.

**Methods:** We built a Bayesian generic NMA model as an extension of the three-level hierarchical model that combine IPD and AD, to incorporate both RCT and NRS evidence in four different ways. Namely we compared: (a)naïve approach, (b) a model that uses NRS as prior information; (c,d) two different bias adjustment models. These models were used to analyze a network of 3 pharmacological interventions and placebo for patients diagnosed with relapsing remitting multiple sclerosis. The dataset consists of 2 AD and 4 IPD, for a total of 4181 patients, enrolled in either phase III RCTs or a cohort study (SMSC). The models were implemented in R.

**We** conducted a network meta-regression with age as a covariate. Across studies, we assumed the relative treatment effects are exchangeable and the covariate effects are fixed.

**Results:** The four models described above were compared to a model that considered evidence for RCTs only. Dimethyl fumarate's and natalizumab's posterior estimates agree to a large extent. For glatiramer acetate, penalization of AD studies as high or unclear risk of bias produced variable estimates of its effect.

**Conclusions:** The inclusion of RCT and NRS evidence in NMA is needed to consider all the available evidence. Arbitrarily ignoring RCT or NRS may lead to biased results and misleading conclusions.

**Acknowledgement:** The HTx project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825162.

# SESSION OC2D

## OC2D-4 Component Network Meta-Analysis Including Individual Participant Data and Summary Aggregate Data

**Ellesha Smith[1], Laura Gray[1], Suzanne Freeman[1], Keith Abrams[2], Stephanie Hubbard[1]**
1 Health Sciences, University of Leicester, United Kingdom
2 University of York, United Kingdom

**Background:** In many settings, particularly public health, interventions are 'complex' meaning that they are comprised of multiple components. Component network meta-analysis (CNMA) is an evidence synthesis method developed to identify combinations of intervention components that are potentially most effective, including combinations that have not been evaluated in previous research.
**Methods:** To identify which interventions should be recommended for particular sub-populations of patients, existing CNMA methods were adapted to include covariates using individual participant data (IPD), where it was available, and summary aggregate data (SAD). The method was applied to a Cochrane Collaboration systematic review dataset that investigated interventions for promoting the use of safety practices for preventing childhood poisonings at home. The interventions in this dataset were made up of the following components: usual care (UC), education (Ed), free or low cost equipment (Eq), installation (In) and home safety inspection (HSI).
**Results:** A network meta-analysis of the SAD identified the Ed+Eq+HSI intervention to be the most effective. The CNMA approach allowed the effect of each component and combinations of components to be estimated. The adapted CNMA method including IPD and SAD has the potential to reduce the uncertainty in the component effect estimates compared to the CNMA using only SAD, minimise ecological bias and identify the intervention combination that has the potential to be the most effective for specific subgroups of the population.
**Conclusions:** This research will demonstrate how evidence on complex interventions can be synthesised to provide better information to policy decision makers for making evidence-based recommendations to sub-populations and direct future research. CNMA has the potential to identify potential intervention combinations that have not been tested in trials but may be more effective than those that have.

## OC2D-5 Bayesian multivariate network meta-analysis model for the difference in restricted mean survival times

**Xiaoyu Tang, Ludovic Trinquart**
Department of Biostatistics, Boston University School of Public Health, Boston MA, United States

Network meta-analysis (NMA) is essential for medical decision-making. NMA enables inference for all pairwise comparisons between healthcare interventions available for the same indication, by using both direct evidence and indirect evidence. In randomized trials with time-to event outcome data, such as lung cancer trials, conventional NMA methods rely on the hazard ratio and the proportional hazards assumption, and ignore the varying follow-up durations across trials. We introduce a novel multivariate NMA model for the difference in restricted mean survival times (RMST). Our model synthesizes all the available evidence from multiple time points simultaneously and borrows information across time points via within-study covariance and between-study covariance for the differences in RMST. We analytically derived the within-study covariance and estimated the model under the Bayesian framework. We evaluated our model by conducting a simulation study. Our multiple-timepoint model yields lower mean squared error over the conventional single-timepoint model at all time points, especially when the availability of evidence decreases. We illustrated the model on a network of randomized trials of second-line treatments of advanced non-small-cell lung cancer. Our multiple-timepoint model yielded increased precision and detected evidence of benefit at earlier timepoints as compared to the single-timepoint model. Our model has the advantage of providing clinically interpretable measures of treatment effects.



# SESSION OC2E

SESSION OC2E — Some thought on research in biostatistics

## OC2E-1 Phases of methodological research in biostatistics – A proposal to increase transparency and reproducibility

**Georg Heinze[1], Anne-Laure Boulesteix[2], Michael Kammer[1], Tim Morris[3], Ian R. White[3]**
1 Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems (CeMSIIS), Medical University of Vienna, Austria
2 Department of Medical Information Processing, Biometry and Epidemiology (IBE), Ludwig-Maximilians-University of Munich, Germany
3 MRC Clinical Trials Unit at UCL, London, United Kingdom

When a new data analysis method is applied in clinical research, it has usually already undergone some evaluations to demonstrate its suitability and possibly optimality or robustness. The stages of establishing a new method cover theoretical justification, investigations of analytical and asymptotic properties and numerical efficiency of estimation algorithms, and performance evaluations in simulated and real-world data sets. Here we propose to frame this process into a series of 'phases' in analogy to the well-established drug development pipeline. Using examples, we explain our proposal and discuss implications for the practice of biostatistical research.
In Phase I a method's 'safety' may be under study: does it produce a meaningful result in some simple cases? Phase II may cover typical proof-of-concept studies in which a method's inventors demonstrate 'efficacy', i.e., that their method outperforms others under idealised conditions. Phase III studies could be protocol-based methodological comparisons, with clearly defined in- and exclusion criteria for the domain of application, and could even include a blinding mechanism to separate the roles of data generator and data analyst in simulations. Phase IV studies, finally, may cover the 'post-marketing surveillance' to develop guidelines for when to use and when to avoid a new method, and to understand its pitfalls. In both Phase III and IV studies, disclosure statements should clarify if and how a method's inventors were involved in the design and conduct of the comparison.
While this classification is an initial attempt and needs further development, we are confident that the transparency and reproducibility of methodological research may greatly benefit from such a systematic process of development and evaluation. In particular, it may emphasize the importance of well-conducted Phase III or Phase IV studies in methodological research. Moreover, it may make transparent biases towards methods favoured by the study authors, avoid over-optimistic interpretation, and raise awareness of the value of further comparison studies. By way of conclusion, we propose that the phase of research should become an essential attribute of methodological research studies in clinical biostatistics.

## OC2E-2 The p-value conundrum: how can a Bayesian analysis help? A case study in reproductive and maternal-fetal medicine

**Karla Hemming[1], Pedro Melo[1], Rong Luo[2], Monica Taljaard[3], Arri Coomarasamy[1]**
1 University of Birmingham, United Kingdom 2 University of Ottawa, Canada 3 Ottawa Hospital Research Institute, Canada

There is a natural desire to interpret result from randomised controlled trials (RCTs) in a definitive way. Thus, overall interpretation of results from RCTs tend to focus on whether the result is statistically significant, even while attempting to abide by good practice reporting recommendations to provide estimated effect sizes along with confidence intervals. In the evaluation of treatment effects with binary primary outcomes, any small change in effect will often be important. Yet, RCTs will have limited power to detect these small effects due to statistical uncertainty. This means that drawing conclusions based on statistical significance becomes problematic. Firstly, there is the possibility of conflating a non-significant result with evidence of no difference – possibly leading to abandoning treatments that might actually work (overly definitive interpretation). Secondly, there is the risk of placing too much emphasis on statistical significance and conflating a result which is statistically not significant, yet suggestive of an effect, with an inconclusive result – perhaps leading to a failure to adopt treatments in a timely way (overly cautious interpretation). Bayesian estimates of the probability of the treatment having a positive effect offer a means to resolve this conundrum.
In this presentation we report the findings from a review where we systematically apply Bayesian methods to a sample of randomised trials in one particular field of medicine. To this end, we report results of a systematic review of a contemporary sample of two-arm individually randomised superiority trials in reproductive and maternal-fetal medicine, published in high-impact general medical and specialty journals between January 2015 and December 2020. For each identified RCT we express the treatment effect for the primary outcome using relative differences (with 95% confidence interval) using a frequentist method and contrast this with a reanalysis using a Bayesian approach (with uninformative priors) to obtain the Bayesian posterior probability of a positive impact. We demonstrate how the Bayesian approach can resolve the p-value conundrum and lead to qualitative different conclusions in an important area of medicine. We offer practical guidance for journal editors, reviewers, statisticians and trialists to avoid misinterpreting trial results.



# SESSION OC2E

## OC2E-3 On the marginality principle, ratios and interactions

**Tim P. Morris[1], Maarten van Smeden[2], Andrew Althouse[3], Darren Dahly[4], Tra My Pham[1]**
1 MRC Clinical Trials Unit, UCL, London, United Kingdom
2 Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, Netherlands
3 Division of General Internal Medicine, University of Pittsburgh, United States
4 HRB Clinical Research Facility Cork, University College Cork, Ireland

The marginality principle states that, for estimation of interactions or other higher-order effects, it is illegitimate to exclude their main (or otherwise lower-order) effects. Some accounts state a second aspect of the principle: for estimation of main (or otherwise lower-order) effects, it is illegitimate to ignore their interactions or any other higher-order effects. For both aspects it is assumed to be a priori not known that the effects are zero. Many statisticians obey the first aspect, even if they do not know the marginality principle by name, but the second is either not known or considered secondary.
Variables derived as the ratio of two measured variables commonly appear in research data. Examples include body mass index, total cholesterol to HDL, waist–hip ratio, left ventricular ejection fraction, and heart rate. Using the example of a regression model that includes a ratio as a covariate, we show how the two aspects of the marginality principle can be understood as two sides of the same coin. Using some of the listed examples of ratios, we argue that adhering to the first aspect while ignoring the second can lead to absurd modelling choices. The example generalises directly to interactions and less directly to polynomials. We conclude that for statistical models including continuous covariates: 1) awareness of both aspects of the marginality principle avoids making arbitrary default modelling choices; 2) it will be necessary to violate the principle in some way for many real modelling tasks; 3) the terms it is reasonable to omit will depend on context, but may be the 'lower-order' terms.

## OC2E-4 An extension to reporting guidelines for systematic reviews of prediction model studies (TRIPOD-SRMA)

**Kym I.E. Snell[1], Brooke Levis[1], Thomas P.A. Debray[2], Lotty Hooft[2], Paula Dhiman[3], Johannes B. Reitsma[2], Karel G.M. Moons[2], Gary S. Collins[3], Richard D. Riley[1]**
1 Centre for Prognosis Research, School of Medicine, Keele University, United Kingdom
2 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Netherlands
3 Centre for Statistics in Medicine, University of Oxford, United Kingdom

**Introduction:** Guidelines exist for reporting the development, validation and updating of risk prediction models (TRIPOD), and for the reporting of systematic reviews (PRISMA). However, no specific guidance exists for reporting systematic reviews of prediction models which can have different aims, ranging from identifying prediction models through to comparing predictive performance of models. Therefore, existing reporting guidelines require modification to be more suitable for reporting systematic reviews and meta-analyses of prediction model studies.
**Objectives:** To develop an extension to the TRIPOD reporting guidelines, specific to systematic reviews of prediction model studies.
**Methods:** Existing reporting guidelines were reviewed. Relevant guideline items were combined and assessed for suitability by two researchers, considering the different aims of systematic reviews: i) identification of prediction models within a broad clinical field, ii) identification of prediction models for a target population, iii) identification of prediction models for a particular outcome, iv) assessing the performance of a particular prediction model, and v) comparison of prediction models (in terms of predictive performance), where aims iv) and v) may also include meta-analysis. Item suitability and wording were discussed within the working group and a draft extension to TRIPOD was produced. Online Delphi surveys are currently being conducted, involving researchers with experience in systematic reviews and prediction modelling to provide feedback on the proposed items. Results from the Delphi surveys will inform the selection and wording of items in TRIPOD-SRMA.
**Results:** PRISMA and TRIPOD-Cluster (submitted) were identified as the most relevant reporting guidelines. They contained many overlapping items; while PRISMA contained some items specific to systematic reviews, TRIPOD-Cluster contained some items specific to prediction models. Items from both guidelines were combined, resulting in many items being merged and modified, while other items specific to model development or individual participant data were removed. Results from the Delphi survey will be presented and a complete draft of our extension (TRIPOD-SRMA) proposed.
**Conclusions:** TRIPOD-SRMA is an extension of existing reporting guidelines to provide more tailored guidance for reporting systematic reviews of prediction model studies.



# SESSION OC2E

## OC2E-5 Reference based multiple imputation for trials – What's the right variance and how to estimate it?

**Jonathan W. Bartlett**
Department of Mathematical Sciences, University of Bath, United Kingdom

Reference-based multiple imputation (MI) such as jump to reference has become a popular approach to handling missing data in clinical trials. One feature of these methods is that estimates of variance using Rubin's rules are larger than the true repeated sampling variance, raising the question of whether use of Rubin's rules is still appropriate. Recently Cro et al have argued that it is, since this results in so called information anchoring – the variance is (approximately) the same as under an MAR analysis.
I argue that reference-based imputation methods are not truly information anchored, when information is judged in terms of repeated sampling variability, and that the repeated sampling variance is the 'right' one. That this variance reduces as the proportion of missing data increases is a logical consequence of strong assumptions made by reference-based methods. The question of which variance is correct is critical, as it materially affects the statistical power of clinical trials using this method.
For estimating the true repeated sampling variance, I describe and illustrate a simple to apply and computationally efficient combination of bootstrapping and MI. I present simulation results demonstrating the performance of this approach for reference-based imputation with both repeatedly measured continuous endpoints and recurrent event endpoints. Finally, I argue that if true repeated sampling information anchoring is desired, new methods must be devised that satisfy this criterion.

# SESSION OC2F

## SESSION OC2F — Methods for analysis of electronic health records

### OC2F-1 Immortal time bias for life-long conditions in retrospective observational studies using electronic health records

Freya Tyrer[1], Krishnan Bhaskaran[2], Mark J. Rutherford[1]
1 Department of Health Sciences, University of Leicester, United Kingdom
2 Department of Non-Communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, United Kingdom

**Background:** Immortal time bias is common in real-world observational studies using electronic health records but is typically described for pharmacoepidemiology studies where there is a delay between cohort entry and treatment initiation.
**Aims:** For this study, immortal time bias is described in the context of electronic health record observational studies where the exposure is a life-long condition/disability.
**Methods:** Using intellectual disability as an example, one million patients from the UK primary care database (Clinical Practice Research Datalink [CPRD]) linked with national deaths data (Office for National Statistics) were selected to compare four different approaches to handling immortal time bias and their impact on life expectancy in patients (aged 10+ years) with and without the exposure of interest (intellectual disability) from 2000 to 2019. The four approaches were: (i) treatment of immortal time as observational time; (ii) exclusion of immortal time before date of first exposure diagnosis; (iii) exclusion of immortal time before assumed date of entry (by the clinician) of first exposure diagnosis; and (iv) treatment of exposure as a time-dependent measure. Smoothed life expectancy curves were estimated using flexible parametric survival models, stratified by exposure status.
**Results:** When not included in cohort entry criteria (Method 1), disproportionately high life expectancy was observed for the exposed population over the earliest calendar period compared with later periods. This effect attenuated but remained when date of diagnosis was incorporated into entry criteria (Method 2 and 4). Setting cohort entry to assumed date of entry of diagnosis (Method 3) resulted in a substantial loss of subjects and person-time and provided a poor proxy measure for date of entry of diagnosis in this CPRD cohort.
**Conclusions:** Immortal time bias presents a significant problem for comparing studies of life-long conditions / disabilities using electronic health record data. Implications of the findings and recommendations are discussed.



SESSION OC2F

### OC2F-2 Identifying high-risk groups for BMI change using electronic health records from 2.3 million adults

Michail Katsoulis[1,2], Alvina G. Lai[1,2], Karla Diaz-Ordaz[3], Manuel Gomes[1,2], Laura Pasea[1,2], Amitava Banerjee[1,2], Spiros Denaxas[1,2], Kostas Tsilidis[4], Pagona Lagiou[5], Gesthimani Misirli[6], Krishnan Bhaskaran[3], Goya Wannamethee[1,2], Richard Dobson[1,2], Rachel Batterham[1,2], Dimitra-Kleio Kipourou[3], Thomas Lumbers[1,2], Nick Wareham[7], Claudia Langenberg[7], Harry Hemingway[1,2]
1 Institute of Health Informatics, University College London, United Kingdom
2 Health Data Research UK, University College London, United Kingdom
3 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom
4 Imperial College London, United Kingdom
5 Harvard School of Public Health, Boston MA, United States
6 Hellenic Health Foundation, Athens, Greece
7 University of Cambridge, United Kingdom

Despite the urgent need to develop new and targeted strategies for population approaches to obesity prevention, current policy has largely been informed by cross sectional studies that do not allow identification of population groups at highest risk of BMI gain. In this study, we calculated longitudinal changes in BMI over 1-, 5- and 10-years and investigated transition between BMI categories using population-based electronic health records of 2,328,477 adults in England (1998- 2016) from CALIBER.
To estimate the 1, 5 and 10-year BMI change we selected at random one pair of measurements per individual within window intervals of 6 months to 2 years, 4 years to 6 years and 8 years to 12 years respectively. The problem of missing values arises because not all individuals had two BMI measurements within the specific windows. We assumed that the missingness mechanism for BMI change was missing not at random (MNAR) and applied multiple imputation with delta adjustment [1]. More specifically, we applied multiple imputation and we then added delta values to our imputed datasets of BMI change. The assumption under the calculation of the delta values was that that the average 10-year BMI estimates per age group and sex from CALIBER, after multiple imputation, would be the same with the corresponding estimates from an annual cross-sectional survey, the Health Survey from England. We then utilised logistic regression models to estimate the relationship of age, sex, social deprivation, ethnicity and region with BMI change.
Youngest age was the most important sociodemographic factor for BMI change, The odds ratio of transitioning from normal weight to overweight or obesity in the youngest (18-24 years) compared to oldest (65-74 years) individuals was 4.22 (3.85-4.62)), from overweight to obesity 4.60 (4.06-5.22), and from non-severe to severe obesity 5.87 (5.23-6.59). Among the youngest adult, socially deprived men 72% transitioned from normal weight to overweight and 68% from overweight to obesity over 10 years.
Multiple imputation using delta adjustment provides a flexible and transparent means to impute missing data under MNAR mechanisms, especially when these delta values can be calculated, using information from another study.

**Reference:** *[1] Leacy FP, Floyd S, Yates TA. White IR.Analyses of Sensitivity to the Missing-at-Random Assumption Using Multiple Imputation With Delta Adjustment: Application to a Tuberculosis/HIV Prevalence Survey With Incomplete HIV-Status Data. Am J Epidemiol. 2017;185(4):304-15.*

# SESSION OC2F

## OC2F-3 Multiple imputation of sporadically-missing continuous time data by Brownian bridge stochastic interpolation

**Colin C. Everett[1], Rebecca Walwyn[1], Julia M. Brown[1], Leonid V. Bogachev[2], Jeanne Houwing-Duistermaat[2], Deborah D. Stocken[1]**
1 Clinical Trials Research Unit, University of Leeds, United Kingdom
2 School of Mathematics, University of Leeds, United Kingdom

**Context:** The LIBERATES randomised controlled trial utilised continuous glucose monitoring (CGM) sensors to collect primary outcome data. The CGM aims to collect high-frequency data at regularly spaced intervals, several times per hour over a period of weeks, potentially generating over 1300 observations within a 2-week period. As was expected with wearable devices, there were many intervals in the final dataset with no observed data, occurring at different times of day and of differing lengths, with an unsynchronised resumption of the data stream. Conventional statistical analyses rely on single imputation methods that make often-implausible assumptions (such as linear or constant trends, ignoring cyclical trends within the data) and don't account for inherent extra uncertainty. Multiple imputation is an established approach to analysis of data that is partially missing, but methods to impute longitudinal data are generally suited to simple rectangular datasets with discrete assessment time points. Although stream data may be summarised to a single measure – or a series of repeated measures, since data streams are partially-observed, imputing at the summary level may lead to implausible imputations, incompatible with the data observed.
**Objective:** 1. To demonstrate multiple forms of the Brownian Bridge stochastic interpolation approach to multiply-impute plausible paths across intervals of missing continuous data. 2. To start from a simple Brownian Bridge moving to more complex interpolation incorporating additional time-varying information derived from the input dataset.
**Methods:** By treating the underlying glucose values as a continuous-time stochastic process, missing data in the stream was addressed using Brownian Bridge interpolation. Multiple plausible paths were created to allow for the inherent additional uncertainty due to missing data. The method is developed by incorporating more observed data from the dataset, allowing for volatility varying over a 24-hour cycle, different volatility profiles per participant and differing drift patterns per participant over a 24-hour cycle.
**Results:** Comparisons are made, using the LIBERATES dataset, between simpler single imputation methods, such as last observation carried forward and simple linear interpolation, and the Brownian Bridge interpolation methods. Limitations relating to the data collection are highlighted, including distinguishing between partially and completely missing outcome measures, and measurement limits.

## OC2F-4 Handling missing data from wearable devices in clinical trials

**Mia Tackney[1], Derek G. Cook[2], Daniel Stahl[3], Khalida Ismail[4], Elizabeth Williamson[1], James Carpenter[1,5]**
1 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom
2 Population Health Research Institute, St George's, University of London, United Kingdom
3 Department of Biostatistics & Health Informatics, King's College London, United Kingdom
4 Department of Psychological Medicine, King's College London, United Kingdom
5 MRC Clinical Trials Unit, University College London, United Kingdom

Accelerometers and other wearable devices are increasingly being used in clinical trials to provide an objective measure of the impact of an intervention on physical activity. These devices measure physical activity on very fine intervals of time called epochs, which are typically aggregated to provide daily or weekly step counts. In this setting, missing data is common as participants may not wear the device as per protocol, or the device may fail due to low battery or water damage. However, there is no consensus on key issues in handling missing data from such devices. Controversy remains for even the most basic aspects, such as how to determine whether a measurement is missing or not.
We propose an analysis framework that uses wear time to define missingness on the epoch and day level, and propose a multiple imputation approach, at an aggregated level, which treats partially observed daily step counts as right censored. This flexible approach allows the inclusion of auxiliary variables as well as sensitivity analysis to be performed. We illustrate its application to the analysis of exercise trials including the 2019 MOVE-IT trial.
**References:** *Ismail K, Stahl D, Bayley A, Twist K, Stewart K, Ridge K, Britneff E, Ashworth M, de Zoysa N, Rundle J, Cook D, Whincup P, Treasure J, McCrone P, Greenough A, Winkley K. Enhanced motivational interviewing for reducing weight and increasing physical activity in adults with high cardiovascular risk: the MOVE IT three-arm RCT. Health Technol Assess. 2019 Dec;23(69):1-144. doi: 10.3310/hta23690. PMID: 31858966; PMCID: PMC6943381.*

# SESSION OC2F

## OC2F-5 Using wristwear device to assess impact of COVID19 lockdown on physical activity

**Agus Salim[1,2,3], Christian Brakenridge[1,4], David Dunstan[1,4], Neville Owen[1,5]**
1 Baker Heart and Diabetes Institute, Melbourne VIC, Australia
2 Melbourne School of Population and Global Health, The University of Melbourne VIC, Australia
3 School of Mathematics and Statistics, The University of Melbourne VIC, Australia
4 Mary Mackillop Institute for Health Research, Australian Catholic University, Melbourne VIC, Australia
5 Centre for Urban Transitions, Swinburne University of Technology, Hawthorn VIC, Australia

Wearable activity trackers such as Fitbit have grown in popularity in recent times. Although less accurate than research-grade accelerometer, Fitbit can nonetheless produce very useful data for measuring physical activities. During the last 12 months, various lockdown measures have been introduced in different parts of the world in attempts to slow the spread of the virus. While arguably effective at slowing the pandemic, COVID19 lockdown has had profound impacts on physical and psychological well-being of those impacted.
Using Fitbits step-count data collected over multiple individuals at 1-minute interval, I assess the impact of lockdown on physical activity using segmented generalized linear model (GLM) that allows for differences in mean level, diurnal and seasonal trends of physical activity before and after lockdown. Results from both linear and negative binomial (NB) regressions will be presented and compared. The results show there is significant between-individual heterogeneity of the impact of lockdown, with an overall negative impact of lockdown on physical activity.
In the second part of the talk, I will highlight several statistical issues inherent in the Fitbits data, including strong autocorrelation and measurement errors and propose methods to calibrate FitBits data using data from research-grade accelerometer.

# SESSION OC2G

## SESSION OC2G — Analysis of gene expression and omics data

### OC2G-1 Prostate cancer intratumor heterogeneity assessment by depth measures analysis on imaging texture features

**Lara Cavinato[1], Alessandra Ragni[2], Francesca Ieva[2], Martina Sollini[3], Francesco Bartoli[4], Paola Erba[4]**
1 Department of Mathematics, Milan Polytechnic, Italy
2 Department of Mathematics, Milan Polytechnic, Italy
3 Department of Biomedical Sciences, Humanitas University, Milan, Italy
4 Department of Translational Research and New Technology in Medicine, University of Pisa, Italy

**Background and Motivations:** Personalized treatment has become a crucial point of modern medicine. Specifically, in patients with cancer, the optimization of therapeutic decision based on prognostic risk assessment is essential. At this regard, preliminary findings have shown imaging-derived biomarkers for spatial intratumor heterogeneity to be fundamental in understanding tumor severity and evolution, impacting on pre-treatment clinical-pathological prognosis [1]. However, a consensus about quantitative definition of heterogeneity has not yet been reached and, with it, informed clinical decisions cannot be implemented. Although quantitative tumor characterization from tomographic PET/CT imaging data inspection, namely radiomics, is catching on, redundancy and high dimensionality of imaging biomarkers prevent its translation to medical practice, calling for agnostic and lossless feature transformation, resulting in an exhaustive lesions texture profiling.
**Objective:** In this work, we propose a depth-based metrics for lesion profiling in order to quantitively define intratumor heterogeneity in patient with metastatic prostate cancer.
**Statistical methods:** 84 patients with multi-metastatic recurrent prostate cancer enrolled in a clinical trail have been analyzed. All patients underwent whole-body [$^{18}$F]FMCH PET/CT for restaging, lesions were segmented and radiomic texture features were extracted form ROI. The 37 radiomic features were sliced into 6 semantic groups and Mahalanobis depth has been computed on each group [2]. Each lesion ended up to be described by a 6-dimentional vector, namely its radiomic profile. Lesions profiles were clustered according to a k-means algorithms (k=2) and clusters were tagged as risk classes. Finally, patients' intratumor heterogeneity was evaluted in terms of concordance of lesions risk factor within patients.
**Results:** Preliminary results showed that texture analysis of [$^{18}$F]FMCH PET/CT with depth measures is able to overcome redundancy and dimentionality issues with no information loss, in order to characterize intratumor heterogeneity in patients with recurrent prostate cancer. The ultimate goal is to provide an imaging biomarker for risk stratification, thus guiding patients' treatment decision making.
**References:** *[1] Sala, E., et al. "Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging." Clinical radiology 72.1 (2017): 3-10.*
*[2] Liu, Regina Y., Jesse M. Parelius, and Kesar Singh. "Multivariate analysis by data depth: descriptive statistics, graphics and inference,(with discussion and a rejoinder by liu and singh)." The annals of statistics 27.3 (1999): 783-858.*



SESSION OC2G

### OC2G-2 Investigating Down syndrome by integrating methylation and glycomics using supervised PO2PLS

**Zhujie Gu[1], Said el Bouhaddani[1], Jeanine Houwing-Duistermaat[1,2,3], Hae-Won Uh[1]**
1 Department of Data Science and Biostatistics, UMC Utrecht, Netherlands
2 Department of Statistics, University of Leeds, United Kingdom
3 Department of Statistical Sciences, University of Bologna, Italy

**Background:** Down syndrome (DS) is a condition that leads to premature or accelerated aging in affected subjects. They develop diseases that are typically observed at a higher age. Studies at the molecular level of DS have reported several alterations in methylation and glycomics. However, these studies were conducted on each omics level separately, overlooking the relationship between omics levels. Joint analysis of methylation and glycomics in the context of DS is needed to gain insight from a multi-omics perspective. Our motivating datasets were measured on 29 DS patients and their healthy siblings and mothers.
**Aim:** We aim to investigate the premature/accelerated aging in DS by analyzing methylation and glycomics jointly, and identify relevant CpG sites and glycans, using our newly proposed method supervised probabilistic O2PLS (supervised PO2PLS).
**Method:** For dimension reduction of high-dimensional correlated omics data, we consider two-way orthogonal partial least squares (O2PLS), which constructs a few joint latent variables that explain the covariance, while taking into account the heterogeneity between the two omics datasets. However, O2PLS does not model the outcome variable, i.e., the DS status, thus needs additional steps to link the latent variables to DS. We propose a probabilistic framework, supervised PO2PLS that combines omics integration and outcome regression in one model. It constructs joint latent variables that explain the covariance between omics and also the variance in the outcome variable. On the 'global' level, it allows for statistical inference on the relationship between the omics datasets, and between the omics data and the outcome as well. On the individual variable level, the assigned weights make it easy to identify important features. All the parameters in the model are estimated using maximum likelihood, taking into account the omics data and the outcome variable simultaneously. A simulation study to evaluate the performance of supervised PO2PLS and results of the DS data analysis will be presented.
**Conclusion:** To conclude, we study aging in DS by considering methylation and glycomics data together. Our proposed method that jointly analyzes multiple omics data with outcome variable may provide new insight into the underlying mechanism of premature/accelerated aging at multi-omics level.

### OC2G-3 Model selection characteristics when using MCP-Mod for dose-response gene expression data

**Julia Duda, Franziska Kappenberg, Jörg Rahnenführer**
Faculty of Statistics, TU Dortmund University, Germany

In the context of drug development, understanding the dose-response relationship of a candidate drug is crucial to determine a target dose of interest. Classical approaches in clinical Phase II dose-finding trials rely on pairwise comparisons between doses and placebo.
A methodological improvement to this is the MCP-Mod (Multiple Comparison Procedure and Modeling) approach, originally developed for Phase II trials (Bretz et al., 2005). MCP-Mod combines multiple comparisons with modeling approaches in a multistage procedure. First, for a set of pre-specified candidate models, it is tested if any dose-response signal is present. Second, considering models with detected signal, either the best model is selected to fit the dose-response curve or model averaging is performed. We extend the scope of application for MCP-Mod to in-vitro gene expression data and assess its characteristics regarding model selection for concentration-gene expression curves. Precisely, we apply MCP-Mod on single genes of a high-dimensional gene expression data set, where human embryonic stem cells were exposed to eight concentration levels of the compound valproic acid (VPA). As candidate models, we consider the sigmoid Emax, linear, quadratic, Emax, exponential and beta model. Through simulations we investigate the impact of omitting one or more models from the candidate model set to uncover possibly redundant models and to evaluate the precision and recall rates of selected models.
Our results clearly support the consideration of various dose-response models when analyzing dose-dependent gene expression data. These include, in particular, the often-neglected non-monotone models such as the quadratic model. Measured by the AIC, all models perform best for a considerable number of genes. For less noisy measurements the popular sigmoid Emax model is frequently selected. For more noisy data, often simpler models like the linear model are selected, but mostly without relevant performance advantage compared to the second-best model. It is also noticeable that the commonly used standard Emax model has an unexpected low performance.

# SESSION OC2G

## OC2G-4 Information sharing across genes for improved parameter estimation in dose-response curves

**Franziska Kappenberg, Jörg Rahnenführer**
Faculty of Statistics, TU Dortmund University, Germany

Determining the right dose is a critical goal in drug-development, from the pre-clinical phase (e.g. in toxicology) to phase 3 studies. In order determine an optimal dose of interest without being restricted to the experimental doses, a common approach is to fit a parametric curve to the dose-response data.
In toxicological assays and for human in vivo data, increasing the number of considered doses or the numbers of replicates or patients per dose yields a higher quality of the fitted curve, but causes critical additional costs. However, technologies for measuring high-dimensional gene expression data are well established. Thus, a statistical approach to obtain higher-quality fits of dose-response curves when gene expression is the target of interest is to exploit similarities between high-dimensional gene expression data. This idea can also be called information sharing across genes. Parameters of the fitted curves can be linked, according to either a priori assumptions or estimates of the distributions of the parameters, in a Bayesian framework.
Here, we consider the special case of the sigmoidal 4pLL model for estimating the curves associated with single genes in a toxicological in vitro assay, and we are interested in the concentration at which the half-maximal effect is reached, the EC50. This value can be considered a reasonable indicator for a relevant expression effect of the corresponding gene. We introduce an empirical Bayes method for information sharing across genes in this situation, by modelling the distribution of the EC50 values across all genes. Based on this distribution, for each gene, a weighted mean of the corresponding individually estimated parameter and of the overall mean of the estimated parameters of all genes is calculated, hence shrinking parameter estimates to the overall mean. We evaluate our approach using a simulation study that is based on the structure of a real gene expression dataset. Results show that the Bayesian method works well in terms of reduction of the mean squared error between true underlying value and estimate. Finally, the method is also applied to the real gene expression dataset to demonstrate the influence of the analysis strategy on the results.

## OC2G-5 A Bayesian approach to estimating dynamic models of co-regulated gene expression

**Sara Venkatraman[1], Sumanta Basu[1], Martin Wells[1], Myung-Hee Lee[2]**
1 Department of Statistics and Data Science, Cornell University, Ithaca NY, United States
2 Department of Medicine, Weill Cornell Medical College, New York NY, United States

Time-course gene expression datasets provide insight into the dynamics of complex biological processes, such as immune response, disease progression, and organ development. It is of interest to identify genes with similar expression patterns over time because such genes often share similar biological characteristics. For instance, they may be co-regulated by the same transcription factors. However, identifying genes with similar temporal expression patterns is challenging because gene expression datasets consist of thousands of genes, measured at a small number of time points, and the time dynamics of gene expression are highly nonlinear. We propose a Bayesian approach to estimating ordinary differential equation (ODE) models of gene expression, from which we derive new metrics that capture the similarity in the time dynamics of two genes. These metrics, which are based on the familiar R^2 value, are simple and fast to compute and can be used for generating clusters or networks of closely related genes. The salient feature of our method is that it leverages external biological databases that document known interactions between genes; the method automatically uses this information to define informative prior probability distributions on the parameters of the ODE model. This ultimately encourages genes with known relationships to receive higher similarity scores, and allows us to infer the functionality of under-studied genes whose time dynamics are very similar to more well-studied ones. We also derive optimal, data-driven shrinkage parameters that balance the ODE model's fit to both the data and to external biological information. Using real gene expression datasets collected from fruit flies, we demonstrate that our approach produces clusters and networks of genes with clear biological interpretations. By doing so, our method is able to reduce the dimensionality of gene expression datasets and reveal new insights about the dynamics of biological systems.



# SESSION OC3A

SESSION OC3A — Propensity score in causal studies

## OC3A-1 Multiple imputation in propensity score matching: obtaining correct confidence intervals

**Corentin Ségalas[1], Clémence Leyrat[1], James Carpenter[1,2], Elizabeth Williamson[1]**
1 Medical Statistics Department, London School of Hygiene and Tropical Medicine, United Kingdom
2 MRC Clinical Trials Unit, University College London, United Kingdom

Propensity score matching is a popular method for handling confounding in observational analyses of observational data. This method builds a matched dataset, sometimes discarding many subjects from the final analysis, and where baseline covariates are balanced using the balancing properties of the propensity score. From this matched dataset, an average treatment effect can be estimated and under some conditions, it is an unbiased estimate of the true causal treatment effect.
One non negligible challenge when applying propensity score matching procedure is the presence of missing data. A classic approach in this case is to use multiple imputation to build several imputed datasets. From each of these imputed datasets, propensity scores are computed and matching is done leading to several completed matched datasets. From each of these matched datasets, an average treatment effect is estimated. Then, these treatment effects are aggregated using Rubin's rules (Rubin, 1987) to compute an aggregated average treatment effect and its variance.
Previous work by Reiter et al., undertaken in the area of measurement error, highlighted a previously unrecognized phenomenon: that using individuals (or units) to develop an imputation model who are not subsequently included in the analysis can result in over-coverage of confidence intervals obtained using Rubin's rules. We show, via simulation studies, that using a sample to perform multiple imputation and then matching, thereby potentially discarding a substantial number of individuals from the final analysis, leads to this phenomenon. We find the coverage of standard application of Rubin's rules to be often > 99%. A simulation study to evaluate the Reiter's procedure has been conducted and gives very satisfying results with most of the coverage rates around the nominal value of 95%. Finally, this result has been illustrated through an application to real data on lung cancer.

**References:** *[1] Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, Inc. [2] Reiter, J. P. (2008). Multiple Imputation When Records Used for Imputation Are Not Used for Disseminated for Analysis. Biometrika 95, 933–946*

# SESSION OC3A

## OC3A-2 Effectiveness of screening colonoscopy in reducing colorectal cancer incidence: emulated target trials from German claims data

**Malte Braitmaier[1], Sarina Schwarz[2], Bianca Kollhorst[1], Vanessa Didelez[1,3], Ulrike Haug[2,4]**

1 Department of Biometry and Data Management, Leibniz Institute for Prevention Research and Epidemiology – BIPS GmbH, Bremen, Germany
2 Department of Clinical Epidemiology, Leibniz Institute for Prevention Research and Epidemiology – BIPS-GmbH, Bremen, Germany
3 Department of Mathematics and Computer Science, University of Bremen, Germany
4 Department of Human and Health Sciences, University of Bremen, Germany

**Introduction:** Observational studies suggest a strong effect of screening colonoscopy in reducing both colorectal cancer (CRC) incidence and mortality. The preventive effect appears more pronounced for distal vs. proximal CRC (Brenner et al. 2014), but there is conflicting evidence regarding the size of this difference. Interpretation of available observational studies reporting site-specific effects is often hampered by low sample size or a statistical analysis that does not explicitly address avoidable biases, while ongoing randomized controlled trials (RCTs) on screening colonoscopy are not powered to assess site-specific effects.
**Methods:** The emulation of target trials is a framework for analysing observational data aiming at causal conclusions, when an RCT is not available or desirable. Explicitly describing the target study protocol ensures a clear formulation and communication of the research question and avoids self-inflicted biases, such as immortal time bias. Building on and extending the approach of García-Albéniz et al. (2016), we emulate target trials using health claims data from Germany, where screening colonoscopy is routinely offered to individuals aged ≥55 years (the database GePaRD covers 20% of the German population). In contrast to García-Albéniz et al. (2016), our database includes individuals younger than 65 years, and we assess overall CRC incidence, but also proximal and distal CRC incidence by estimating event-specific cumulative incidence functions (CIFs). This is implemented using flexible pooled logistic models, avoiding an implausible proportional hazards assumption. CIFs will be compared between subjects undergoing screening at baseline and subjects not undergoing screening at baseline, corresponding to an intention-to-screen effect. Confounding will be adjusted for by inverse probability of treatment weighting. Confidence intervals will be estimated, using subject-level bootstrapping to account for repeated inclusion of subjects.
**Results:** We will estimate covariate-adjusted event-specific cumulative incidence curves (CIFs), assessing the screening effect over 11 years of follow-up. Moreover, we will discuss the strengths and weaknesses of using the emulation of target trials framework in assessing screening effectiveness from observational data.
**References:** *[1] García-Albéniz X, Hsu J, Bretthauer M, Hernán MA (2016) Effectiveness of Screening Colonoscopy to Prevent Colorectal Cancer Among Medicare Beneficiaries Aged 70 to 79 Years – A Prospective Observational Study; Annals of Internal Medicine; 166(1): 18-26 [2] Brenner H, Stock C, Hoffmeister M (2014) Effect of Screening Sigmoidoscopy and Screening Colonoscopy on Colorectal Cancer Incidence and Mortality: Systematic Review and Meta-Analysis of Randomized Controlled Trials and Observational Studies; BMJ; 348*



SESSION OC3A

## OC3A-3 Variance estimators for weighted and stratified linear dose-response function estimators using generalized propensity score

**Valérie Garès[1], Guillaume Chauvet[2], David Hajage[3]**

1 Rennes University, INSA, CNRS, IRMAR - UMR 6625, France
2 Rennes University, ENSAI, CNRS, IRMAR - UMR 6625, France
3 HP, Hôpital Pitié-Salpêtrière, Département de Santé Publique, Centre de Pharmacoépidémiologie, Paris, France

Propensity score methods are widely used in observational studies for evaluating marginal treatment effects. The generalized propensity score (GPS) is an extension of the propensity score framework, historically developed in the case of binary exposures, for use with quantitative or continuous exposures. In this paper, we proposed variance estimators for treatment effect estimators on continuous outcomes. Dose-response functions (DRF) were estimated through weighting on the inverse of the GPS, or using stratification. Variance estimators were evaluated using Monte Carlo simulations. Despite the use of stabilized weights, the variability of the weighted estimator of the DRF was particularly high, and none of the variance estimators (a bootstrap-based estimator, a closed-form estimator especially developed to take into account the estimation step of the GPS, and a sandwich estimator) were able to adequately capture this variability, resulting in coverages below the nominal value, particularly when the proportion of the variation in the quantitative exposure explained by the covariates was large. The stratified estimator was more stable, and variance estimators (a bootstrap-based estimator, a pooled linearized estimator, and a pooled model-based estimator) more efficient at capturing the empirical variability of the parameters of the DRF. The pooled variance estimators tended to overestimate the variance, whereas the bootstrap estimator, which intrinsically takes into account the estimation step of the GPS, resulted in correct variance estimations and coverage rates. These methods were applied to a real data set with the aim of assessing the effect of maternal body mass index on newborn birth weight.

## OC3A-4 Confounder selection strategies targeting stable treatment effect estimators

**Wen Wei Loh[1], Stijn Vansteelandt[2,3]**

1 Department of Data Analysis, Ghent University, Belgium
2 Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium
3 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom

**Clinical research problem and statistical challenges:** Inferring the causal effect of a treatment on an outcome in an observational study requires adjusting for observed baseline confounders to avoid bias. However, adjusting for all observed baseline covariates, when only a subset are confounders of the effect of interest, is known to yield potentially inefficient and unstable estimators of the treatment effect. Furthermore, it raises the risk of finite-sample bias and bias due to model misspecification. For these stated reasons, confounder (or covariate) selection is commonly used to determine a subset of the available covariates that is sufficient for confounding adjustment.
**Objective:** In this article, we propose a confounder selection strategy that focuses on stable estimation of the treatment effect. In particular, when the propensity score model already includes covariates that are sufficient to adjust for confounding, then the addition of covariates that are associated with either treatment or outcome alone, but not both, should not systematically change the effect estimator.
**Statistical** Methods: The proposal, therefore, entails first prioritizing covariates for inclusion in the propensity score model, then using a change-in-estimate [1] approach to select the smallest adjustment set that yields a stable effect estimate. The proposal therefore explicitly assesses the stability of the treatment effect estimator across different (nested) covariate subsets as a selection criterion.
**Results:** The ability of the proposal to correctly select confounders, and to ensure valid inference of the treatment effect following data-driven covariate selection, is assessed empirically and compared with existing methods using simulation studies. We demonstrate the procedure using three different publicly available datasets commonly used for causal inference.
**Conclusion:** The proposal was demonstrated empirically to yield approximately valid inference following a data-driven selection of covariates through the combined use of (i) double selection for prioritizing the covariates, (ii) stability-based assessment to select covariates for confounding adjustment, and (iii) randomization inference using full matching to control the type I error when testing the null of no (individual) treatment effect.
**Reference:** *[1] Greenland S, Daniel R, Pearce N. Outcome modelling strategies in epidemiology: traditional methods and basic alternatives. Int J Epidemiol. 2016;45(2):565-575.*

# SESSION OC3A

## OC3A-5 Causal inference for combining RCTs and observational studies: methods comparison and medical application

**Bénédicte Colnet[1]*, Imke Mayer[2]*, Guanhua Chen[3], Awa Dieng[4], Ruohong Li[5,6], Gaël Varoquaux[1], Jean-Philippe Vert[4], Julie Josse[7], Shu Yang[8]**

1 Université Paris-Saclay, INRIA, France
2 Centre d'Analyse et de Mathématique Sociales, École des Hautes Études en Sciences Sociales, PSL University, Paris, France
3 Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison WI, United States
4 Google Research, Brain team, Paris, France
5 Department of Biostatistics, Indiana University School of Medicine, Indianapolis IN, United States
6 Richard M. Fairbanks School of Public Health, Indianapolis IN, United States
7 Inria, Sophia-Antipolis, Valbonne, France
8 Department of Statistics, North Carolina State University, Raleigh NC, United States
* First co-authors

The simultaneous availability of observational and experimental data for the same medical question of a treatment effect is at the same time an opportunity and a theoretical and methodological challenge. In this work we address the question of how to leverage the advantages and information and how to address the shortcomings of both data sources to improve the validity and scope of the treatment effect estimates. This work is motivated by the analysis of a large prospective database counting about over 20,000 severely traumatized patients in France and a multi-centered international randomized controlled trial (RCT) studying the effect of tranexamic acid administration on mortality among patients with traumatic brain injury.
We first discuss identification and estimation methods that improve generalizability of RCTs using the representativeness of observational data. Classical estimators include weighting, difference between conditional outcome models, and doubly robust estimators, especially calibration weighting. We then discuss methods that combine RCTs and observational data to improve the (conditional) average treatment effect estimation, handling possible unmeasured confounding in the observational data. We compare the methods with extensive simulations and highlight the very good behaviour of calibration weighting. Additionally, we propose an implementation of the different methods to provide analysis pipelines for reproducible data analyses. Finally, we propose to combine both structural causal model and potential outcomes frameworks to provide a complete workflow to analyze both data sources.
The analysis shows that both RCT and the observational data conclude on a zero effect of the drug. The same conclusion is obtained when generalizing the effect of the RCT on the observational Traumabase data while taking into account the distributional shift. In the proposed analysis we also discuss additional challenges such as missing values and mixed data and propose several solutions to tackle them.

# SESSION OC3B

SESSION OC3B — COVID19 modelling

## OC3B-1 Factors involved in COVID-19 prognosis of patients hospitalized in Campania Region. Findings from COVOCA Study

**Vittorio Simeon[1], Pia Clara Pafundi[2], Paolo Chiodini[1], Raffaele Galiero[2], Alfredo Caturano[2], Erica Vetrano[2], Ferdinando Carlo Sasso[2]; on behalf of COVOCA study group Investigators**

1 Department of Physical and Mental Health and Preventive Medicine, University of Campania "Luigi Vanvitelli", Naples, Italy
2 Department of Advanced Medical and Surgical Sciences, University of Campania "Luigi Vanvitelli", Naples, Italy

Italy has been the first Western country heavily affected by COVID-19 and among pioneers of pandemic's clinical management. Although investigated, the association between pre-existing comorbidities and clinical outcome remains controversial [1]. Identification of patients at the highest risk seems mandatory to improve the outcome. We thus aimed to identify comorbidities/clinical conditions upon admission associated with in-hospital mortality in Campania Region (Italy) during first peak.
COVOCA is a multicentre retrospective observational cohort study, involving 18 COVID Centers, with data from patients who completed their hospitalization between March-June 2020. The primary endpoint was in-hospital mortality. Data were described both in the overall population and by considering comorbidity status (0 to +3 sum of comorbidities). The association between presence/absence of single comorbidity for each possible crossing, was assessed calculating the phi coefficient. Univariable/multivariable logistic regression models were performed to evaluate association between in-hospital mortality and exposure variables. Comparison between models, in order to evaluate model's improvement and significance, was performed applying the Likelihood-ratio test for nested models or Bayesian Information Criterion (BIC) for non-nested models, preferring lower BIC models. As sensitivity analysis, Firth's correction by van Smeden M et al. was used to increase beta-estimates' accuracy [2]. Cumulative incidence function (CIF) showed cumulative failure rates over time due to in-hospital mortality, with discharge as competing event.
Among 618 COVID-19 hospitalized patients included (62% males, mean age 60 yrs.), 143 in-hospital mortality events were recorded (CIF 23%). At multivariable analysis, male sex (OR 2.63, 95%CI 1.42-4.90), Chronic Liver Disease (OR 5.88, 95%CI 2.39–14.46) and malignancies (OR 2.62, 95%CI 1.21–5.68) disclosed an independent association with a poor prognosis, as well as need for NIV ventilation or intubation. Higher Glasgow-Coma-Score values were instead associated with a better prognosis. Sensitivity analysis further enhanced these findings.
Mortality of patients hospitalized for COVID-19 appears strongly affected by clinical conditions on admission and comorbidities. Originally, we observed a very poor outcome in subjects with chronic liver diseases, alongside with an increase of hepatic damage.
Our findings allow to underline the fundamental importance of early identification of high-risk patients from hospitalization to improve the pandemic's clinical management.

**References:** *[1] Wang B et al. Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis. Aging 2020; 12(7):6049-6057. [2] van Smeden M, de Groot JA, Moons KG, Collins GS, Altman DG, Eijkemans MJ, Reitsma JB. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. BMC Med Res Methodol. 2016 Nov 24;16(1):163. doi: 10.1186/s12874-016-0267-3.*

# SESSION OC3B

## OC3B-2 Laplace approximations for fast Bayesian inference of the time-varying reproduction number under misreported epidemic data

**Oswaldo Gressani[1], Christel Faes[1], Niel Hens[1,2]**

1 Interuniversity Institute for Biostatistics and statistical Bioinformatics, Data Science Institute, Hasselt University, Belgium

2 Centre for Health Economics Research and Modelling Infectious Diseases, Vaccine & Infectious Disease Institute, University of Antwerp, Belgium

In epidemic models, the effective reproduction number $R_t$ is of central importance to dynamically assess the transmission mechanism of an infectious disease and to orient health intervention strategies. Publicly shared data during an outbreak often suffers from two sources of misreporting (underreporting and delay in reporting) that should not be overlooked when estimating $R_t$. The main statistical challenge in models that intrinsically account for a misreporting process lies in the joint estimation of the time-varying reproduction number and the delay/underreporting parameters. Existing Bayesian methods typically rely on Markov chain Monte Carlo (MCMC) [1] that are extremely costly from a computational perspective. We propose a much faster alternative based on Laplace-P-splines (LPS) [2] that combines Bayesian penalized B-splines for flexible and smooth estimation of $R_t$ and Laplace approximations to selected posterior distributions. Assuming a known generation interval distribution, the incidence at a given calendar time is governed by the epidemic renewal equation and the delay structure is specified through a composite link framework. Laplace approximations to the conditional posterior of the spline vector are obtained from analytical versions of the gradient and Hessian of the log-likelihood, implying a drastic speed-up in the computation of posterior estimates. Furthermore, the proposed LPS approach can be used to obtain point estimates and approximate credible intervals for the delay and underreporting parameters. Simulation of epidemics with different combinations for the underreporting rate and delay patterns (one-day, two-day and weekend delays) show that the proposed LPS methodology delivers fast and accurate estimates and highlights the added value from a computational point of view. Finally, we conclude by illustrating the use of LPS on a real case study of an epidemic outbreak.

**References:** *[1] Azmon, A., Faes, C., & Hens, N. (2014). On the estimation of the reproduction number based on misreported epidemic data. Statistics in medicine, 33(7), 1176-1192. [2] Gressani, O., & Lambert, P. (2021). Laplace approximations for fast Bayesian inference in generalized additive models based on P-splines. Computational Statistics & Data Analysis, 154, 107088.*

## OC3B-3 Evaluating the effectiveness of locally supported contact tracing on NHS test and trace for Covid19

**Pantelis Samartsidis[1], Shaun Seaman[1], Charlotte Anderson[2], Abbie Harrison[2], Andre Charlett[2], Gareth Hughes[2], Isabel Oliver[2], Daniela De Angelis[1,2]**

1 MRC Biostatistics Unit, University of Cambridge, United Kingdom

2 Public Heath England, United Kingdom

In the UK, the NHS Test & Trace programme (TT) was developed to ensure that individuals who test positive for Covid19 and their close contacts are notified that they must self-isolate in order to stop further spread of the virus. Since August 2020, several local authorities have introduced local tracing partnerships (LTP) to assist TT. In this work, we are interested in evaluating the impact that LTP had on the effectiveness of TT in terms of case completion, timeliness (within 48 hours) of case completion and average number of contacts obtained. Further, we are interested in identifying effect modifiers, that is variables that affect the magnitude of the effect of LTP within each area.

Causal inference in the setting outlined above is typically carried out through factor analysis (FA) or synthetic control methods. However, these methods are not appropriate for count data, particularly when the counts are low, as in our context. Further, some of these methods can only be implemented on a single outcome at a time and thus do not allow sharing of information between the outcomes. To overcome these limitations, we propose a Bayesian multivariate FA model for mixed outcomes and show how this model can be adapted to account for effect modification.

Application of our methods on the motivating NHS TT dataset provides valuable insights regarding the effectiveness of LTP.



# SESSION OC3B

## OC3B-4 Accounting for time-dependant confounding variables in mechanistic ODE model: simulations and application to a vaccine trial

**Marie Alexandre, Mélanie Prague, Rodolphe Thiébaut**

Department of Public Health, Bordeaux University, Inserm, Inria SISTM, France

Mechanistic models based on ordinary differential equations (ODE) represent an alternative approach for causal modelling (Commenges J.R.Statist.Soc.B 2009). The model is defined with three parts: i) a structural model defined by compartments that are interacting ii) a statistical model defining how the model parameters are varying across units/individuals iii) an observation model relating the observed quantities to compartments.

It can be used to analyze the within-host response to vaccine in experimental studies. A structural mathematical model is defined for the dynamics of the virus which is infecting susceptible cells that are producing viral particles once infected. Then, a statistical model is defined to explain the variation of parameters between individuals which could be explained by explanatory variables (X) or captured through random effects. The explanatory variables can be fixed (e.g. experimental groups) or time varying such as the immunological markers measuring the response to the vaccine. The parameters (fixed effects and variances of random effects) are estimated using standard approaches (stochastic EM algorithm). This approach has been applied to a recent study evaluating a vaccine against SARS-Cov2 in 18 macaques (6 being vaccinated) who were experimentally infected. The estimation of model parameters using the repeated measurements of viral load showed a decrease of the viral infectivity by 99% in the vaccinees. The next step was to explain the decrease of viral infectivity by various immunological markers repeatedly measured over time such as the neutralizing and binding antibody titres as time-varying explanatory variables (X). However, this approach for analyzing the influence of X over some of the model parameters is not taking into account that a given model compartment (e.g. the virus) may itself influences X. Ignoring this relationship between the virus and X may lead to a biased estimates of the effect of X. An alternative approach is to model X as a compartment of the structural model which gives the opportunity of capturing an effect of V over X and vice versa. We are exploring the impact of the two approaches for the validity of the estimates through simulations before applying it on the real dataset.

## OC3B-5 Interventions to control nosocomial transmission of SARS-CoV-2: a modelling study

**Thi Mui Pham[1]*, Hannan Tahir[1]*, Janneke H.H.M. van de Wijgert[1,2], Bastiaan Van der Roest[1], Pauline Ellerbroek[3], Marc J.M. Bonten[1,4], Martin C.J. Bootsma[1,5], Mirjam E. Kretzschmar[1]**

1 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Netherlands

2 Institute of Infection, Veterinary, and Ecological Sciences, University of Liverpool, United Kingdom

3 Department of Internal Medicine, University Medical Center Utrecht, Utrecht University, Netherlands

4 Department of Medical Microbiology, University Medical Center Utrecht, Utrecht University, Netherlands

5 Mathematical Institute, Utrecht University, Netherlands

* These authors contributed equally to this work.

**Background:** Emergence of more transmissible SARS-CoV-2 variants requires more efficient control measures to limit nosocomial transmission and maintain healthcare capacities during pandemic waves. Yet, the relative importance of different strategies is unknown.

**Methods:** We developed an agent-based model and compared the impact of personal protective equipment (PPE), screening of healthcare workers (HCWs), contact tracing of symptomatic HCWs, and HCW cohorting on nosocomial SARS-CoV-2 transmission. The model was fit on hospital data from the first wave in the Netherlands (February until August 2020) and assumed that HCWs used 90% effective PPE in COVID-19 wards and self-isolated at home for seven days immediately upon symptom onset. We accounted for a variable infectiousness and sensitivity of the diagnostic test from time since infection. Intervention effects on the effective reproduction number (R), HCW absenteeism and the proportion of infected individuals among tested individuals (positivity rate) were estimated for a more transmissible variant.

**Results:** Introduction of a variant with 56% higher transmissibility increased – all other variables kept constant – R from 0.4 to 0.65 (+63%) and nosocomial transmissions by 303%, mainly because of more transmissions caused by pre-symptomatic patients and HCWs. Compared to baseline, PPE use in all hospital wards (assuming 90% effectiveness) reduced R by 85% and absenteeism by 57%. Screening HCWs every three days with perfect test sensitivity reduced R by 67%, yielding a maximum test positivity rate of 5%. Screening HCWs every three or seven days assuming time-varying test sensitivities reduced R by 9% and 3%, respectively. Contact tracing reduced R by at least 32% and achieved higher test positivity rates than screening interventions. HCW cohorting reduced R by 5%. Sensitivity analyses for 50% and 70% effectiveness of PPE use did not change interpretation.

**Implications:** In response to the emergence of more transmissible SARS-CoV-2 variants, PPE use in all hospital wards might be considered to effectively prevent nosocomial transmission. Regular screening and contact tracing of HCWs may also increase the effectiveness of control strategies, but critically depend on the sensitivity of the diagnostic test used.

# SESSION OC3C

## SESSION OC3C — Dynamic prediction

### OC3C-1 Individual dynamic prediction of clinical endpoint from large dimensional longitudinal biomarker history: a landmark approach

**Anthony Devaux[1], Robin Genuer[1,2], Karine Pérès[1], Cécile Proust-Lima[1]**
1 Inserm BPH U1219, University of Bordeaux, France
2 INRIA Bordeaux Sud-Ouest, Talence, France

The individual data collected throughout patient follow-up constitute crucial information for assessing the risk of a clinical event, and eventually for adapting a therapeutic strategy. Joint models and landmark models have been proposed to compute individual dynamic predictions from repeated measures to one or two markers. However, they hardly extend to the case where the complete patient history includes much more repeated markers possibly. Our objective was thus to propose a solution for the dynamic prediction of a health event that may exploit repeated measures of a possibly large number of markers. We combined a landmark approach extended to endogenous markers history with machine learning methods adapted to survival data. Each marker trajectory is modeled using the information collected up to landmark time, and summary variables that best capture the individual trajectories are derived. These summaries and additional covariates are then included in different prediction methods. As we need to handle a possibly large dimensional history, we rely on machine learning methods adapted to survival data, namely including regularized regressions and survival random forests, to predict the event from the landmark time, and we show how they can be combined into a superlearner. Performances of the prediction tools are evaluated by cross-validation using estimators of Brier Score and the area under the Receiver Operating Characteristic curve adapted to censored data. We demonstrate in a simulation study the benefits of machine learning survival methods over standard survival models, especially in the case of numerous and/or non-linear associations between the predictors and the event. We illustrate the methodology in a public health context with the prediction of death in the general elderly population at different ages using multiple markers of aging (depressive symptoms, cognitive functions, dependency, ...). Our methodology enables the prediction of an event using the entire individual longitudinal history, even when the number of repeated markers is large. Although introduced with mixed models for the repeated markers and survival methods for a single right censored time-to-event, the same methodology can be used with any other appropriate modeling technique for the markers and can be easily extended to competing risks setting.

**References:** *Ishwaran H, Kogalur UB, Blackstone EH and Lauer MS (2008). Random survival forests. The Annals of Applied Statistics, 2(3), 841-860., Van Houwelingen HC (2007). Dynamic Prediction by Landmarking in Event History Analysis. Scandinavian Journal of Statistics, 34(1), 70-85.*



# SESSION OC3C

### OC3C-2 Comparison of multiple dynamic predictive accuracies

**Clémence Moreau[1], Jérémie Riou[2,3], Marine Roux[1]**
1 HIFIH, UPRES 3859, SFR 4208, Angers University, France
2 MINT, UMR INSERM 1066, CNRS 6021, Angers University, France
3 Methodology and Biostatistics Department, Delegation to Clinical Research and Innovation, Angers University Hospital, France

In the clinical environment, due to recent technological advances, more and more information can be collected during the patient follow-up. Dynamic models are particularly well adapted for the analysis of this type of data, allowing some potential changes in the follow-up to be taken into account. In particular, this makes it possible to obtain more accurate predictions by updating the available information throughout the patient monitoring. Blanche et al. [1] have developed some mathematical tools to quantify and compare the effectiveness of dynamic predictions. Dynamic versions of the Area Under the ROC Curve (AUC) and the Brier score are used for quantification and some tests are provided for comparison. Nevertheless, only two predictions can be compared, which may be too restrictive in a clinical context. Here, we propose a new procedure, based on dynamic AUC or Brier score of Blanche et al. [1], which allows multiple comparisons. First, we want to assess if at least one of the accuracy of the considered dynamic predictions differs from the others at a fixed prediction horizon time. Under the null, we proved that our statistical tests converge to a gamma law. Then, if the test is significant, post-hoc tests were conducted to find where the differences occurs, which entails a multiplicity issue. To address this point, based on Blanche et al. [2], the Shaffer's procedure was used to strongly control the Family-Wise Error Rate (FWER). Performances of our new testing procedure were assessed by simulations. For two predictions, a power closed to Blanche et al. [1] was reached. In all studied scenarios, the FWER control was checked. Moreover, a motivating application in hepatology will be presented. We aim at identifying the most appropriate biomarker to predict liver-related complications of patients with liver fibrosis. The new procedure will select it among a set of candidate biomarkers while controlling the probability to make at least one false discovery. Finally, this work allows to compare more than two dynamic predictive accuracies and will available through an R package as soon as these results are published.

**References:** *[1] Blanche P, Proust-Lima C, Loubère L, Berr C, Dartigues JF, Jacqmin-Gadda H. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. Biometrics 2015; 71(1): 102–113. [2] Blanche P, Dartigues JF, Riou J. A closed max-t test for multiple comparisons of areas under the ROC curve. Biometrics 2020.*

### OC3C-3 Spatio-temporal score driven modeling of resting state fMRI data

**Francesca Gasperoni[1], Alessandra Luati[2], Lucia Paci[3], Enzo D'Innocenzo[2]**
1 MRC Biostatistics Unit, University of Cambridge, United Kingdom
2 Department of Statistical Sciences, University of Bologna, Italy
3 Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Milan, Italy

Resting state functional Magnetic Resonance Imaging (R-fMRI) data are receiving an increasing attention both in the clinical and the statistical research. Indeed, R-fMRI signals represent the neuronal activity that is intrinsically generated by the brain (Fox and Raichle, 2007) and have been already described as the candidate tool for getting deeper insights on neural spontaneous activations that cannot be explained by external stimuli or structural connectivity. However, statistical models capable of detecting spontaneous activation are still limited. Furthermore, the available studies and a relevant part of the fMRI literature relies on the often-untested Gaussian assumption, Eklund et al. (2016). In this work, we provide a novel statistical model that goes beyond the Gaussian assumption and addresses the complex nature of R-fMRI data at the same time. Specifically, we introduce a spatial simultaneous autoregressive score driven model with multivariate Student-t distributed errors. The model specification lies on a novel spatio-temporal filter that delivers robust estimates of the time varying blood oxygenation level dependent (BOLD) signal. As a model by-product, we develop a procedure for detecting spontaneous activations, based on the assumption that they correspond to residuals peaks (in line with the clinical literature) of a possibly heavy tailed distribution. It is important to highlight that the proposed model can collapse to a classical spatial autoregressive (SAR) model with Gaussian distributed errors, as the Student-t degrees of freedom parameter tends to infinity. Inference estimation is based on the method of maximum likelihood and asymptotic theory is developed. We evaluate the whole procedure through an extensive simulation study. To conclude, we apply the proposed model on the R-fMRI data coming from the pilot study of the Enhanced Nathan Kline Institute-Rockland Sample project. The data consists of multi-subject brain imaging data (fMRI and Diffusion Tensor Imaging, DTI) collected on 70 Region of Interests based of the Desikan Atlas. We exploited the information on structural connectivity by defining a subject-specific spatial weight matrix based on DTI. We run subject-specific analysis and we show the obtained results via dynamic activations brain images.

**References:** *Eklund, A., T. E. Nichols, and H. Knutsson (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proceedings of the National Academy of Sciences 113 (28), 7900-7905. Fox, M. D. and M. E. Raichle (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nature Reviews Neuroscience 8 (9), 700.*

# SESSION OC3C

## OC3C-4 Breast cancer risk prediction in mammography screening cohorts: an approach based on modeling tumour onset and growth

**Rickard Strandberg, Kamila Czene, Per Hall, Keith Humphreys**
Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

Mammography screening programmes are aimed at reducing mortality due to breast cancer by detecting tumours at an early stage. There is currently interest in moving away from the age-based screening programmes, and towards personalised screening based on individual risk factors. To accomplish this, risk prediction models for breast cancer are needed to determine who should be screened, and when.
We use a Swedish cohort to predict the short-term risk of breast cancer, based on a number of established risk factors, by using a (random effects) continuous growth model. It jointly models breast cancer tumour onset, tumour growth rate, symptomatic detection rate, and screening sensitivity. Unlike existing breast cancer prediction models, this approach can account for each woman's individual screening history in the prediction.
In addition to predicting the short-term risk of breast cancer, this model can make separate predictions regarding specific tumour sizes, and the mode of detection (e.g. detected at screening, or through symptoms between screenings). It can also predict how these risks change depending on whether or not a woman will attend her next screening. In our study, we predict that the probability of having a tumour less than 10mm diameter when detected is increased by 140%, on average, if a woman in the cohort attends their next screening. This indicates that the model can also be used to evaluate the short-term benefit of screening attendance, on an individual level.

## OC3C-5 Accounting for improvements in survival when developing risk prediction models in a competing risks setting

**Sarah Booth[1], Richard D. Riley[2], Joie Ensor[2], Paul C. Lambert[1,3], Mark J. Rutherford[1]**
1 Biostatistics Research Group, Department of Health Sciences, University of Leicester, United Kingdom
2 Centre for Prognosis Research, Research Institute for Primary Care and Health Sciences, Keele University, Newcastle, United Kingdom
3 Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

**Introduction:** Risk prediction models are often developed with data from patients diagnosed across a long time period. If there have been improvements in survival over this time, not accounting for the temporal trend can lead to predictions which over-estimate the risk for recently diagnosed patients.
**Methods:** Temporal recalibration was proposed to address this issue. This method involves first developing the risk prediction model using the standard approach and then using delayed entry techniques to re-estimate the baseline in a recent time window with the most recent data. This allows improvements in baseline survival to be captured. Here we show an example of how this method can be applied in a competing risks setting by fitting a cause-specific hazard model to each cause of death and temporally recalibrating each model separately. This allows more up-to-date cause-specific cumulative incidence functions to be estimated for each cause of death, improving the risk predictions for new patients. This process is illustrated using an example of survival following a diagnosis of colon cancer, where the event of interest is death from colon cancer and the competing event is death from other causes. These models are fitted using cancer registry data from the United States Surveillance, Epidemiology and End Results (SEER) database.
**Results:** Using the standard approach and not accounting for the improvements in baseline survival led to predictions which over-estimated the risk of death from cancer and other causes for more recently diagnosed patients. However, the calibration of the risk predictions for these patients was improved by using temporal recalibration.
**Conclusion:** Temporal recalibration can easily be applied in a competing risks setting by updating the baseline of each cause-specific hazard model to take account of improvements in survival in each cause. This can lead to more up-to-date and accurate risk predictions for patients who are currently being diagnosed.
**References:** *Booth, S.; Riley, R. D.; Ensor, J.; Lambert, P. C.; Rutherford, M. J. Temporal recalibration for improving prognostic model development and risk predictions in settings where survival is improving over time. International Journal of Epidemiology 2020; 49(4):1316-1325.*

# SESSION OC3D

SESSION OC3D — Meta-analysis for prediction models

## OC3D-1 COVID-PRECISE: A living methodological review of prediction models for diagnosis and prognosis of covid-19

**Wynants Laure[1,2], Ben Van Calster[2,3], Gary S. Collins[4,5], Richard D. Riley6, Ewout W. Steyerberg[3], Georg Heinze[7], Luc J.M. Smits[1], Karel G.M. Moons[8,9], Maarten van Smeden[8], *the COVID-PRECISE group (Precise Risk Estimation to optimise covid-19 Care for Infected or Suspected patients in diverse sEttings)***
1 Department of Epidemiology, CAPHRI Care and Public Health Research Institute, Maastricht University, Netherlands
2 Department of Development and Regeneration, KU Leuven, Belgium
3 Department of Biomedical Data Sciences, Leiden University Medical Centre, Netherlands
4 Centre for Statistics in Medicine, University of Oxford, United Kingdom
5 NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital, UK
6 Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Newcastle, UK
7 Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria
8 Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht University, Netherlands
9 Cochrane Netherlands, University Medical Centre Utrecht, Utrecht University, Netherlands

**Objective:** Critically appraise all diagnostic and prognostic models for individualized prediction of covid-19 risk, with a focus on methodology and statistics.
**Methods:** A living systematic review [1] of studies that developed or validated a multivariable covid-19 related prediction model for diagnosis or prognosis purposes, using any combination of predictors including demographic, clinical or imaging input data. Data sources included PubMed and Embase through Ovid, arXiv, medRxiv, and bioRxiv. At least two authors independently extracted data using CHARMS (critical appraisal and data extraction for systematic reviews of prediction model studies); risk of bias was assessed using PROBAST (prediction model risk of bias assessment tool) [1]. See www.covid-precise.org.
**Results:** 37 420 titles were screened, and 170 studies describing 236 prediction models were included. We identified 11 models for identifying people at risk in the general population; 118 diagnostic models for detecting covid-19 (75 were based on medical imaging, 10 to diagnose disease severity); and 107 prognostic models for predicting mortality risk, ICU admission, or other adverse outcomes. The models were build using logistic regression (34%), neural networks/deep learning (32%), tree-based methods (7%), Cox regression (6%), support vector machines (4%), or other methods (17%). Predictive performance of 212 newly developed models was evaluated with internal validation only (53%), external validation (24%), or neither (24%). 24 studies independently validated an existing prediction model. C-indexes range from 0.54 to 0.99. Risk of bias was low in 4 models, unclear in 6 models, and high in 226 models. Most common reasons for high risk of bias were insufficient data for the chosen modeling strategy (70%), inappropriate or incomplete evaluation of discrimination and calibration (68%) and inappropriately dealing with overfitting and optimism (53%).
**Conclusion:** Prediction models for covid-19 are quickly entering the academic literature to support medical decision making at a time when they are urgently needed. Most proposed prediction models are at high risk of bias, and their reported performance is probably optimistic. To date, six promising models warrant further validation. External validation using individual patient data from multiple cohorts is ongoing research by the COVID-PRECISE consortium. The review will be updated before the conference.
**References:** *[1] Wynants Laure, Van Calster Ben, Collins Gary S, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020; 369 :m1328.*
*[2] Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019;170:W1-33.*

# SESSION OC3D

## OC3D-2 A Bayesian model for heterogeneous treatment effects on the additive risk scale in meta-analysis

**Doranne Thomassen, Ewout Steyerberg, Saskia le Cessie**
Biomedical Data Sciences, Leiden University Medical Center, Netherlands

Faced with a newly diagnosed patient, clinicians consider which of the available treatments will provide the largest absolute risk reduction in their individual patient. To answer this question requires statistical methods that quantify heterogeneous 'personalized' treatment effects on the clinically relevant scale.
We propose a Bayesian (meta-)regression model for binary outcomes on the additive risk scale. The model allows treatment effects, covariate effects, interactions and variance parameters to be estimated directly on the scale of clinical interest. The model was applied in single trial analysis, meta-analysis and network meta-analysis of the TherapySelector (TS) dataset, containing 5,842 hepatitis C patients from 20 randomized trials. We compared our model to two other approaches: an alternative additive risk model (Warn et al., 2002)) and a logistic model that transforms predictions back to the natural scale after regression (Chalkou et al., 2020).
Some trials in the TS database have cure rates close to 100%, illuminating the main differences between the approaches. Our model is very sensitive to the effect of treatment at the boundaries of the risk parameter support [0,1]. This can be a strength and a weakness; on the TS data it was helpful. Patients with predicted risks close to 0 or 1 contribute little to posterior precision in the model by Warn et al., making it less suitable for trial arms with ~100% successes. In such cases, the logistic model sometimes produces extreme effect estimates, leading to instability in the network setting. A conceptual advantage of the additive models is that variance parameters capture heterogeneity on the scale of interest. On the other hand, it can be argued that meta-analysis should be done on the scale where there is least heterogeneity, the log(odds) scale for the TS data.
Their respective characteristics make the compared models suitable for different analysis settings. Therefore, our proposed model is a useful addition to the available statistical methods to model heterogeneous treatment effects on the additive risk scale.

**References:** *[1] D. E. Warn, S. G. Thompson, and D. J. Spiegelhalter (2002). Bayesian random effects metaanalysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. Statistics in Medicine, 21(11), 1601–1623.*
*[2] K. Chalkou, E. Steyerberg, M. Egger, A. Manca, F. Pellegrini, and G. Salanti (2020). A twostage prediction model for heterogeneous effects of many treatment options: application to drugs for multiple sclerosis. preprint, 04 2020. URL: https://arxiv.org/abs/2004.13464.*

# SESSION OC3D

## OC3D-3 Assessing risk of bias in individual participant data meta-analyses for prediction model research

**Brooke Levis[1], Kym I.E. Snell[1], Gary S. Collins[2], Thomas Debray[3], Karel G. Moons[3], Johannes B. Reitsma[3], Richard D. Riley[1]**
1 Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Newcastle, UK
2 Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, United Kingdom
3 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Netherlands

**Background:** Assessing risk of bias and applicability (RoB) of included studies is critical for interpreting meta-analysis results. In meta-analyes of multivariable prediction models, RoB can be assessed using PROBAST. However, individual participant data meta-analyses (IPDMAs) differ from aggregate-data MAs in that in IPDMAs, datasets may include additional information, eligibility criteria may differ from the original publications, definitions for predictors and outcomes can be standardized across studies, and analysis methods can be improved. Therefore, a tailored RoB tool may be needed.
**Objectives:** To review how RoB is currently assessed in IPDMAs of multivariable diagnostic or prognostic prediction model studies, and to preliminarily examine PROBAST, with the goal of developing an IPDMA extension (PROBAST-IPD).
**Methods:** We reviewed RoB assessments in IPDMAs of prediction model studies published from January 2018 to May 2020. We then examined how PROBAST items might be evaluated in an IPDMA context, noting which items might be removed, edited, or added; and we hypothesized how results may be incorporated into IPDMA analyses.
**Results:** Twenty-five prediction model IPDMAs were included. We observed that current IPDMAs rarely and inconsistently evaluate RoB of included IPD, and most do not incorporate RoB judgements into analyses. Our findings indicate using PROBAST to assess RoB in the IPD datasets themselves, rather than solely based on study publications. In initial considerations for developing PROBAST-IPD, we propose that certain items need to be evaluated and coded at the participant level (e.g., timing between predictor assessment and outcome determination), whereas others (e.g., quality of assessments tools) may apply uniformly to an included study. Most analysis items (e.g., pre-specification of variables for analysis) are no longer relevant, as the IPDMA researchers perform the analyses themselves. RoB results may be incorporated into analyses by conducting subgroup analyses among studies and participants with overall low RoB or by conducting formal interaction analyses with item-level RoB responses.
**Conclusions:** Development and dissemination of PROBAST-IPD will allow improved RoB assessments in IPDMAs of prediction model studies.

# SESSION OC3D

## OC3D-4 Using meta-analysis for external validation of prediction models in big data, accounting for competing risks

Lucinda Archer[1], Kym Snell[1], Joie Ensor[1], Constantinos Koshiaris[2], James Sheppard[2], Richard Riley[1]
1 Centre for Prognosis Research, School of Medicine, Keele University, Newcastle, United Kingdom
2 Nuffield Department of Primary Care Health Sciences, University of Oxford, United Kingdom

**Background:** In prediction modelling, any event occurring prior to the event of interest, and thus preventing that event from happening, is known as a competing risk. Prediction models must be properly developed to account for competing risks to avoid overestimating the risk for the event of interest. Upon validation of such prediction models, competing risks should also be accounted for to ensure reliable performance estimation.
**Aims:** To describe the methods used to externally validate a competing risk model in a large primary care database; discussing how the competing risk was accounted for in the meta-analysis of each performance statistic across GP practices.
**Methods:** We externally validated a multivariable model predicting Acute Kidney Injury (AKI) at 10 years, while accounting for the competing risk of death, in 3,805,366 eligible patients from the CPRD Aurum database. Missing data were multiply imputed. Meta-analysis techniques were used to examine heterogeneity in model performance across GP practices, where case-mix and outcome prevalence varied. Predictive performance was assessed using calibration plots, and measures of discrimination (D-statistic and Time-dependant AUC), calibration (Observed/Expected ratio) and clinical utility (Net Benefit).
**Results:** Accounting for the competing risk in model predictions required incorporating the log-log baseline Cumulative Incidence Function (CIF) of AKI at 10 years across imputations from the development data. Central to the validation analysis was incorporating the observed CIF from the validation data for comparison. This required calculation of imputation-specific CIFs to calculate performance statistics prior to pooling.
**Calibration** plots additionally required observed CIFs to be generated within subgroups: by magnitude of predicted probabilities for standard plots; and within individual GP practices for practice level calibration. Random-effects meta-analysis of GP level performance estimates showed considerable heterogeneity for calibration, with over-prediction of AKI at 10 years on average. Discrimination, however, was more consistent, with both D-statistic and C-statistic showing low heterogeneity and narrower distributions in funnel plots of value by practice size.
**Conclusions:** When validating competing risks models, the competing event must be accounted for by properly incorporating the CIF for the event of interest in all analyses. Meta-analysis techniques are then helpful to summarise predictive performance across data clusters.

# SESSION OC3E

SESSION OC3E — Multi-state model

## OC3E-1 MSMplus: A dynamic interactive web tool for presentation of multi-state model analysis results

Nikolaos Skourlis[1], Michael J. Crowther[1], Therese M-L. Andersson[1], Paul C. Lambert[2]
1 Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Solna, Sweden
2 Department of Health Sciences, University of Leicester, United Kingdom

**Multi-state** models are used in complex disease pathways to describe a process where an individual moves from one state to the next, taking into account competing states during each transition. In a multi-state setting, there are various measures to be estimated that are of great epidemiological importance. However, increased complexity of the multi-state setting and predictions over time for individuals with different covariate patterns may lead to increased difficulty in communicating the estimated measures. The need for easy and meaningful communication of the analysis results motivated the development of a web tool to address these issues. MSMplus is a publicly available web tool, developed in RShiny, and is primarily targeted to researchers conducting multi-state model analyses. The results from any multi-state model analysis are uploaded to the application in a pre-specified format. Through a variety of user-tailored interactive graphs, the application contributes to an improvement in communication, reporting and interpretation of multi-state analysis results as well as comparison between different approaches. The predicted measures that can be supported by MSMplus include, among others, the transition probabilities, the transition intensity rates, the length of stay in each state, the probability of ever visiting a state and user defined measures. Representation of differences, ratios and confidence intervals of the aforementioned measures are also supported. MSMplus is a useful tool that enhances communication and understanding of multi-state model analyses results. Further use and development of web tools should be encouraged in the future as a means to communicate scientific research.

## OC3E-2 Statistical models for the natural history of breast cancer, with application to data from a Milan cohort study

Laura Bondi[1], Marco Bonetti[2], Denitsa Grigorova[3], Antonio Russo[4]
1 Department of Decision Sciences, Bocconi University, Milan, Italy
2 Department of Social and Political Sciences, Dondena Research Center, and Bocconi Institute for Data Science and Analytics, Bocconi University, Milan, Italy
3 Big Data for Smart Society Institute and Faculty of Mathematics and Informatics, Sofia University, Bulgaria
4 UOC Osservatorio Epidemiologico, ATS-Milano, Italy

We develop a new class of multi-state models for the natural history of breast cancer, where the main events of interest are the start of asymptomatic detectability of the disease and the start of symptomatic detectability. The former kind of detection occurs through screening, while the latter through the insurgence of symptoms.
We develop a cure rate parametric specification that allows for dependence between the times from birth to the two events, and present the results of the analysis of data collected as part of a motivating study from Milan. Participants in the study had a varying degree of compliance to a regional breast cancer screening program. The subjects' ten-year trajectories have been obtained from administrative data collection performed by the Italian national health care system. We first present a tractable model for which we develop the likelihood contributions of the possible observed trajectories, and perform maximum likelihood inference on the latent process.
Likelihood based inference is not feasible for flexible models and we rely on a likelihood-free method, Approximate Bayesian Computation (ABC), for inference on such more flexible models. Issues that arise from the use of ABC for model choice and parameter estimation are discussed, with a focus on the problem of choosing appropriate summary statistics.
The estimated parameters of the underlying disease process allow for the study of the effect of different examination schedules (ages and frequencies for screening examinations) and different adherence patterns on a population of asymptomatic subjects.

# SESSION OC3E

## OC3E-3 Reevaluating dementia incidence trends: The critical role of adequate design and methodology

**Nadine Binder[1], Martin Schumacher[2]**

1 Institute of Digitalization in Medicine, Medical Center – University of Freiburg, Germany
2 Institute of Medical Biometry and Statistics, Medical Center – University of Freiburg, Germany

A seeming decline in dementia incidence in Western nations has been a topic of continuous debate resulting in a recently published analysis of data from seven population-based cohort studies (Wolters et al., 2020). Constructing several nonoverlapping 5-year epochs, the corresponding design and analysis closely follows a framework previously used within the Framingham Heart Study (FHS) cohort. However, we challenged the finding of the FHS cohort on the basis that bias may have resulted from the failure to adequately account for potential disease onset in the period between last observation and death. Re-analyzing the FHS data using spline-based analytic methods, we did not find convincing evidence for a decline in dementia incidence over the epochs (Binder et al 2019). Yet, there is further room for improvement. First, the classification of calendar time into 5-year epochs is both unnecessary and arbitrary; the conclusion of a linear decline in dementia incidence in the FHS data would not have held had e.g., a 4-year follow-up period been used (Binder et al, 2019). Second, two separate cohorts (the 'original' FHS cohort and a cohort of their offspring) were combined for analysis, which may be inappropriate if they differ markedly. A more suitable approach for analyzing the question of how dementia incidence has evolved over time in the FHS would be to consider the two cohorts separately, and to dispense with the epoch structure by analyzing age as the time scale. This results in the analysis of separate generations which are ageing over time and being subject to death in greater numbers over time without replenishment from younger participants. If missing dementia cases due to death lead to a bias, the effect of this would therefore be to underestimate the incidence of dementia cases to a greater extent over time. This problem requires the use of statistical methods based on the illness-death multi-state model, such as spline-based penalized likelihood as employed in our earlier study. We will present the findings of the proposed design and analysis strategy, aiming for a realistic quantification of the dementia incidence trend in the Framingham Heart Study cohort.

**References:** *[1] Wolters FJ, Chibnik LB, Waziry R, et al. Twenty-seven-year time trends in dementia incidence in Europe and the United States: The Alzheimer Cohorts Consortium. Neurology. 2020;95(5):e519-e531. [2] Binder N, Balmford J, Schumacher M. A multi-state model based reanalysis of the Framingham Heart Study: Is dementia incidence really declining? Eur J Epidemiol. 2019;34(11):1075-1083.*

## OC3E-4 Statistical inference for transition probabilities in non-Markov multi-state models subject to both random left-truncation and right-censoring

**Alexandra Niessl[1], Arthur Allignol[1], Jan Beyersmann[1], Carina Müller[2]**

1 Institute of Statistics, Ulm University, Germany
2 Metronomia Clinical Research GmbH, Munich, Germany

The Aalen-Johansen estimator generalizes the Kaplan-Meier estimator for independently left-truncated and right-censored survival data to estimate the transition probability matrix of a time-inhomogeneous Markov model with finite state space. Such multi-state models have a wide range of applications for modelling complex courses of a disease over the course of time, but the Markov assumption may often be in doubt. If censoring is entirely unrelated to the multi-state data, it has been suggested that the Aalen-Johansen estimator still consistently estimates the state occupation probabilities. This approach has been extended to transition probabilities using landmarking, which is, inter alia, useful for dynamic prediction. We complement these findings in three ways. Firstly, we provide a rigorous proof of consistency of the Aalen-Johansen estimator for state occupation probabilities, correcting and simplifying the earlier result. Secondly, delayed study entry is a common phenomenon in observational studies, and we extend the earlier results to multi-state model data also subject to left-truncation. Thirdly, our proof is suggestive of wild bootstrap resampling. Studying wild bootstrap is motivated by the fact that it is desirable to have a technique that works both with non-Markov models subject to random left-truncation and right-censoring and with Markov models where left-truncation and right-censoring need not be entirely random. In our motivating data example, the occurrence and the impact of Methicillin-resistant staphylococcus aureus (MRSA) infection in hospital compared to patients only colonized with MRSA, is investigated using an illness-death multi-state model. Violations of the Markov assumption arise if the time of MRSA infection affects the hazard of end of hospital stay. Patients may have a delayed study entry if a positive laboratory result is only available some time after admission. We use landmarking to compare the residual length of stay of those in the infectious state with those still in the initial state of colonization. We present both the results of the real data example and the results of simulation studies showing that the landmark Aalen-Johansen estimator performs well and the wild-bootstrap provides confidence intervals close to the nominal level.



SESSION OC3E

## OC3E-5 Harmonization of endpoints in ICU Trials using multi-state modelling

**Maja von Cube[1], David Hajage[2], Virginie Lemiale[3], Didier Dreyfuss[4], Stephane Gaudry[5], Jerome Lambert[6]**

1 Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Germany
2 Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, AP-HP; Sorbonne Université, Hôpital Pitié Salpêtrière, Département de Santé Publique, Centre de Pharmacoépidémiologie (Cephepi), CIC-1421, Paris, France
3 Medical Intensive Care Unit, Hôpital Saint Louis, AP-HP, Paris, France
4 Medical Intensive Care Unit, Hôpital Louis Mourier, AP-HP, Paris, France
5 Intensive Care Unit, Hôpital Avicenne, AP-HP, Paris, France
6 INSERM U1153 CRESS, Epidemiology and Clinical Statistics for Tumor, Respiratory, and Resuscitation Assessments (ECSTRRA) Team, Paris, France

Several randomized trials in ICUs assess the effect and/or the ideal timing of organ support therapy (OST, e.g. ventilatory support, renal replacement therapy, or extracorporeal membrane oxygenation). In many cases, the aim is not only to reduce mortality, but also to reduce the duration of OST to prevent iatrogenic consequences. OST can also be part of the outcome of ICU trials assessing the effect of different drugs. For example, most of the randomized trials evaluating Covid-19 treatments have based their primary outcome on the WHO clinical progression scale, an ordinal scale for the different levels of care (including different levels of oxygenation support). These outcomes are evaluated with various methods, like Kaplan-Meier estimation or Cox model for time to event endpoints (e.g. time to improvement), logistic models or proportional odds models for the proportion of patients in a given state, competing risk models for length of stay, or calculation of OST-"free days" for the duration without OST.

Recently, multi-state modelling has been proposed as a simple and direct way to analyze such data, by defining each category as a distinct state. It allows a wide range of estimands important from a patient perspective and also from a hospital administration perspective, for planning of resources. The course of disease can be presented by a stacked probability plot which illustrates the probability of being in each specific state over time. Multi-state modelling accommodates problems of competing endpoints, non-monotonic patient trajectories (e.g. multiple intermittent episodes of OST) and censoring. Based on the states occupation probabilities, mean length of stay in each state, or in relevant combination of different states, can be calculated, and compared between arms.

Our objective is to highlight how multi-state modelling can complete traditional analyses, provide additional insights and unifies heterogeneous analysis strategies between trials. This will be illustrated using data from 2 published randomized controlled trials in ICU settings: a trial comparing early and delayed renal replacement therapy in patients with severe acute kidney injury, and a trial comparing early non-invasive ventilation to oxygen alone in patients with non-hypercapnic acute hypoxemic respiratory failure.

# SESSION OC3F

## SESSION OC3F — Neural network and machine learning

### OC3F-1 Neural networks for survival prediction in medicine: a review and critical appraisal

**Georgios Kantidakis[1,2,3], Audinga-Dea Hazewinkel[1,2,4], Marta Fiocco[1,2,5]**
1 Mathematical Institute (MI), Leiden University, Netherlands
2 Biomedical Data Sciences, Leiden University Medical Center (LUMC), Netherlands
3 Statistics, European Organization for Research and Treatment of Cancer (EORTC), Brussels, Belgium
4 MRC Integrative Epidemiology Unit, Bristol Medical School, United Kingdom
5 Trial and Data Center, Princess Máxima Center for Pediatric Oncology (PMC), Netherlands

**Context:** Prediction models with machine learning are becoming ubiquitous in the medical field. Over the years, an increasing number of algorithms have been developed and adapted to deal with censored data. Here, we consider publications that predicted survival with artificial neural networks (ANN), one of the most widely used machine learning techniques in healthcare applications.
**Objective(s):** A structured overview is presented, which provides a comprehensive understanding of the current literature. We discuss how researchers have used ANN to fit survival data for prediction in the medical field, and critically appraise which aspects of the models should be designed and reported more carefully.
**Methods:** We performed a global search in PubMed, considering articles published in the period 1990-2019. Additional studies were identified using a "local search" to follow citations of citations. Relevant manuscripts were classified as methodological/technical (novel methodology or new theoretical model) or applications. We identified key characteristics of prediction models (i.e. number of patients/predictors, evaluation measures, validation, calibration), and compared ANN's predictive performance to that of the Cox proportional hazards model.
**Results:** Our search yielded 217 studies. Of these, 13 methodological studies and 10 practical applications (7 real-world data, 3 simulations) were considered relevant. We identified two methodological trends; either time was added as part of the input features of the ANN and a single output node was specified, or multiple output nodes were defined for each time interval. The median sample size was 920 patients, and the median number of predictors was 7. Major findings in the 23 studies included poor reporting (e.g., regarding missing data, hyperparameters), use of improper performance measures, as well as inaccurate model development / validation. Calibration was neglected in more than half of the studies. Cox models were not developed to their full potential and claims for the performance of ANN were exaggerated.
**Conclusions:** Light is shed on the current state of art of survival neural networks in medicine. Recommendations are made for the correct application of clinical prediction models with ANN. Limitations are discussed, and future directions are proposed for researchers who seek to develop existing methodology.
**References:** *[1] Ripley, B. D., & Ripley, R. M. (2001). Neural networks as statistical methods in survival analysis. Clinical Applications of Artificial Neural Networks, 237–255. [2] Wang P, Li Y, Reddy CK. (2019). Machine learning for survival analysis: A survey. ACM Computing Surveys (CSUR), 51(6), 1-36. https://doi.org/10.1145/3214306.*

# SESSION OC3F

### OC3F-2 Comparison of imputation methods that solve granularity problem resulting from healthcare structured data integration

**Alexia Sampri[1], Nophar Geifman[1], Philip Couch[1], Niels Peek[1,2]**
1 Division of Informatics, Imaging and Data Sciences, University of Manchester, United Kingdom
2 NIHR Manchester Biomedical Research Centre, The University of Manchester, Manchester Academic Health Science Centre, United Kingdom

**Background:** Disparate collections of data are inevitably heterogeneous and have made aggregation a difficult challenge. In this paper, we focus on the issue of content heterogeneity in data integration due to granularity i.e when some datasets and/or variables include more categories/levels and subsets than others. Traditional approaches map all source datasets to a common data model that includes only low level items, and thus omit all items that vary between datasets.
**Objectives:** Our focus is on the integration of structured data and solving the granularity problem by keeping the highest level items and therefore use all the available information. We assume that each one of these datasets that needed to be integrated consists of a single table; and that each of these datasets describes a disjoint set of entities. Therefore, record linkage is not needed.
**Methods:** From a probabilistic perspective, imperfect alignment of different data sources is not problematic as long as we can derive which information each of the sources provides in answering our research question. In our case we would like to use all the available information across the datasets being integrated. The general idea behind our integration method is that the problem of content heterogeneity, presented as granularity problem, could be translated into a missing value problem and then solved using well established methods (imputation).
**Results:** We perform a simulation study designed to investigate our probabilistic methods in a simpler and general setting to evaluate the suggested method. We also illustrate and solve granularity problem with the proposed probabilistic data integration approaches on dataset examples provided by MASTERplans. MASTERplans aims to improve care for Systemic Lupus Erythematosus patients by taking a precision medicine approach to identifying groups of patients that respond to particular biologic therapies.
**Conclusions:** Our approach insist on the future existence of health data heterogeneity. Our probabilistic data integration approaches are pragmatic because they always provide an answer. Evaluation and application's results show that the suggested probabilistic data integration approaches outperform traditional data integration approach and provide similar results to the true models.

# SESSION OC3F

## OC3F-3 Survival Predictions and Uncertainty Measures with Censored Data

**Elvire Roblin[1,2], Paul-Henry Cournede[2], Stefan Michiels[1]**

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

2 Univeristy of Paris-Saclay, CentraleSupélec, MICS - Laboratory of Mathematics and Computer Science, Paris, France

**Background:** Neural Networks have been increasingly used for prediction problems. As they are flexible non-linear models, they may be relevant when many candidate covariates and complex interactions are to be evaluated.
**Objectives:** This work aims to study neural network models with time-to-event data, using specific ways to handle censorship to study their operating characteristics in a simulation study, to introduce auto-encoders in the models and evaluate uncertainty in model predictions.
**Methods:** We compared survival models based on neural networks with different loss functions: Cox-MLP (Kvamme et al., 2019) uses a loss based on a case-control approximation. DeepHit (Lee et al.,2018) combines a log-likelihood with a ranking loss. DNNSurv (Zhao et al., 2019) uses pseudo observations. We also proposed other ways of computing pseudo-observations. We used random survival forests by Yshwaran et al. (2008) and lasso penalization as a benchmark. We simulated data from the AFT model proposed by Friedman et al. (2001), with 3 different censoring rates (20%, 40%, and 60%). We simulated 100 datasets of 1,000 samples and 20 variables each, with pairwise interactions and non-linear effects of random subsets of these. We built an oracle model for comparison purposes. We further applied the methods to 2 real datasets: the METABRIC breast cancer data set, including 1,960 patients, 6 clinicopathological covariates, and the expression of 1,000 genes and a set of data on lung cancer, consisting of 4,120 patients, 3 clinical variables, and 1,000 genes. We investigated the effect of a pre-training using Variational Auto-encoders (Simidjievski et al., 2019) on model's predictions. We also studied the predictive uncertainty of our models using MC Dropout (Gal et al., 2016) and ensembling methods (Dietterich et al., 2000).
**Results:** In the simulation study, we obtained the highest c-indices and lower integrated Brier score with CoxTime for low censoring and pseudo-discrete with high censoring. On the METABRIC data, different neural networks models obtained comparable 5-year and 10-year discrimination performances, but with slightly lower values than random survival forests and penalized Cox model. Detailed results from the lung cancer data will be shown at the conference.

## OC3F-4 Parametric and non-parametric variable selection methods for predictive modeling with binary response

**Johannes Vey, Dorothea Kronsteiner, Meinhard Kieser**

Institute of Medical Biometry and Informatics, University Hospital Heidelberg, Germany

The demand for statistical models to support medical decision making is increasing, especially when aiming to predict an outcome of interest of an individual patient. Clinical outcomes can be of diagnostic or prognostic nature and might be affected by various patient and disease characteristics. A key concern in building predictive models is to identify influential variables in regard to predicting the outcome as accurately as possible. Many modeling techniques can deal with large numbers of variables. However, the inclusion of irrelevant variables can cause overfitting, introduce noise, and might lead to a decrease in prediction performance. To address this issue, a broad range of variable selection methods has been developed. Widely applied parametric methods are, e.g., stepwise regression and regularised regression. These methods assume a linear relationship between the outcome and the variables, hence their performance might decrease in case of non-linear associations. Non-parametric tree-based methods provide variable selection based on variable importance measures and are able to deal with non-linear relations as well as with highly correlated and interacting variables. While random forests theoretically have a low risk to overfit, the number of iterations and the learning rate of gradient boosting algorithms play a critical role regarding potential overfitting. In general, the choice of one method is not trivial and depends on the between-variable relations in the data.The optimal number of boosting iterations can be estimated by computing the cross-validation prediction error in each iteration. This results in achieving a model with optimal performance and prevents overfitting but does not necessarily perform variable selection. To tackle this issue, we developed a step-by-step approach to perform variable selection using gradient boosting trees. We performed simulation studies to evaluate the mentioned methods regarding the selection of true influential variables and the performance of their final models in clinical relevant scenarios. The main focus was to investigate strengths and weaknesses to provide recommendations for the choice of method in common clinical data scenarios, because the set of selected variables depends on the chosen method. Moreover, the methods were compared using a real data set for predicting neurological improvement as binary response [1].

**Reference:** *1 Weyland et al., 2021. Predictors for failure of early neurological improvement after successful mechanical thrombectomy in the anterior circulation. Stroke. accepted*

# SESSION OC3F

## OC3F-5 Performance measures for assessing machine learning algorithms in clinical trials

**Ian White[1], Louise Brown[1], Matteo Quartagno[1], Carlos Diaz Montana[1], Karla Diaz-Ordaz[2,3], Ashwini Venkatasubramaniam[2], Elizabeth C. George[1], Chris Holmes[2]**

1 MRC Clinical Trials Unit at UCL, London, United Kingdom

2 Alan Turing Institute, London, United Kingdom

3 Medical Statistics Department, LSHTM, London, United Kingdom

Machine learning (ML) methods have great promise but have been little used in clinical trials. They may be of particular benefit in assessing treatment effect heterogeneity, where they aim to find unexpected patient subgroups that may benefit even from treatments that are of little benefit overall. ML methods have previously been applied to identify treatment effect modifiers in clinical trials, but translation of findings into clinical practice has been poor. This lack of clinical impact may be because ML outputs are difficult to interpret by non-ML specialists; it may also be because clinical trials are usually powered for main effects, and thus ML methods can lack power and may find false positives. Simulation studies are needed to evaluate the methods, but it is not clear how to do these. This work explores what the estimands and performance measures should be for future simulation studies. We first consider simple data generating mechanisms where the treatment effect is determined by a single variable. Here, type 1 error rate and power are suitable performance measures, but this setting is too simple to demonstrate the potential of ML methods. For realistically complex data generating mechanisms, one possible estimand is the set of individual treatment effects, and performance can be measured by the mean squared error, averaged over individuals. However, if clinical treatment choices are to be driven by the trial, then some errors are more important than others. We therefore propose as an estimand the subgroup who would benefit from treatment. Performance measures here could be as simple as the sensitivity and specificity of the estimated benefiting subgroup. However, a better performance measure quantifies the clinical benefit of treating according to the ML results, compared with treating according to a simpler rule. We call this clinical utility and illustrate its properties, compared with other performance measures, in a simple simulation study based on the TRACT trial comparing blood transfusion volumes in children with and without fever being treated for severe anaemia in Africa. Initial results show that models with smaller mean squared error tend to have larger clinical utility, but marked reversals can occur.

# SESSION OC3G

## SESSION OC3G — Study designs

### OC3G-1 Using Historical Data to Predict Health Outcomes – The Prediction Design

**Stella Erdmann, Manuel Feißt, Johannes Krisam, Meinhard Kieser**
Institute of Medical Biometry and Informatics, University of Heidelberg, Germany

The gold standard for investigating the efficacy of a new therapy – the randomized controlled trial (RCT) – is costly, time-consuming, and not always feasible in a reasonable time frame. At the same time, huge amounts of available control condition data of previous RCTs or real-world data (RWD) in analyzable format are neglected, if not often completely ignored [1]. To overcome this shortcoming, alternative study designs with more efficient data use would be desirable. Assuming that the standard therapy and its mode of functioning are well known and large amounts of patient data exist, it is possible to set up a prediction model to determine the treatment effect of this standard therapy for future patients. If a new therapy is to be tested against the standard therapy, the vision would be to conduct a single-arm study and use the prediction model to determine the effect of the standard therapy on the outcome of interest for patients receiving only the test treatment, rather than setting up a two-arm study for this comparison. While the advantages of using historical data to estimate the counterfactual are obvious, bias could be caused by confounding or a number of other data issues that could compromise the validity of the nonrandomized comparison [2]. To investigate whether and how such a design – the prediction design – could be used to provide information on treatment effects using existing infrastructure and data sources (historical data from RCTs and/or RWD), we explored the assumptions under which a linear regression model could be used to predict the counterfactual of patients accurately enough to construct a test to assess the treatment effect for normally distributed outcomes. To overcome the implications of violating the model assumptions, the use of robust methods (e.g., "robust linear regression," LASSO) was explored. This was applied to a data set comparing liraglutide and sitagliptin in type II diabetes. Simulations were used to examine the amount of data needed on historical data as well as for the single-arm study. Depending on the amount of available historical data, the sample size could be reduced compared to a conventional RCT.

**References:** *[1] Rosenfeld, N., Y. Mansour, and E. Yom-Tov. Predicting counterfactuals from large historical data and small randomized trials. in Proceedings of the 26th International Conference on World Wide Web Companion. 2017. [2] Ellenberg, S.S. and J.H. Ellenberg, Proceedings of the University of Pennsylvania 12th annual conference on statistical issues in clinical trials: Electronic health records (EHR) in randomized clinical trials—Challenges and opportunities. 2020, SAGE Publications Sage UK: London, England.*



SESSION OC3G

### OC3G-2 Bayes Factors for Equivalence, Non-inferiority, and Superiority Designs Using baymedr

**Maximilian Linde, Don van Ravenzwaaij**
Department of Psychometrics and Statistics, Faculty of Behavioural and Social Sciences, University of Groningen, Netherlands

Biomedical research often seeks to determine the equivalence, non-inferiority, or superiority of an experimental condition (e.g., a new drug) compared to a control condition (e.g., a placebo). The use of frequentist statistical methods, in the form of null hypothesis significance testing (NHST), to analyze data for these types of designs is ubiquitous. However, frequentist inference has several limitations. Among the most critical limitations of NHST are the inability to quantify evidence in favor of the null hypothesis and the necessity of inflexible adherence to a predetermined sampling plan. Bayesian inference remedies these shortcomings and allows for intuitive interpretations. We present the R package baymedr (available at https://cran.r-project.org/web/packages/baymedr/; Linde & van Ravenzwaaij, 2019) and an associated Shiny web application for the computation of Bayes factors for equivalence, non-inferiority, and superiority designs (see also van Ravenzwaaij, Monden, Tendeiro, & Ioannidis, 2019). These two sources allow using either raw data or summary statistics. The R package baymedr and the web application have a focus on user-friendliness and are especially intended for researchers who are not statistical experts but can also be used by (clinical) statisticians. The web application can be utilized without any programming experience. We explain and compare the frequentist and Bayesian conceptualizations of equivalence, non-inferiority, and equivalence designs and showcase baymedr and the associated web application by analyzing raw data and by reanalyzing existing empirical studies using the published summary statistics.

**References:** *[1] Linde, M., & van Ravenzwaaij, D. (2019). baymedr: An R package for the calculation of Bayes factors for equivalence, non-inferiority, and superiority designs. arXiv: 1910.11616 [stat.OT] [2] van Ravenzwaaij, D., Monden, R., Tendeiro, J. N., & Ioannidis, J. P. A. (2019). Bayes factors for superiority, non-inferiority, and equivalence designs. BMC Medical Research Methodology, 19(1), 71. doi: 10.1186/s12874-019-0699-7*

### OC3G-3 Forecast Alzheimer's disease progression to better select patients for clinical trials

**Etienne Maheux, Juliette Ortholand, Igor Koval, Arnaud Valladier, Stanley Durrleman**
Inria, Aramis project-team, Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS, UMR 7225, Sorbonne Université, Paris, France

**Context:** Clinical trials focusing on neurodegenerative diseases face the difficulty of subtle, long-term and individual-specific worsening of the endpoints. Eventually, they end up recruiting heterogeneous patients that prevents from showing any drug effect, especially if the latter is believed to be more effective at a precise disease stage.
**Objective:** Targeting the right patients during trial screening is a way to reduce the needed sample size or conversely to improve the proven effect size.
**Methods:** From Alzheimer's disease (AD) observational cohorts, we selected longitudinal data that matched AD trials (inclusion and exclusion criteria, trial duration and primary endpoint). We modelled EMERGE, a phase 3 trial in pre-clinical AD, and a mild AD trial, using 4 research cohorts (ADNI, Memento, PharmaCog, AIBL) totalling more than 5500 individuals. For each patient, we simulated its treated counterpart by applying an individual treatment effect. It consisted in a linear improvement of outcome for effective decliners, calibrated to match the expected trial effect size. Next, we built a multimodal AD course map that grasps long-term disease progression in a mixed-effects fashion [1] with Leaspy (open-source Python package). We used it to forecast never-seen individuals' outcomes from their screening biomarkers. Based on these individual predictions, for each trial we selected a clinically relevant [2] sub-group of screened patients. Finally, we compared the effective sample size that would had been needed for the trial, with and without our selection. The dispersion was evaluated using a bootstrap procedure.
**Results:** For all investigated setups and cohorts, our selection enabled to decrease the needed sample size. In particular, in the EMERGE (resp. mild AD) trial, selecting patients with a predicted CDR-SoB change between 0.5 and 1.5 points per year (resp. MMSE change between 1 and 2 points per year) reduced the sample size by 38.2 ± 3.3 % (resp. by 38.9 ± 2.2%).
**Conclusions:** In AD clinical trials, using our forecasts of individual outcomes from multimodal screening assessments as an extra inclusion criterion allows to better control trial population and thus to reduce the needed sample size for a given treatment effect.

**References:** *[1] Schiratti J-B, Allassonniere S, Colliot O, Durrleman S. A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. In Journal of Machine Learning Research (JMLR) 18(1):4840-4872. 2017. [2] Galasko DR, Gould RL, Abramson IS, Salmon DP. Measuring cognitive change in a cohort of patients with Alzheimer's disease. In Stat Med 19(11-12):1421-1432. 2000.*

# SESSION OC3G

## OC3G-4 Design effects and analysis considerations for the split-mouth design with an unequal numbers of sites per patient

**Caroline Kristunas[1], Karla Hemming[2], Thomas Dietrich[3,4]**
1 Institute of Clinical Sciences, University of Birmingham, United Kingdom
2 Institute of Applied Health Research, University of Birmingham, United Kingdom
3 Birmingham Community Healthcare Trust, Birmingham, United Kingdom
4 Department of Oral Surgery, The School of Dentistry, University of Birmingham, United Kingdom

**Background:** Split-mouth studies are a common variant of within-person randomised trials, often conducted in dentistry. In a split-mouth design, sites (usually teeth) within the same person, are randomised to intervention or control. A split-mouth design has many similarities to a cluster-cross over design. Current sample size formulae for split-mouth designs assume all patients provide an equal number of sites. Having a varying number of sites per patient is similar to varying cluster sizes in a cluster trial which can accommodate varying cluster sizes. Analysis of split-mouth studies, prima-facie, seems similar to that of a cluster trials, for example, by including random effects for patients. However, in contrast to cluster trials, split-mouth studies have very small cluster sizes (number of sites per patient), many recruit only a small number of patients, and much larger correlation coefficients are often observed. For these reasons, the performance of both sample size and analysis methods for split-mouth studies is unknown.
**Objectives:** To determine an appropriate design effect for split-mouth studies with an unequal number of sites per patient and evaluate the performance of common methods for the analysis of split-mouth studies, when the number of patients and sites per patient are small.
**Methods:** Data were simulated from a linear mixed effects model for split-mouth studies varying in size (overall and average sites per patients), variability in sites per patients and intra-patient correlation coefficient. The sample size was estimated using design effects adapted from those used in cluster trials. Analyses were conducted using a linear mixed-effects regression, generalised estimating equations and cluster-level summaries, with the inclusion of small sample corrections. For each scenario, the power and coverage were assessed.
**Findings:** We present, using a motivating example, a simple design effect that can account for an unequal number of sites per patient in split-mouth studies and make recommendations for the analysis of split-mouth studies with a small number of patients and sites per patients.

# SESSION OC3G

## OC3G-5 Challenges of using big health data to identify patterns of anxiety and depression in multimorbid population

**Roma Puronaitė[1,2,3], Audronė Jakaitienė[1,3], Kristina Švaikevičienė[2,3], Greta Burneikaitė[2,3], Justas Trinkūnas[3,4], Vytautas Kasiulevičius[2,3], Edita Kazėnaitė[2,3]**
1 Faculty of Mathematics and Informatics, Institute of Data Science and Digital Technologies, Vilnius University, Lithuania
2 Faculty of Medicine, Vilnius University, Lithuania
3 Vilnius University Hospital Santaros Klinikos, Lithuania
4 Department of Information Systems, Vilnius Gediminas Technical University, Lithuania

**Clinical research problem:** Patients with multimorbidity tend to have less continuity of care than patients with a single condition. Unfortunately, multimorbidity is a highly prevalent condition and one in three patients with multimorbidity is having a coexisting mental health condition. Existing disease surveillance systems have not been used optimally to understand multimorbidity and its effects, or to guide effective action [1]. Therefore, investigation of mental health between multimorbid patients is very important.
Statistical challenges: We used health administrative data collected for administrative purposes not for primary but for secondary use of health research.
**The objective and statistical methods:** The aim of this work was to use a big health-administrative database to assess the frequency of anxiety and depression in patients with multimorbidity. For this purpose Lithuanian National Health Insurance Fund under the Ministry of Health administrative health data of 1 254 167 subjects with multimorbidity covering the period from 2014 till 2019 was analyzed. We used hierarchical clustering and exploratory factor analysis for cross-sectional phenotype identification.
**Results and Conclusions:** Patterns of anxiety and depression were identified and an unexpectedly small percentage of anxiety (3.9%) and depression (8.1%) were found among analyzed multimorbid patients. These findings may be related to mental health stigma and may be associated with the unwanted disease diagnostic code and other related causes. Even if general trends can be seen, conclusions need to be drawn carefully. Researchers planning to use data not only on their primary collection purpose, need to understand limitations of the database and data generating mechanism and collaborate with biostatisticians, clinicians and other experts. In addition, other study designs may be useful in further analyzing the clinical research problem.
**References:** *[1] Pearson-Stuttard, J., Ezzati, M., & Gregg, E. W. (2019). Multimorbidity—a defining challenge for health systems. The Lancet Public Health, 4(12), e599-e600.*

# SESSION OC4A

## SESSION OC4A — Missing data in causal studies

### OC4A-1 Sensitivity to MNAR dropout in clinical trials: use and interpretation of the Trimmed Means Estimator

**Audinga-Dea Hazewinkel[1,2], Jack Bowden[2,3], Kaitlin Wade[1,2], Tom Palmer[1,2], Nicola Wiles[4], Kate Tilling[1,2]**
1 Population Health Sciences, Bristol Medical School, University of Bristol, UK
2 Medical Research Council Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, UK
3 Exeter Diabetes Group (ExCEED), College of Medicine and Health, University of Exeter, UK
4 Centre for Academic Mental Health, Population Health Sciences, Bristol Medical School, University of Bristol, UK

Missing data is a common feature in randomized controlled trials (RCTs) and may result in biased inference. The impact of missing data depends on the missingness mechanism and the analysis model. When data are missing completely at random (MCAR) or missing at random (MAR), estimates from a complete case analysis (CCA) or multiple imputation (MI) will generally be unbiased. Outcome values may be missing not at random (MNAR), if patients with extreme outcome values are more likely to drop out (e.g., due to perceived ineffectiveness of treatment, or adverse effects). In such scenarios, CCA and MI estimates will be biased.
It is impossible to statistically verify whether data are MAR or MNAR. To increase confidence in the primary results, current practice recommends testing the robustness of the model assumptions by performing sensitivity analyses under plausible alternative assumptions. We propose using the trimmed means (TM) estimator [1] as a sensitivity analysis for clinical trial data with outcome value dropout, when there is cause to suspect an MNAR dropout mechanism.
The TM estimator operates by setting missing values to the most extreme value, and then "trimming" away equal fractions of both treatment groups, estimating the treatment effect using the remaining data. The TM estimator relies on two assumptions, which we term the "strong MNAR" and "location shift" assumptions. We derive formulae for the bias resulting from the violation of these assumptions for normally distributed outcomes, and demonstrate how these formulae can be used to inform sensitivity analyses.
We applied our method in a sensitivity analysis of the CoBalT RCT [2], which compares the effectiveness of cognitive behavioural therapy (CBT) as an adjunct to pharmacotherapy versus usual care in 469 patients with treatment resistant depression. Results were consistent with a beneficial CBT treatment effect. The MI estimates were closer to the null than the CCA estimate, whereas the TM estimate was further from the null.

**References:** *[1] Permutt T, Li F. Trimmed means for symptom trials with dropouts. Pharm Stat. 2017;16(1):20-28.*
*[2] Wiles NJ, Thomas L, Turner N, et al. Long-term effectiveness and cost- effectiveness of cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: follow-up of the CoBalT randomised controlled trial. Lancet Psychiatry. 2016;3(2):137- 144.*



# SESSION OC4A

### OC4A-2 Comparison of two causal inference methods for multiple treatments in clinical research on observational data

**François Bettega[1], Sébastien Bailly[1], Clémence Leyrat[2]**
1 HP2, University Grenoble Alpes, France
2 Department of Medical Statistics, London School of Hygiene & Tropical Medicine, United Kingdom

In medical research, a common and reliable way to show the causal link between exposure and an outcome is randomized clinical trial. However randomized clinicals trials are not always feasible to estimate causal effects, but a suitable analysis of observational data can lead to the estimation of causal effects. An increasingly popular method is inverse probability of treatment weighting (IPTW) estimator which is largely applied on binary exposure. However, while there is a rich literature on IPTW for binary exposure, its extensions to multiple exposures are scarce.
IPTW relies on the assessment of a weight for each individual, there are multiple methods to estimate these weights. The aim of this study is therefore to compare two of these methods: a simple multinomial logistic regression and generalized boosted models (GBM), to assess advantages and drawbacks of GBM for multiple treatment causal inference. Indeed, GBM are proven to provide more stable weights estimations than parametric models in simulations studies. Additionally, GBM do not require linearity assumption compared to logistic regression. In this work, we assess weight distribution by treatment groups and covariable balance after weighting. We also evaluate whether weighting methods affect treatment effect estimates. To do this, we use weighted linear regression to estimate the causal treatment effect and compare the point estimates and their variances between weighting the two methods. These results are also contrasted with the results of a crude and adjusted (multivariable regression) analyses. We compare both methods on a national prospective cohort of sleep. Obstructive sleep apnoea syndrome (OSAS) is a major health concern with multiples consequences, especially on patient's quality of life. Continuous positive airway pressure (CPAP), the first-line therapy for OSAS, is highly effective in terms of symptom improvement but depends on the level of adherence. Little is known on the impact of CPAP adherence on OSAS subjective sleepiness assessed using the Epworth scale.
We show that distributed weights were similar with both methods, but the computational time was longer for GBM, which gives a substantial advantage to multinomial regression weight estimators.

### OC4A-3 Handling missing data for causal effect estimation in longitudinal cohort studies using Targeted Maximum Likelihood Estimation: a simulation study

**S. Ghazaleh Dashti[1], Katherine J. Lee[1,2], Julie A. Simpson[3], Ian R. White[4], John B. Carlin[1,2], Margarita Moreno-Betancur[1,2]**
1 Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Melbourne VIC, Australia
2 Clinical Epidemiology and Biostatistics Unit, Department of Paediatrics, University of Melbourne, Melbourne VIC, Australia
3 Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne VIC, Australia
4 MRC Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, UK

Causal inference from longitudinal cohort studies plays a pivotal role in epidemiologic research. One of the available methods for estimating causal effects is Targeted Maximum Likelihood Estimation (TMLE), which is a doubly robust method, combining a model for the outcome and a model for the exposure and only one of the two has to be consistent to obtain unbiased estimates. It also offers asymptotically valid confidence intervals even when these models are fitted using machine learning approaches, which allow the relaxation of parametric assumptions. However, it is unclear how missing data should be handled when using TMLE with machine learning, which is problematic given that missing data are ubiquitous in longitudinal cohort studies and can result in biased estimates and loss of precision if not handled appropriately. We sought to evaluate the performance of currently available approaches for dealing with missing data when using TMLE. These included complete case analysis, an extended TMLE method in which a model for the outcome missingness mechanism is incorporated in the procedure, the missing indicator method for missing covariate data, and multiple imputation (MI) using standard parametric approaches or machine learning algorithms to concurrently handle missing outcome, exposure and covariate data. Based on motivating data from the Victorian Adolescent Health Cohort Study, we conducted a simulation study to evaluate the performance (bias and precision) of these approaches for estimation of the average causal effect. We considered a simple setting, where the exposure and outcome were generated from main-effects regression models, and a complex setting, where the models also included two-way and higher order interactions. Our results aim to provide guidance for handling missing data in a range of missingness scenarios depicted using causal diagrams. We illustrate the practical value of these findings in an example examining the effect of adolescent cannabis use on young adulthood mental health.

# SESSION OC4A

## OC4A-4 Estimands in clinical trials: making the hypothetical strategy concrete

**Camila Olarte Parra[1], Rhian Daniel[2], Jonathan Bartlett[3]**
1 Department of Mathematical Sciences, University of Bath, United Kingdom
2 Division of Population Medicine, Cardiff University, United Kingdom
3 Department of Mathematical Sciences, University of Bath, United Kingdom

The ICH E9 guidelines addendum introduces the term intercurrent event to refer to events that happen after randomisation and that can either preclude the outcome of interest or affect the interpretation of the treatment effect. The addendum proposes 5 options for sensible targets of inference (i.e. 5 estimands) in the presence of intercurrent events, but does not suggest statistical methods for their estimation. In this talk, we focus on estimands defined using the hypothetical strategy, where the treatment effect is estimated under the hypothetical scenario in which we (somehow) intervene to prevent the intercurrent event from occurring. To estimate a hypothetical estimand, we consider methods from causal inference (G-computation and inverse probability of treatment weighting) and missing data (multiple imputation and mixed models). We establish that certain 'causal inference estimators' are identical to certain 'missing data estimators'. These links may help those familiar with one set of methods but not the other. Moreover, they allow us to transparently show using potential outcome language the assumptions missing data methods are relying on to estimate hypothetical estimands. We also present Monte Carlo simulations that provide evidence of the performance of the methods in different settings including varying rates of occurrence of the intercurrent event, intercurrent events happening at different time points during follow-up and the presence of the intercurrent event affecting the outcome.

## OC4A-5 Incorporating baseline covariates to validate surrogate endpoints with a constant biomarker under control arm

**Emily Roberts, Michael Elliott, Jeremy M. G. Taylor**
Department of Biostatistics, University of Michigan, Ann Arbor MI, United States

A surrogate endpoint S in a clinical trial is an outcome that may be measured earlier or more easily than the true outcome of interest T. In this work, we extend causal inference approaches to validate such a candidate surrogate using potential outcomes. The causal association paradigm assesses the relationship of the treatment effect on the surrogate with the treatment effect on the true endpoint. Using the principal surrogacy criteria, we utilize the joint conditional distribution of the potential outcomes T, given the potential outcomes S. Let S(z) and T(z) refer to the endpoint values had the treatment, possibly counter-factually, been assigned to level z. We build upon previous models for the joint distribution of potential outcomes that assume multivariate normality among the endpoints S(0), S(1), T(0), T(1) under a binary treatment. In particular, our setting of interest allows us to assume the surrogate under the placebo, S(0), is zero-valued, and we incorporate baseline covariates. Having rich baseline patient characteristic data may improve the quality of the surrogacy assessment. First, conditioning on covariates may improve the plausibility of conditional independence assumptions, and second, it allows us to make inference about whether there are subgroups of the population for whom the quality of the surrogate varies. We develop Bayesian methods to incorporate conditional independence and other modeling assumptions and explore their impact on the assessment of surrogacy. We demonstrate our approach via simulation and data that mimics an ongoing study of a muscular dystrophy gene therapy where the primary outcome is a continuous functional score. Since muscular growth and deterioration from disease have major impact on mobility, both baseline ambulatory ability (which is measured pre-treatment) and age are important to take into consideration to evaluate surrogacy. Based on our simulations, our validation method suggests that the proposed surrogate, micro-dystrophin expression, will only be valid for a subgroup of younger patients (four years of age). The trial of interest will also include a cross-over portion where placebo subjects receive the experimental treatment mid-trial, and we consider modeling these additional endpoints in the validation framework.

# SESSION OC4B

SESSION OC4B — Cluster randomized trials

## OC4B-1 Sample size calculation for stepped wedge cluster randomized trials with multiple levels of clustering

**Kendra Davis-Plourde[1], Monica Taljaard[2,3], Fan Li[4,5]**
1 Department of Internal Medicine, Yale University, New Haven CT, United States
2 Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa ON, Canada
3 School of Epidemiology and Public Health, University of Ottawa, Ottawa ON, Canada
4 Department of Biostatistics, Yale University, New Haven CT, United States
5 Center for Methods in Implementation and Prevention Science, Yale University, United States

The stepped wedge cluster randomized trial is an attractive design for evaluating health services delivery or policy interventions. In this design, clusters start in the control condition and gradually cross over to the treatment based on a schedule dictated by random assignment. Outcomes may be assessed on the same individuals over time (i.e., a cohort design) or different individuals (i.e., a cross-sectional design). A key consideration in this design is that sample size calculation and analysis must account for within-period as well as between-period intracluster correlations; cohort designs have additional correlations due to repeated measures on the same individuals. While numerous methods have been developed to account for within- and between-period intracluster correlations with a single level of clustering during each time period, few methods are available to accommodate multiple levels of clustering. Our objectives were to develop computationally-efficient sample size procedures that recognize within-period and between-period intracluster correlations in stepped wedge trials with more than two levels of clustering. Focusing on three levels of clustering and assuming equal cluster-period sizes, we consider three variants, depending on whether each level is treated as a cross-sectional or closed-cohort design. We introduce an extended block exchangeable matrix to characterize the correlation structures both within- and between-clusters in each cluster-period and develop convenient sample size expressions that depend on this correlation structure. With a continuous outcome, we show the sample size expression depends on the correlation structure only through two eigenvalues of the extended block exchangeable matrix. For binary outcomes under a mixed effects framework, we develop a sample size expression based on a first-order Taylor approximation. We conduct simulation studies to examine the finite-sample properties of the proposed sample size algorithms and demonstrate the application of the proposed methods using the Washington State Expedited Partner Therapy trial: a multilevel stepped wedge trial that randomized local health jurisdictions (level 4) consisting of clinics (level 3) and observed patients (level 2) with respect to their Chlamydia infection status (level 1).

# SESSION OC4B

## OC4B-2 Stepped-wedge cluster randomised trials with binary outcomes and small numbers of clusters: a case study

**Christina Easter[1], Jennifer Thompson[2], Karla Hemming[1]**
1 Institute of Applied Health Research, The University of Birmingham, UK
2 Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, UK

The stepped-wedge cluster randomised trial remains a novel study design yet is becoming a more popular design choice. Randomisation is at the level of the clusters, with clusters randomly transitioning from control to treatment at a set number of 'steps', and then remaining exposed to the treatment for the duration of the study. The average number of clusters randomised is 17 [Inter-quartile range, 8-38], just over 50% have a binary primary outcome, often with low prevalence, and more than 50% are analysed using generalised linear mixed models (GLMM) (Martin, 2018). CONSORT reporting guidelines recommend both absolute (e.g. risk difference) and relative (e.g. risk ratio) measures of treatment effects are reported. Methods of analysis therefore need to allow estimation of relative and absolute measures of effect for binary outcomes, possibly with low prevalence, and with a small number of clusters.
In linear mixed models, both maximum likelihood and restricted maximum likelihood methods produce a downward bias in estimated variance parameters; and standard Wald tests do not provide nominal levels of coverage, when there are a small number of clusters. Small sample corrections, including Satterthwaite and Kenward-Roger corrections, are therefore recommended in the setting of parallel cluster trials. These corrections are sometimes used with GLMMs despite their performance being less well documented. In the setting of logistic regression, alternative simple corrections to the degrees of freedom might be sufficient (Li & Redden, 2015). To our knowledge there has been no evaluation of small sample corrections of GLMM with binomial distribution and log or identity link (to report relative risks and risk differences), a setting where model convergence if often problematic with low prevalence.
In this talk, we illustrate the choice of methods available for data-analysis of a stepped-wedge trial conducted in 18 intensive care units (the clusters) with a binary outcome with low prevalence, considering the choice of small sample corrections, degrees of freedom corrections, and availability in standard statistical software. We illustrate that whilst desirable to report risk differences, models often fail to converge. This illustrative case study will form the prelude to a simulation study investigating these properties more widely.

**References:** *[1] Li, P. & Redden, D. T. (2015) Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analyzing binary outcome in small sample cluster-randomized trials. BMC Med Res Methodol, 15, 38. [2] Martin, J. (2018) Advancing knowledge in stepped-wedge cluster randomised controlled trials. Ph.D. thesis, University of Birmingham.*

## OC4B-3 Inference for the treatment effect in longitudinal cluster randomized trials when treatment effect heterogeneity is ignored

**Rhys Bowden, Andrew Forbes, Jessica Kasza**
Department of Epidemiology and Preventive Medicine, Monash University, Melbourne VIC, Australia

Longitudinal cluster randomized trials, such as the stepped wedge, can sometimes have a treatment whose effect varies between clusters, often known as treatment effect heterogeneity. Treatment effect heterogeneity is not usually accounted for in outcome regression models, perhaps due to the additional complexity of doing so. Until now, the effect of failing to account for treatment effect heterogeneity when it is present has only be studied in a limited set of scenarios, via simulation.
In this work, we provide an analytical approximation for the impact of failing to include treatment effect heterogeneity, in particular on the variance of the treatment effect estimator when outcomes are continuous. We use this to highlight and explain what influence the design and design parameters such as numbers of clusters, number of time periods and number of observations have on the error introduced by this form of model misspecification.



# SESSION OC4B

## OC4B-4 Cluster randomised trials and a small number of clusters: Analysis method for a binary outcome

**Jennifer A. Thompson[1], Clemence Leyrat[2], Katherine Fielding[1], Richard Hayes[1]**
1 International Statistics and Epidemiology Group, London School of Hygiene & Tropical Medicine, UK
2 Medical Statistics Department, London School of Hygiene & Tropical Medicine, United Kingdom

Cluster randomised trials (CRTs) are often designed with a small number of clusters, but it is not clear which analysis methods are optimal when the outcome is binary.
There are three types of analysis: cluster-level analysis (CL), generalised linear mixed models (GLMM), and generalised estimating equations with sandwich variance (GEE). We conducted a broad simulation study to determine (i) whether these approaches maintain acceptable type-one error, if so (ii) which methods are most efficient, and (iii) the impact of non-normality of cluster means on these approaches.
We simulated CRTs with 8-30 clusters in total, mean cluster-size from 10-1000, varying and common cluster-size, control arm prevalence of 10% or 30%, intracluster correlation coefficient from 0.001-0.1, and cluster means following a normal, gamma, or uniform distribution. We ran 1000 repetitions of each scenario. We analysed each dataset with weighted and unweighted CL; GLMM with adaptive Gauss-Hermite quadrature and restricted pseudolikelihood; GEE with Kauermann-and-Carroll and Fay-and-Graubard sandwich variance using independent and exchangeable working correlation matrices. All methods compared test statistics to a t-distribution with degrees of freedom (DoF) as clusters minus cluster-level parameters. For GLMM pseudolikelihood, we also calculated Satterthwaite and Kenward-Rogers DoF.
Unweighted CL maintained type-one error<6.4% in 854/864(99%) scenarios. GLMM pseudolikelihood with clusters minus parameters DoF controlled type-one error in 853/864(99%) scenarios. Other DoF were more conservative. Fay-and-Graubard GEE with independent working correlation matrix controlled type one error in 808/864(94%) scenarios. Exchangeable correlation results were similar. Other methods had poorer type-one error control. Cluster-mean distribution did not affect analysis method performance.
GEE had the least power. Compared to CL, with 20 or more clusters, GLMM tended to have greater power with varying cluster-size but similar power with common cluster-size. With fewer clusters, GLMM had less power with common cluster-size, similar power with medium variation in cluster-size, and greater power with large variation in cluster-size.
We recommend that CRTs with ≤30 clusters and a binary outcome use an unweighted CL or restricted pseudolikelihood GLMM both with DoF clusters minus cluster-level parameters. The methods and findings are illustrated by application to a CRT of an intervention to increase adherence to Tuberculosis medication.

# SESSION OC4B

## OC4B-5 Under what conditions do open-cohort cluster RCTs provide improvements over conventional designs? A simulation study

**Laura E. Marsden[1], Andrew Copas[2], Amanda Farrin[1], Rebecca Walwyn[1]**
1 Leeds Institute of Clinical Trials Research, University of Leeds, United Kingdom
2 MRC Clinical Trials Unit at University College London, United Kingdom

**Background:** DCM-EPIC [1], a care home cluster-randomised trial (CRT), had ~45% unavoidable drop-out of residents after 16 months. Institutions such as care homes, schools and hospitals can be viewed as 'open cohorts', because individuals move in and out of them over time. There are currently two established designs for parallel-group CRTs in open cohorts where outcomes are measured repeatedly over time. Closed-cohort (CC) designs recruit individuals at baseline who are followed over time. (Repeated) cross-sectional (R-CS) designs allow recruitment post-randomisation, sampling one or more cross-sections of individuals at different time points. CC designs can assess individual change over time but are limited by the drop-out of individuals, which introduces missing data and bias, and affects generalizability. R-CS designs are more robust to drop-out but can only provide population-wide inference at specific time points. Although DCM-EPIC was designed as CC, high levels of drop-out warranted a design change during the trial. We propose the open-cohort (OC) CRT design as a potential solution. In this hybrid of the existing designs, a sample of individuals in each cluster are followed over time, with further recruitment replacing individuals who drop out. The OC-CRT design could be attractive to trialists as change at both individual and population-level can be estimated.
**Objectives:** To determine whether the OC-CRT design provides improved precision and bias compared to the existing designs over a range of study parameters and realistic complications.
**Methods:** Open cohort data will be simulated using various longitudinal multilevel models. Study parameters to be varied include the design, ICC, number/size of clusters and number of follow-ups, amongst others. Complications include the level of selection bias from post-randomisation recruitment, drop-out mechanism, turnover rate of individuals and more.Datasets will be analysed using two models; Kasza's single-timescale model [2] and a new extension, which includes an additional timescale.
**Results:** Simulation results are under review and will be presented.
**Conclusion:** Open-cohort designs have the potential to be superior to existing CRT designs when clusters have a moderate to high turnover of individuals, as in care homes, and can address a wider range of research questions.
**References:** *[1] Surr, C.A., Holloway, I., Walwyn, R.E., Griffiths, A.W., Meads, D., Kelley, R., Martin, A., McLellan, V., Ballard, C., Fossey, J., Burnley, N., Chenoweth, L., Creese, B., Downs, M., Garrod, L., Graham, E.H., Lilley-Kelley, A., McDermid, J., Millard, H., Perfect, D., Robinson, L., Robinson, O., Shoesmith, E., Siddiqi, N., Stokes, G., Wallace, D. and Farrin, A.J., 2020. Dementia Care Mapping™ to reduce agitation in care home residents with dementia: the EPIC cluster RCT. Health technology assessment (Winchester, England), 24(16), p.1. [2] Kasza, J., Hooper, R., Copas, A. and Forbes, A.B., 2020. Sample size and power calculations for open cohort longitudinal cluster randomized trials. Statistics in medicine, 39(13), pp.1871-1883.*

# SESSION OC4C

SESSION OC4C — Meta-analysis

## OC4C-1 Comparison of frequentist and Bayesian methods for two-arm borrowing of historical data

**Jinran Zhan, Nigel Stallard**
Warwick Medical School, University of Warwick, Coventry, United Kingdom

The slow progress of drug development and the high costs associated with clinical trials urgently call for more innovative clinical trial design and analysis methods to reduce development costs and patient burden. To address this problem, a potential strategy could be to supplement data from a current clinical trial with existing data from relevant historical studies. This so-called extrapolation or borrowing is particularly valuable when the recruitment of patients is difficult due to ethical, logistical or financial reasons, for example in trials in paediatric or rare diseases.
The main issue associated with the use of historical data is the potential for inflation of the type I error rate. This means that it is important to choose the right extrapolation method, ensuring that the amount of strength borrowed from the historical study is appropriate and is adjusted to the agreement between the two trials with the aim of increasing the power of the current trial whilst at the same time controlling the type I error rate.
A number of frequentist and Bayesian statistical methods have been proposed for borrowing historical control-arm data [1]. However, there is relatively little research on borrowing information from both the control and treatment arms of a single historical two-arm trial [2]. In this work, we extend static and dynamic borrowing methods proposed for the control-arm borrowing, including the test-then-pool, Bayesian power prior, commensurate prior and meta-analytic-predictive prior methods, to the setting of two-arm borrowing. These methods are then evaluated in simulation studies investigating a two-arm trial with a binary outcome to find appropriate borrowing parameters whilst optimising the trade-off between type I error and power.
Our simulation studies show that the degree of type I error inflation is mainly affected by the historical rate difference. Dynamic borrowing approaches are shown to offer better control of the type I error inflation over a wide range of scenarios, with the choice of borrowing parameters playing an important role.
**References:** *[1] Viele, K., Berry, S., Neuenschwander, B. et al. Use of historical control data for assessing treatment effects in clinical trials. Pharmaceutical Statistics. 2014; 13: 41-54. https://doi.org/10.1002/pst.1589*
*[2] Feißt, M, Krisam, J, Kieser, M. Incorporating historical two-arm data in clinical trials with binary outcome: A practical approach. Pharmaceutical Statistics. 2020; 19: 662– 678. https://doi.org/10.1002/pst.2023*

# SESSION OC4C

## OC4C-2 Implications of Analyzing Time-to-Event Outcomes as Binary in Meta-analysis

**Theodosia Salika, Rebecca Turner, David Fisher, Jayne Tierney, Ian R. White**
MRC Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, UK

**Background:** Systematic reviews and meta-analysis of time-to-event outcomes are frequently published within the Cochrane Database of Systematic Reviews (CDSR), however, these outcomes are handled differently across meta-analyses [1]. They can be analysed on the hazard ratio (HR) scale or can be dichotomized and analysed as binary outcomes using effect measures such as odds ratios (OR) [2]. We investigated the impact of reanalysing meta-analyses from the CDSR that used these different scales and using individual participant data (IPD).
**Methods:** We extracted two types of meta-analysis data from the CDSR either recorded in a binary form (A) or in binary form together with observed minus expected ("O-E") and variance ("V") statistics (B). We explored how results for time-to-event outcomes originally analysed as binary on an OR scale (A) change when analysed using the complementary log-log (clog-log) link on a HR scale. For the data originally analysed as HRs (B), we compared these results to analysing them as binary on a HR scale using the clog-log link or using a logit link on an OR scale. Additionally, using IPD meta-analyses, we compared results from analysing time-to-event outcomes as binary on an OR scale to analysing on the HR scale using the clog-log link, the log-rank approach or a Cox proportional hazards model.
**Results:** For both data types within the CDSR, approximately 19% of meta-analyses provided significant results under one scale and non-significant results under the other. Results from the log-rank approach and Cox proportional hazards model were almost identical; situations under which the clog-log link performed better than logit link and vice versa were apparent, indicating that the correct choice of the method does matter. Differences between scales arise mainly from the following reasons: (1) high event probability, (2) differences in between-study heterogeneity, (3) increased within-study standard error in the OR relative to the HR analyses, (4) percentage censoring, and (5) follow-up time.
**Conclusions:** We identified that dichotomising time-to-event outcomes may be adequate for low event probabilities and short term outcomes but not for high event probabilities; these findings provide guidance on the appropriate methodology that should be used when conducting such meta-analyses.

**References:** *[1] Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials [Electronic Resource]. 2007;8:16. [2] Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions: John Wiley & Sons; 2019.*

## OC4C-3 Exploring non-linear treatment-covariate interactions at multiple time points using multivariate IPD meta-analysis

**Miriam Hattle, Melanie A. Holden, Danielle van der Windt, Richard D. Riley**
School of Medicine, Keele University, Newcastle, United Kingdom

**Background:** Personalised medicine refers to how we tailor treatment decisions to each patient conditional on their characteristics. This requires research to identify interactions between treatment effect and patient-level covariates. An Individual Participants Data (IPD) meta-analysis allows us to better explore such interactions but challenges arise when included trials have multiple and missing follow-up time points and we aim to examine continuous covariates with potentially non-linear associations.
**Objectives:** To develop and apply a two-stage multivariate IPD meta-analysis model to estimate non-linear treatment-covariate interactions across multiple time points using IPD from multiple randomised trials with a continuous outcome.
**Method:** In the first stage, in each study separately, we model non-linear interactions by restricted cubic spline functions across multiple time points jointly, using longitudinal linear models to account for participant-level correlation in each trial. Knots are forced to be in the same location in each trial. In the second stage, we pool the study-specific spline function parameter estimates from all time-points simultaneously, using a multivariate meta-analysis that accounts for their within-study and between-study correlations.We apply the model to IPD from a large dataset of 31 trials that investigated covariates that interact with the effect of exercise interventions for the treatment of knee and/or hip osteoarthritis (STEER OA).
**Results:** The proposed method allows borrowing strength across multiple time points, and can handle participants and trials that are missing information at some time points. The results allow graphical displays of study-specific and summary non-linear interactions to help disseminate findings to clinicians and patients. In our application, baseline pain and baseline functional activity are found to have a non-linear interaction with the treatment effect on pain and function at 3 months and 12 months. This was masked when only considering linear trends.
**Conclusion:** Given IPD from multiple randomised trials we recommend exploring non-linear interactions across multiple time points using a two-stage multivariate IPD meta-analysis to account for correlations both at an individual level and across multiple time points.



# SESSION OC4C

## OC4C-4 A comprehensive framework for 'deft' (within-trials) interactions in meta-analysis

**Peter J. Godolphin, Ian R. White, Jayne F. Tierney, David J. Fisher**
MRC Clinical Trials Unit at University College London, Institute of Clinical Trials and Methodology, London, UK

A key question for meta-analyses is to reliably assess whether treatment effects vary across different patient groups. Traditionally, these interactions have been estimated using approaches known to induce aggregation bias, so we previously recommended a 'deft' (within-trials) approach to provide unbiased estimates for binary or ordered-categorical patient-level treatment-covariate interactions [1]. However, patients, clinicians and policy-makers also need to know the relative and absolute size of the overall treatment effect within each covariate subgroup, to target treatments appropriately. In this presentation, we extend the 'deft' methodology to a fully flexible framework to (1) estimate 'deft' interactions for covariates with multiple levels; (2) estimate a set of subgroup-specific treatment effects consistent with the 'deft' interactions; and (3) incorporate heterogeneity into the estimation of both interactions and subgroup effects, considering four distinct heterogeneity structures. These methods require relatively little information and can be applied to aggregate (or "published") source data, as well as individual patient data (IPD); and as such have wide practical application. We demonstrate a straightforward implementation in Stata with the existing user-written package "mvmeta". In a recent aggregate data meta-analysis investigating the effect of corticosteroids on mortality among critically-ill COVID-19 patients [2], we applied our methodology to a binary covariate: whether patients received invasive mechanical ventilation (IMV) at randomisation. Although a 'deft' interaction test was reported (p=0.0084), the published subgroup-specific effect sizes (IMV: OR=0.69, 95% CI 0.55 to 0.86; No IMV: OR=0.41, 95% CI 0.19 to 0.88) were at risk of aggregation bias. Using our methodology, we estimated subgroup effects, compatible with the 'deft' interaction, under a common-effect model (IMV: OR=0.73, 95% CI 0.58 to 0.92; No IMV: OR=0.19, 95% CI 0.07 to 0.50). In a further IPD example in lung cancer, we apply our methodology to a covariate with three levels, and derive absolute differences in survival. We compare the results and interpretations of our four different approaches to modelling heterogeneity, discuss the impact of trials which only contribute to a single subgroup, and propose recommendations for best practice.

**References:** *[1] Fisher DJ, Carpenter JR, Morris TP, Freeman SC, Tierney JF. Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach? BMJ 2017;356:j573. [2] The WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group. Association between administration of systemic corticosteroids and mortality among critically ill patients with COVID-19: a meta-analysis. JAMA 2020;324(13):1330-41*

## OC4C-5 Inclusion of real world data in surrogate endpoint evaluation: a Bayesian meta-analytic approach

**Anastasios Papanikos[1,2], Sylwia Bujkiewicz[1]**
1 Biostatistics Group, Department of Health Sciences, University of Leicester, Leicester, UK
2 GlaxoSmithKline R&D Centre, GlaxoSmithKline, Stevenage, UK

Surrogate endpoints play an important role in drug development when they can be used to measure treatment effect early compared to the final clinical outcome and to predict clinical benefit or harm. Meta-analysis provides a useful framework for combining evidence from multiple studies and can be used to evaluate a relationship between treatment effects on a surrogate endpoint and a final outcome.Traditionally, data from randomised controlled trials (RCTs) have been used to evaluate surrogate relationships as they achieve high internal validity. However, when few RCTs are available, meta-analysing sparse RCT data may affect the evaluation of surrogate endpoints as the estimates of the parameters describing a surrogate relationship can be obtained with considerable uncertainty and poor accuracy. In such circumstances, the inclusion of observational cohort studies (OBCs) can help to obtain more precise estimates of the parameters describing surrogate relationships as well as more precise predictions of clinical benefit. This can be crucial when policy decisions need to be made based on a surrogate endpoint and further experimentation may be lengthy or unfeasible due to budget constrains. In this paper, a new method for combining evidence from different sources is proposed to improve the evaluation of surrogate endpoints in circumstances where RCTs offer limited evidence. The method extends a model proposed by Begg and Pilote [1] to the bivariate case and allows for adjusting for systematic biases across different types of designs. This is important as the limited internal validity of OBCs can introduce bias to the estimates of the parameters describing surrogate relationships and potentially affect the evaluation of surrogate endpoints. A simulation study was carried out to assess the proposed method in various scenarios. We also applied the method to a data example in advanced colorectal cancer investigating the impact of combing RCTs with OBCs on the evaluation of progression-free-survival (PFS) as a surrogate endpoint of overall survival (OS). The inclusion of OBCs in the meta-analysis improved the evaluation of PFS as a surrogate endpoint of OS, resulting in reduced uncertainty around the estimates of the parameters describing the surrogate relationships and around the predicted effects on OS.

**Reference:** *[1] Colin B Begg and Louise Pilote. A model for incorporating historical controls into a meta-analysis. Biometrics, pages 899-906, 1991.*

# SESSION OC4D

## SESSION OC4D — Prediction model for omics data

### OC4D-1 Tailored Bayesian variable selection for risk prediction modelling under unequal misclassification costs

➜ *StCA AWARD WINNER*

**Solon Karapanagiotis[1], Oscar Rueda[1]*, Sach Mukherjee[1,2], Paul D.W. Kirk[1]*, Paul J. Newcombe[1]***
1 MRC Biostatistics Unit, University of Cambridge, United Kingdom
2 German Center for Neurodegenerative Diseases (DZNE), Germany
* Joint senior authors

**Background:** Risk prediction models are a crucial tool in healthcare. They are often constructed using methodology which assumes the costs of different classification errors are equal. However, in many healthcare applications, this assumption is not valid, and the differences between misclassification costs can be quite large. For instance, in a diagnostic setting, the cost of misdiagnosing a person with a life-threatening disease as healthy may be larger than the cost of misdiagnosing a healthy person as a patient. As a result, Tailored Bayes (TB) was proposed as a principled, simple and widely applicable umbrella framework to incorporate misclassification costs into Bayesian modelling [1]. Using both simulations and real data the authors showed that the TB approach allows us to "tailor" model development with the aim of improving performance in the presence of unequal misclassification costs.
**Objective:** To extend the TB framework by incorporating a variable selection procedure, a ubiquitous challenge in statistical modelling, especially, with the rise of high-dimensional data.
**Method:** We incorporate the TB approach into a hierarchical sparse regression framework and apply it to the METABRIC cohort (n = 1787). We investigate the clinical utility of already identified genes when their effects are modelled jointly, alongside routinely used clinicopathological covariates to predict 5-year risk of relapse in breast cancer. In total, we search over 1501 covariates. We compare the results between the TB and standard Bayesian (SB) modelling.
**Results and Conclusions:** We show that TB favours smaller models (with fewer covariates) compared to SB, whilst performing better or no worse than SB. This pattern was seen both in simulated and real data. This allows more parsimonious explanations for the data at hand. In addition, we show the ranking of covariates changes when we take misclassification costs into consideration. This has implications for risk prediction models since smaller models may result in lower data collection costs and different covariates used in further downstream analysis, for instance in genetic fine-mapping and related applications.
**Reference:** *[1] Solon Karapanagiotis, Umberto Benedetto, Sach Mukherjee, Paul DW Kirk, and Paul J Newcombe. Tailored Bayes: a risk modelling framework under unequal misclassification costs. Under review, 2021.*

### OC4D-2 Feature selection in multivariate varying-coefficient mixed models for drug response prediction

**Zhi Zhao[1,2], Manuela Zucknick[1], Tero Aittokallio[1,2,3]**
1 Oslo Centre for Biostatistics and Epidemiology (OCBE), University of Oslo, Norway
2 Institute for Cancer Research, Oslo University Hospital, Norway
3 Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland

Large-scale pharmacogenomic datasets often include multiple anti-cancer drugs, different cancer tissue types and heterogeneous multi-omics data. There are several challenges to model these data, such as those posed by correlated responses between multiple drugs, heterogeneity both between multiple tissues and between multi-omics data. We propose a multivariate varying coefficient mixed model which uses our IPF-tree-lasso method (Zhao and Zucknick, 2020) to take into account the drug- drug similarities and heterogeneity between multi-omics data. Importantly, the novel model employs random effects and varying coefficients to capture the underlying heterogeneity between multiple tissue samples. Simulation studies show that our proposed model improves the accuracy of drug response predictions and feature selection when comparing with existing lasso-type methods. We demonstrate the practical performance of our approach on a large preclinical pharmacogenomic study, the Cancer Therapeutics Response Portal (CTRP), where the model predicted the sensitivity of ca. 500 cancer cell line samples to ca. 200 drugs using ca. 10000 genomic features of the cell lines, including gene expression, copy number variation and point mutations.
**Reference:** *Zhao, Z. and Zucknick, M(2020). Structured penalized regression for drug sensitivity prediction. Journal of the Royal Statistical Society: Series C (Applied Statistics) 69, 525-545.*



SESSION OC4D

### OC4D-3 Variational Bayes for Model Averaging for Multivariate models using Compositional Microbiome predictors

**Alex Lewin, Darren Scott**
London School of Hygiene and Tropical Medicine, United Kingdom

High-throughput technology for molecular biomarkers produces multivariate data exhibiting strong correlation structures, and thus should be analysed in an integrated manner. Bayesian models are strongly suited to this aim. A particular case of interest is microbiome data, which is inherently compositional, and thus imposes a constraint on model space. A Bayesian model is presented for multivariate analysis of high-dimensional outcomes and high-dimensional predictors, including compositional microbiome predictors. The model includes sparsity in feature selection for predictors and covariance selection. A model averaging approach is taken to ensure robust selection of predictors. A hybrid Variational Bayes - Monte Carlo computational approach (following Ye et al. 2020) is used for the compositional data updates.
**Reference:** *Ye et al. 2020, Statistics and Computing volume 30, 887–905.*

### OC4D-4 Fast marginal likelihood estimation of penalties for group-adaptive elastic net

**Mirrelijn M. van Nee[1], Tim van de Brug[1], Mark A. van de Wiel[1,2]**
1 Epidemiology and Data Science, Amsterdam University Medical Centers, Netherlands
2 MRC Biostatistics Unit, University of Cambridge, United Kingdom

Nowadays, clinical research routinely uses omics, such as gene expression, for predicting clinical outcomes or selecting markers. Additionally, so-called co-data are often available, providing complementary information on the covariates, like groups of genes corresponding to pathways. Elastic net is widely used for prediction and covariate selection. Group-adaptive elastic net learns from co-data to improve prediction and selection, by penalising important groups of covariates less than other groups. Existing methods are, however, computationally expensive. Here we present a fast method for marginal likelihood estimation of group-adaptive elastic net penalties for generalised linear models. The method uses a low-dimensional representation of the Taylor approximation of the marginal likelihood and its first derivative for group-adaptive ridge penalties, to efficiently estimate these penalties. Then we show by using asymptotic normality of the linear predictors that the marginal likelihood for elastic net models may be approximated well by the marginal likelihood for ridge models. The ridge group penalties are then transformed to elastic net group penalties by using the variance function. The method allows for overlapping groups and unpenalised variables. We demonstrate the method in a cancer genomics application. The method substantially decreases computation time while outperforming or matching other methods by learning from co-data.

### OC4D-5 Improving model performance estimation in high-dimensional data settings by using learning curves

**Jeroen M. Goedhart, Thomas L.T. Klausch, Mark van de Wiel**
Department of Epidemiology and Biostatistics, Amsterdam University Medical Centers, Netherlands

In high-dimensional prediction settings, i.e. when p > n, it remains challenging to estimate the test performance (e.g. AUC). Especially for medical applications, e.g. predicting whether a certain therapy is successful, this should be done reliably. Arguably, the most widely used method is conventional K-fold cross-validation, which aims to balance between enough samples to learn the model and estimate its performance. We show that combining estimates from a trajectory of subsample sizes, rendering a learning curve [1], leads to several benefits. Firstly, use of a smoothed learning curve can improve the performance estimate compared to 10-fold cross-validation. Secondly, a still growing- or saturating learning curve indicates whether or not additional samples will boost the prediction accuracy. Thirdly, comparing the trajectories of different learners gives a more complete picture than doing so at one sample size only, which we demonstrate by evaluating a lasso-, ridge-, and random forest model. Fourthly, the learning curve allows computation of a lower confidence bound for the performance. Standard cross-validation produces very wide confidence bounds due to the small amount of test samples and the correlation structure between different training- and test splits. The learning curve finds a better trade-off between training- and test sample sizes, which leads to sharper bounds. This confidence bound is proven to be valid. We show coverage results from a simulation, and compare those to a state-of-the-art technique based on asymptotics [2]. Finally, we demonstrate the benefits of our approach by applying it to several classifiers of tumor location from blood platelet RNAseq data.
**References:** *[1] Mukherjee et al (2003). Estimating dataset size requirements for classifying DNA microarray data. Journal of computational biology, 10, 119-142 [2] LeDell, E. et al. (2015). Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. Electronic journal of statistics, 91, 1583-1607*

# SESSION OC4E

## SESSION OC4E — Competing risks and multi-state models

### OC4E-1 Bayesian inference for the direct approach for competing risk modeling with Gompertz distribution

**Tiemoko Kenneth Marie Galboni[1,2], Eric Agodio Bernard Dabone[1,2], Serge Manituo Aymar Somda[1,2]**
1 Université Nazi Boni, UFR Sciences et Techniques, Bobo-Dioulasso, Burkina Faso
2 Centre de Calcul, Centre MURAZ, Bobo-Dioulasso, Burkina Faso

The direct approach for competing risk modeling of survival data was proposed by Jeong and Fine (Jeong and Fine, 2006) to model simultaneously the cumulative incidence of several events individuals in a cohort are at risk of. This is a good alternative to the Fine and Gray method which were modeling one specific event in presence of risk of events which can prevent its realization. However, there are problems in estimating the maximum likelihood of this model. The form of the likelihood does not allow to determine global minima for most of the algorithms available in the software. In this article, we therefore propose a Bayesian inference approach, which will allow a better approach of the parameters.
Three different approaches were presented to evaluate our approach. We first used the Jeffrey's non-informative prior. The second prior distribution was the Zellener's maximal data information prior (MDIP). Finally, we tried independent gamma distributions. The three models were applied to simulated data. They were compared with a maximum likelihood estimation.
The maximum likelihood inference for the Gompertz competing risk does not offer guarantee of reaching the global minimum, thus providing the right estimations of the parameters. When the right prior is chosen, this estimation presents more accurate results. This version of the model can now be disseminated and even proposed in the more advanced statistical packages to permit good estimation of long-term survival models with multiple events.

**References:** *Jeong, J.-H. and Fine, J. P. (2006) 'Direct parametric inference for the cumulative incidence function', Applied Statistics, 55(2).*

### OC4E-2 Analysis of competing risks data using restricted mean time lost ➜ *CFCD AWARD*

**Hongji Wu[1], Hao Yuan[1], Zijing Yang[1], Yawen Hou[2], Zheng Chen[1],***
1 Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou, China
2 Department of Statistics, College of Economics, Jinan University, Guangzhou, China

In clinical and epidemiological studies, hazard ratios are often applied to compare treatment effects between two groups for survival data. For competing risks data, the corresponding quantities of interest are the subdistribution hazard ratio (SHR). However, the clinical applications and practices of SHR still have some limitations related to model assumptions and clinical interpretation. Therefore, an alternative statistic, restricted mean time lost (RMTL) [1-3], has been recommended for its intuitive and simplicity interpretation. However, the published researches of RMTL seem lack of robustness and completeness in statistical inference and practical application. Thus, we propose a new estimation and hypothetical test and sample size estimator based on the difference in RMTL (RMTLd). The simulation results show that the RMTLd test has robust statistical performance (both type I error and power). Meanwhile, the RMTLd-based sample size can approximately achieve the predefined power level. The results of the example analyses also verify the performance and acceptability of the RMTLd test. From the perspectives of clinical interpretation, application conditions and statistical performance, we recommend that the RMTLd be reported with the SHR when analyzing competing risks data and that the RMTLd even be regarded as the primary outcome when the proportional hazard assumption fails.

**References:** *[1] Andersen PK. Decomposition of number of life years lost according to causes of death. Stat Med. 2013;32(30):5278-85. [2] Zhao L, Tian L, Claggett B, et al. Estimating Treatment Effect With Clinical Interpretation From a Comparative Clinical Trial With an End Point Subject to Competing Risks. JAMA Cardiol. 2018;3(4):357-358. [3] Lyu J, Hou Y, Chen Z. The use of restricted mean time lost under competing risks data. BMC Med Res Methodol. 2020;20(1):197.*



SESSION OC4E

### OC4E-3 Impact of competing event in COVID-19 clinical data analysis ➜ *CFCD AWARD*

**Hao Yuan[1], Hongji Wu[1], Zijing Yang[1], Yawen Hou[2], Zheng Chen[1],***
1 Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou, China
2 Department of Statistics, College of Economics, Jinan University, Guangzhou, China

**Objective:** Many coronavirus disease 2019 (COVID-19) trials' researchers have calculated the sample sizes and tested hypotheses based on single time-to-event methods and selected clinical improvement or recovery as event of interest while death as right censoring. However, the sample sizes and the conclusions may be misleading.
**Statistical methods:** To compare competing risks methods with single time-to-event methods in calculating sample sizes and testing hypotheses at different competing event rates, we calculated the sample sizes and tested hypotheses based on eight reconstructed clinical trial datasets using competing risks methods (sub-distribution hazard, SHR and restricted mean time lost difference, RMTLd) and single time-to-event methods (hazard ratio, HR and restricted mean survival time difference, RMSTd). Monte Carlo simulations were conducted to compare differences in sample sizes and powers between competing risks methods and single time-to-event methods under different competing event rates.
**Results:** In four COVID-19 trials, the sample sizes based on competing risks methods were all higher than those based on single time-to-event methods. In the trials of Sharples and Imazio, the conclusions drawn based on the results from competing risks methods and those drawn based on the results from single time-to-event methods may be opposite. The simulation results show that the powers based on competing risks methods increase rapidly as the competing event rate increases. If powers were calculated based on the sizes of hazard ratios and competing risks methods, they might not reach the target and decrease as the competing event rate increases. In similar to COVID-19 studies, competing risks methods are recommended to calculate the sample size and test hypothesis while the event of interest and the competing event cannot be treated as a composite event, nor can the competing events be right censored.

**References:** *[1] Lyu J, Hou Y, Chen Z. The use of restricted mean time lost under competing risks data. BMC Med Res Methodol. 2020;20:197. [2] Wu H, Yuan H, Chen Z. Implementation of an alternative method for assessing competing risks: restricted mean time lost. 2021. Under Review*

### OC4E-4 Parametric Landmark estimation of the transition probabilities in survival data with multiple events

**Gustavo Soutinho[1], Luís Meira-Machado[2], Pedro Oliveira[1]**
1 EPIUnit, Institute of Health Public of University of Porto, Portugal
2 Department of Mathematics, University of Minho, Portugal

The estimation of transition probabilities is of major importance in the analysis of survival data with multiple events. These quantities play an important role in the inference in multi-state modeling providing in a simple and summarized manner long-term predictions of the process. Recently, de Uña-Álvarez and Meira-Machado (2015) proposed nonparametric estimators based on subsampling, also known as ladmarking, which have already proved to be more efficient than other nonparametric estimators in case of strong violation of the Markov condition. However, as the idea behind the landmarking is to use specific portions of data when the subsample sizes are reduced or in the presence of heavily censored data this may lead to higher variability of the estimates. To avoid the high variability of the nonparametric landmark estimator proposed by the de Uña-Álvarez and Meira-Machado (2015), we introduce parametric estimators for the transition probabilities that are also based on subsampling. We have considered several flexible distributions to handle this issue appropriately. One of the proposed approaches, which provides good results, with high flexibility, is based on the generalized gamma distribution.
Results of simulation studies confirm the good behavior of the proposed methods. We also illustrate and compare the new methods to the nonparametric landmark estimator through a real data set on colon cancer.

**References:** *[1] 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, 364–375. [2] Meira-Machado, L. and Sestelo M. (2019). Estimation in the progressive illness-death model: A nonexhaustive review. Biometrical Journal, 61:245–263, 2019. doi: 10.1002/bimj.201500038.*

# SESSION OC4E

## OC4E-5 Phase I/II dose-finding design for right censored toxicity endpoints with competing disease progression

**Anaïs Andrillon, Sylvie Chevret, Lucie Biard**
INSERM U1153 team ECSTRRA, Université de Paris, France

**Background:** The growing interest in new classes of anti-cancer agents, such as molecularly-targeted therapies (MTAs) and immunotherapies drugs with modes of action different from those of cytotoxic chemotherapies has changed the dose-finding paradigm. In particular, dose-finding designs should be able to handle the frequent late-onset toxicities by defining prolonged observation windows. In this setting, it is likely that the observation of late-onset toxicity endpoints may be precluded by trial discontinuation due to disease progression, defining a competing event to toxicity. Specific trial designs with prolonged observation windows, where dose-finding is modelled using survival models to handle right-censored endpoints in a competing risks framework, appear particularly suited.
**Objectives:** To propose a phase I/II dose-finding design using survival models for censored endpoints allowing the outcomes to be delayed and handling possible informative censoring by considering a competing-risks framework.
**Methods:** In these competing risks framework, we defined the cause-specific hazard for dose-limiting toxicity (DLT) and progression, both assumed exponentially-distributed and we estimated model parameters using Bayesian inference. For dose-finding, we targeted the cumulative incidences which are sub-distribution functions of time-to-DLT and time-to-progression. Given an observation window, the objective is to recommend the dose that minimizes the progression cumulative incidence, among an acceptable set of doses with DLT cumulative incidence inferior to a target threshold. In addition, we propose a nonparametric benchmark approach for evaluation of dose-finding designs with right-censored time-to-event endpoints. Design operating characteristics were evaluated in a simulation study, notably in terms of correct dose selection and safety, including sensitivity analysis to time-varying hazards of events and to different patient accrual schemes.
**Results:** The performance of the proposed methods was consistent with the complexity of scenarios as assessed by the nonparametric benchmark. We found that the proposed design present desirable operating characteristics, in particular in cases of non-negligible hazard of progression competing with DLT, compared to other existing phase I/II designs.
**Conclusion:** We propose a framework for seamless phase I/II trials targeting the subdistribution cumulative incidences of toxicity and progression for dose-finding, using working models for censored data. It allows prolonged observation windows resulting in administrative censoring and competing risks.



# SESSION OC5A

SESSION OC5A — Bayesian clinical trial analysis

## OC5A-1 Bayesian joint modeling of a bivariate toxicity for dose-regimens in early phase oncology

**Emma Gerard[1,2], Sarah Zohar[1], Moreno Ursino[1,3], Marie-Karelle Riviere[2]**
1 Inserm, Centre de Recherche des Cordeliers, Université de Paris, Sorbonne Université, Paris, France
2 Biostatistics department, Sanofi R&D, Chilly-Mazarin, France
3 F-CRIN PARTNERS platform, AP-HP, Université de Paris, France

**Context:** Most phase I trials in oncology aim to find the maximum tolerated dose (MTD) based on the occurrence of dose-limiting toxicities (DLTs). A DLT is a binary toxicity defined from multiple toxicity types and grades. Varying the dose-regimen, defined as the combination of the dose and the administration schedule, may reduce the risk of some types of DLT but may have a different effect on other toxicities. In a motivating trial, dose-regimens defined with intra-patient dose-escalation were administered to the patients to reduce the risk of cytokine release syndrome (CRS) while the effect on other toxicities ($DLT_o$) was unclear (NCT03594955).
**Objective:** The aim of the work was to propose a Bayesian method to evaluate the maximum tolerated dose-regimen (MTD-regimen) by modeling the DLT as a bivariate binary outcome to differentiate CRS from $DLT_o$. This method was developed to be applied at the end of the trial, once all data have been collected.
**Methods:** A Bayesian dose-regimen assessment method was used to model the CRS with the entire dose-regimen by incorporating pharmacokinetic and pharmacodynamic (PK/PD) modeling as proposed by Gerard et al (2020). Then, a Bayesian cumulative model was developed to model the $DLT_o$ with the dose-regimen, as no pharmacodynamic assumption was known. Finally, we considered three approaches to model the joint distribution of CRS and $DLT_o$: assuming independence between toxicities, adding a correlation parameter via copula modeling and via conditional modeling.
**Results:** Through an extensive simulation study, we observed that our joint approaches improved the proportions of selecting the true MTD-regimen in most scenarios compared to the recommendation of the dose-allocation method implemented (modified continual reassessment method), with a difference from 6.9% to 11.5%. Our joint approaches could also predict the DLT probabilities of new dose-regimens that were not tested in the study and that could be investigated in further stages of the trial.
**Conclusion:** We proposed a joint modeling approach to evaluate the effect of dose-regimens on two types of toxicity where one type of toxicity could be related to a PD biomarker, while no such assumption could be raised for the other one.
**References:** *Gerard E., Zohar S, Thai HT., Lorenzato C., Riviere MK. and Ursino M. (2020) Bayesian dose-regimen assessment in early phase oncology incorporating pharmacokinetics and pharmacodynamics. Arxiv preprint.*

## OC5A-2 Interim analysis of a clinical trial using the predictive probability of success based on a surrogate endpoint

**Loïck Vidot[1], Ronan Fougeray[1], Gaëlle Saint-Hilary[1,2]**
1 Department of Clinical Biostatistics, Servier, France
2 Department of Mathematics, Politecnico di Torino, Italy

The Predictive Probability of Success (PPoS) of a future clinical trial is a key quantitative tool for decision-making in drug development. It is generally derived from prior knowledge and evidence on the primary endpoint collected from previous clinical trials. The methodology to calculate the PPoS of a future trial based on historical data on surrogate endpoints was recently proposed. However, because pharmaceutical industries are willing to speed up drug developments and to make decisions to continue or to stop clinical trials as early as possible, interim analyses based on surrogate endpoints are raising interest in clinical trial designs. In this context, we extended the recently proposed methodology to study designs where an interim analysis based on a surrogate endpoint is set. An informative prior, called surrogate prior, is derived from (1) the information on the surrogate endpoint observed at the interim analysis and (2) the joint distribution of the surrogate and primary endpoints, estimated using a meta-analytic approach on past clinical trials. If available, the information on the primary endpoint at the interim analysis could be combined with the surrogate prior to generate the PPoS. Then, at the interim analysis, the futility rule is based on a pre-defined level of PPoS. This methodology was investigated in a phase III oncology study in colorectal cancer where Overall Survival is the primary endpoint and Progression-Free Survival data might be used at an interim futility analysis. We present the operating characteristics of the design in different settings, considering the amount of available information at the time of the interim analysis and potential prior data conflicts between the surrogate prior and the available evidence on the primary endpoint.

# SESSION OC5A

## OC5A-3 Trials of Vaccine Efficacy for COVID-19: Inferential and Practical Issues

**Stephen Senn**
Consultant Statistician, Edinburgh, United Kingdom

Several trials of vaccine efficacy for COVID-19 have now reported and attracted much interest in the media. Analysis of these trials raises a number of inferential issues. These will be discussed using the protocols, results and publications of these trial but paying particular attention to those by Moderna, Pfizer/BioNTech and Astra-Zeneca/Oxford.
For example, a number of claims regarding efficacy according to dose, dose interval and virus strain have been made on uncontrolled comparisons. It can be shown that if concurrent control is respected and strata and trials are fitted and the comparisons made 'honestly', the uncertainty is much greater than might naively be supposed.
Other inferential issues are to what extent statistical information is carried by cases only and what practical difference, if any, a Bayesian approach (such as was adopted by Pfizer/BioNTech) makes to interpreting effects and what a useful scale is for reporting results.
Practical matters include whether a 2:1 randomisation, such as was employed by AstraZeneca/Oxford is a good idea and also whether and under what conditions it is logical from a public health point of view to stretch dosing intervals in order to allow more subjects to receive a first dose given that some infections may occur between first and second dose. Various graphical approaches to understanding these issues will be presented.

## OC5A-4 Incorporating multiple parameters from historical controls using the meta-analytic-predictive (MAP) prior

**Hongchao Qi, Dimitris Rizopoulos, Joost van Rosmalen**
Department of Biostatistics, Erasmus University Medical Center, Rotterdam, Netherlands

**Background:** The meta-analytic-predictive (MAP) prior was proposed to incorporate information from comparable historical controls in the design and analysis of a new trial by assuming exchangeability of the new and historical trials. Analysis of covariance (ANCOVA) is often used to analyze data of clinical trials with a pretest-posttest design. In ANCOVA, both the intercept and the baseline effect influence the treatment effect estimate. However, previous implementations of the MAP mainly focused on the between-study variation in a single parameter, often the intercept or the mean outcome.
**Objective:** To extend the MAP prior to account for the between-study variation in multiple parameters, and to illustrate this approach in clinical trials with an ANCOVA model.
**Method:** The MAP prior was extended to allow for the between-study variation in the intercept and the baseline effect in ANCOVA, as well as the correlation between these parameters. To quantify the amount of information, prior effective sample sizes (ESS) were calculated using the variance ratio method. Different priors for the between-study variation were compared in terms of the prior ESS and the estimated treatment effect. The method was illustrated using data of six clinical trials conducted by the UC San Diego Alzheimer's Disease Cooperative Study (ADCS).
**Results:** The MAP prior based on the historical controls yielded approximately normal informative priors for the intercept and the baseline effect with prior ESS of 17 and 43, respectively. The MAP only slightly improved the precision of the estimated treatment effect in the ANCOVA model (posterior standard deviations reduced by 1.5%), but larger improvements were observed for the estimated intercept and baseline effect. The results were robust to different priors for the between-study variation.
**Conclusions:** The MAP can be extended to an ANCOVA model where more than one model parameter may vary across trials. Estimation of the between-study variation in multiple model parameters seems feasible even with a limited number of historical trials. However, the gains of using historical data in terms of required sample size and the precision of the estimated treatment effect may be restricted.



SESSION OC5A

## OC5A-5 Bayesian hierarchical modeling for MedDRA coded adverse events in RCTs

**Alma Revers, Michel Hof, Koos Zwinderman**
Epidemiology and Data Science, Amsterdam Medical Centers, Netherlands

Patients participating in randomized controlled trials (RCTs) often report a wide range of different adverse events (AEs) during trial. MedDRA is a hierarchical standardization terminology to structure reporting of AEs. The lowest level (i.e. Preferred Terms (PT)) is a single type of medical events and higher levels aggregate specific lower levels (i.e. Higher Level Terms (HLT), Higher General Level Terms (HLGT), System Organ Class (SOC)). MedDRA has a multiaxial structure where a single lower level could be aggregated in multiple higher levels.
Most of these AEs are uncommon and may occur once or twice in a patient. In general, power calculations of RCTs are not focused on AEs and we observe very low incidence rates. As a result, there is limited statistical power to detect rare AEs, leading to a high rate of false negatives. Therefore AE data of higher levels of the MedDRA structure are reported; these have higher incidence, but are less informative since contain a spectrum of AEs.
We propose hierarchical Bayesian models for identifying MedDRA coded AE relative risks (RRs) and odds ratios (ORs). Our model allows estimation of ORs and RRs at all levels and can deal with the multiaxial structure of AEs. Following other authors we started out by specifying a hierarchical binomial model. To account for multiple occurrences of specific AEs we specified a hierarchical Poisson model. To incorporate the hierarchical and multiaxial structure, the parameters from the Poisson or binomial distributions were sampled from weighted normal distributions of the HLT-level. These parameters were then sampled from distributions from higher levels up to the SOC-level. A full Bayesian model was specified using noninformative prior distributions of the means and variances of the AE-logORs and -logRRs. Models were implemented in the Stan-language and run through the rstan-package in R. When PTs occurred in only few patients, our models did not converge. Such PTs were aggregated and modeled using binomial/Poisson distributions at the HLT-level with suitable adjustment of the mean logOR/logRR.
We illustrate our model with AE-data from a large RCT (n=2658) and we compare results with other methods for analyzing AEs.

# SESSION OC5B

## SESSION OC5B — Prediction by Machine learning

### OC5B-1 Individual risk prediction: comparing Random Forests with Cox proportional-hazards model by a simulation study

**Valia Baralou*, Natasa Kalpourtzi*, Giota Touloumi**
Department of Hygiene, Epidemiology & Medical Statistics, National and Kapodistrian University of Athens, School of Medicine, Athens, Greece
* Equally contributed

With big data becoming more widely available in healthcare, Machine Learning algorithms such as Random Forest (RF) that ignores time-to-event information and its extension Random Survival Forest (RSF) [1] are used for individual risk prediction as an alternative approach to Cox proportional-hazards (Cox-PH) regression model. Our objective was to systematically evaluate and compare RF and Cox-PH models' predictive performance. Cox-PH, RSF with two split criteria [log-rank (RSF-LR), log-rank score [2] (RSF-LRS)] and RF were applied and evaluated through an extensive simulation study based on Athens Multicenter AIDS cohort study (AMACS) data. Several simulation scenarios were set up assuming different associations between the predictors and the outcome [linear (L), linear and interactions (LI), non-linear (NL), non-linear and interactions (NLI)], different sample sizes (500, 1000, 5000), censoring rates (50%, 75%, 93%), hazard functions (increasing, decreasing, constant) and number of predictors (7 or 15), leading to 216 scenarios in total. To evaluate the performance of the methods, equal-sized training and testing datasets were independently generated and time-dependent area under curve (AUC), C-index and integrated Brier score (IBS) were calculated. To reduce the variability of the model performance estimates, as a sensitivity analysis, testing datasets of 10000 sample size were generated. In all scenarios, RF had by far the worst performance among all models considered. In scenarios with low number of events (NE<250) as well as under linearity assumption, Cox-PH outperformed RSF models by 3% on average. Both methods performed similarly in LI scenarios when NE increased above 500. In NL and NLI scenarios, RSF-LR performed better than Cox-PH regression when NE≥250, resulting in approximately up to 2% increase in performance improvement. No notable differences in models' performance were observed among different hazard functions. Sensitivity analysis confirmed our results lowering variability especially in scenarios with small training datasets (≤1000). When applied to real data, models that incorporated survival time performed better. Although RSF models are a promising alternative to conventional regression methods as data complexity increases, they require much larger datasets for training. In time-to-event analysis, it is important to use RF models that consider the survival time.

**References:** *[1] Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Ann Appl Stat. 2008;841–60.*
*[2] Hothorn T, Lausen B. On the exact distribution of maximally selected rank statistics. Comput Stat Data Anal. 2003;43(2):121–37.*



# SESSION OC5B

### OC5B-2 Predicting patient mortality from large sets of diagnosis codes using logistic regression and machine learning

**Thomas E. Cowling[1], David Cromwell[1], Alexis Bellot[2], Linda D. Sharples[3], Jan van der Meulen[1]**
1 Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, UK
2 Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK
3 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, UK

**Background:** Electronic healthcare records are increasing in volume and scope, presenting growing opportunities to use large sets of predictors and model their relationships with more flexible methods. Machine learning approaches have been used to model interactions between many diagnosis codes in large datasets of electronic healthcare records. No previous studies have directly compared regression and machine learning approaches for predicting patient outcomes from large sets of individual International Classification of Diseases (ICD) codes.
**Objective:** To compare the performance of logistic regression and boosted trees for predicting patient mortality from large sets of diagnosis codes in electronic healthcare records.
**Study Design and Setting:** We analysed national hospital records and official death records for patients with myocardial infarction (n=200,119), hip fracture (n=169,646), or colorectal cancer surgery (n=56,515) in England in 2015-17. One-year mortality was predicted from patient age, sex, and socioeconomic status, and 202 to 257 International Classification of Diseases 10th Revision codes recorded in the preceding year or not (binary predictors). Performance measures included the c-statistic, scaled Brier score, and several measures of calibration.
**Results:** One-year mortality was 17.2% (34,520) after myocardial infarction, 27.2% (46,115) after hip fracture, and 9.3% (5,273) after colorectal surgery. Optimism-adjusted c-statistics for the logistic regression models were 0.884 (95% CI 0.882, 0.886), 0.798 (0.796, 0.800), and 0.811 (0.805, 0.817). The equivalent c-statistics for the boosted tree models were 0.891 (95% CI 0.889, 0.892), 0.804 (0.802, 0.806), and 0.803 (0.797, 0.809). Model performance was also similar when measured using scaled Brier scores. All models were well calibrated overall.
**Conclusion:** In large datasets of electronic healthcare records, logistic regression and boosted tree models of numerous diagnosis codes predicted patient mortality comparably. Our results suggest that there is little or no advantage to using machine learning rather than regression approaches in this particular study context.

### OC5B-3 Predicting individual life years lost due to cancer using pseudo-observations with random forest, in the absence of cause of death information

**Dimitra-Kleio Kipourou[1], Aris Perperoglou[2], Bernard Rachet[1], Aurelien Belot[1]**
1 Inequalities in Cancer Outcomes Network (ICON), Department of Non-Communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, UK
2 School Mathematics, Statistics and Astrophysics, Newcastle University, UK

Competing risk analyses are essential for making inferences about a specific disease and relevant methods that account for more than one event should be applied. Application to population-based registry data involves additional methodological challenges due to the absence of reliable information on the cause of death. Excess hazard methodology is widely applied to such settings allowing cause-specific inference in terms of net survival, crude probability of death and life-years lost (LYL). LYL are of particular interest due to their easy interpretation and communication to non-statistical audience. Here, we show how to estimate the LYL due to a specific cause using the pseudo-observation approach combined with excess hazard methodology. Jack-knife pseudo-observations for LYL are computed for each individual regardless of their initial censoring status at one time-point. The complete set of (continuous) pseudo-observations can be subsequently modelled with conventional generalized models or with a variety of machine learning tools ranging from random forests to support vector machines and neural networks. In this study, we illustrate this method (using pseudo-observations with random forest) on English lung cancer data with the ultimate aim to predict the individual LYL due to cancer based on numerous variables, including comorbidities and clinical characteristics.

# SESSION OC5B

## OC5B-4 A comprehensive comparison of approaches for the calibration of probability machines

**Francisco M. Ojeda[1], Yao Hu[2], Alexandre Thiéry[2], Stefan Blankenberg[1,3], Christian Weimar[4,5], Matthias Schmid[6], Andreas Ziegler[1,2,7]**

1 Department of Cardiology, University Heart & Vascular Center Hamburg, Germany
2 Cardio-CARE, Medizincampus Davos, Switzerland
3 German Center for Cardiovascular Research, Partner Site Hamburg/Kiel/Lübeck, Germany
4 BDH-Klinik Elzach, Germany
5 Institute for Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Germany
6 Institute of Medical Biometry, Informatics and Epidemiology, University of Bonn, Germany
7 School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, South Africa

Statistical prediction models have gained popularity in applied research. One challenge is the transfer of the prediction model to a different population which may be structurally different from the model for which it has been developed. An adaptation to the new population can be achieved by calibrating the model to the characteristics of the target population, for which numerous calibration techniques exist. In view of this diversity, we performed a systematic evaluation of various popular calibration approaches used by the statistical and the machine learning communities. Focusing on models for two-class probability estimation, we provide a review of the existing literature and present the results of a comprehensive analysis using both simulated and real data. The calibration approaches are compared with respect to their empirical properties and relationships, their ability to generalize precise probability estimates to external populations and their availability in terms of easy-to-use software implementations. Calibration methods that estimated one or two slope parameters in addition to an intercept consistently showed the best results in the simulation studies. Calibration on logit transformed probability estimates, i.e., the linear predictor or two log transformed probability estimates generally outperformed calibration methods on non-transformed estimates. In case of structural differences between training and validation data, re-estimation of the entire prediction model should be outweighted against sample size of the validation data. We recommend regression-based approaches using transformed probability estimates where at least one slope is estimated in addition to an intercept for updating probability estimates in validation studies.



# SESSION OC5B

## OC5B-5 Methodological conduct of clinical prediction models using machine learning methods in oncology needs to be improved

**Paula Dhiman[1], Jie Ma[1], Constanza Andaur Navarro[2], Benjamin Speich[3], Garrett Bullock[1], Johanna Aa Damen[2], Shona Kirtley[1], Lotty Hooft[2], Richard D. Riley[4], Ben Van Calster[5], Gary S. Collins[1], Karel Gm Moons[2]**

1 Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, United Kingdom
2 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Netherlands
3 University Hospital Basel, University of Basel, Switzerland
4 Keele University, Newcastle, United Kingdom
5 KU Leuven, Netherlands

**Context:** Clinical prediction models are widely used in oncology for medical decision making. Using modern modelling methods, such as machine learning (ML), to improve prediction is a rapidly growing area of research. ML is often portrayed as offering many advantages, such as flexible modelling and ability to analyse 'big', non-linear and high dimensional data. These promises are yet to be realised and there is concern that prediction models developed using ML use poor and inefficient methodology and are at high risk of bias.
**Objective:** To assess methodological conduct and the risk of bias of studies that develop clinical prediction models using ML (as defined by primary study authors) in the field of oncology.
**Methods:** We conducted a systematic review of prognostic clinical prediction models developed using ML, published during 2019. We extracted data on study design, sample size, data pre-processing, hyperparameter tuning and other analysis methods, and items for risk of bias assessment. Primary outcome was risk of bias assessed using the Prediction model Risk Of Bias ASsessment Tool (PROBAST) per model that was developed and validated.
**Results:** We identified 2,922 publications and excluded 2,860 due to study design and publication type. We reviewed full-text of 62 publications; 48 development-only studies and 14 development with validation studies. 152 models were developed overall, with a median of 2 models (range:1 to 6) developed per publication. Most prevalent ML models were classification trees (18%), logistic regression (18%), random forest (15%) and neural networks (12%).
**Development** of 84% of models and validation of 51% of developed models were found to be high risk of bias. Bias in the analysis domain was the largest contributor to the overall high risk of bias during model development. Sample size was justified in 5 publications, a median of 16 predictors (range:4-33788), 647 participants (range:20-582398) and 195 events (range:7-45797) were used to develop the models. 45% of studies used split sample to internally validate their models.
**Conclusion:** Most prediction models developed using ML in oncology were at high risk of bias, largely due to analysis methods. Urgent methodological guidance is needed to improve the quality of these models.

# SESSION OC5C

## SESSION OC5C — Screening and Diagnostic studies

### OC5C-1 The Natural History of Invasive Breast Cancers Detected in the Scandinavian Mammography Screening Programs: A Cohort Study

**Per-Henrik Zahl**
Department of Mental and Physical Health, Norwegian Institute of Public Health, Oslo, Norway

**Background:** The prevailing theory is that a large proportion of mammography-detectable tumors have very long lead-times and that tumors never regress. Two staggered cohort studies have previously suggested that almost all incidence increase when screening with mammography is due to detection of small tumors which natural fate is to regress before they become clinical disease [1,2]. Here we study alternative methods to estimate lead-time and cancer regression.
**Materials** and Methods: Prospective cohort study of 375,064 Swedish, 127,064 Norwegian and 149,266 Danish women invited to a first time mammography in 1986-89, 1996-97 and 2008-9, respectively. The proportions of tumors with lead-time over 1 year (women under age 50) and over 2 years (aged 50-69) are estimated. We study if tumors accumulate in the breast in the absence of screening. Regression is also studied by comparing the prevalence increase to incidence increases in succeeding screening rounds.
**Results:** RR after start of annual screening was 1.33 (95% CI:1.21-1.47;P<0.0001) for women aged 40-49 years in Sweden and with no prevalence peak. Thus, maximum lead-time is 1 year for these tumors. About 20% (95% CI:0.11-0.29;P<0.0001) and 29% (95% CI:0.26-0.33;P<0.0001) of tumors in the prevalence screening in Norway and Denmark of women aged 50-69 years had lead-time over 2 years. There is no evidence that slow-growing tumors accumulate in the absence of screening mammography under age 65 when comparing age-specific prevalence peaks to succeeding screening rates. A staggered cohort analysis of the introduction of public screening in Denmark yielded RR=1.10 (95% CI:1.05-1.17;P=0.0004) also suggesting that tumors do not accumulate in the absence of screening.
**Conclusion:** There is no evidence of any large reservoir of breast tumors with long lead-times. The incidence increase when screening is too large to be solely explained by early diagnosis – many mammography detected tumors must regress. Adjustment for long lead-time is not justified when calculating overdiagnosis, and lead-time bias is an exaggerated problem in cancer epidemiology.

**References:** *[1] Zahl,Mæhlen,Welch. The natural history of invasive breast cancers detected by screening mammography. Arch Intern Med 2008;168:2311-6. [2] Zahl,Gøtzsche,Mæhlen. Natural history of breast cancers detected in the Swedish mammography screening program; a cohort study. Lancet Oncology 2011;12:1118-24.*



SESSION OC5C

### OC5C-2 Meta-analysis of dichotomous and polytomous diagnostic tests without a gold standard

**Enzo Cerullo[1,2], Hayley E. Jones[3], Terry Quinn[4], Nicola J. Cooper[1,2], Alex J. Sutton[1,2]**
1 Biostatistics Research Group, Department of Health Sciences, University of Leicester, United Kingdom
2 Complex Reviews Support Unit, University of Leicester University of Glasgow, United Kingdom
3 Population Health Sciences, Bristol Medical School, University of Bristol, United Kingdom
4 Institute of Cardiovascular and Medical Sciences, University of Glasgow, United Kingdom

Standard methods for the meta-analysis of diagnostic tests without a gold standard are limited to the analysis of dichotomous tests. Multivariate probit models are used to analyze correlated binary data, and can be extended to multivariate ordered probit models to model polytomous (i.e. non-binary) data. Within the context of an imperfect gold standard, they have previously been used for the analysis of dichotomous and polytomous diagnostic tests in a single study and for the meta-analysis of dichotomous tests. We developed a hierarchical, semi-ordered latent class multivariate probit model for the meta-analysis of polytomous and dichotomous diagnostic tests without a gold standard. Our model enables the synthesis of data from studies reporting accuracy at varying thresholds, and can accommodate a hierarchical partial pooling model on the conditional within-study correlations, which allow us to obtain summary estimates of joint test accuracy. Dichotomous tests use binary probit likelihoods and polytomous tests use ordered probit likelihoods. We fitted the models using Stan, which uses a state-of-the-art Hamiltonian Monte Carlo algorithm. In the first case study, we applied the models to a dataset in which studies evaluated the accuracy of tests, and combinations of tests, for deep vein thrombosis. We also compared our results to the original study, which assumed a perfect reference test. We found that modelling the polytomous test (the Wells score) as dichotomous and conducting stratified analyses at each threshold resulted in substantial bias in the reference test (ultrasonography) and the other index test under evaluation (the D-Dimer). Furthermore, we found substantial imperfect standard bias (over 10% difference in the joint test accuracy of the Wells and D-Dimer) in the original analysis, which assumed a perfect gold standard. In the second case study, we applied the methods to a dataset for clinical dementia, where studies reported accuracy from one to 16 distinct thresholds, and compared our results to stratified analyses. Our results suggest that the original analysis underestimated the sensitivity of the Mini-Mental State Examination (MMSE) by over 10% and overestimated the specificity. We discuss limitations and possible ways to improve scalability by making use of recently proposed algorithms.

### OC5C-3 Comparison of methods for the linear combination of biomarkers under Youden Index optimisation criterion

**Rocío Aznar-Gimeno[1], Luis Mariano Esteban[2], Sergio Sabroso[3], Ángel Borque Ferrando[4], Gerardo Sanz[5], Rafael del-Hoyo-Alonso[1]**
1 Instituto Tecnológico de Aragon, ITAINNOVA, Zaragoza, Spain
2 Escuela Universitaria Politécnica de La Almunia, Universidad de Zaragoza, Spain
3 Spanish National Cancer Research Centre (CNIO), Madrid, Spain
4 Department of Urology, Hospital Universitario Miguel Servet, IIS-Aragon, Zaragoza, Spain
5 Department of Statistical Methods & Institute for Biocomputation & Physics of Complex Systems-BIFI, University of Zaragoza, Spain

In clinical practice, it is common to have information on multiple biomarkers for disease diagnosis. Combining them all into a single marker is a common and widespread practice and often provides a better diagnostic yield.The formulation of algorithms for the estimation of binary classificatory models that maximise the AUC has been a widely explored line of research. The Youden index is a statistical metric also widely and successfully used in several clinical studies and serves as a summary for diagnosis. However, unlike the AUC, the study and exploration of methods that optimise the Youden index has not received sufficient attention in the literature. The aim of our study was to propose a new step by step algorithm to combine continuous biomarkers that maximise the Youden index, and additionally, to explore, evaluate and compare with other methods. Three methods are based on Pepe and Thompson's empirical search [1] (our proposed stepwise approach, Yin and Tian's stepwise approach [2] (SWD) and the Min-max approach [3] (MM)) and three numerical search methods based on derivatives (the logistic regression, a parametric approach under multivariate normality and a non-parametric kernel smoothing approach (KS)). To compare the performance of these methods, a comprehensive simulation study was performed and also analysed on two real data sets (Duchenne Muscular Dystrophy Dataset and Prostate Cancer Dataset). The simulated data analysed cover a wide range of scenarios regarding the probability distribution of biomarkers (normal, non-normal), the discrimination ability between biomarkers (similar or different) and the correlation between them, considering from small to large sample sizes. The results obtained show that our proposed stepwise approach outperforms all other algorithms in most scenarios. In general, it is followed by KS and SWD. The MM algorithm is the worst performer in most scenarios, except in normal biomarker scenarios, with the same means and different covariance matrix for the diseased and non-diseased population, where it outperforms other algorithms.

**References:** *[1] Pepe MS, Thompson ML. Combining diagnostic test results to increase accuracy. Biostatistics 2000; 1(2): 123-140. [2] Yin J, Tian L. Optimal linear combinations of multiple diagnostic biomarkers based on Youden index. Statistics in medicine 2014; 33(8): 1426-1440. [3] Liu C, Liu A, Halabi S. A min–max combination of biomarkers to improve diagnostic accuracy. Statistics in medicine 2011; 30(16): 2005-2014*

# SESSION OC5C

## OC5C-4 Single and multiple imputation combined with missing indicators in clinical prediction models: a simulation study

**Rose Sisk, Matthew Sperrin, Niels Peek, Glen P. Martin**
Centre for Health Informatics, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, United Kingdom

**Background & Aims:** Clinical prediction models (CPMs) allow the communication of risk between patients and caregivers, based on current patient and clinical characteristics. The development and validation of CPMs is a complex process, especially in the presence of missing data, which is especially common in clinical data. Multiple imputation is often heralded as the gold standard for imputing missing data in both causal inference work and prediction modelling studies, but due to practical limitations can be difficult to implement in clinical practice, where CPMs are applied. Key limitations include the requirement for access to the development data and computational power at prediction time, which are often not available.We therefore aim to consider whether regression imputation could offer a promising alternative to multiple imputation in the context of prediction, since it is a deterministic process requiring only access to the imputation model to impute missing data at the point of care. Both MI and RI both rely on the assumption that data are missing at random, which is often a dubious assumption in clinical research data, particularly within electronic health records where data collection is opportunistic. We therefore also consider whether the inclusion of missing indicators in combination with MI and RI can improve the predictive performance of models developed under informative missingness. We setup a simulation study to allow us to explore these ideas in more detail.
**Results:** We assessed ideal and pragmatic performance (Wood et al., 2015) of models developed and validated using both MI and RI, with and without missing indicators included as predictors. We found that under MAR and MNAR structures, the inclusion of a missing indicator can indeed improve the calibration and discrimination of models developed and applied in the presence of missing data. We also found that regression imputation could offer a practical alternative to MI where it is not possible to apply MI at the point of prediction.
**Conclusion:** We advocate the careful use of regression imputation and missing indicators in the development and validation of clinical prediction models, where the missing data are assumed to be informative with respect to patient condition.

**References:** *Wood, A. M., Royston, P., & White, I. R. (2015). The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data. Biometrical Journal, 57(4), 614–632. https://doi.org/10.1002/bimj.201400004*

## OC5C-5 Diagnosing Latent Class Analysis for Analyzing Diagnostic Tests in the Absence of a Gold Standard

**Alfred K. Keter[1,2,3], Lutgarde Lynen[1], Alastair Van Heerden[2], Els Goetghebeur[3], Bart K.M. Jacobs[1]**
1 Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium
2 Centre for Community Based Research, Human Sciences Research Council, South Africa
3 Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium

Diagnostic tests play a key role in disease control. The absence of a gold standard however hampers estimation of disease prevalence and misclassification errors (ME) of available imperfect tests. With a set of such test results jointly available for a sample of patients, latent class analysis (LCA) allows for correct estimates of these parameters under restrictions. For unknown (latent) true disease status, LCA assumes test ME are independent within each class and constant across subpopulations. These assumptions are violated when serious comorbidity affects the targeted disease risk and/or ME rates. We examine implications of envisaged model violations on the working likelihood estimators and through simulation focusing on population prevalence, sensitivity and specificity as target estimands. We derive if and when results are still reliable and consider how a simple and more comprehensive adapted conditional version of LCA may alleviate problems. We support our results with finite sample simulations mimicking a setting of passive case finding of presumptive pulmonary tuberculosis (PTB) patients with or without HIV comorbidity, before applying the methods on a case study. Based on realistic sensitivities and specificities of five commonly used diagnostic tests for PTB (any TB symptom, digital chest X-ray, CRP, Xpert MTB/RIF & culture), we simulated test results in samples of various sizes with different PTB prevalence across HIV subgroups. We thus generated independent test results within the true PTB latent classes conditional on HIV. For different numbers of tests (5 versus 3), we performed Bayesian LCA for working models with or without constant PTB prevalence and ME across the HIV subgroups. The working model with different PTB prevalence but constant ME across HIV subgroups generated substantially more biased PTB prevalence and ME than when ignoring comorbidity all together. With three tests, all models produced poor coverage. Reassuringly, models with five tests allowing for different ME yielded largely unbiased total population prevalence with acceptable coverage. Standard LCA is not robust to model violation through heterogonous ME across subpopulations, especially with fewer tests. Well-chosen covariate-specific adaptations can alleviate the problem.

**Acknowledgement:** *This project is part of the EDCTP2 programme supported by the European Union (RIA2018D-2489 TB TRIAGE+)*

# SESSION OC5D

SESSION OC5D — Bayesian Joint models for longitudinal data and time-to-event

## OC5D-1 Bayesian multilevel nonlinear joint model to characterize the variability in the response to immunotherapy

**Marion Kerioui[1,2,3,4], Julie Bertrand[1], René Bruno[5], François Mercier[6], Solène Desmée[2], Jérémie Guedj[1]**
1 Université de Paris, INSERM IAME, F-75018 Paris, France
2 Université de Tours, Université de Nantes, INSERM SPHERE, UMR 1246, Tours, France
3 Institut Roche, Boulogne-Billancourt, France
4 Genentech/Roche, Clinical Pharmacology, Paris, France
5 Genentech/Roche, Clinical Pharmacology, Marseille, France
6 F. Hoffmann-La Roche AG, Biostatistics, Basel, Switzerland

**Background:** The association between survival and tumor dynamics, assessed by the Sum of the Longest Diameters (SLD) of the target lesions, has brought a lot of attention from statisticians, with the goal to anticipate the outcome of clinical trials and/or identify most at risk's patients [1]. However, SLD is an aggregate measure result, which sums up the dynamics of several lesions that can have different dynamics. Moreover, these lesions can be located in different organs, and hence play may be differently associated with survival. Whether this heterogeneity is exacerbated by immunotherapy, and could be associated with survival, has been suggested.
**Objectives:** Here we aimed to quantify, in a large population of individuals with advanced urothelial cancer, i) the impact of tumor dynamics on survival in different organs ii) the intra-patient variability and its association with treatment response.
**Methods:** We analyzed the tumor dynamics from a phase 3 clinical trial (IMVigor211) of 900 patients randomized between immunotherapy (Atezolizumab) and chemotherapy treatments. We developed nonlinear parametric joint models to describe the SLD dynamics in 5 different locations of the body (lymph, lung, liver, bladder, other) and quantify their marginal impact on survival. Then, we developed a Bayesian multilevel joint model, where the individual lesions dynamics (up to 5 per individual) were modeled using a nonlinear mixed effect model with an additional level of random effects. Inference was done using HMC algorithm in Stan [2].
**Results:** We observed a great variability in tumor dynamics across organs, with different impacts on survival. In particular, liver tumor dynamics was strongly associated with survival compared to other locations. Considering proper association between organ-specific tumor kinetics and survival instead of one single association with all organs significantly improved the survival data fit. Thanks to a great amount of data (2133 target lesions), we expect to demonstrate a larger intra-patient variability under immunotherapy compared to chemotherapy.
**Conclusion:** This approach allowed to characterize inter and intra-patient variability in response to immunotherapy, which might help to early identify the most at-risk patients in a perspective of personalized medicine.

**References:** *[1] Tardivon C, Desmée S, Kerioui M, et al. Association Between Tumor Size Kinetics and Survival in Patients With Urothelial Carcinoma Treated with Atezolizumab: Implication for Patient Follow-Up. Clin Pharmacol Ther. 2019;106(4):810-820. doi:10.1002/cpt.1450. [2] Kerioui M, Mercier F, Bertrand J, et al. Bayesian inference using Hamiltonian Monte-Carlo algorithm for nonlinear joint modeling in the context of cancer immunotherapy. Stat Med. Published online 8 October 2020:sim.8756. doi:10.1002/sim.8756*

# SESSION OC5D

## OC5D-2 Multistate inference based on longitudinal/competing risks joint modeling under misclassified cause of failure

**Christos Thomadakis[1], Loukia Meligkotsidou[2], Constantin T. Yiannoutsos[3], Giota Touloumi[1]**
1 Department of Hygiene and Epidemiology, National and Kapodistrian University of Athens, Greece
2 Department of Mathematics, National and Kapodistrian University of Athens, Greece
3 Department of Biostatistics, Indiana University Fairbanks School of Public Health Indianapolis, Indianapolis IN, USA

Thomadakis et al. (2019) proposed joint modeling of a disease marker through a linear mixed model (LMM) and competing risks using cumulative incidence functions (CIFs), with the CIFs dependent on the "true" marker values to remove measurement error. The generalized odds rate transformation was adopted, with the proportional subdistribution hazards model being a special case. In HIV studies, patients receiving antiretroviral therapy may die or disengage from care (competing risks). The CD4 count, a longitudinally measured marker, is a predictor of clinical outcomes, which suggests joint analysis. However, biases can occur as many deaths may be incorrectly classified as disengagements from care, especially in studies from resource-constrained countries.
We extend [1] to account for failure cause misclassification through double sampling, where the true failure cause is ascertained in a small random sample of individuals initially classified as disengaged from care, using a Bayesian MCMC procedure. We also estimate multistate probabilities defined jointly by marker and competing-risk data. Based on the assumed joint model, we derive posterior samples for (a) probabilities of being event-free and "true" marker values being in predefined intervals (states) and (b) population-averaged CIFs. Both (a) and (b) are re-estimated by baseline marker state. A simulation study is performed assuming the true failure causes are available in 20% of patients. Marker data are generated by an LMM, with two scenarios for the CIFs: (i) a proportional odds rate model and (ii) a proportional subdistribution hazard model. Under each scenario, (i) and (ii) are fitted yielding estimates with small biases and good coverage rates (93-97%). Multistate/transition probability estimates are nearly unbiased even under misspecified survival submodels. The proposed models are applied to data from the East Africa IeDEA cohort study. It is estimated that only 29.2% of deaths are correctly classified, leading to significant adjustments in the mortality estimates. Mortality rates are substantially higher at lower initial CD4 counts.
We have extended a flexible CIF-based joint modeling approach to account for potential failure misclassification and derive multistate/transition probabilities. Our approach is particularly useful when the effect of the marker on the failure probabilities is of primary interest.

**References:** *[1] Christos Thomadakis, Loukia Meligkotsidou, Nikos Pantazis, Giota Touloumi; Joint modeling of longitudinal and competing-risks data using cumulative incidence functions for the failure submodels, ISCB 2019 Leuven, Belgium*

## OC5D-3 Bayesian Predictive Model Averaging for Joint Model of Survival and Longitudinal Data: Application to an Immunotherapy Trial

**Zixuan Yao, Satoshi Morita**
Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine, Japan

In many clinical researches, multiple biomarkers are repeatedly measured over time, so that physicians are clearly aware of patients' conditions during the follow-up. Predicting a patient's future survival status based on such recorded longitudinal information is of great interest, known as the dynamic individualized prediction by jointly modeling longitudinal and survival data. From a predictive viewpoint, better results can generally be obtained by averaging over the candidate models to account for the model uncertainty, compared with selecting the single best model. In this study we apply a Bayesian predictive model averaging approach to the dynamic prediction in the context of the joint model analysis, which evaluates fitted models using estimated out-of-sample prediction accuracy. In extensive simulation studies across a broad range of situations, we examine the operating characteristics of the proposed method in terms of the predictive performance, by measuring the calibration and discrimination abilities of the dynamic predictions. We discuss the strengths and limitations of the proposed model averaging approach in comparison with the single-model based method through an application to an ovarian cancer immunotherapy clinical trial as well as the simulation studies. It is suggested that the proposed modeling framework can provide generally more precise predictions on survival probabilities, which could help subsequent medical decision making process.



# SESSION OC5D

## OC5D-4 A Joint Model for Multiple Longitudinal Outcomes, Recurrent and Terminal Events using CF Patient Registry Data

**Pedro Miranda Afonso[1], Dimitris Rizopoulos[1], Anushka Palipana[2,3], John P. Clancy[4], Rhonda D. Szczesniak[2,3], Eleni-Rosalina Andrinopoulou[1]**
1 Department of Biostatistics, Erasmus University Medical Center, Netherlands
2 Department of Pediatrics, Cincinnati Children's Hospital Medical Center, United States
3 Department of Mathematical Sciences, University of Cincinnati, United States
4 Cystic Fibrosis Foundation, United States

Cystic fibrosis (CF) is an inherited disease primarily affecting the lungs and gastrointestinal tract. Thick and infected mucus in the patient's airways lead to recurrent acute respiratory events known as pulmonary exacerbation (PE); thereby worsening lung function. It is of clinical interest to investigate the association between the risk of PE with lung function and nutritional decline, as direct positive associations between lung function and nutritional status have been reported. Previous work has been limited to continuous longitudinal markers and time-to-first PE, thereby neglecting subsequent occurrences and other survival outcomes [1]. This was mainly due to the unavailability of appropriate and robust statistical software. Our primary goal is to simultaneously investigate the association between the risk of PE, lung function decline (FEV1), evolution of the patient's growth and nutritional status (e.g., BMI), and the risk of lung transplant or death using all available U.S. CF Foundation (CFF) patient registry data. We intend to explore different forms of association between the longitudinal markers and the events of interest. We propose a joint modeling framework accommodating multiple longitudinal markers, a recurrent event process, and a terminal event. The terminal outcome accounts for informative censoring due to lung transplantation or death from respiratory failure. Novel elements of our approach, compared to previously proposed joint models for recurrent events, are: (i) allowance for multiple longitudinal markers with different distributions, (ii) specifying various functional forms to link these markers with the risk of a recurrent event and the risk of the terminating event, and (iii) accommodation of discontinuous intervals of risk, and the time can be defined in terms of the gap or calendar timescale. The developed model will be available in the R statistical package JMbayes2.
Analysis of all recurrent events with multiple biomarkers enhances our understanding of risks posed by PEs. Full MCMC algorithm implementation in C++ enables model fit in a timely fashion, despite its complexity. The proposed multivariate joint model affords the opportunity to make more efficient use of all available CFF registry data. It thereby brings new insights into CF disease progression and contributes to monitoring and treatment strategies.

**References:** *[1] Andrinopoulou ER, Clancy JP, Szczesniak RD. Multivariate joint modeling to identify markers of growth and lung function decline that predict cystic fibrosis pulmonary exacerbation onset. BMC Pulm Med. 2020 May 19;20(1):142. doi: 10.1186/s12890-020-1177-z. PMID: 32429862; PMCID: PMC7236487.*

## OC5D-5 Joint Modeling of Incomplete Longitudinal Data and Time-to-Event Data

**Yuriko Takeda, Toshihiro Misumi, Kouji Yamamoto**
School of Medicine, Yokohama City University, Japan

Clinical studies often collect longitudinal and time-to-event data for each subject. Joint modeling is a powerful methodology for evaluating the association between these data. The existing models, however, have not sufficiently addressed the problem of missing data which are commonly encountered in longitudinal studies. In most cases, analysis methods are based on the assumption of MNAR or MAR, and sensitivity analyses are performed to assess the robustness of findings to plausible alternative assumptions about the missing data. When we cannot determine whether missingness is MNAR or MAR on the collected data, a robust model corresponding to both MAR and MNAR missing mechanism assumption is needed. Shared parameter model is one of the model-based approaches to dealing with missing data in longitudinal studies. This model is considering that the relationship between outcome and missingness models is connected by means of common random effects. In this presentation, we introduce a novel joint model with shared random effects for incomplete longitudinal data and time-to-event data. Our proposed joint model consists of three submodels: a linear mixed model for the longitudinal data, a Cox proportional hazard model for the time-to-event data, and a Cox proportional hazard model for the time-to-drop-out from the study. By simultaneously estimating the parameters included in these submodels, the biases of estimators are expected to decrease under both missing mechanisms, MAR and MNAR. The proposed model is estimated by Bayesian approach, and we evaluate the performance of our method through Monte Carlo simulation studies. The results from simulation studies indicate that our proposed model provides less biased results with respect to the association parameter of longitudinal data and time-to-event data comparing with the existing joint model.

# SESSION OC5E

## SESSION OC5E — Cure and mixture models

### OC5E-1 RECeUS: Ratio Estimation of Censored Uncured Subjects for Studying Sufficient Follow-Up in Studies of Long-Term Survivors ➜ *StCA WINNER*

**Subodh Selukar[1], Megan Othus[2]**
1 Department of Biostatistics, University of Washington, United States
2 Department of Biostatistics and Biomathematics, Fred Hutchinson Cancer Research Center, United States

The need to model a cure fraction, the proportion of a cohort not susceptible to the event of interest, arises in a variety of contexts including tumor relapse in oncology. Existing methodology assumes that follow-up is long enough for all uncured subjects to have experienced the event of interest at the time of analysis, and researchers have demonstrated that fitting cure models without sufficient follow-up leads to bias. However, this assumption is rarely testable in practice; for example, no diagnostic tests exist to determine if a cancer patient is cured of the disease. Limited statistical methods exist to evaluate sufficient follow-up, and they can exhibit poor performance and lead users to falsely conclude sufficient follow-up, leading to bias, or to falsely claim insufficient follow-up, possibly leading to additional, costly data collection. The goal of this project is to develop a new quantitative statistic to evaluate whether cure models may be appropriate to apply to censored data. Specifically, we propose that the proportion of censored uncured subjects in a study can be used to evaluate cure model appropriateness. We implement this via maximum likelihood estimation. Asymptotic and simulation results demonstrate that the statistic can be estimated in finite samples with desirable statistical properties. We also apply the method to two oncology examples to compare the approach with existing methods.

### OC5E-2 Multiple imputation for survival analysis with missing data and a cure fraction: a study of osteosarcoma

**Maria Quelhas[1], Marta Fiocco[1,2,3], Eni Musta[4]**
1 Mathematical Institute, Leiden University, Netherlands
2 Department of Biomedical Data Science, Leiden University Medical Centre, Netherlands
3 Princess Maxima Centre for Pediatric Oncology, Netherlands
4 Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Netherlands

**Background:** Nowadays many cancer patients get cured after treatment. Hence, distinguishing between prognostic factors with curing or life-prolonging effect is of great interest in clinical research. However, some of these covariates are often not fully observed. In childhood osteosarcoma, histologic response to pre-operative treatment is known to have a strong effect on survival while the benefits of intensified chemotherapy remain unclear. Previous osteosarcoma studies have treated all patients as uncured and have excluded those with missing histologic response. This approach might affect the results.
**Objectives:** The study aims at developing and analyzing innovative methods for incorporating observations with missing covariates into the analysis of survival data with cured subjects. This is statistically challenging because of the complex model structure and the latent cure status.
**Methods:** A mixture cure model is considered, assuming that patients are either cured or uncured. In particular, the cure status and the failure times are modeled respectively through a logistic and a Cox regression model. Several methods based on the multiple imputation approach are explored for handling missing covariates, incorporating outcome information in the imputation model. Compared to the method discussed by Beesley et al., the proposed approach is more general because it allows for inclusion of different covariates in the two model components. Existing procedures, are used to estimate the logistic-Cox model. The proposed methods are evaluated through an extensive simulation study and are used to analyze osteosarcoma data from MRC BO06/EORTC 80931 clinical trial.
**Results:** The developed methodology allows to simultaneously consider cured patients and partially missing covariates. It leads to smaller variance of the estimates and larger power to detect significant effects. This study shows for the first time that histologic response has a strong prognostic value for the cure status and the effect of intensified chemotherapy is strongly related to histologic response.
**Conclusions:** The complex nature of the disease and the limitation present in the data ask for more advanced statistical techniques. This study shows that imputation of missing covariates while account- ing for cured patients provides more accurate interpretative and forecasting tools in osteosarcoma research and other oncological studies.
**References:** *[1] Beesley L.J., Bartlett J.W., Wolf G.T., and Taylor J.M. (2016). Multiple imputation of missing covariates for the Cox proportional hazards cure model. Statistics in medicine. 35(26): 4701-17 [2] Musta E., Patilea V., and Van Keilegom, I. (2020). A presmoothing approach for estimation in mixture cure models. arXiv preprint arXiv:2008.05338*



# SESSION OC5E

### OC5E-3 A simulation analysis of reliability and robustness of a cancer cure model accounting for extra non-cancer mortality

**Laura Botta[1], Juste Goungounga[2], Riccardo Capocaccia[3], Gaelle Romain[2], Marc Colonna[4,5], Gemma Gatta[1], Olayidé Boussari[6], Valérie Jooste[2,5]**
1 Evaluative Epidemiology Unit, Department of Preventive and Predictive Medicine, Fondazione IRCCS "Istituto nazionale dei Tumori", Milan, Italy
2 Registre Bourguignon des Cancers Digestifs; Centre Hospitalier Universitaire de Dijon Bourgogne; INSERM, UMR1231; Université de Bourgogne Franche-Comté, Dijon, France
3 Editorial Board, Epidemiologia e Prevenzione, Milan, Italy
4 Isere Cancer Registry, Centre Hospitalier Universitaire Grenoble, , France
5 FRANCIM, French Network of Cancer Registries, , Toulouse, France
6 Fédération Francophone de Cancérologie Digestive (FFCD), Département de Méthodologie, INSERM, UMR1231; Université de Bourgogne Franche-Comté, Dijon, France

**Introduction:** The proportion of cancer patients cured of the disease is estimated with standard cure models assuming they have the same risk of death as the general population [1]. Cured patients, however often maintain an extra risk of dying compared to the overall population due to other causes than cancer [2]. The aim of this work was to develop and validate an extended cure model incorporating a patients' relative risk (a) of death from other causes compared to that observed in the general population.
**Methods:** We extended mixture cure models considering Weibull relative survival of the uncured by adding a relative risk (a) multiplying the observed mortality in the general population. The parameters were estimated using maximum likelihood method for individual data and unweighted least square for grouped data. We evaluated the standard and the extended cure models through a simulation study, assessing their performances when all the assumptions are valid and their robustness when survival of uncured patients did not follow a Weibull distribution or extra non-cancer death risk was dependent of age at diagnosis or it randomly varied across patients. Two scenarios, simulating lung and breast cancer survival patterns, were varied by different relative risks (a) and lengths of follow-up. 1000 samples of different sizes (500- 20,000 cases each) were generated. The models were also applied to colon cancer FRANCIM real data.
**Results:** When the assumptions were satisfied, both the extended cure models estimated correctly the parameters and their standard errors providing excellent coverage in all scenarios, although maximum likelihood coverage outperformed unweighted least square. The standard model underestimated π by 7% when a= 1.2, and by 40% when a=2.0. Age effect on the cure fraction was heavily overestimated. For reasonable deviations from the assumption parameter estimates appeared fairly robust, with relative difference from the true values in the range ±10%. Applied to real male colon cancer data, the extended models estimated (a) around 1.2 and π around 56% higher than that of the conventional model 52%.
**Conclusions:** The present analysis suggests that conventional indicators overestimate cancer-specific death and underestimate cure fraction and net survival of cancer patients.
**References:** *[1] Lambert PC, Thompson JR, Weston CL, et al. Estimating and modelling the cure fraction in population-based cancer survival analysis. Biostatistics 2007; 8: 576–594. [2] Phillips N, Coldman A, McBride M. Estimating cancer prevalence using mixture models for cancer survival. Stat Med 2001; 21: 1257-1270.*

# SESSION OC5E

## OC5E-4 An extension of Fellegi-Sunter record linkage model for mixed-type data with application to SNDS

**Thanh-Huan Vo[1,2], Guillaume Chauvet[3], André Happe[4], Emmanuel Oger[4], Stéphane Paquelet[2], Valérie Garès[1]**

1 Univ Rennes, INSA Rennes, CNRS, IRMAR - UMR 6625, Rennes, France
2 IRT b-com - Institut de Recherche Technologique b-com, France
3 Univ Rennes, ENSAI, CNRS, IRMAR - UMR 6625, Rennes, France
4 EA 7449 REPERES, France

Probabilistic record linkage is a process of combining data from different sources, when such data refer to common entities and that identifying information is not available. Fellegi and Sunter proposed a probabilistic record linkage framework that takes into account multiple non-identifying information but is limited to simple binary comparison between matching variables. In our work, we propose an extension of this model especially when matching data contains different types of variables (binary, categorical and continuous). We developed a mixture of discrete distribution for handling low prevalence categorical matching variables, and a mixture of hurdle gamma distribution for handling continuous matching variables. The maximum likelihood estimates for model parameters are obtained by means of the Expectation Conditional Maximization (ECM) algorithm. Through a Monte Carlo simulation study, we evaluated both the posterior probability estimation for a record pair to be a match and the prediction of matched record pairs. The first simulation results indicate that the proposed methods perform well as compared to existing methods. The next step will be to apply the proposed method to real datasets, which aim to find corresponding patients in SNDS (Système National des Données de Santé) and GETBO (Groupe d'étude de la Thrombose de Bretagne Occidentale) register data.

# SESSION OC5F

SESSION OC5F — Longitudinal data analysis

## OC5F-1 Multiple imputation approaches for handling incomplete three-level data with time varying cluster memberships

**Rushani Wijesuriya[1,2], Margarita Moreno-Betancur[1,2], John B. Carlin[1,2,3], Anurika P. De Silva[3], Katherine J. Lee[1,2]**

1 Department of Paediatrics, Faculty of Medicine Dentistry and Health Sciences, The University of Melbourne, Parkville, Australia
2 Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Parkville, Australia
3 Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Australia

Three-level data structures arising from repeated measures on individuals who are clustered within larger units are common in clinical and population health studies. An additional complexity arises when individuals move between clusters over the course of the study resulting in a cross-classified data structure, which needs to be accounted for in the analysis. In these studies, missing data are also a common occurrence. Multiple imputation (MI) is a popular approach for handling missing data, but its validity depends on appropriate tailoring to the target analysis. In the context of cross-classified data, this means that the three-level structure and the time-varying cluster membership should be appropriately accommodated in the imputation model. While three-level data can be handled by either adapting single- and two-level MI approaches using dummy indicators and/or imputing repeated measures in wide format, or using a three-level MI approach, the implementability and comparability of these approaches in the context of cross-classified structures remain unclear. We conducted a simulation study to evaluate MI approaches handling incomplete cross-classified data when the substantive analysis uses a cross-classified random effects model. The simulation design was based on a longitudinal cohort study estimating the effect of depressive symptoms on the academic performance of students over time, clustered by time-varying school. The simulations were conducted under various missing data mechanisms and strengths of cluster correlation. The approaches evaluated included ad-hoc methods, ignoring the time-varying nature of cluster membership by taking the first or the most common cluster; pragmatic extensions of single-level and two-level MI approaches within the joint modelling (JM) and the fully conditional specification (FCS) framework; and a three-level FCS MI approach specifically developed for handling cross-classified data. We also compare the approaches in the longitudinal cohort case study. Simulation results indicated that the FCS implementations performed well in terms of bias and precision while the JM approaches performed poorly. The results in the case study were in line with simulations, with the JM approaches resulting in comparatively different estimates for the variance components than the FCS approaches.

# SESSION OC5F

## OC5F-2 Separation in Marginal Logistic Regression Models

**Angelika Geroldinger[1], Rok Blagus[2,3], Georg Heinze[1]**

1 Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria
2 Institute for Biostatistics and Medical Informatics, University of Ljubljana, Slovenia
3 Faculty of Sports, University of Ljubljana, Slovenia

Clustered or longitudinal data are frequently encountered in clinical research, e.g. when multiple study centers collect data or when measurements are performed multiple times per individual. With a binary outcome of interest, extensions of the logistic regression model to the context of correlated data are commonly applied. One popular option are marginal logistic regression models which relate the population-averaged log-odds to a linear combination of explanatory variables. Using generalized estimating equations (GEE) for fitting the marginal model results in consistent coefficient estimates even if the within-unit associations are notspecified correctly.
If the binary outcome can be perfectly predicted by a linear combination of the explanatory variables, i.e. the data are 'separated', then the marginal model fitted by GEE will not give finite coefficient estimates. Similarly, with independent data, logistic regression fails to converge if the data are separated. A popular solution with independent data is to resort to Firth's penalized logistic regression (FL), which was originally proposed to reduce the bias in maximum likelihood coefficient estimates. We found that the stabilizing property of FL can be transferred to the analysis of correlated data in a pragmatic way: noting that FL is equivalent to maximum likelihood estimation of an appropriately augmented data set, we suggest to first perform logistic regression ignoring the correlation structure of the data in order to create the corresponding augmented data set. The marginal model can then be fitted by GEE on this augmented data set, stabilizing the coefficient estimates. We illustrate the performance of our proposed method analysing clustered data from a study on the occurrence of hematological complications in implant dentistry, where most patients underwent multiple implant procedures. The data were separated as there were no complications for one level of a risk factor. Furthermore, we present a simulation study comparing our method to a recently published approach integrating Firth's penalty in GEE in a more rigorous way. Interestingly, this latter method suffered from severe non-convergence problems for some data structures. Our approach performed better with respect to convergence and has the advantage of being easily implementable in any software where FL is available.

## OC5F-3 A geometric Brownian motion model with non-normal random effect for the prediction of the growth of abdominal aortic aneurysms

**Robin Ristl[1], Martin Posch[1], Christoph Neumayer[2], Christine Brostjan[2], Wolf Eilenberg[2]**

1 Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Austria
2 Department of General Surgery, Division of Vascular Surgery, Medical University of Vienna, Austria

Patients with abdominal aortic aneurysms (AAA) require regular monitoring of aneurysm size, and, once the AAA has exceeded a critical diameter of 50-55 mm, a surgical intervention is indicated to avoid rupture. We propose a statistical model for the growth of AAAs to assess the risk that the AAA exceeds the critical size in a given time period and to find determinants of fast versus slow growth. The growth of AAAs is characterised by (1) growth rates that increase with increasing aneurysm size, (2) right-skewed heterogeneity in growth rates between patients and (3) heterogeneity in growth rates within a patient across time. Further, (4) patients present at different stages of their disease at the initial diagnoses and a prediction model should be applicable regardless of disease stage. We regarded the growth of AAAs as a stochastic process and developed a parametric prediction model that is based on geometric Brownian motions with log-normally distributed growth rates which are modelled as random effect. The model parameters include the mean growth rate (which may depend on co-variates), a scale factor quantifying within-patient variability in growth and the random-effect variance quantifying between-patient variability. In contrast to models for AAA growth proposed in the current literature, the stochastic growth model accounts for all considered characteristics (1)-(4) and in addition has a self-consistency property. A model fitting routine using maximum likelihood with a Laplace approximation was implemented in R. Also, methods to calculate the model-based distribution function and the quantile function of AAA size at a given time-point depending on the current AAA size were implemented. The model was fit to a longitudinal data set of 87 patients with a median follow-up time of 1.8 years and a total of 522 AAA diameter measurements. The comparison of prediction intervals across time with observed growth curves showed high agreement, and leave-one-out cross validation revealed that the distribution of diameters at the last visit was accurately predicted based on the individual initial diameters. An online-calculator for predictions using the fitted model was made available. The model may also be applied to other clinically relevant growth or deterioration processes.



SESSION OC5F

## OC5F-4 Predicting Patient Risk for Adverse Drug Events in Health Care Claims Data using Functional Targets

**Mariam R. Rizkallah[1,2], Louis Dijkstra[1], Roland Linder[3], Adalbert F.X. Wilhelm[4], Iris Pigeot[1], Ronja Foraita[1]**

1 Biometry & Data Management, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Germany
2 Computer Science & Electrical Engineering, Jacobs University Bremen, Germany
3 Analytics & Insights, Corporate Development, Techniker Krankenkasse, Hamburg, Germany
4 Psychology & Methods, Jacobs University Bremen, Germany

Adverse drug events (ADEs) represent a burden on health care systems. Detecting ADE signals in pharmacovigilance is mainly achieved using spontaneous reporting systems. Longitudinal data offer a promising alternative due to their high-volume and high-resolution nature. Particularly, health care claims databases contain information on prescriptions, diagnoses and demographic risk factors.
We present a novel strategy for predicting ADEs based on health care claims data, in which we group drug and disease predictors according to their biological functional targets (FTs), i.e., pathways of molecular targets (e.g., receptors). We hypothesize that drugs and diseases involved with an FT are more likely to lead to the ADEs associated with that FT. Exploiting such knowledge may better explain the relationships between predictors, allow for utilizing the full spectrum of longitudinal data, and increase methods' predictive power.
We compared the predictive performance of four methods in three settings: 1) FT-grouping, 2) grouping according to the WHO drug/disease classification, and 3) no grouping (ng). We compared machine learning methods: random forests (RF; ng) and block forests [BF; grouping (g)], and regression methods: the LASSO (g + ng) and an extension of the adaptive rank truncated product (ARTP; g), a method used in genetic applications, to enable outcome prediction.
We applied the strategy to the German Pharmacoepidemiological Research Database, which contains claims data from ~24 million insurants, to predict gastrointestinal bleeding (GIB) and intracranial bleeding (ICB), two known ADEs of direct oral anticoagulants (DOACs). We analyzed data from the years 2015-2016, and created two matched subcohorts (1:4) of adult insurants diagnosed with GIB (N=64,720) or ICB (N=34,600). We controlled for age, sex, region of residence and time-to-event. Performance evaluation measures included the area under the precision-recall curve.
**For** both subcohorts, the LASSO, RF and BF comparably outperformed the ARTP. For GIB, the BF with FT-grouping ranked the DOACs group higher than WHO-grouping did. For ICB, BF with FT-grouping ranked etiology pathways the highest. FT-based grouping may better describe ADE risk profiles, however, dataset size, and FT-group sizes and overlap affect prediction. Regression-based methods, e.g., the group LASSO, are challenged by data dimensions and require scalable implementations.

# SESSION OC5G

## SESSION OC5G — Missing data & measurement error

### OC5G-1 Multiple imputation for missing data in case-cohort studies: simulation and case study

**Melissa Middleton[1,2], Cattram Nguyen[1,2], Margarita Moreno-Betancur[1,2], John Carlin[1,2], Katherine J. Lee[1,2]**
1 Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Australia
2 Department of Paediatrics, University of Melbourne, Australia

**Background:** Case-cohort studies are useful when exposure data are expensive to collect on the full cohort. In case-cohort studies, a random subcohort is selected with probability < 1 and exposure collected only on the subcohort and all cases. The unequal sampling probabilities in case-cohort studies are accounted for during analysis through inverse probability weighting. Missing data is commonly addressed by multiple imputation (MI), but valid use requires compatibility between the imputation and analysis models. When unequal sampling probabilities are incorporated into the analysis, compatibility requires the probabilities also be accommodated for during imputation. It is unclear how to best apply MI in case-cohort studies to address missing covariates in order to achieve compatibility. This study assessed the performance of various approaches to implement MI in the context of a case-cohort study, in which the target analysis was a weighted model estimating either a risk ratio or odds ratio.
**Methods:** A simulation study was conducted with missingness in two covariates, motivated by a case-cohort investigation within the Barwon Infant Study (BIS). MI methods considered were ignoring weights, including an interaction between the outcome (as a proxy for weight groupings) and all other analysis variables, and imputing separately for cases and controls. Factors such as the proportion of incomplete observations, missing data mechanism and subcohort selection probabilities were varied to assess performance of MI methods. A weighted complete case analysis (CCA) was performed on the subcohort and cases with complete covariate information for comparison. MI methods were also applied to a subset of the BIS data.
**Results:** There was similar performance in terms of bias and efficiency in both estimates across all MI methods, with expected improvements compared with the CCA. For all methods, an expected increase in precision as the subcohort selection probability increased was observed. These results were consistent with the case study.
**Conclusions:** Our results suggest that the use of MI to handle missing data is more efficient than a CCA. How weighting is included in the imputation model makes little difference in the analysis of case-cohort studies potentially because there are only two weight classes in this setting.

### OC5G-2 Application of three level multiple imputation in national surveys

**Nidhi Menon[1], Alice Richardson[2], Hwan-Jin Yoon[2]**
1 Biological Data Science Institute, Australian National University, Canberra, Australia
2 Statistical Consulting Unit, Australian National University, Canberra, Australia

The theory of multiple imputation requires the sampling design be incorporated in the imputation process. Not accounting for complex sample design features, such as stratification and clustering, during imputations can yield biased estimates from a design-based perspective. Most datasets in public health research show some form of natural clustering (individuals within households, households within the same district, patients within wards, etc.). Cluster effects are often of interest in health research. Missing values can occur at any level in multilevel data, but there is limited guidance on extending multiple imputation to impute variables captured at three levels or more. This paper implements and extends the Gelman and Hill approach for imputation of missing data at multiple levels by including aggregate forms of individual-level measurements to impute for missing values at higher levels. In our study, we use the fourth National Family Health Survey (NFHS-4) data of India, to implement our extensions of Multiple Imputation using Chained Equations to impute for missing values in three level data structures. The dataset is naturally hierarchical with children nested within mothers who are further nested within households, and aims to identify maternal and household level predictors of anaemia in children in India. This study is novel in its approach in imputing for missing data in the third level (households) in the NFHS survey for India.



SESSION OC5G

### OC5G-3 Profiles of COVID 19-Hematological patients: Franco-Brazilian observational cohort study

**Lilith Faucheux[1], Lucas Bassolli[2], Sylvie Chevret[1], Vanderson Rocha[2]**
1 INSERM-UMR1153, Paris, France
2 Faculdade de Medicina da Universidade de São Paulo, Pacaembu, Brazil

**Background:** The coronavirus disease 2019 (COVID-19) pandemic that began in China, has rapidly spread to the rest of the world. The range of disease presentation is large, from asymptomatic and low severity cases up to severe life-threatening forms. Getting further insights in patient profiles of COVID-19 in immunodeficient patients is of particular interest. A bi-national cohort of 263 patients affected by both COVID-19 and an hematological disease, from France and Brazil, was analyzed.
**Objectives:** We first aimed at identifying COVID-19-hematological patient profiles using data-driven, "unsupervised", methods. A secondary objective was to identify the interest of using a semi-supervised procedure to obtain a partition associated with patient survival, compared to the unsupervised procedure.
**Methods:** Patient profiles were obtained from continuous variables (age, number of comorbidities, biological measurements), and an archetype variable derived from symptoms and underlying hematological disease characteristics using a generalized low rank model. Learning methods adapted to the presence of missing data were used for unsupervised and semi-supervised learning [1] for a survival endpoint [2].
**Results:** Both methods identified two clusters with a few unclassified patients. While the unsupervised clusters differed notably in terms of comorbidities and age, the semi-supervised partition differed from the former (as measured by an ARI of 0.29), additionally distinguishing patients on biological variables such as creatinine level, and with an increased prognostic value. Actually, the 30-day survival rate was 77.1% in the cluster of young patients with low C-reactive Pr, D-dimers, LDH and creatinine levels, compared to 46.7% in the second cluster. The partition and patient age achieved additive prognostic information.
**Conclusion:** The interest of the semi-supervised method was shown, allowing to identify two highly different prognostic profiles of patients with hematological disease and COVID-19, mostly based on age, comorbidities, and biological evidence of inflammation.

**References:** *[1] Faucheux, L., Resche-Rigon, M., Curis, E., Soumelis, V. & Chevret, S. Clustering with missing and left-censored data: A simulation study comparing multiple-imputation-based procedures. Biometrical Journal (2021).*
*[2] Faucheux, L., Soumelis, V. & Chevret, S. Multiobjective semisupervised learning with a right-censored endpoint adapted to the multiple imputation framework (Submitted).*

# SESSION OC5G

## OC5G-4 Prediction of cancer incidence in areas without registries using proxy and registry data

**Edouard Chatignoux[1], Laurent Remontet[2,3], Marc Colonna[4,5], Isabelle Albert[6], Zoé Uhry[1,2]**

1 Santé publique France, Saint-Maurice, France
2 Service de biostatistique, Hospices civils de Lyon, France
3 Université Lyon 1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, Lyon, France
4 Registre des cancers de l'Isère, Grenoble, France
5 Réseau français des registres de cancer Francim, France
6 UMR MIA 518, INRAE-AgroParisTech-Univsertsité Paris-Saclay, Paris, France

**Background:** In France, cancer registries only cover about 20% of the population. To predict incidence at local levels, one may use imperfect proxy data, correlated to incidence, such as health care databases or mortality data. To do so, we developed a calibration model based on modelling the ratio between proxy (P) and incidence (I) observed in the registries area.
**The aims of this study** are to 1) present a global methodology to predict incidence at local level, involving three steps: calibration model, assessment of predictions quality and spatial smoothing for disease mapping 2) show the properties of the estimators derived from the calibration model and 3) illustrate the application of this methodology to predict cancer incidence at the district-level in France (départements).
**Material and methods:** In the registries area, the ratio between the number of patients from the proxy source and incident cases is modelled by age using a Poisson mixed model. This model provides i) a smooth P/I age ratio f(a) and ii) an estimation of the district-variability of the ratio. For a new district, prediction are derived by age using the P numbers divided by f(a). Predictions follow a lognormal distribution, with variances depending on the variability of the ratio. The properties of the predictions were evaluated through realistic simulations. The whole methodology was applied to predict cancer incidence in France over the 2007–15 period for 24 sites, using several health care indicators and mortality.
**Results:** The calibration model provided unbiased estimations of number of incident cases; coverage rates of the 95% predictions intervals ranged from 91 to 96%. Incidence prediction was of sufficient quality for 27/34 solid sex-sites combinations and only 2/8 sex-haematological malignancies combinations. Mapping of smoothed predicted incidence provided a clear picture of the main contrasts in incidence.
**Conclusion:** Our calibration approach is an adequate tool to predict local cancer incidence, using proxy measures and registries data, when registries cover only part of the territory. Future developments are oriented toward a joint modelling of incidence and proxys processes. It offers an interesting perspective to provide predictions at smaller geographic scales.



SESSION OC5G

## OC5G-5 What is the real prevalence of hypertension in France?

**Edouard Chatignoux[1], Valérie Olié[1], Christophe Bonaldi[1], Amélie Gabet[1], Clémence Grave[1], Jacques Blacher[2]**

1 Santé publique France, Sainte-Maurice, France
2 Centre de diagnostic et de thérapeutique, Hôtel Dieu, Université de Paris, France

**Background:** Hypertension (HT), defined by permanent high blood pressure (BP) level, is a leading modifiable risk factor for cardiovascular and renal diseases. In practice, HT is diagnosed as BP (systolic/diastolic) exceeding a threshold level (140/90mmHg). Due to variability of BP measures within one's individual, it is recommended to use BP measurements during several visits for diagnosis of HT.
However, in epidemiological studies, BP is frequently measured during a single visit. In such designs, direct count of patients with hypertension neglects within-person variability and bias estimates of HT prevalence.
The aim of our study was to provide factors taking into account the different components of BP measures variability (between individuals, between visits and between measures) to correct bias in HT prevalence estimation in epidemiological studies. The method was applied to estimate HT prevalence in France in 2015.
**Methods:** We used data from the Nhanes III study, in which patient's BP was measured in three visits with three measures per visit. For each gender and type of BP, components of BP variance were modelled with random effects model. Variances of random effects were allowed to vary with age and modelled using penalized splines. Models were estimated in a Bayesian framework, using Hamiltonian Monte-Carlo (Stan software).
Components of variance allowed calculation of factors to correct estimates of HT prevalence in epidemiological studies. The method was applied to data from the Esteban study, where three standardized BP measurements were performed during a single clinical exam.
**Results:** The shape of the components of BP variability varied greatly with age, with different patterns for systolic and diastolic BP. Variability of BP was driven by between visit and between individual variances, between measures variability being much lower. The raw prevalence of HT in the Esteban study reached 31.4%. After correction for BP variability, estimates decreased to 28.0%. By applying these corrected proportions to the French adult population, the number of hypertensive patients reached 15,300,000 in 2015.
**Conclusion:** Correction of within individual BP variability to estimate HT prevalence from a single measure could avoid a substantial over-estimation of the prevalence of hypertension in the population.

# SESSION OC6A

## SESSION OC6A — Causal inference

### OC6A-1 Incident and prevalent-user designs and the definition of study time origin in pharmacoepidemiology: a systematic review

**Kim Luijken[1], Judith J. Spekreijse[2], Maarten van Smeden[1,3], Helga Gardarsdottir[4,5,6], Rolf H.H. Groenwold[1,7]**

1 Department of Clinical Epidemiology, Leiden University Medical Center, Netherlands
2 Department of Paediatrics, Diakonessenhuis, Utrecht, Netherlands
3 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Netherlands
4 Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Netherlands
5 Department of Clinical Pharmacy, Division Laboratories, Pharmacy and Biomedical Genetics, University Medical Center Utrecht, Utrecht University, Netherlands
6 Faculty of Pharmaceutical Sciences, University of Iceland, Reykjavik, Iceland
7 Department of Biomedical Data Sciences, Leiden University Medical Center, Netherlands

**Background:** Guidelines for observational comparative effectiveness and drug safety research recommend implementing an incident-user design whenever possible, since it reduces the risk of selection bias in exposure effect estimation compared to a prevalent-user design. The uptake of these guidelines has not been studied extensively.
**Methods:** We reviewed 89 observational effectiveness and safety cohort studies published in six pharmacoepidemiological journals in 2018 and 2019. We developed an extraction tool to assess how frequently incident-user and prevalent-user designs were implemented. Specifically, we extracted information about the extent to which start of follow-up, treatment initiation and moment of meeting eligibility criteria were aligned in the study design.
**Results:** Of the 89 studies included, 40% implemented an incident-user design for both the active arm and the comparator arm, while 13% implemented a prevalent-user design in both arms. Of the 40 studies implementing a prevalent-user design in at least one treatment arm, 4 provided a rationale for including prevalent users. The start of follow-up, treatment initiation and moment of meeting eligibility criteria were aligned in both treatment arms in 22% of studies. We provided examples of studies that minimized the risk of introducing bias due to left truncation, immortal time, or a time lag.
**Conclusions:** Almost half of the included studies followed the recommendation to implement an incident-user design. Rationales for implementing a prevalent-user design were scarcely reported. Information on the extent to which start of follow-up, treatment initiation and moment of meeting eligibility criteria were aligned by design was often incomplete. We recommend that the choice for a particular study time origin is explicitly motivated to enable assessment of validity of the study.



SESSION OC6A

### OC6A-2 A framework for meta-analysis of studies with baseline exposure through standardized survival curves

**Joris Hautekiet[1,2], Els Goetghebeur[2,3]**

1 Cancer Centre, Sciensano, Belgium
2 Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium
3 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden

Meta-analyses of survival studies assessing the effect of a well specified exposure at baseline, ideally combine measures of the 'causal' effect(s) seen in different studies into one overall meaningful measure. Standard forest plots of hazard ratios however mix apples and oranges into an ill interpretable cocktail. To enhance interpretation and transportability, one must confront between study heterogeneity in more than the usual dimensions. While randomized studies expect different covariate distributions between studies, observational studies additionally encounter baseline covariate heterogeneity between exposure groups within studies. There is the issue of study duration and hence maximum follow-up time. Also, the nature of possibly informative or explainable censoring must be addressed. Last but not least there is the complicating role of the non-collapsible hazard ratio typically presented in forest plots as the study-specific and overall summary of survival contrasts between exposure groups. With many features ignored, implausible and hidden underlying assumptions may critically affect interpretation of the pooled effect measure. We propose several different forms of standardized survival curves over a fitting time horizon to derive estimates of distinct target estimands most relevant in specific settings. Corresponding forest plots then show standardized risk differences at well-chosen time points. Alternatively post hoc summaries may contrast standardized survival curves by hazard ratios or restricted mean survival time in this two-step approach. We discuss advantages and disadvantages of direct standardization besides transportability of results for meta-analyses of randomized treatments, observed treatments as well as patient-specific exposure characteristics. Our case study analyzes individual patient data of six studies to evaluate how the adjusted effect of tumor marker p16 on overall survival of squamous anal cancer patients may help guide treatment policies.

### OC6A-3 Causal inference in practice: two case studies in nephrology

**Christine Wallisch[1], Maria C. Haller[1,2], Michael Kammer[1,3], Georg Heinze[1], Rainer Oberbauer[3], Susanne Strohmaier[1,4]**

1 Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria
2 Department of Medicine III, Nephrology, Hypertension, Transplantation, Rheumatology, Geriatrics, Elisabethinen Hospital, Ordensklinikum Linz, Austria
3 Division of Nephrology and Dialysis, Department of Medicine III, Medical University of Vienna, Austria
4 Department of Epidemiology, Center for Public Health, Medical University of Vienna, Austria

For certain medical research questions, randomization is unethical or infeasible and then causal effects are estimated from observational data. If confounding and exposure status are time-dependent, this requires sophisticated methodology such as target trial emulation involving longitudinal matching methods. Here we discuss issues in two examples from nephrology aiming to quantify survival benefits of (first/second) kidney transplantation compared to remaining on dialysis and never receiving an organ, across ages and across times since waitlisting.
We analyzed data from the Austrian Dialysis and Transplant Registry comprising patients on dialysis and waitlisted for a kidney transplant with repeated updates on patient characteristics and waitlisting status. As often with registry data, a tricky task was data management, i.e. dealing with inconsistencies and incompleteness. Data availabilities also had to be taken into consideration when deciding on the most relevant causal effect that could be identified and estimated. We adapted the approaches of Gran et al. (2010) and Schaubel et al. (2006) by emulating a series of target trials, where each trial was initiated at the time of a transplantation (relative to time of first/second waitlisting). Transplanted patients contributed to the treatment group while patients with current active waitlisting status were classified to the control group. Controls were artificially censored if they were transplanted at a later time and their transplantation then initiated a further trial of the series. Estimation of inverse probability of treatment and censoring weights (IPTWs, IPCWs) to achieve exchangeability and to account for artificial censoring required many practical decisions. We applied pooled logistic regression adjusted for time-varying patient characteristics to estimate IPTWs and trial specific Cox proportional hazards models to compute yearly updated IPCWs. The marginal effect and the effect conditional on age and duration of waitlisting expressed on different scales (hazard ratios, survival probabilities and restricted mean survival times) were obtained from Cox models weighted by the product of IPTWs and IPCWs fitted to the stacked data set of all trials. The bootstrap approach to compute confidence intervals additionally increased computation time. We critically discuss our proposed solution and possible alternatives and show results from our nephrological case studies.

# SESSION OC6A

## OC6A-4 Machine learning, G-computation and small sample sizes: a simulation study

**Arthur Chatton[1,2], Florent Le Borgne[1,2], Maxime Léger[1,3], Rémi Lenain[1,4], Yohann Foucher[1,5]**
1 UMR INSERM 1246 SPHERE, Université de Nantes, Université de Tours, Nantes, France
2 IDBC/A2COM, Pacé, France
3 Centre Hospitalier Universitaire d'Angers, Angers, France
4 Centre Hospitalier Universitaire de Lille, Lille, France
5 Centre Hospitalier Universitaire de Nantes, Nantes, France

Drawing causal inferences from cohorts consisting of some hundred patients is challenging, especially when the data-generating mechanism is unknown. While machine learning approaches are increasingly used in prediction, their applications for causal inference are more recent. However, several causal inference methods involve a prediction step where such data-adaptive approaches could be helpful.
We propose an approach combining machine learning and G-computation1 to estimate the causal effect of a binary exposure on a binary outcome. We evaluated and compared, through a simulation study, the performances of penalized logistic regressions, neural network, support vector machine, boosted classification, regression trees, and an ensemble method called super learner2. We proposed six different scenarios, including various sample sizes and relationships between covariates, binary exposure, and binary outcome. In a cohort of 252 patients, we have also illustrated the application of these approaches used to estimate the effect of barbiturates prescribed during the first 24h of an episode of intracranial hypertension.
We reported that, used in a G-computation approach to estimate the probabilities of the individual counterfactual outcomes, the super learner tended to outperform other approaches in terms of bias and variance, especially for small sample sizes. Support vector machine also resulted in performant properties, albeit the mean bias was slightly higher compared to the super learner.
We show empirically that, contrary to a preconception, using a machine learning approach to draw causal inferences can be valid even for a sample constituted by one hundred subjects, as in the majority of medical studies. The G-computation with the super learner is available in the R package RISCA.

**References:** *[1] Hernán M, Robins JM. Causal Inference: What if? Boca Raton: Chapman & Hall/CRC; 2020. [2] Naimi AI, Balzer LB. Stacked generalization: an introduction to super learning. European Journal of Epidemiology. 2018 May;33(5):459–464.*

## OC6A-5 Valid Uncertainty Interval for the Average Causal Effect in a High-dimensional Setting

**Niloofar Moosavi, Xavier de Luna, Tetiana Gorbach**
USBE/Statistics, Umeå University, Sweden

During the last years, a great extent of work has been done on constructing confidence intervals for average causal effect parameters that are uniformly valid over a set of data generating processes even when high-dimensional nuisance models are estimated by post-model-selection or machine learning estimators. These developments assume that all the confounders are observed to ensure point identification. We contribute by showing that valid inference can be obtained in the presence of unobserved confounders and high-dimensional nuisance models. We thus propose uncertainty intervals, which allow for nonzero confounding bias. The later bias is specified and estimated and is function of the amount of unobserved confounding allowed for. We show that valid inference can ignore the finite sample bias and randomness in the estimated value of confounding bias by assuming that the amount of unobserved confounding is small relative to the sample size; the latter is formalized in terms of convergence rates. An interpretation is that more confounders are collected as the sample size grows. Simulation results are presented to illustrate finite sample properties, and the application of the results is illustrated with a study, where we fit the effect of regular food intake on health.

# SESSION OC6B

SESSION OC6B — Adaptive clinical trial design

## OC6B-1 Adaptive designs for three-arm gold-standard non-inferiority trials

**Jan Meis, Maximilian Pilz, Meinhard Kieser**
Institute of Medical Biometry and Informatics, University of Heidelberg, Germany

A common criticism of non-inferiority trials comparing an experimental treatment to an active control is that they may lack assay sensitivity. This denotes the ability of a trial to distinguish an effective treatment from an ineffective one. The 'gold-standard' non-inferiority trial design circumvents this concern by comparing three groups in a hierarchical testing procedure. First, the experimental treatment is compared to a placebo group in an effort to show superiority. Only if this succeeds, the experimental treatment is tested for non-inferiority against an active control group. Ethical and practical considerations require sample sizes of clinical trials to be as large as necessary, but as small as possible. These considerations come especially pressing in the gold-standard design, as patients are exposed to placebo doses while the control treatment is already known to be effective.
Group sequential trial designs are known to reduce the expected sample size under the alternative hypothesis. In their pioneer work, Schlömer and Brannath (2013) show that the gold-standard design is no exception to this rule. They calculate approximately optimal rejection boundaries for the group-sequential gold-standard design using sample size allocation ratios of the optimal single-stage design. We extend their work by relaxing the constraints put on the group allocation ratios and allowing for futility stops at interim. Using this new design, we will present methods to calculate power and type I error conditional on the observations at interim. These extensions tackle a pressing issue in the traditional gold-standard design: Disappointing results at interim might make a positive result in the final analysis very unlikely. In such scenarios, stopping a trial for futility or reassessing the required sample size whilst controlling the conditional type I error rate can minimize the risk of further patients being unnecessarily exposed to ineffective treatment. Besides the extended design options, we analyze different choices of optimality criteria. The above considerations suggest that the null hypothesis also plays a vital role in the assessment of the gold-standard design. Therefore, optimality criteria that incorporate the design performance under the alternative and the null hypothesis are introduced.

**Reference:** *[1] Schlömer, P., & Brannath, W. (2013). Group sequential designs for three‐arm 'gold standard'non‐inferiority trials with fixed margin. Statistics in Medicine, 32(28), 4875-4889.*

## OC6B-2 A comparison of estimation methods adjusting for selection bias in adaptive enrichment designs with time-to-event endpoints

**Fulvio Di Stefano[1], Matthieu Pannaux[2], Anne Correges[2], Stephanie Galtier[2], Veronique Robert[2], Gaelle Saint-Hilary[1,2]**
1 Dipartimento di Scienze Matematiche (DISMA) Giuseppe Luigi Lagrange, Politecnico di Torino, Italy
2 Department of Clinical Biostatistics, Servier, France

Adaptive enrichment designs in clinical trials have been developed to enhance drug developments. They permit, at interim analyses during the trial, to select the sub-population that benefits the most from the treatment. This results in improvements both in terms of resources and ethics, by reducing the number of patients receiving non effective treatments. Because of this selection, the naive maximum-likelihood estimation of the treatment effect, commonly used in classical randomized controlled trials, is biased. In the literature, several methods have been proposed to obtain a better estimation of the treatments' effects in such contexts. To date, most of the works have focused on normally endpoints, and some estimators have been proposed for time-to-event endpoints but they have not all been compared side-byside. In this work, we conduct an extensive simulation study, inspired by a real case-study in Heart Failure, to compare the maximum-likelihood estimator (MLE) with an unbiased estimator, shrinkage estimators and bias-adjusted estimators for the estimation of the treatment effect with time-to-event data. The performances of the estimators are evaluated in terms of bias, variance and mean squared error. Based on the results, we recommend using the unbiased estimator and the single-iteration bias-adjusted estimator: the former completely eradicates the bias, but is highly variable with respect to a naive estimation; the latter is less biased than the MLE estimator and only slightly more variable.

# SESSION OC6B

## OC6B-3 A stochastically curtailed two-arm randomized phase II trial design for binary outcomes

**Martin Law[1], Michael J. Grayling[2], Adrian P. Mander[3]**
1 Hub for Trials Methodology Research, Medical Research Council Biostatistics Unit, University of Cambridge, UK
2 Population Health Sciences Institute, Newcastle University, UK
3 College of Biomedical and Life Sciences, Cardiff University, UK

Randomised controlled trials are considered the best way of assessing a new treatment. However, in the early phases of assessing cancer treatments, where treatment outcome is binary, trials are often single-arm. Consequently, the results from the new trial then have to be compared to results from another source. This can make it more difficult to estimate how well the new treatment works, as the two sets of results may come from groups that are different in various ways. Though a number of reasons exist for choosing a single-arm trial, the primary reason is that single-arm designs require fewer participants than their randomised equivalents. Therefore, the development of randomised trial designs that reduce sample size is of value to the trials community. This talk introduces a randomised two-arm binary outcome trial design that greatly reduces expected sample size compared to typical randomised trial designs. Sample size is reduced in two ways: firstly, the design uses stochastic curtailment, which means allowing the possibility of stopping a trial before the final conclusions are known with certainty. Secondly, the proposed design uses randomised blocks, which allows investigators to control the number of interim analyses, the points in the trial at which early stopping may occur. This proposed design is compared with existing designs that also use early stopping, in terms of maximum and expected sample size, through the use of a loss function. Comparisons are also made using an example from a real trial. The comparisons show that in many cases, the proposed design is superior to existing designs. Further, the proposed design may be more practical than existing designs, by allowing a flexible number of interim analyses. One existing design produces superior design realisations when the anticipated response rate is low. However, when using this existing design, the probability of incorrectly concluding that a treatment works can be higher than expected if the response rates are different to what was anticipated. Therefore, when considering randomised designs in cancer trials or for any trial with a binary outcome, we recommend the proposed approach over other designs that allow early stopping.

## OC6B-4 Fast-tracking clinical trial innovations – a COVID-19 silver lining

**Alyssa M. Vanderbeek, Zhulin Yin, Christina Yap**
Clinical Trials and Statistics Unit, Institute of Cancer Research, United Kingdom

**Background:** Platform trials can save time, resources, and speed decision-making by offering the flexibility to add new treatment options to an ongoing trial as they become available. But as a modern approach to research, their use has been limited to a concise number of trials mostly in oncology. COVID-19 has provided a ripe opportunity to change this. High profile platform trials RECOVERY and Solidarity have received considerable favorable attention for their efficient conduct and success in identifying improved treatments. However, the full extent of platform design adoption in COVID-19 has not yet been assessed. We present a rapid review to systematically explore how platform trials are operating in the fight against COVID-19.
**Methods:** We conducted searches in PubMed, ClinicalTrials.gov, and the Cytel COVID-19 Clinical Trials Tracker between October 21st and November 4th 2020. Platform trials were defined by their explicitly stated flexibility to add future arms. Selection of relevant trials was validated on a sample between two independent reviewers.
**Results:** Forty-five platform trials in COVID-19 were registered in the first 10 months of 2020. Of these, 41 (91%) incorporate adaptive features and 13 (29%) clearly state a Bayesian approach. Multiple trials have tested the same therapies, and justification for this is not clear in all cases. REMAP-CAP has opened 15 treatment arms for COVID-19, while remaining trials have had between 2 and 9 active arms. RECOVERY has both opened and closed 3 arms to date. Twenty trials (44%) have publicly shared their protocols; only three provide full statistical analysis plans. Fifteen trials (33%) have committed to sharing individual patient data (IPD).
**Conclusions:** The huge surge of platform design use in COVID-19 is promising for the future of clinical trial design and conduct. Future work may examine what unique challenges these trials face during conduct compared to more standard designs. COVID-19 is a global tragedy, but the experience gathered from the rapid and widespread implementation of platform trials will greatly accelerate our understanding and adoption of these innovative designs.



# SESSION OC6B

## OC6B-5 Adaptive clinical trials with selection of composite endpoints and sample size reassessment

**Marta Bofill Roig[1,2], Guadalupe Gómez Melis[2], Martin Posch[1], Franz Koenig[1]**
1 Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Vienna, Austria
2 Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, Spain

For randomized clinical trials where a single, primary, binary endpoint would require unfeasibly large sample sizes, composite endpoints are widely chosen as the primary endpoint. For example, in cardiovascular trials, it is usual to consider the composite of death, fatal myocardial infarction (MI), and hospitalization as primary endpoint, and death or fatal-MI as main secondary endpoint. Despite being commonly used, composite endpoints entail challenges in designing and interpreting results. Given that the components may be of different relevance and have different effect sizes, the choice of components must be made carefully (EMA, 2017). Especially, sample size calculations for composite binary endpoints depend not only on the anticipated effect sizes and event rates of the composite components but also the correlation between them (Bofill & Gómez, 2019). However, information on the correlation between endpoints is usually not reported in the literature which can be an obstacle for sound trial design.
We consider two-arm randomized controlled trials with a primary composite binary endpoint and an endpoint that consists only of the clinically more important component of the composite endpoint. We propose a trial design that allows an adaptive modification of the primary endpoint based on blinded information obtained at an interim analysis. Especially, we consider a decision rule to select between a composite endpoint and its most relevant component as primary endpoint. The decision rule chooses the endpoint with the lower estimated required sample size. Additionally, the sample size is reassessed using the estimated event rates and correlation, and the expected effect sizes of the composite components. We investigate the statistical power and significance level under the proposed design through simulations. We show that the adaptive design is equally or more powerful than designs without adaptive modification on the primary endpoint. Besides, the targeted power is achieved even if the correlation is misspecified at the planning stage while maintaining the type I error. All the computations are implemented in R and illustrated by means of a cardiovascular trial.

**References:** *[1] Bofill Roig, M., Gómez Melis, G. (2019). A new approach for sizing trials with composite binary endpoints using anticipated marginal values and accounting for the correlation between components. Statistics in Medicine, 38(11), 1935-1956.*
*[2] European Medicines Agency. Guideline on multiplicity issues in clinical trials. June 2017.*

# SESSION OC6C

## SESSION OC6C — Joint models & 2-stage approach

### OC6C-1 Marginalized two-part joint modeling of longitudinal semi-continuous responses and survival data: with application to medical expenses

**Mohadeseh Shojaei Shahrokhabadi[1], Ding-Geng (Din) Chen[1], Sayed Jamal Mirkamali[2]**
1 Department of Statistics, University of Pretoria, Pretoria, South Africa
2 Department of Mathematics, Faculty of Sciences, Arak University, Arak, Iran

The distribution of medical cost data is generally right-skewed, with a substantial number of zero values. Data censorship due to incomplete follow-up and death of participants is quite common. It is important to consider the potential dependence of survival status and medical cost where this censorship is death-related. Dismissing cost data in survival analysis can lead to a biased estimation of the parameters. Despite the wide use of conventional two-part joint models (CTJMs) to capture zero inflation, they are limited to conditional interpretations of the regression coefficients in the model's continuous part. In the present study, we propose a marginalized two-part joint model (MTJM) to analyze semi-continuous longitudinal cost data and survival data. The aim of this study was to extend the joint modeling approach to handle marginal inferences about the covariate effects on the average expenditures. We conducted a simulation study to evaluate the performance of the proposed MTJM compared to the CTJM. To illustrate the properties of the MTJM, we applied the model to a set of real electronic health record data recently collected in Iran. We found that the MTJM yields a smaller standard error and root mean square error of estimates. Furthermore, we identified a significant positive correlation between cost and survival confirmed the simulation results. In summary, the MTJM is a better alternative to the CTJM for handling marginal inferences for death-related censoring.

**References:** *[1] Rustand D, Briollais L, Tournigand C, Rondeau V. Two-part joint model for a longitudinal semicontinuous marker and a terminal event with application to metastatic colorectal cancer data. Biostatistics. 2020. [2] Smith VA, Neelon B, Preisser JS, Maciejewski ML. A marginalized two-part model for longitudinal semicontinuous data. Statistical Methods in Medical Research. 2017;26(4):1949-68.*

### OC6C-2 Healthy life expectancy computation using the Item Response Theory combined with a joint modeling approach

**Cécile Proust-Lima, Viviane Philipps, Ahmadou Alioum, Karine Pérès**
Univ. Bordeaux, INSERM, BPH, U1219, F-33000 Bordeaux, France

Healthy life expectancy and more broadly life expectancy without specific conditions constitute key epidemiological quantities for public health policies, notably to estimate the associated needs and to organise and plan health and welfare systems. In some contexts, the target condition is a specific disease diagnosed at a certain time, and an illness-death multi-state approach can be employed to derive such quantities from prospective cohort data. In other contexts, such as the disability-free life expectancy, the conditions are defined as the occurrence of one or various limitations that are intermittently assessed at cohort visits using items from measurement scales. We propose an original approach to compute epidemiological quantities when the target conditions are defined according to the items of a measurement scale repeatedly measured over time. We rely on the Item Response Theory (IRT) that we extend to handle repeated item measures and association with a time-to-event. Specifically, (i) the underlying dimension measured by the repeated items is modeled over time using a mixed model, (ii) ordinal and continuous items are considered using appropriate measurement equations, and (iii) association with the time-to-event is modeled within the shared random effect framework. From the Maximum Likelihood Estimates of this IRT joint model (obtained within lcmm R package), we derive quantities such as life expectancy without certain items limitations or expected duration with intermediate levels of item limitations. The methodology is validated using simulations and contrasted with the multi-state modeling approach. Given the major challenge of aging and associated dependency, we apply this methodology to the analysis of the limitations in basic and instrumental activities of daily living (ADL and IADL). Using the long-term data of French population-based aging cohorts, we provide estimations of life expectancy under different scenarios of dependency, for instance life expectancy without severe limitation in any basic ADL or expected duration with mild to moderate dependency (i.e., IADL limitation but no ADL limitation). Compared to multi-state models, this methodology has the advantage of integrating all the available information of the measurement scale and its underlying structure, and naturally handles interval censoring.



SESSION OC6C

### OC6C-3 Modelling the effect of longitudinal latent toxicity profiles on survival: an application to osteosarcoma

**Marta Spreafico[1,2], Francesca Ieva[1], Marta Fiocco[2]**
1 MOX – Department of Mathematics, Politecnico di Milano, Milan, Italy
2 Mathematical Institute, Leiden University, Leiden, Netherlands

**Background:** In many clinical applications involving longitudinal data, the interest lies in analysing the evolution of similar latent profiles related to subgroups of individuals rather than in studying their observed attributes. Since latent variables can be considered as the outcomes of a stochastic latent process which determines the evolution of observed responses, these characteristics may reflect patients' quality-of-life, including valuable information related to patient's health status and disease progression.
**In** cancer trial, the analysis of longitudinal chemotherapy data is a complex task due to the presence of negative feedbacks between exposure to cytotoxic drugs and treatment toxicities. Models for time-to-event able to deal with the longitudinal nature of toxicity evolution during treatment are necessary, still not well developed.
**Objectives:** The main purpose is to study the evolution of longitudinal latent toxicity progression and their association with time-to-event outcome. New methods to reconstruct the longitudinal latent profiles of toxicity evolution by means of Latent Markov Models (LMM) and Functional Data Analysis (FDA) are developed. Inclusion into survival models is further discussed.
**Methods:** Data from MRC-BO06/EORTC-80931 randomised controlled trial for osteosarcoma treatment are used. Non-haematological chemotherapy-induced toxicity for nausea, mucositis, ototoxicity, infection, neurological and cardiac toxicity are registered at each cycle and graded according to the Common Terminology Criteria for Adverse Events. First, a LMM approach is applied to identify and reconstruct the latent profiles over chemotherapy cycles in terms of observed toxicity grades. FDA techniques to represent the latent characteristics in terms of dynamic functions are exploited. Through dimensionality reduction methods, the functional latent profiles are finally included into survival models.
**Results & Conclusions:** The proposed methodology allows (i) to move from complex chemotherapy data to different subgroups of individuals with similar patterns of toxicity progression by identifying the latent states, (ii) to reconstruct the longitudinal latent toxicity profiles in a tailored way and (iii) to quantify the association with patient's survival. This approach represents a novelty for osteosarcoma treatment, providing new insights for childhood cancer. Thanks to its flexibility, this procedure could be tailored to other fields according to the needs of the study.

# SESSION OC6C

## OC6C-4 Joint model versus linear mixed model to analyze longitudinal data of health-related quality of life in cancer clinical trials

**Célia Touraine[1,2,3], Benjamin Cuer[1,2], Thierry Conroy[4,5], Beata Juzyna[6], Sophie Gourgou[1,2], Caroline Mollevi[1,2,3]**
1 Cancer Institute of Montpellier, Univ Montpellier, France
2 National Platform Quality of Life and Cancer, France
3 IDESP, Inserm, Univ Montpellier, France
4 Lorraine Institute of Oncology, France
5 APEMAC, Univ Lorraine, France
6 R&D Unicancer, France

Health-related quality of life (HRQoL) is increasingly used as an endpoint in cancer randomized clinical trials. In palliative settings, the observation of the longitudinal HRQoL outcome is often truncated by death. HRQoL and risk of death being hardly independent, unbiased HRQoL estimates cannot be obtained unless modeling the death process jointly with the HRQoL outcome, for instance with a joint model (JM). However, the linear mixed model (LMM) remains the usual strategy to analyze HRQoL longitudinal data. The aim of this work is to highlight the situations where the LMM fails, and to measure the size and direction of the biases and their consequences on the clinical conclusions.

We have first compared the most frequently used LMM, namely the random intercept and slope model, and its corresponding JM on data of patients suffering from metastatic pancreatic cancer. HRQoL was assessed using the EORTC QLQ-C30 questionnaire and we focused on six dimensions of interest. From this application, we have also derived assertions on the situations where biases arise from the LMM, then confirmed and complemented through a simulation study.

Death acting as informative dropout, larger the risk of death in a given arm, larger the bias of the slope governing the HRQoL trajectory in this arm. Thus, if death is associated with poor HRQoL, the LMM will be too optimistic. In the application, the JM found that (compared to the LMM) HRQoL in the control arm improved less (global health status/QoL, role functioning, fatigue, pain), deteriorated more (physical functioning) or deteriorated instead of improved (social functioning). If survival is similar between arms, the two slopes will be equally biased by the LMM, maintaining the right slope difference. However, if survival is better in the experimental arm because of the direct treatment effect and/or a better HRQoL (indirect treatment effect), the HRQoL slope will be less biased in this arm, and thus the arm-by-time interaction effect will be biased in disfavor of the experimental arm.

Based on real and simulated data, this work reveals when, how and why the LMM produces biased estimates and shows the interest of using instead a JM.

# SESSION OC6C

## OC6C-5 Joint modelling of the temporal relationships between multivariate longitudinal markers of Alzheimer's disease progression and clinical endpoints

**Anaïs Rouanet[1], Bachirou Taddé[2], Cécile Proust-Lima[1]**
1 Inserm U1219 – Bordeaux Population Health Research Center, France
2 IQVIA, France

**Background:** Alzheimer's disease and related dementias (ADRD) are a major public health issue with substantial economic and social costs. Many neuropathological processes have been identified in the preclinical progression of ADRD including protein accumulation in the brain, regional brain atrophies and cognitive dysfunction with repercussions on daily life. Yet, the underlying mechanisms linking these processes over time are still poorly understood. The wealth of biomarker data measured repeatedly over time now available in aging cohorts offers an unprecedented opportunity to better understand the complex pathways of ADRD, essential to improve prevention strategies and individual healthcare. However, current statistical methods are not designed to identify a structure of temporal dependence between repeated biomarkers.

**Objective:** We propose an original joint modelling framework to describe the complex dynamics of the different processes involved in ADRD progression along with the ADRD diagnosis and death, and apprehend their temporal relationships.

**Methods:** The longitudinal submodel is a dynamic causal model which combines the multivariate mixed model theory with difference equations to explain the future change over time of each dimension according to the history of the others (Bachirou et al., 2020). The longitudinal model accommodates various natures of biomarkers (continuous, ordinal). The association of the biomarkers with ADRD diagnosis and death are described via a shared random effect joint modelling approach. Inference is performed in the Maximum Likelihood framework within the CInLPN R package.

**Results:** The methodology is applied on a French epidemiological cohort on cognitive ageing, the 3-City study. In particular, we aim to disentangle the temporal relationships between three major drivers of ADRD natural history, cerebral atrophy, cognition and functional autonomy, in link with the two major clinical events in ADRD progression: diagnosis and death. This methodology and application envision to help understand the complex interplay between the biomarkers of ADRD, and identify relevant targets which may slow down the progression at each stage of the disease.

Although primarily motivated by the ADRD study, this dynamic causal joint model constitutes a novel tool to investigate temporal associations between repeated markers with potential notably in longitudinal mediation analyses.

**Reference:** *[1] Bachirou et al., Biometrics, 2019, https://doi.org/10.1111/biom.13168.*

# SESSION OC6D

## SESSION OC6D — Time-to-event methods for non-proportional hazards

### OC6D-1 Methods for analyzing time-to-event endpoints in immuno-oncological trials with delayed treatment effects

**Rouven Behnisch, Johannes Krisam, Meinhard Kieser**
Institute of Medical Biometry and Informatics, University of Heidelberg, Germany

In cancer drug research and development, immunotherapy plays an ever more important role. A common feature of immunotherapies is a delayed treatment effect which is quite challenging when dealing with time-to-event endpoints [1]. In case of time-to-event endpoints, regulatory authorities often require a log-rank test, which is the standard statistical method. The log-rank test is known to be most powerful under proportional-hazards alternatives but suffers a substantial loss in power if this assumption is violated. Hence, a rather long follow-up period is required to detect a significant effect in immunooncology trials. For that reason, the question arises whether methods exist that are more susceptible to delayed treatment effects and that can be applied early on to generate evidence anticipating the final decision of the log-rank test to reduce the trial duration without inflation of the type I error. Such alternative methods include, for example, weighted log-rank statistics with weights that can either be fixed at the design stage of the trial, or chosen based on the observed data, tests based on the restricted mean survival time, survival proportions, accelerated failure time (AFT) models, or additive hazard models.
We evaluate and compare these different methods systematically with regard to type I error control and power in the presence of delayed treatment effects. Our systematic simulation study for type I error includes aspects such as different censoring rates and types, different times of delay, and different failure time distributions. For a preliminary power evaluation, simulations based on published data from the CHECKMATE trials were performed. First results show that most methods achieve type I error rate control and that, by construction, the weighted log-rank tests which place more weight on late time points have a greater power to detect differences when the treatment effect is delayed. It is furthermore investigated whether and to what extent these methods can be applied at an early stage of the trial to predict the decision of the log-rank test later on.

**References:** *[1] T. Chen (2013): Statistical issues and challenges in immuno-oncology. Journal for ImmunoTherapy of Cancer 1:18*

### OC6D-2 Non-proportional hazards in immuno-oncology: Is an old perspective needed?

**Dominic Magirr**
Advanced Methodology and Data Science, Novartis Pharma AG, Switzerland

A fundamental concept in two-arm non-parametric survival analysis is the comparison of observed versus expected numbers of events on one of the treatment arms (the choice of which arm is arbitrary), where the expectation is taken assuming that the true survival curves in the two arms are identical. This concept is at the heart of the counting-process theory that provides a rigorous basis for methods such as the log-rank test. It is natural, therefore, to maintain this perspective when extending the log-rank test to deal with non-proportional hazards, for example, by considering a weighted sum of the "observed - expected" terms, where larger weights are given to time periods where the hazard ratio is expected to favor the experimental treatment. In doing so, however, one may stumble across some rather subtle issues, related to difficulties in the interpretation of hazard ratios, that may lead to strange conclusions. An alternative approach is to view non-parametric survival comparisons as permutation tests. With this perspective, one can easily improve on the efficiency of the log-rank test, while thoroughly controlling the false positive rate. In particular, for the field of immuno-oncology, where researchers often anticipate a delayed treatment effect, sample sizes could be substantially reduced without loss of power.



SESSION OC6D

### OC6D-3 Weighted hazard ratio for time to event endpoints under non proportional hazards

**Bharati Kumar, Jonathan Bartlett**
Department of Mathematical Sciences, University of Bath, United Kingdom

Non-proportional hazards (NPH) have been observed in confirmatory clinical trials with time to event outcomes. Under NPH, the hazard ratio does not stay constant over time and the log-rank test is no longer the most powerful test. The weighted log-rank test (WLRT) has been introduced to deal with the presence of non-proportionality in clinical trial data. The WLRT allows for different weights to be assigned at time points where we expect a higher treatment effect in a particular time of the study. For example, in a delayed treatment effect setting, we may down-weight the early time points and hence emphasise the latter part of the survival curve. We will focus our attention on the WLRT and the complementary Cox model based on time-varying treatment effect proposed by Lin and Leon [1]. The model incorporates treatment effect β weighted by the effect adjustment factor *A(t)* which is essentially the weight function *w(t)* scaled by max *(w(t))*. The estimate derived from the model provides a description of the treatment effect over time where β is interpreted as the maximal treatment effect. For this study, two scenarios representing the NPH pattern were simulated. The WLRT and the complementary Cox model were applied to the simulated data. In the diminishing treatment effect scenario, the NPH are simulated in a way such that the chosen weight function would be optimal from a testing perspective. However, the proposed model [1] gives biased estimates and overestimates the treatment effect. According to the Lin & Leon suggestion of estimation, the treatment effect appears to diminish more rapidly than the real time profile of the treatment effect. This leads to a large difference between the estimated treatment effect and the true treatment effect. We conclude that the proposed weighted log-rank test [1] and the complementary Cox model based on time-varying treatment effect cannot be recommended as the primary analysis tool in a randomised controlled trial.

**References:** *Lin, R. S. & Leon, L. F., 2017. Estimation of treatment effects in weighted log-rank tests. Contemporary Clinical trials Communications, Volume 8, pp. 147-155.*

### OC6D-4 Comparison of statistical methods for estimating time-varying treatment effect on adverse event risk

**Liliane Manitchoko[1], Jacques Bénichou[1,2], Anne Thiébaut[1]**
1 High-Dimensional Biostatistics for Drug Safety and Genomics, CESP, Inserm U 1018, Université Paris-Saclay, UVSQ, France
2 Department of Biostatistics, Rouen University Hospital, France

**Background:** In pharmacoepidemiology, assessing the effect of drug exposure on the risk of an adverse event is challenging because exposure can vary over time and its effect can be complex. Cohort and nested case control (NCC) designs are widely used in this context. However, the complexity introduced into the exposure-response relationship by the need to account for the temporal variation of exposure makes it necessary to evaluate the properties of the resulting estimators.
**Methods:** We simulated 1000 prospective cohorts of 5000 individuals for fixed or time-varying exposure with a unique change (from "unexposed" to "exposed") during follow-up. We varied exposure prevalence, hazard ratios of event associated with exposure and proportions of subjects who experienced the event (cases). The resulting data for the entire cohort were analyzed using conventional Cox model with time-invariant or time-dependent covariates. For the NCC design, k controls (k up to 20) were selected for each case and analysis relied on conditional logistic regression.
**Results:** In all scenarios, the cohort design had small relative bias (<5%), was more precise and had greater power than the NCC design. The NCC design displayed negative bias that tended to increase with higher proportion of events for both types of exposure and higher hazard ratio, while it decreased with lower exposure prevalence for fixed exposure and lower hazard ratio. It also decreased with increasing number of controls per case (from more than 20% for one control to less than 8% for 5 controls or more), but remained substantial relative to the cohort design. These methods were applied to estimate the risk of breast cancer associated with menopausal hormone therapy (MHT) in 38,092 women of the E3N cohort, of whom 45.1% were considered exposed to MHT (ever users) at baseline and 5.9% were diagnosed with breast cancer during follow-up. Differences observed between the two designs were consistent with simulated data.
**Conclusion:** Although the NCC design may help refine cohort analyses, results from this design should be interpreted with caution given its potential limitations. More complex exposures (e.g., time-varying with multiple changes or decreasing hazard ratios) are under evaluation.

# SESSION OC6D

## OC6D-5 Evidence Synthesis of Time-To-Event Outcomes in the Presence of Non-Proportional Hazards

**Suzanne Freeman[1], Nicola Cooper[1], Alex Sutton[1], Michael Crowther[1,2], James Carpenter[3,4], Neil Hawkins[5]**
1 Departent of Health Sciences, University of Leicester, United Kingdom
2 Department of Medical Epidemiology & Biostatistics, Karolinska Institutet, Sweden
3 MRC Clinical Trails Unit at UCL, University College London, United Kingdom
4 Department of Medical Statistics, London School of Hygiene & Tropical Medicine, United Kingdom
5 Health Economics & Health Technology Assessment, University of Glasgow, United Kingdom

**Introduction:** In the world of evidence-based medicine systematic reviews and meta-analyses are often considered as providing the strongest and highest quality evidence for treatment recommendations. Time-to-event outcomes are often synthesised using effect measures from Cox proportional hazards models assuming a constant hazard ratio over time. However, where treatment effects vary over time an assumption of proportional hazards is not always valid. Any bias will be further compounded if studies vary markedly in duration of follow-up. Several methods have been proposed for synthesising time-to-event outcomes in the presence of non-proportional hazards. However, guidance on choosing between these methods and the implications for downstream decision making is lacking.
**Methods:** Through application to a network of individual patient data from melanoma trials reporting overall survival we compared five methods for estimating treatment effects from time-to-event outcomes which relax the proportional hazards assumption. We compared a two-stage network meta-analysis model synthesising restricted mean survival time with four one-stage approaches: piecewise exponential, fractional polynomial, Royston-Parmar and an accelerated failure time generalised gamma model. All models were fitted in the Bayesian setting. To assess the performance of these methods under varying amounts of non-proportionality we conducted a simulation study. Individual patient data was generated from a mixture Weibull distribution assuming a treatment-time interaction. We assessed performance using bias, mean-square error and coverage.
**Results:** All models fitted the melanoma data reasonably well with some differences in the survival curves and initial variation in the treatment rankings. However, from 18 months onwards, all models consistently ranked the same treatment as the most effective. The simulation study demonstrated the potential for different conclusions from different modelling approaches.
**Conclusions:** The generalised gamma, piecewise exponential, fractional polynomial, Royston-Parmar and two-stage restricted mean survival time models can all accommodate non-proportional hazards and differing lengths of trial follow-up within an evidence synthesis of time-to-event outcomes. Further work is needed in this area to extend the simulation study to the network meta-analysis setting and provide guidance on the key considerations for informing model choice for the purposes of clinical decision making.

# SESSION OC6E

SESSION OC6E — Relative survival and net benefit estimation

## OC6E-1 Additional benefit method assessment for time-to-event endpoints – A comparison of ESMOs and IQWiGs approach

**Christopher Büsch, Johannes Krisam, Meinhard Kieser**
Institute of Medical Biometry and Informatics, University of Heidelberg, Germany

**Background:** New cancer treatments are often promoted as major advances after a significant phase III trial. Therefore, a clear and unbiased knowledge about the magnitude of the clinical benefit of newly approved treatments is important to assess the amount of reimbursement from public health insurance of new treatments. To perform these evaluations, two distinct "additional benefit assessment" methods are currently used in Europe. The European Society for Medical Oncology (ESMO) developed the Magnitude of Clinical Benefit Scale Version 1.1 classifying new treatments into 5 categories using a dual rule considering the relative and absolute benefit assessed by the lower limit of the 95% HR confidence interval or the observed absolute difference in median treatment outcomes, respectively [1]. As an alternative, the German IQWiG compares the upper limit of the 95% HR confidence interval to specific relative risk scaled thresholds classifying new treatments into 6 categories [2]. Until now, these methods have only been compared empirically and further research is required to understand the differences between the benefit assessment methods.
**Methods:** We evaluate and compare the two methods in a simulation study with focus on time-to-event outcomes. The simulation includes aspects such as different censoring rates and types, incorrect HRs assumed for sample size calculation, informative censoring, and different failure time distributions. Since no "placebo" method reflecting a true (deserved) maximal score is available, different thresholds of the simulated treatment effects are used as alternatives. The methods' performance is assessed via ROC curves, sensitivity / specificity, and the methods' percentage of achieved maximal scores.
**Results and Conclusion:** The results of the first step of a comprehensive comparison between the methods indicate that IQWiG's method is usually more conservative than ESMO's dual rule. Moreover, in some scenarios such as quick disease progression or incorrect assumed HR, IQWiG's method is too liberal compared to ESMO. Nevertheless, further research is required, e.g. investigation of the methods' performance under non-proportional hazards. In addition, the American Society of Clinical Oncology (ASCO) has developed another method for the assessment of additional benefit of new treatments using the HR point estimate, which remains to be compared to IQWiG and ESMO.
**References:** *[1] N.I. Cherny, U. Dafni et al. (2017): ESMO-Magnitude of Clinical Benefit Scale version 1.1. Annals of Oncology, 28:2340-2366 [2] G. Skipka, B. Wieseler et al. (2016): Methodological approach to determine minor, considerable, and major treatment effects in the early benefit assessment of new drugs. Biometrical Journal, 58:43-58*

# SESSION OC6E

## OC6E-2 Obtaining long-term stage-specific relative survival estimates in the presence of incomplete historical stage information

**Rachael Stannard[1], Paul C. Lambert[1,2], Mark J. Rutherford[1]**
1 Biostatistics Research Group, University of Leicester, United Kingdom
2 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden

Long-term estimates of cancer survival which are both up-to-date and stage-specific are valuable to researchers, clinicians and patients alike. However, the completeness of recording for cancer stage is historically poor in some registries, making it challenging to provide long-term stage-specific survival estimates. It is well-known that stage-specific survival differences are driven by differences in prognosis shortly after diagnosis and that the impact of cancer on survival is also greatest in the short-term. Hence, estimated survival metrics using period analysis [1] (analysing follow-up information for a recent time window to provide up-to-date survival estimates) are unlikely to be sensitive to the imputed values for historical stage data.
Our aim was to use period analysis and multiple imputation together with flexible parametric models (FPMs) on the excess hazard scale [2] to estimate up-to-date stage-specific long-term survival metrics, when historical stage information is largely incomplete. We further intended to show the lack of sensitivity to extreme assumptions about the historical stage distribution.
We used data from the Surveillance, Epidemiology, and End Results (SEER) Program for three cancer sites with varying prognoses and high stage completeness: lung, colon and melanoma. The period window for analysis was set to 2015-2017, and we further defined a pre-window 2012-2015 to evaluate various scenarios. We artificially inflated the proportion of missing stage information prior to the period window (in some scenarios conditional on stage) to mimic other registry settings. Our standard imputation model included sex, age, grade, the event indicator and Nelson-Aalen cumulative hazard estimate [2]. Four separate scenarios were imposed for individuals diagnosed before the pre-window, with some aiming to emulate the most extreme imputed stage distributions possible. In each case, a FPM was fitted on the excess hazard scale and the differences in stage-specific survival metrics were assessed. Estimates were also obtained from non-parametric approaches for validation. There was little difference between the 10-year marginal stage-specific relative survival, regardless of the assumed stage distribution for the historical stage data. We show that when conducting a period analysis, multiple imputation can be used to obtain stage-specific long-term estimates of relative survival, even when the historical stage information is largely incomplete.

**References:** *[1] Brenner H, Gefeller O. An alternative approach to monitoring cancer patient survival. Cancer. 1996;78(9):2004-2010. [2] Falcaro M, Nur U, Rachet B, Carpenter J. Estimating Excess Hazard Ratios and Net Survival When Covariate Data Are Missing. Epidemiology. 2015;26(3):421-428.*

# SESSION OC6E

## OC6E-3 A latent class model for the estimation of the excess mortality hazard for correcting inaccurate background mortality

**Robert Darlin Mba[1], Nathalie Graffféo[1], Valérie Jooste[2], Anne-Marie Bouvier[2], Roch Giorgi[3]**
1 Aix Marseille Univ, INSERM, IRD, SESSTIM, SESSTIM Marseille, France
2 Burgundy Digestive Cancer Registry, Dijon University Hospital, UMR 1231 INSERM – University of Burgundy, Dijon, France
3 Aix Marseille Univ, APHM, INSERM, IRD, SESSTIM, Hop Timone, BioSTIC, Marseille, France

**Context:** Net survival is the survival that would be observed if only deaths from the studied disease were considered. In the absence of known cause of death, net survival is estimated through excess mortality by splitting the observed mortality into excess and background mortalities. The latter is obtained from general population lifetables, usually stratified on a limited number of covariates. Specifically, lifetables do not always include some covariates that may influence the background mortality, whether they are observable or not in the data. This can lead to biased estimates of their effects on the excess mortality. To address this issue, regression models have been proposed to estimate excess mortality [1-2]. However, they only consider the case of a single missing covariate observable in the data.
**Objective:** To propose a latent class approach for modelling excess mortality by correcting inaccurate background mortality when one or more covariates are missing in the lifetable, whether they are observable or not in the data. Additionally, it allows to characterize unobserved (latent) subgroups of patients.
**Method:** The latent class model includes two modelling components: a multinomial logistic regression model for the latent class membership, and an excess hazard model with multiplicative parameters according to the profiles of the latent classes. We assessed performance of this model through simulation studies and compared them to other models [1-2]. We considered plausible scenarios from an epidemiological standpoint based on one or more missing covariates in the lifetable, these covariates being potentially unobservable in the data.
**Results:** Compared to both models developed in the presence of a single observable covariate in the data and missing from the lifetable, the proposed model showed comparable performance (biases close to 0, similar mean square errors). It remained robust in the other scenarios where more covariates are missing. Applied to population-based colon cancer registry data, the proposed model provided estimates of excess hazard ratio in each latent class.
**Conclusion:** In case of insufficiently stratified lifetables, a latent class model performed well in estimating excess mortality. Moreover, it yields an *a posteriori* classification of patients allowing a better description of their epidemiological profiles.

**References:** *[1] C Touraine, N Graffféo, R Giorgi and the CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. Stat Methods Med Res 2020, 29(1):122-36. [2] RD Mba, JA Goungounga, N Graffféo, R Giorgi and the CENSUR working survival group. Correcting inaccurate background mortality in excess hazard models through breakpoints. BMC Med Res Methodol 2020, 20(268).*

## OC6E-4 Robust statistical inference for the matched net benefit and win ratio

**Roland Matsouaka**
Duke University, Durham NC, United States

As alternatives to the time-to-first-event analysis of composite endpoints, the net benefit (NB) and the win ratio (WR) – which assess treatment effects using prioritized component outcomes based on clinical importance – have been proposed. However, statistical inference of NB and WR relies on large-sample assumptions, which can lead to an invalid test statistic and inadequate, unsatisfactory confidence intervals, especially when the sample size is small or the proportion of wins is near 0 or 1. For this talk, we will show how to address these limitations in a paired-sample design. We first introduce a new test statistic under the null hypothesis of no treatment difference. Then, we present new ways to estimate the confidence intervals of NB and WR. The confidence interval estimations use the method of variance estimates recovery (MOVER). The MOVER combines two separate individual-proportion confidence intervals into a hybrid interval for each estimand of interest. We assess the performance of the proposed test statistic and MOVER confidence interval estimations through simulation studies. The results show that the MOVER confidence intervals are as good as the large-sample confidence intervals when the sample is large and the proportions of wins is bounded away from 0 and 1. Moreover, the MOVER intervals outperform their competitors when the sample is small or the proportion of wins is near the boundaries 0 and 1. We illustrate the method (and its competitors) using three examples from randomized clinical studies.

# SESSION OC6E

## OC6E-5 A unifying framework for flexible excess hazard modeling with applications in cancer epidemiology

**Alessia Eletti[1], Giampiero Marra[1], Manuela Quaresma[2], Rosalba Radice[3], Francisco Javier Rubio[4]**
1 Department of Statistical Science, University College London, United Kingdom
2 Department of Non-Communicable Diseases Epidemiology, London School of Hygiene & Tropical Medicine, UK
3 Faculty of Actuarial Science and Insurance, Business School, City, University of London, United Kingdom
4 Department of Mathematics, King's College London, United Kingdom

Excess hazard modeling has become the preferred tool in population-based cancer survival research as it overcomes drawbacks of both the traditional overall survival setting as well as of the so-called cause-specific setting. This is achieved by assuming an additive decomposition of the overall hazard into two components: the excess hazard, e.g. due to the cancer of interest, and the population hazard due to all other causes of death.
Despite its widespread use, a unifying framework supporting virtually any application fully implemented in a ready to use package is not, to the best of our knowledge, available in the literature. We thus propose a solution which allows for excess hazard modeling in addition to traditional overall survival analysis. The general nature of our approach extends to two further levels. The first is that the link-based additive model proposed allows for many types of covariate effects. In particular, any type of smoother and modeling spatial effects. Estimation is achieved using a carefully structured efficient and stable penalized likelihood algorithm.
The proposed model is then extensively evaluated through a battery of simulation studies in which it is also compared with the analogous state-of-the-art framework (Fauvernier et al., 2019, JRSSC, 68(5),1233–1257), implemented in the R package survPen. Two practical applications in breast, colon and lung cancer epidemiology are presented and highlight the benefits of using this flexible framework. The model which includes non-linear time-dependent effects and spatial effects is, in fact, found to be superior to that with only linear effects. Additional insight on socio-demographic inequality in cancer survival is also found by including spatial effects. Some theoretical properties are discussed. The proposed approach is readily available in the R package GJRM.

# SESSION OC6F

SESSION OC6F — Deep learning

## OC6F-1 Cluster analysis on emergency COVID-19 data: A result-based multiple imputation for missing data

**Halehsadat Nekoee Zahraei[1,2], Allison Gilbert[3], Anh Nguyet Diep[1], Renaud Louis[2], Alexandre Ghuysen[3,4], Anne-Françoise Donneau[1]**
1 Biostatistics Unit, Department of Public Health, University of Liège, Belgium
2 Department of Pneumology, GIGA, University of Liège, Belgium
3 Emergency Department, University Hospital Center of Liège, Liège, Belgium
4 Medical Simulation Center, Public Health Department, University of Liège, Liège, Belgium

**Background** and Objective: In 2020, hospitals have been confronted with an influx of COVID-19 confirmed patients. Grouping patients based on clinical features could help clinicians to identify a structure of patients who needs more attention. This study considers cluster analysis to identify different clinical phenotypes with similar properties while accounting for the presence of missing data. Several frameworks exist for handling missing data in cluster analysis, in this study, a new perspective was introduced for multiple imputation in cluster analysis that focused on the result of clustering.
**Method:** To handle the uncertainty of missing values, *m* imputed datasets were generated. The model-based clustering strategy was applied on the imputed datasets. Based on BIC criterion, the best method and the best number of groups were defined for all imputed datasets. The most repetitive number of groups and types was fixed. In the next step, cluster analysis was re-applied on *m* imputed datasets by the fixed number of clusters and type. The results of the statistical analysis were reported for each of the groups in imputed datasets. According to Rubin's rules, in the pooled step, the final results were combined by mean and the statistical inferences were applied by considering between and within variance.
**Results:** The performance of the proposed framework was compared in several scenarios. The proposed method with 20 clinical features was performed on 628 confirmed COVID-19 patients who presented at University Hospital of Liege from March to May 2020. Based on model-based clustering and BIC criterion for multiple imputation, the patients were classified into four clusters. The rate of hospitalization in Cluster2 with older patients was higher than those in Cluster1. The oldest patients were assigned to Cluster3 and Cluster 4. The rate of comorbidity was almost close to 100% in Cluster 4 and percentage of infectious disease in cluster3 was less than Cluster4; however, Cluster3 had a higher rate of hospitalization than Cluster4.
**Conclusions:** The proposed method handled cluster analysis on missing data by multiple imputations. Also, the present study identified four clusters of patients confirmed with COVID-19 and the corresponding rate of hospitalization based on clinical features.

# SESSION OC6F

## OC6F-2 Statistical Power for Single Cell Representations

**Martin Treppner[1,2], Harald Binder[1,2]**
1 Institute of Medical Biometry & Statistics, Faculty of Medicine & Medical Center - University of Freiburg, Germany
2 Freiburg Center for Data Analysis and Modelling, University of Freiburg, Germany

One of the most common applications of single-cell RNA-sequencing (scRNA-seq) experiments is to discover groups of cells with a similar expression profile in an attempt to define cell identities. The similarity of these expression profiles is typically examined in a low-dimensional latent space, which can be learned by deep generative models such as variational autoencoders (VAEs). However, the quality of representations in VAEs varies greatly depending on the number of cells under study, which is also reflected in the assignment to specific cell identities. We propose a strategy to answer what number of cells is needed so that a pre-specified percentage of the cells in the latent space is well represented.
We train scVI, a VAE that has been adapted to scRNA-seq data, on a varying number of cells and evaluate the quality of the learned representation by use of the estimated log-likelihood lower bound of each cell. The distribution arising from the values of the log-likelihoods are then compared to a permutation-based distribution of log-likelihoods. We generate the permutation-based distribution by randomly drawing a small subset of cells before training the VAE and permuting the expression values of each gene among these randomly drawn cells. By doing so, we ensure that the overall structure of the latent representation is preserved, and at the same time, we obtain a null distribution for the log-likelihoods. We then compare log-likelihood distributions for different numbers of cells. We also harness the properties of VAEs by artificially increasing the number of samples in small datasets by generating synthetic data and combining them with the original pilot datasets.
We illustrate the performance of our approach with two application examples. First, we show an application for sample size planning on a subset of the Tabula Muris dataset. Second, we show how to estimate statistical power for latent representations when studying cell subtypes using the PBMC dataset from 10x Genomics. We show that such analyses can potentially improve the reliability of downstream analyses such as cell identity detection.

## OC6F-3 Effects of Interactions and Dataset Size in Neural Networks and Other Machine Learning Approaches

**Alexandre Bailly[1,2], Corentin Blanc[1,2], Elie Francis[1], Thierry Guillotin[1], Fadi Jamal[3], Béchara Wakim[4], Pascal Roy[2]**
1 Research and Development Lab, Everteam Software, France
2 Service de Biostatistique-Bioinformatique, Hospices Civils de Lyon, Lyon, France; Université de Lyon, Lyon, France; Université Lyon 1, Villeurbanne, France; Équipe Biostatistique-Santé, Laboratoire de Biométrie et Biologie Évolutive, CNRS UMR 5558, Villeurbanne, France
3 izyCardio, France
4 Mediapps innovation SA, France

Machine Learning and Deep Learning are powerful models able to predict the presence of a disease in a patient. A good predictive performance of such a model depends on the size of the training dataset and the complexity of the relationships between the presence of the disease and the variables included in the model. Usually, this complexity is unknown. Several datasets were simulated to allow comparing the effects of the two factors in linear prediction models. Datasets were constructed to represent a situation similar to that of the Framingham study in which the purpose of the model was to predict the presence of coronary disease. Three dataset sizes and three complexity levels were considered. A few logistic regressions (penalized and non-penalized) and neural networks (with different numbers of hidden layers) were trained on the simulated datasets. All models were evaluated by cross-validation with accuracy, F1-score, and area under the ROC curve as criteria. Logistic regressions were trained with terms of various interaction orders. Comparisons showed that logistic regressions were less influenced by the size of the dataset but needed the right interaction terms to achieve good performance. On the contrary, neural networks were more sensitive to the dataset size but did not need interaction terms to make good predictions. In the most complex scenarios, neural networks performed better than logistic regressions without interaction terms. However, logistic regressions with the right interaction terms provided better results.



SESSION OC6F

## OC6F-4 CamemBERT word-embedding for Information Extraction in a Biomedical Context

**Corentin Blanc[1], Alexandre Bailly[1], Elie Francis[1], Thierry Guillotin[1], Fadi Jamal[2], Béchara Wakim[3], Pascal Roy[4]**
1 Research and Development Lab, Everteam Software, France
2 izyCardio, Lyon, France
3 Mediapps Innovation SA, Lyon, France
4 Service de Biostatistique-Bioinformatique, Hospices Civils de Lyon, Lyon, France; Université de Lyon, Lyon, France; Université Lyon 1, Villeurbanne, France; Équipe Biostatistique-Santé, Laboratoire de Biométrie et Biologie Évolutive, CNRS UMR 5558, Villeurbanne, France

Deep Learning has totally changed the landscape of Natural Language Processing (NLP) over the last few years. Very large word-embedding models pre-trained on colossal amounts of data (such as BERT) are now used to perform NLP tasks. However, these models are rarely trained on French biomedical data because these data are difficult to obtain due to the European General Data Protection Regulation. Within this context, we wondered whether these word-embedding models are able to perform French NLP tasks in the biomedical field in which they are not yet trained. The French language model CamemBERT was used in two different approaches (feature-based and fine-tuning) and three recurrent networks (Recurrent Neural Network, Long Shot-Term Memory, and Bidirectional Long Short-Term Memory) in a biomedical Natural Language Understanding task. This task aims to extract directly socio-demographic and biomedical information from patient speech. For each approach-network combination, 50 bootstraps were performed to compare the average performance in terms of precision, recall, and F1 score. The study showed that extracting socio-demographic and biomedical information was much easier with CamemBERT than without it whatever the approach and the network. The fine-tuning approach had a slight advantage over the feature-based approach. Bidirectional Long Short-Term Memory performed better than Long Short-Term Memory and Recurrent Neural Network. Interestingly, CamemBERT performed satisfactorily without previous training in the biomedical field.

## OC6F-5 Interpretable effect estimates of semi-structured predictors in deep distributional regression models

**Lucas Kook[1,2], Lisa Herzog[1,2,3], Oliver Duerr[4], Torsten Hothorn[1], Susanne Wegener[3], Beate Sick[1,2]**
1 Epidemiology, Biostatistics & Prevention Institute, University of Zurich, Switzerland
2 Institute for Data Analysis and Process Design, Zurich University of Applied Sciences, Switzerland
3 Department of Neurology, University Hospital Zurich, Switzerland
4 Institute for Optical Systems, Konstanz University of Applied Sciences, Germany

Interpretable effect estimates and reliable predictions in applications that rely on unstructured data, such as images or electronic health records, require flexible modeling approaches. Deep-learning based models have proven exceptional in prediction tasks and seamlessly incorporate unstructured data, while lacking interpretability. Classical regression models provide readily interpretable effects, but are constrained to tabular data, which are typically obtained via feature extraction in the case of image data. In this work, we propose a flexible class of deep distributional regression models which trade off flexibility and interpretability of model parameters by additively decomposing the contribution of prognostic and predictive features on a predefined interpretational scale, such as the odds-scale. The backbone of these models are parametric transformation models, which estimate the entire conditional distribution of the outcome given the features by decomposing this conditional distribution into an a priori chosen, parameter-free target distribution and a parametric transformation function. The transformation function is parameterized via several neural networks, which are jointly trained by minimizing the negative log-likelihood induced by the transformation model. The benefits of deep transformation models in prediction tasks have been demonstrated before. However, interpretable effect estimates require concrete model architectures, which have received much less attention. For instance, co-linear features between tabular predictors and the image have to be circumvented by orthogonalizing image effects from tabular features. We discuss how to interpret the resulting image effects and regression coefficients in practice. We apply deep distributional regression to prediction of functional outcome after acute ischemic stroke, measured on an ordinal scale, and based on MRI data and routinely collected clinical parameters at baseline, such as age, blood pressure and time-to-first-image. Although we use observational data in this example, data from RCTs, for instance, allow estimation of predictive treatment effects using the very same models. The proposed models unite the benefits of deep learning and regression models by being able to capture non-linear effects of tabular features and the image as a whole, while still possessing interpretable model components.

# SESSION OC7A

## SESSION OC7A — Causal inference in survival analysis

### OC7A-1 Marginal structural Cox models in nested case-control studies with time-varying treatments

**Yoshinori Takeuchi, Yasuhiro Hagiwara, Yutaka Matsuyama**
Department of Biostatistics, School of Public Health, Graduate School of Medicine, The University of Tokyo, Japan

Marginal structural Cox proportional hazards models (Cox-MSMs) are frequently employed in survival analyses in both clinical and epidemiological studies to estimate the causal effects of time-varying treatments [1]. However, the fitting procedures of Cox-MSMs to nested case-control (NCC) sampling data have not been straightforward, which may be due to the difficulties in calculating the inverse probability weights.
In this study, we propose a method to fit Cox-MSMs using inverse probability weighting under NCC sampling to obtain consistent estimators of the causal effects of time-varying treatments.
Here, we consider an observational (full) cohort where subjects had received a particular treatment of interest or control treatment at each follow-up start date, but treatment changes can occur during follow-up. After NCC sampling, we additionally sample subjects whose treatment statuses had changed at any time point and subjects whose observations were censored before the administrative censoring time to calculate the inverse probability of treatment and censoring weights. We obtain the estimates of causal treatment effects by maximizing the weighted Cox partial likelihood. In our method, we must assemble both treatment and covariate histories only for NCC samples (i.e., cases and controls) and additional samples. The confidence interval of effect estimates is then calculated based on the robust variance estimator for the NCC sampling data. Our method can employ additional matching by baseline covariates and counter matching by treatment history [2].
A simulation study was conducted to evaluate the finite sample performance of the proposed method in the presence of treatment-confounder feedback compared with an ordinary analysis of an NCC study that employs Cox models using the information just before outcome occurrence (ordinary NCC analysis) and a Cox-MSM fitted to a full cohort (MSM-full). The results showed that our proposed method and MSM-full provided negligibly biased estimates of the causal effect of treatment while the ordinary NCC analysis produced biased estimates. When compared with simple random sampling, both additional matching and counter matching improved the efficiency of the proposed method.
We also applied the proposed method to a pharmacoepidemiological study that examined the effect of antihyperlipidemics on hyperglycemia incidence using a Japanese medical claims database.

**References:** *[1] Hernán MA, Brumback B, Robins JM. Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments. J Am Stat Assoc. 2001;96(454):440-8.*
*[2] Borgan O. Nested Case-Control Studies: A Counting Process Approach. In: Borgan O, Breslow NE, Chatterjee N, Gail MH, Scott A, Wild CJ, eds. Handbook of Statistical Methods for Case-Control Studies. Boca Raton, FL: CRC Press; 2018 329-349.*



SESSION OC7A

### OC7A-2 Benefit-based organ allocation in liver transplantation

**Ilaria Prosepe[1], Ben F.J. Goudsmit[2], Dries E. Braat[2], Nan van Geloven[1], Hein Putter[1]**
1 Department of Biomedical Data Sciences, LUMC, Netherlands
2 Department of Surgery, LUMC, Netherlands

The field of liver transplantation (LT) is dealing with an ever-increasing shortage of donor organs. Since donor livers are scarce, patients are placed on a waiting list (WL). Unfortunately, many cannot be treated on time. To deal with this scarcity, WL patients are prioritized by medical urgency. Although this strategy prevents WL deaths, it also denies livers to (less sick) patients that could have a larger gain in life-years. Thus, donated organs are not distributed optimally. Benefit-based organ allocation is proposed to resolve this ineffective use of deceased-donor livers. Benefit-based allocation not only considers the WL mortality, but it also accounts for the post-operative outcome, which is a critical addition to the current system [1].
More formally, benefit is here estimated as the average difference in restricted mean survival time attributable to LT under a certain allocation strategy [2]. Hence, to estimate the benefit, the following two quantities are contrasted: the post-transplant survival and the survival in a "counterfactual" world where transplantation never happens. Methods from causal inference, in particular inverse probability of censoring weighting, are used to deal with the dependent censoring caused by transplantation. The post-transplant survival model is estimated given the patients' history until transplantation and the organ condition. The pre-treatment model is estimated via landmark analysis methods that are able to combine the two time axes of this problem: the time since first eligibility for transplantation and the calendar time. The first is relevant due to the time-dependent nature of MELD-score (i.e. one of the strongest predictors), while the latter is important because donor organs become available at specific calendar dates. The effects of the different allocation schemes are quantified by a Monte Carlo simulation. By simulating the influx of both patients on the waiting list and incoming donor livers, it is possible to compare the number of life-years saved for the current allocation scheme and for a benefit-based allocation scheme.

**References:** *[1] Schaubel DE, Guidinger MK, Biggins SW, et al. Survival benefit-based deceased-donor liver allocation. Am J Transplant. 2009;9(4 PART 2):970-981. doi:10.1111/j.1600-6143.2009.02571.x [2] Gong Q, Schaubel DE. Estimating the average treatment effect on survival based on observational data and using partly conditional modeling. Biometrics. 2017 Mar;73(1):134-144. doi: 10.1111/biom.12542.*

### OC7A-3 Extending multistate models with g-computation to evaluate the effect of treatment timing

**Nan van Geloven[1], Matea Skypala[2], Tina Nane[2], Saskia le Cessie[1,3]**
1 Department of Biomedical Data Sciences, Leiden University Medical Center, Netherlands
2 Department of Applied Mathematics, Delft University of Technology, Netherlands
3 Department of Clinical Epidemiology, Leiden University Medical Center, Netherlands

The decision when to intervene rather than allow nature to take its course is a critically important question in medicine. To avoid unnecessary treatment, a 'wait-and-see' approach is commonly advised where a natural recovery of the patient is awaited before starting treatment. There is however a huge lack of information on the impact of the exact duration of such a delay period on the outcome of patients. In the absence of trials, large observational studies are the best source of data to study the effect of timing of treatment. In this work we develop causal methods that -under certain assumptions- allow evaluating treatment timing using such observational data sources. We combine multistate modelling with g-computation to target the counterfactual cumulative proportion of recovered patients for different delay periods. The multistate model describes the speed at which patients transition between the disease, treatment and recovery states. It uses Cox proportional hazards models for each of the transitions, employing the clock-reset time scale. In contrast to common use of multistate models that describes *actual* observed speeds of transitions, we here use the model to evaluate *counterfactual* outcomes. In particular, we evaluate the effect of different delay periods by predicting the cumulative percentage of recovered patients had -counter to the fact- all unrecovered patients transitioned to treatment at the same wait time. This is done by using g-computation methodology on the transition model from treatment to recovery where the wait time at which patients entered the treatment state is one of the covariates. Uncertainty is quantified by bootstrapping. We apply the developed methodology to study the effect of delaying the start of intrauterine insemination treatment in a cohort of 1896 couples with unexplained subfertility. We estimate the expected cumulative proportion of pregnant woman 1.5 years after diagnosis for treatment strategies where couples start treatment if not yet pregnant after 0, 3, 6 or 9 months. Our method allows contrasting the expected number of additional pregnancies with the expected number of extra treatments, allowing a well-informed evaluation of different delay strategies.

# SESSION OC7A

## OC7A-4 Causal assessment of surrogacy for time-to-event endpoints using meta-analytic data

**Quentin Le Coënt[1], Catherine Legrand[2], Virginie Rondeau[1]**

1 Department of Biostatistics, Bordeaux Population Health Research Center (INSERM U1219), Université de Bordeaux, France

2 ISBA/LIDAM, Université catholique de Louvain, Louvain-la-Neuve, Belgium

With the ongoing development of treatment procedure and the increase of survival in oncology, clinical trials based on endpoint such as overall survival can require long follow-up times in order to observe enough events to ensure sufficient statistical power. This increase of the follow-up times can compromise the feasibility of the trial. The use of surrogate endpoints can thus be attractive for these trials. However, in order to yield valid conclusion of the treatment efficacy regarding the true endpoint based on the sole observation of the surrogate the latter must have been previously statistically validated as being a good surrogate. In this work we propose an alternative approach for surrogate validation when both the surrogate and the true endpoint are time-to-event. This approach is based on the causal framework of mediation analysis and is developed for meta-analytic data. It uses a joint regression model for the hazard functions of both endpoints. The mediation analysis enables one to study the decomposition of the total effect of the treatment on the true endpoint into a direct effect and an indirect effect through the surrogate. This decomposition is defined using a mediation formula for counterfactual outcomes. The surrogate can then be deemed validated if the ratio of indirect effect over total effect is sufficiently high. The meta-analytic nature of the data is taken into account in the model by using shared random effects at the individual and trial levels. The indirect effect of the treatment on the true endpoint through the surrogate is allowed as the composition of a direct effect of the treatment on the surrogate and a direct effect of the surrogate on the true endpoint. The estimation of the parameters of this model is carried out through likelihood maximization. The estimators are then used to compute the time-dependent ratio of the indirect over total effect of the treatment on the true endpoint. We designed a simulation study to evaluate the estimators. We applied this method for the assessment of the desease-free survival as a surrogate of the overall survival for adjuvant chemoterapy in the context of resectable gastric cancers.

## OC7A-5 Estimating treatment effects on survival in an entirely treated cohort: negative controls in longitudinal data

**Ruth Keogh**

Department of Medical Statistics, London School of Hygiene & Tropical Medicine, United Kingdom

Treatments are sometimes introduced for all patients in a particular cohort. When an entire cohort of patients receives a treatment it is not straightforward to estimate its effect (the treatment effect in the treated) because there are no directly comparable untreated patients. Attempts can be made to find a suitable control group, (e.g. historical controls), but underlying differences between the treated and untreated can result in bias. The application that motivates this relates to a disease-modifying treatment in cystic fibrosis, ivacaftor, which has been available for everyone in the UK with a specific CF-causing genetic mutation since 2012. Its impact on several health outcomes has been demonstrated in randomized controlled trials, but it is of interest to understand the causal effect of ivacaftor on survival, which has not been possible to investigate in trials due to short term follow-up.

Observational data provide the opportunity to estimate causal effects of treatments. A strong assumption made in most such investigations is that of positivity, meaning individuals have a probability strictly less than one of receiving the treatment. When the entirely cohort of interest is treated the positivity assumption is violated, meaning that 'standard' causal inference methods cannot be applied. We show how negative control outcomes combined with an extension of difference-in-differences analysis to survival outcomes can be used to assess bias in treatment effect estimates and obtain unbiased estimates under certain assumptions, making use of longitudinal data on treatment use and covariates. Causal diagrams and the potential outcomes framework are used to explain the methods and assumptions. We will discuss the use of different underlying analysis models, including Cox regression and Aalen's additive hazards model.

The methods are applied to the motivating example using longitudinal data from the UK Cystic Fibrosis Registry. Negative control outcomes observed in patients who do not receive the treatment are used to enable estimation of the causal effect of ivacaftor on survival, including patients eligible for the new treatment but before its availability and patients with an ineligible genotype. The methods described provide a framework for robustly estimating the effects of new disease-modifying treatments on survival.



# SESSION OC7B

SESSION OC7B — Clinical trial design and sample size calculation

## OC7B-1 Analysis and sample size calculation for a survival model conditional on a binary surrogate endpoint

**Samuel Kilian, Johannes Krisam, Meinhard Kieser**

Institute of Medical Biometry and Informatics, University of Heidelberg, Germany

The primary endpoint in oncology is usually overall survival, where differences between therapies may only be observable after many years. To avoid withholding of a promising therapy, preliminary approval based on a surrogate endpoint is possible [1]. The approval can be confirmed later by assessing overall survival within the same study. Then, the correlation between surrogate endpoint and overall survival has to be taken into account for sample size calculation and analysis. This relation can be modeled by means of a conditional survival model which was proposed by Xia et al. [2]. They investigated the correlation and assessed power of the logrank test but did not develop methods for statistical testing, parameter estimation, and sample size calculation.

In this talk, a new statistical testing procedure based on the conditional model and Maximum Likelihood estimators for its parameters will be presented. An asymptotic test for survival difference will be given and an approximate sample size formula will be derived. Furthermore, an exact test for survival difference and an algorithm for exact sample size determination will be provided. Type I error rate, power, and required sample size for both newly developed tests will be determined exactly. These characteristics will be compared to those of the logrank test.

It will be shown that for small sample sizes the asymptotic parametric test and the logrank test exceed the significance level. For a given sample size, the power of the asymptotic and the exact parametric test is similar, whereas the power of the logrank test is considerably lower in some situations. The sample size needed to attain a prespecified power is comparable for the asymptotic and the exact parametric test, but considerably higher for the logrank test in some situations.

We conclude that the presented exact test performs well under the assumptions of the conditional model and is a better choice than the asymptotic parametric test or the logrank test, respectively. Furthermore, the talk will give some insights in performing exact calculations for parametric survival time models, thus facilitating the planning, conduct, and analysis of oncology trials with the option of accelerated approval.

**References:** *[1] Joshua D Wallach, Joseph S Ross & Huseyin Naci (2018). The US Food and Drug Administration's expedited approval programs: Evidentiary standards, regulatory trade-offs, and potential improvements, Clinical Trials, Vol. 15(3) 219–229, DOI: 10.1177/1740774518770648 [2] Yi Xia, Lu Cui & Bo Yang (2014). A Note on Breast Cancer Trials With pCR-Based Accelerated Approval, Journal of Biopharmaceutical Statistics, 24:5, 1102-1114, DOI: 10.1080/10543406.2014.931410*

# SESSION OC7B

## OC7B-2 Dangers of wrongly assuming linearity in a trial sample size calculation when treatment affects rate of change

**Katy E. Morgan[1], Ian R. White[2], Chris Frost[1]**
1 Medical Statistics Department, LSHTM, United Kingdom
2 MRC clinical trials unit, UCL, United Kingdom

For certain conditions, such as multiple sclerosis, treatments might aim to lessen deterioration over time. A trial outcome in this setting could be a continuous measure (e.g. Expanded Disability Severity Scale, EDSS) recorded on multiple occasions. Such outcomes could be analysed using a mixed-effects model with random slopes and intercepts for participants, with the treatment effect being the difference between the slopes over time in the treated and placebo groups. Thus, an effective treatment will slow the mean rate of deterioration compared to participants receiving control. A sample size can be obtained by estimating the necessary variances and covariances from a pre-existing dataset, e.g. an observational study conducted in a similar setting. These estimates can be used in standard sample-size formulae for mixed-effects models.
A random slopes model assumes trajectories are linear, and therefore any treatment effect increases linearly over time, but this may not be the case. Our simulation study assesses what effect a non-linear trajectory and a treatment effect not proportional to time have on the proposed trial's nominal power. Simulations were based on EDSS data from the MS-STAT trial[1]. We used four trajectories – steady decline (i.e. linear trajectory), early and late decline (based on exponential curves), and intermediate decline (inverse logit) – and simulated 5000 observational studies of 1000 people each, to calculate trial sample sizes. Trials of the appropriate size were then simulated, with each of the following treatment effects: none (to assess Type I error), proportional to time, non-proportional. Preliminary results suggest that, provided the treatment effect is proportional to time, powers are generally close to nominal regardless of whether trajectories are linear or not. However, when the treatment effect is not proportional to time, its estimate from a random slopes model can be badly biased, leading to powers far from nominal levels. These errors can be exacerbated when the length of the proposed trial differs from that of the observational study used to plan the sample size. In such cases, it might be more appropriate to use a model that allows trajectories to vary freely over time instead of the random slopes model.

**References:** *[1] J Chataway et al., "Effect of high-dose simvastatin on brain atrophy and disability in secondary progressive multiple sclerosis (MS-STAT): a randomised, placebo-controlled, phase 2 trial", Lancet 2014; 383: 2213–21, http://dx.doi.org/10.1016/ S0140-6736(13)62242-4*

## OC7B-3 Estimating sample size for biomarker-strategy designs with survival endpoints

**Derek Dinart[1,2], Virginie Rondeau[1,3], Carine Bellera[1,2]**
1 Bordeaux Population Health Center, INSERM U1219, Bordeaux, France
2 Clinical Research and Clinical Epidemiology Unit, Institut Bergonie, Comprehensive Cancer Center, Bordeaux, France
3 Biostatistic team, University of Bordeaux, France

In response to the rapid growth of precision medicine and the number of molecules entering the drug development pipeline, several study designs including the biomarker-strategy design (BSD) have been proposed. Contrary to traditional designs, the emphasis here is on the comparison of treatment strategies and not on the treatment molecules as such. Patients are assigned to either a biomarker-based strategy (BBS) arm where biomarker-positive patients receive an experimental treatment or a non-biomarker-based strategy (NBBS) arm where patients receive a treatment regardless of their biomarker status. We examined several designs of BSD according to the biomarker assessment and the treatment received in NBBS arm and used frailty survival models to analyse them. Depending on the limits and specificity of each, we proposed statistical models that best described each design. We thus developed a partially clustered frailty model (PCFM) for the (standard) case where the biomarker status is only known in BBS arm. The PCFM allows us to account for the complex structure of BSD that may consider clustering only in one arm. In addition, we proposed an approach to calculate sample size for survival data relying on PCFM. We also proposed statistical tests to measure the overall strategy effect as well as the biomarker-by-strategy interaction effect. We conducted extensive simulations to assess, for each design, the robustness and performances of the different statistical models. We also performed power analysis and sample size estimation to compare the performance of PCFM to more traditional frailty models, and provided guidelines on the use of BSD in survival analysis. We also performed power analyses to compare the performance of PCFM to more traditional frailty models.



# SESSION OC7B

## OC7B-4 Multidimensional Go/No-Go decision rules

**Cornelia Ursula Kunz, Frank Fleischer**
Biostatistics + Data Sciences Corp., Boehringer Ingelheim Pharma GmbH & Co. KG, Germany

Within clinical drug development, early phase trials are an integral component of the development plan. They constitute the generation of data that provide information on key elements such as proof of principle, proof of concept and dose selection prior to the commencement of confirmatory trials. Therefore, the focus of early phase trials should be on generating data that provides meaningful information regarding critical decision making in particular regarding whether to conduct the confirmatory phase. A strict control of the type I error in a confirmatory sense is usually not needed for this purpose. Instead go/no-go decision criteria are needed that ensure certain probabilities for correct decision making. A statistical go/no-go is a pre-defined, quantifiable criterion to allow for decision making given the observed results in the trial. In order to improve the probability for a correct conclusion, it is often desirable to base the decision on more than one endpoint.
Two possible methods for deriving a decision criterion when having multiple endpoints are the intersection and the union of the individual go/no-go areas for the separate endpoints. For the intersection area, a go is achieved by all endpoints passing their individual go-criterion while for the union area, it is sufficient if at least one endpoint has crossed the individual go-criterion. While both methods are easy to implement, they have also some disadvantages which may lead to non-intuitive decision making. We investigate different ways of defining the go-, no-go- and consider-region based on more than one endpoint. The different approaches are compared to each other regarding their properties. Results are illustrated using a real data example.

## OC7B-5 Online control of the False Discovery Rate in platform trials

**Sonja Zehetmayer, Franz König, Martin Posch**
Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria

When testing multiple hypotheses, the control of a suitable error rate is desirable even in exploratory trials. For such applications, e.g., the control of the False Discovery Rate (FDR) has been proposed. The FDR is defined as the expected proportion of false positive rejections among all rejections. Conventional methods to control the FDR, e.g., the Benjamini-Hochberg procedure, assume that all p-values are available at the time point of test decision. In perpetual platform trials however, treatment arms can enter and leave the trial at any time during its conduct. Therefore, the number of hypotheses is not fixed in advance and the hypotheses are not tested at once, but sequentially. Recently, the concept of online control of the FDR was introduced [1], where hypothesis tests and test decisions can be performed in a sequential manner and there application to platform trials has been proposed [2].
We investigate different procedures proposed by, e.g., Javanmard and Montanari, 2018, to control the online FDR in the setting of platform trials. This includes methods distributing the alpha level unequally among the hypotheses of interest and increasing the local significance level in case of recent discoveries. The results depend sensitively on prior distributions of effects sizes, e.g., whether true alternatives are uniformly distributed over time or not. We optimize design parameters depending on prior distributions to maximize operating characteristics such as the overall power, which is the proportion of rejected alternatives among all alternatives. Furthermore, we investigate the impact on error rates by including both concurrent and non-concurrent control data. By including the latter the power can be increased, but the control of the FDR may be negatively affected in case of time trends. Finally, we show how the procedures can be extended to allow for interim analyses with the option of early stopping for individual hypotheses.

**References:** *[1] Javanmard, A, and Montanari, A (2018). Online Rules for Control of False Discovery Rate and False Discovery Exceedance. Annals of Statistics, 46(2): 526-554. [2] Robertson, DS, and Wason, JMS (2018) Online control of the false discovery rate in biomedical research. arXiv preprint (https://arxiv.org/abs/1809.07292)*

# SESSION OC7C

## SESSION OC7C — Frailty model and recurrent events

### OC7C-1 A non-mixture cure model with frailty correction for inaccurate background mortality to estimate time-to-cure

**Juste Goungounga[1,2], Olayide Boussari[2,3], Valerie Jooste[1,2,4,5]**
1 Registre Bourguignon des cancers digestifs, Université de Bourgogne, France
2 INSERM, UMR1231, France
3 Département de méthodologie, Fédération Francophone de Cancérologie Digestive, Dijon, France
4 Centre Hospitalier Universitaire de Dijon, Bourgogne, France
5 Réseau Français des Registres des Cancers (FRANCIM), Faculté de médecine, Toulouse, France

**Context:** Cure models are used in population-based studies to estimate net survival and its asymptotic value, the cure fraction. Net survival, the survival that would be observed if the studied disease (e.g. cancer) were the only possible cause of death, is estimated by splitting the observed mortality into two forces: one due to cancer (excess mortality) and one due to other causes (expected mortality). Usually, the latter is drawn from the general population life tables but this assumption may not hold.
**Objectives:** To propose a non-mixture cure model accounting for inaccuracy of general population life tables as other causes mortality tables.
**Methods:** Boussari et al. previously proposed a non-mixture cure model which allows direct estimation of the time-to-cure as the time from which the excess mortality becomes null. In the present work, we allow the cancer patients' expected mortality to be different from the general population observed mortality. To account for this non-comparability bias, we proposed two models assuming that the expected mortality equals the population mortality multiplied by either a constant parameter or a frailty term, with estimations based on maximum likelihood method.We assessed the performance of the three models in a simulation study designed to mimic real data. The simulation scenarii were built by varying the cure fraction (low, medium and high), the time-to-cure (early, late), the sample size, the censoring process as well as the magnitude of non-comparability effect (null, moderate, severe) and its variability (null, narrow, wide). We applied these three models to colon cancer data from French cancer registries (FRANCIM).
**Results:** In the simulation study, the three models performed equally when the comparability assumption held. In presence of heterogeneous non-comparability effect, the frailty rescaled cure model outperformed the constant rescaled cure model while meaningful biases were observed with Boussari model. In the application, the frailty cure model provided the lowest AIC and BIC and the non-comparability was meaningful in the two new cure models, for the colon cancer data.
**Conclusions:** We recommend the proposed models for the estimation of net survival, cure proportion and time-to-cure. An R package implementing these models will be available soon.
**References:** *[1] Boussari O, Bordes L, Romain G, Colonna M, Bossard N, Remontet L, Jooste V. Modeling excess hazard with time-to-cure as a parameter. Biometrics. 2020. doi: 10.1111/biom.13361*
*[2] Rubio FJ, Rachet B, Giorgi R, Maringe C, Belot A. On models for the estimation of the excess mortality hazard in case of insufficiently stratified life tables. Biostatistics. 2019 May 28;22(1):51–67.*



SESSION OC7C

### OC7C-2 A family of discrete random effect distributions for modelling bivariate time-to-event data

**Maximilian Bardo[1], Niel Hens[2,3], Steffen Unkel[1]**
1 Department of Medical Statistics, University Medical Center Goettingen, Germany
2 I-BioStat, Data Science Institute, Hasselt University, Diepenbeek, Belgium
3 Centre for Health Economics Research and Modelling Infectious Diseases, Vaccine & Infectious Disease Institute, University of Antwerp, Antwerp, Belgium

Random effect models for time-to-event-data, also known as frailty models, provide a conceptually simple and appealing way of modelling individual heterogeneities resulting from factors which may be difficult or impossible to measure; an example is heterogeneity induced by genetics or through environmental exposure.
The frailty is usually assumed to have a continuous distribution. In some areas of application,
however, the frailty distribution might have considerable discrete impacts, such as the (unobserved) number of sexual partners for sexually transmitted diseases. In others, the true frailty distribution might be multimodal and the most crucial differences in time-to-event data might be explained by the clusters' distance to the different high-density areas of the frailty distribution. In such scenarios, a discrete frailty model might be more appropriate for capturing important differences in individual heterogeneities.
Hence, we model a family of discrete frailty distributions, originally introduced by Farrington et al. (2012), and provide implementations and applications for both bivariate current-status and right-censored, possibly left-truncated data. We suggest an interpretation of the discrete frailties as being ordered latent risk categories. As we often encountered, only a few risk categories representing most of the probability mass, this facilitates the interpretation of the proposed model. Furthermore, our estimation algorithm is designed such that the frailty distribution and the distribution of the conditional survival time is chosen by the data among a set of distributions. We analyze clustered time-to-event data of diabetic patients at 'high-risk' of retinopathy with one eye being randomly assigned to laser treatment. We assume that the frailty distribution is stratified by type I and type II diabetes. Preliminary data analysis suggests, that type II patients have higher risk of loss of vision than type I diabetic patients at each distinct risk-category at any time. This is balanced, however, by more type II diabetic patients being in the lowest risk-category.
**References:** *[1] Farrington C. P., Unkel S. and Anaya-Izquierdo K. (2012): The relative frailty variance and shared frailty models, Journal of the Royal Statistical Society Series B, Vol. 74, pp. 673-696.*

### OC7C-3 A general approach to fit flexible hazard regression models with multiple random effects

**Hadrien Charvat[1], Aurélien Belot[2]**
1 Division of International Collaborative Research, Center for Public Health Sciences, National Cancer Center, Japan (presenting author)
2 Department of Non-Communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, United Kingdom

Excess mortality hazard regression models are now widely used when the cause of death is unknown or unreliable, especially in the context of population-based cancer registry data. Given the hierarchical structure of such data (patients nested within geographical areas, hospitals, etc.), the assumption of independence between survival times, on which standard survival modelling procedures rely, might not be adequate. We recently proposed an R package, named mexhaz, to fit flexible excess mortality hazard models accounting for the unobserved between-cluster heterogeneity by including a normally distributed random intercept defined at the cluster level. However, it might sometimes be necessary to also allow for between-cluster variations in the effect of one or several covariates.The objective of this work is thus to present an extension of our model that allows the user to fit flexible (excess) hazard regression models with multiple Gaussian random effects. The logarithm of the baseline hazard is modelled by a B-spline of time, and non-linear and time-dependent effects of covariables can be included. Parameter estimation is based on a likelihood maximisation procedure implemented in the R software. Multivariate adaptive Gauss-Hermite quadrature is used to approximate the cluster-specific marginal likelihood. The performance of the approach is demonstrated through a simulation study assessing the impact of different scenarios (varying the number of clusters, the cluster size, the variances of the random effects and their correlation) on parameter estimation. We then illustrate the use of the appproach on population-based cancer registry data by analysing the between-registry variation of the effect of covariates such as gender and year of diagnosis on survival from cancer.

# SESSION OC7C

## OC7C-4 Handling recurrent events in the context of clinical trials

**Florence Gillaizeau, Sébastien Cambier, Marie-Cécile Fournier, Sébastien Leuillet**
Department of Statistics and BioInformatics, Biofortis Mérieux NutriSciences, France

**Introduction:** In clinical trials, the effectiveness of a product can be assessed by the decrease of harmful events for the patient. While the primary endpoint is frequently limited to the first event, the literature shows that it is more precise to consider all events, either by counting them, or by modeling recurrent events with methods extending traditional survival approaches.
**Methods:** A review of methods to describe or model recurrent events in clinical research was carried out. The recommendations of regulatory health authorities and guidelines (ICH, EFSA, Consort) have also been revised regarding the analysis of adverse events. For this work, we have retained four generalized linear models for the analysis of the number of adverse events (Poisson regression, negative binomial, zero-inflated Poisson, and zero-inflated negative binomial) and four models for the analysis of adverse events as recurrent events (Andersen-Gill, Prentice William Peterson, marginal and frailty models). These approaches were implemented with SAS software and evaluated on data from three clinical trials where adverse events corresponded to primary efficacy or safety objectives.
**Results:** The four generalized linear models studied for the analysis of the number of adverse events differed by handling overdispersion in the data. On our applications, the negative binomial distribution was the most appropriate distribution, correctly describing the heterogeneity of the data, but this could not be improved with zero-inflated models. In the literature, simulation studies showed that methods for recurrent events make the best use of all the data available and improve the precision of the analyzes compared to the methodology considering only the first event. On our applications, the Andersen-Gill and frailty models were the most relevant methods and gave similar results in terms of significance and goodness of fit.
**Conclusion:** The choice of the method for analyzing adverse events depends on the objectives of the study, the relationship between events over time and the heterogeneity between individuals. This literature review and analyses on real data have resulted in a decision tree that will be applied for the statistical analyzes of our next clinical trials with recurrent events.

# SESSION OC7C

## OC7C-5 Retro-prospective modelling of recurrent events

**Gaëlle Munsch[1], Manal Ibrahim-Kosta[2,3], Louisa Goumidi[3], Pierre-Emmanuel Morange[2,3], David-Alexandre Trégouët[1], Hélène Jacqmin-Gadda[1]**
1 Institut National pour la Santé et la Recherche Médicale (INSERM), Unité Mixte de Recherche en Santé (UMR_S) 1219, Bordeaux Population Health Research Center, University of Bordeaux, France
2 Laboratory of Haematology, La Timone Hospital, Marseille, France
3 INSERM UMR_S 1062, Nutrition Obesity and Risk of Thrombosis, Center for CardioVascular and Nutrition research (C2VN), Aix-Marseille University, Marseille, France

Risk factors analysis of recurrent events is a research topic frequently encountered in biomedical domains. When the event of interest occurs before the inclusion in the study, the information is generally not used in statistical analyses aimed at identifying risk factors for a recurrent event. Indeed, to avoid selection bias, only post-inclusion information is used, although the integration of pre-inclusion data could increase the statistical power when the explanatory variables are non-time-dependent. We here propose a weighted survival model allowing to study both prospective and retrospective events. This work is motivated by the analysis of venous thromboembolism (VTE) recurrence risk factors in the MARTHA cohort.
The MARTHA study was initially designed to investigate VTE risk factors. From 1994 to 2010, 1,473 patients were recruited at La Timone Hospital in Marseille. Among them, 402 had already had several VTE. Between 2013 and 2018, 780 patients were contacted to gather information on post-inclusion VTE, which led to the identification of 152 recurrences. Our total sample consisted of 1,473 individuals including 554 recurrences. The studied variables were gender, VTE family history and age at first VTE (before or after 50). The association between these variables and risk of VTE recurrence was modelled through a weighted Cox model. The weights were calculated from the inverse of the survival probability up to the date of recurrence data collection using a delayed-entry Cox model applied to the available death event.
In the prospective analysis of 780 subjects including 152 recurrences, male gender ($HR=1.8\pm0.18, p=1.10^{-3}$) and presence of VTE family history ($HR=1.6\pm0.17, p=8.10^{-3}$) were significantly associated with an increased risk of recurrence, but not age ($HR=1.1\pm0.18, p=0.76$). The retro-prospective analysis allowed to refine these observations. The regression coefficients obtained were consistent and the standard deviations almost divided by 2: $HR=1.7\pm0.09$ ($p=2.10^{-9}$), $HR=1.2\pm0.09$ ($p=0.04$) and $HR=1.3\pm0.10$ ($p=0.02$), respectively.
The proposed methodology enables to optimize the analysis of non-time-dependent risk factors of a recurrent event by integrating both pre- and post-inclusion data. This modelling has an immediate field of application in the context of genetic association studies on the risk of a disease's recurrence where studied DNA polymorphisms are fixed at birth.

# SESSION OC7D

## SESSION OC7D — Functional data analysis

### OC7D-1 Trajectory clustering using mixed classification models

**Amna Klich[1], René Ecochard[2], Fabien Subtil[3]**
1 Service de Biostatistique-Bioinformatique, Hospices Civils de Lyon, France
2 Université de Lyon, France
3 CNRS UMR 5558, LBBE, Équipe Biostatistique Santé, Villeurbanne, France

Trajectory classification has become frequent in clinical research to understand the heterogeneity of individual trajectories. The standard classification model for trajectories assumes no between-individual variance within groups. However, this assumption is often not appropriate, and so the error variance of the model may be overestimated, leading to a biased classification. Hence, two extensions of the standard classification model were developed through a mixed model. A first one considers an equal between-individual variance across groups, and a second one considers unequal between-individual variance. Simulations were performed to evaluate the impact of these considerations on the classification. The simulation results showed that the first extended model gives a lower misclassification percentage (with differences up to 50%) than the standard one in case of a true variance between individuals inside groups. The second model decreases the misclassification percentage compared with the first one (up to 11%) when the between-individual variance is unequal between groups. However, these two extensions require higher number of repeated measurements to be adjusted correctly.
Using human chorionic gonadotropin trajectories after a curettage for hydatidiform mole, the standard classification model mainly classified trajectories according to their level whereas the two extended models classified them according to their pattern, leading to more clinically relevant groups.
In conclusion, for studies with a non-negligible number of repeated measurements, the use, in first instance, of a classification model that considers equal between-individual variance across groups rather than a standard classification model, appears more appropriate. A model that considers unequal between-individual variance may find its place thereafter.

### OC7D-2 Bayesian concurrent functional regression for sensor data

**Beatrice Charamba, Andrew J. Simpkin**
School of Mathematics, Statistics and Applied Mathematics, National University of Ireland Galway, Ireland

**Introduction:** Functional data analysis (FDA) methods have recently been developed to analyse several variables measured repeatedly and concurrently over a domain such as time in a cohort of individuals. However, many FDA methods require data to be measured regularly, with data being collected at the same fixed times for all individuals. Often, with studies in humans, there tend to be missing data. Some studies focus on sparse but regular data and some focus on dense but irregular data. Of those who focus on both sparse and irregular data, only a few have readily available software to implement their methods and they use only complete case data for modelling, and hence some information is lost. We have developed a Bayesian model for function-on-function regression in the situation of sparse and irregular data which uses all the data for modelling and easily obtain inferences.
**Methods:** A simulation study was performed to compare the Bayesian model with other methods available in software to determine which performs best when there are sparse and irregular data. Four functions for the parameter were considered (linear, exponential, fifth order polynomial and sinusoidal). Missingness was induced in four ways, 10%, 20% and 40% missing at random as well as missing chunk of data. Three sample sizes were considered, 50, 100 and 250. Number of observations per individual considered were, 50,100, 300 and 1000. Bayesian model was then applied to concurrently measured glucose (every 5 minutes for 1 week) and electrocardiogram (ECG) data (every 10 minutes for 1 week) in a cohort of n = 17 type 1 diabetics. All models were fitted using R v 3.5.
**Results:** The Bayesian model is competitive with other models particularly in complex (fifth order polynomial and sinusoidal functions) and irregular data. Its performance drops in the linear and exponential functions. It was found the Bayesian model is robust to missingness compared to other models.
**Conclusion:** For irregular sensor data with missingness we recommend the use of this Bayesian functional regression model.



SESSION OC7D

### OC7D-3 ERNEST: A Representation Learning-based Cross-Subject EEG Channel Selection Algorithm

**Michela Carlotta Massi[1,2], Francesca Ieva[1,2,3]**
1 MOX – Laboratory for Modeling and Scientific Computing, Department of Mathematics, Politecnico di Milano, Italy
2 CADS – Center for Analysis Decisions and Society, Human Technopole, Milan, Italy
3 CHRP – Center for Healthcare Research and Pharmacoepidemiology, Bicocca University, Milan, Italy

**Background:** EEG is a non-invasive powerful technology that finds applications in several research areas. Currently, most EEG systems require subjects to wear several electrodes (channels) on the scalp. However, this induces longer clinical trial set-up times, hinders EEG-based assistive technologies' portability, and penalizes any Statistical or Machine Learning effort to decode EEG recordings by adding noisy and redundant information and rising computational complexity. One way to increase the signal-to-noise ratio and aid classification is combining Channel Selection (CS) with feature extraction. However, EEG signals' strong inter-subject variability led most of the existing literature to focus on subject-specific CS.
**Aims:** We propose **ERNEST**, *EEG* ***E****mbedde****R*** *a****N****d chann****E****l* ***S****elec****T****or*, a novel method for cross-subject CS.
**Methods:** Irrespectively of the task, each statistical unit is represented by a *trial* (recording session) matrix, where each row is a high-dimensional signal from one of the channels, associated to a binary target variable. ERNEST is designed to learn channel-specific 1-Dimensional Convolutional Neural Networks (1D-CNN) to embed signals grouped by electrode in a latent space of small dimensionality that maximizes intra-class separability. Then, it builds a unique representation of each trial by concatenating the channels' embedding into a *trial vector* from which it selects the most relevant channels to perform classification. For CS, ERNEST exploits a Deep Learning-based Feature Selection algorithm readapted from [1]. After training, ERNEST transfers the parametrized subgroup of selected channel-specific 1D-CNNs to embed new signals, obtaining *trial vectors* of small dimensionality and high predictive power that can be fed to any classifier.
**Experiments:** Cross-subjectivity can be *group-dependent* (testing on new trials of the same group of subjects used for CS) or *fully subject-agnostic* (testing on new subjects). We tested ERNEST in both frameworks on an Event-Related Potential (ERP) experiment discriminating between signals recording reactions to visual stimuli, and a patient classification task separating alcoholics from controls.
**Results:** Results are promising, especially compared to the limited literature on cross-subject CS. For instance, in *fully subject-agnostic* patients' classification ERNEST yields 0.858±0.018 AUROC with 5 channels out of 64, whereas selecting 10 channels for ERP achieves 0.951±0.012 and 0.724±0.024 AUROC in *group-dependent* and *subject-agnostic* frameworks respectively.
**References:** *[1] M. C. Massi, F. Ieva, F. Gasperoni and A. M. Paganoni, (2019) "Minority class feature selection through semi-supervised deep sparse autoencoders," MOX Report, vol. 38/2019.*

# SESSION OC7D

## OC7D-4 Linear statistical models and ridge regression used in shape index calculation on human face

**Stanislav Katina[1,2], Vojtěch Šindlář[1]**
1 Institute of Mathematics and Statistics, Masaryk University, Brno, Czech Republic
2 Institute of Computer Science of the Czech Academy of Sciences, Prague, Czech Republic

Spatial interpolation and smoothing is usually done for one surface. In our case, we have random samples of such surfaces represented by human faces captured by stereo-photogrammetry and characterised by about 150,000 points. These points are triangulated by about 300,000 triangles. The number of points is extremely high for the purpose of statistical analyses, therefore the 3D coordinates of (semi) landmarks on curves or surface patches sufficiently characterising the shape have to be automatically identified and this simplified model comprising about 1000 points is then used in further statistical modelling in functional data analysis setting. The identification of (semi)landmarks is a complex process during which B-splines, P-splines and thin-plate splines are used together with the measures of local surface topology, including principal curvatures and shape index. Shape index is calculated in R using different linear regression models and ridge regression model (allowing more flexibility for regression coefficients) of *z* coordinates on *x* and *y* coordinates, i.e. quadratic with interaction without/with intercept, cubic with interaction of *x* and *y* without/with intercept (without/with other interactions), and similar models of higher order. The estimates of regression coefficients related to the quadratic terms and their interaction are elements of Weingarten matrix from which the principal curvatures are calculated. These models are applied on sufficiently large neighbourhood of all points in local 3D coordinate system. Since the measures of local surface topology represent principal guide in estimating locations of ridge and valley curves across the face, we aim to compare different regression models used in shape index calculation on faces of patients with facial palsy and healthy controls. We suggest to use quadratic or cubic linear regression model or ridge regression model with interaction of the first order without intercept.

**Acknowledgment:** *The work was supported (partly) by RVO:67985807 and MUNI/A/1615/2020.*

**References:** *[1] Vittert L, AW Bowman, S Katina. A hierarchical curve-based approach to the analysis of manifold data. The Annals of Applied Statistics 13,4 (2019): 2539–2563. [2] Katina S, et al. The definitions of three-dimensional landmarks on the human face: an interdisciplinary view. Journal of anatomy 228.3 (2016): 355–365.*

## OC7D-5 Unsupervised classification of ECG signals via FDA to look for different patterns of variation among patients

**Annamaria Porreca**
Department of Economic Studies, University "G.d'Annunzio", Chieti-Pescara, Italy

Research on electrocardiogram records (ECG) often aims on supervised classification. Based on recorded signals, the main purpose is to create a classifier to recognize healthy people and new patients affected by specific heart disease. The reason is that the dataset on heart disease is often composed of records of heart signals over time and the labels of groups for each patient. Generally, the latter is a binary variable such as "Disease" and "Healthy": however, it's also possible to deal with more than two possible modalities of the grouping variable. Therefore, this kind of dataset pushes researchers to use this information and deal with supervised classification problems rather than unsupervised classification. Thus, rarely, scholars worry about finding specific patterns in ECG signals regardless of the known labels. This study's basic idea is to exploit functional data analysis (FDA) and unsupervised classification (i.e. clustering) to look for additional information in the data. The main aim of using these two approaches is to find patterns and different types of variability between the know groups. The starting point is to treat ECG signals using FDA. Because the former can be seen as a function in the time domain, the most intuitive approach is to treat these curves as functions of time as single objects. In the case of a binary outcome, having healthy and non-healthy people's curves, we can calculate the functional mean within each group. Nevertheless, this approach would not provide us with exhaustive information. Thus, we propose to use functional clustering to detect specific patterns within known groups. Clearly, in FDA, different metrics and approaches can be used for unsupervised classification. This study focuses on the functional k-means based on the functional principal components' semi-metric (FPCs). We present the results of this approach applied on a dataset composed of 200 patients recorded during one heartbeat. The dataset distinguishes patients with a regular heartbeat and people with a diagnosis of Myocardial Infarction. Our approach shows that within the "disease" group, we can highlight different patterns of non-healthy people and thus different type of Myocardial Infarction subclasses.

# SESSION OC7E

**SESSION OC7E** Selection and validation of prediction models

## OC7E-1 To tune or not to tune, a case study of ridge logistic regression in small or sparse datasets

**Hana Šinkovec[1], Georg Heinze[1], Rok Blagus[2], Angelika Geroldinger[1]**
1 Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria
2 Institute for Biostatistics and Medical Informatics, University of Ljubljana, Slovenia

For finite samples with binary outcomes penalized logistic regression such as ridge logistic regression (RR) has the potential of achieving smaller mean squared errors (MSE) of coefficients and predictions than maximum likelihood estimation. There is evidence, however, that RR is sensitive to small or sparse data situations, yielding poor performance in individual datasets. In such low-dimensional settings Firth's correction (FC) may be preferable. Motivated by an endometrial cancer study relating histological grading to three risk factors we demonstrate that the results provided by RR strongly depend on the choice of complexity parameter. However, estimating this parameter from the data by minimizing some measure of the out-of-sample prediction error or information criterion may result in arbitrary optimized values. We elaborate this issue further by performing a comprehensive simulation study, investigating the performance of RR in terms of coefficients and predictions and compare it to FC. In addition to RR where complexity parameter is estimated from the data, we also consider pre-specifying the degree of shrinkage according to some meaningful prior assumptions about true effects. Our benchmark is defined by oracle models that show what the best possible performance of RR could be if the true underlying data generating mechanism was known. We show that complexity parameter values optimized in small or sparse datasets are negatively correlated with optimal values and suffer from substantial variability which translates into large MSE of coefficients and large variability of calibration slopes. Therefore, applying tuned RR in such settings is problematic. In contrast, if the degree of shrinkage is pre-specified, accurate coefficients and predictions can be obtained even in non-ideal settings such as encountered in the context of rare outcomes or sparse predictors.

## OC7E-2 Propensity-based standardization to enhance the interpretation of prediction model discrimination

**Valentijn M.T. de Jong[1], Jeroen Hoogland[1], Karel G.M. Moons[1], Richard D. Riley[2], Tri-Long Nguyen[3], Thomas P.A. Debray[1]**
1 Julius Center for Health Sciences and Primary Care, UMC Utrecht, Netherlands
2 Centre for Prognosis Research, Keele University, United Kingdom
3 Section of Epidemiology, University of Copenhagen, Denmark

Interpreting model discrimination estimates in external validation studies is often challenging, because differences in discrimination may arise from invalid model coefficients and differences in (the distribution of) population characteristics. Hence, it may be unclear whether a developed prediction model may benefit from local revisions. This is particularly relevant in pooled data sets, where prediction models can be externally validated in multiple data sets. We aimed to disentangle the effects of differences in case-mix and invalid regression coefficients, to allow for the identification of reproducibility of model predictions (and predictor effects) in a target population.
We propose propensity-weighted measures of the concordance statistic that are standardized for case-mix differences between the development and validation samples. The propensity scores are derived with (multinomial) membership models that predict the originating samples of observations. We illustrate our methods in an example on the validation of eight diagnostic prediction models for the detection of deep vein thrombosis (DVT) that may aid in the diagnosis of patients suspected of DVT in 12 external validation data sets. We assess our methods in a simulation.
In the illustrative example, summary estimates of the meta-analysis of discrimination performance were not greatly affected by standardization. However, standardization substantially reduced the between-study heterogeneity of concordance, which indicated that variation in model concordance could partially be attributed to differences in case-mix, rather than invalid coefficients. In the simulation, only the propensity-score method estimated with splines produced unbiased estimates of concordance in the target population, whereas an unweighted approach and a propensity method with linear terms did not. Propensity score-based standardization may facilitate the interpretation of (heterogeneity in) prediction model performance across external validation studies, thereby guiding model updating strategies. These propensity score models should allow for non-linear effects to capture differences in sample variation.

# SESSION OC7E

## OC7E-3 A novel score for prognostic index assessment with event-free survival outcome

**Paola M.V. Rancoita**
University Centre of Statistics in the Biomedical Sciences (CUSSB), Vita-Salute San Raffaele University, Italy

In many clinical settings (especially, in cancer studies), a challenging goal consists in the definition of a prognostic index able to subdivide patients into groups with different event-free survival curve, on the basis of a subset of clinical variables. In order to be used in the practice, the identified classification must represent "clinically relevant" groups so that physicians may assign patients to different treatments or managements, depending on the level of prognosis. Thus, a good prognostic classification should satisfy the following properties: (1) the groups must correspond to "well separated" event-free survival curves, (2) the order of the prognostic levels must be retained in all cohort of the same clinical setting, (3) the groups must be reliable/robust in terms of size, (4) the classification should be characterized by a good survival prediction accuracy. Commonly applied scores for the assessment of a prognostic index (e.g. the Brier score [1] or the c-index [2]) actually evaluate only one or two of these characteristics.

In order to have a more comprehensive evaluation of a prognostic index, we defined a new measure of separation (called Expected SEParation, ESEP) which represents, from the theoretical point of view, the expected difference between the survival times of any two patients, given that they belong to groups with "consecutive" levels of prognosis. From its definition, ESEP is able to evaluate the properties (1-3). Hence, for a complete assessment of a prognostic index, ESEP must be used together with an error measure of survival prediction (such as the Brier score). From the estimation point of view, ESEP relies on the estimation of the restricted mean survival time, whose usage is suggested even in case of non-proportional hazards assumption. Overall, ESEP outperformed many other scores proposed in the literature on a wide range of simulated data, being able to better identify wrong prognostic classifications, even in case of small sample size and/or high percentage of censored data. The same behaviour was also observed when comparing different prognostic classifications on public real data, such as the ones of the German Breast Cancer Study Group 2.

**References:** *[1] Graf, E., Schmoor, C., Sauerbrei, W., Schumacher, M. (1999): Assessment and comparison of prognostic classification schemes for survival data. Statistics in Medicine, 18(17-18), 2529–2545.*
*[2] Pencina, M.J., D'Agostino, R.B. (2004): Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Statistics in Medicine, 23(13), 2109–2123.*



# SESSION OC7E

## OC7E-4 Minimum sample size for external validation of a clinical prediction model with a binary or time-to-event outcome

**Richard D. Riley[1], Thomas Debray[2], Gary S. Collins3,4, Lucinda Archer[1], Joie Ensor[1], Maarten Van Smeden[2], Kym I.E. Snell[1]**
1 Centre for Prognosis Research, School of Medicine, Keele University, Staffordshire, United Kingdom
2 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Netherlands
3 Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, United Kingdom
4 NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, United Kingdom

**Background:** In prediction model research, external validation is needed to examine an existing model's performance using data independent to that for model development. Current external validation studies often suffer from small sample sizes and consequently imprecise predictive performance estimates.
**Aims and methods:** In this talk, we propose how to determine the minimum sample size needed for a new external validation study of a prediction model for a binary or time-to-event outcome. Our calculations aim to precisely estimate calibration (Observed/Expected and calibration slope), discrimination (C-statistic) and clinical utility (net benefit). For each measure, we propose closed-form and iterative solutions for calculating the minimum sample size required [1]. These require specifying: (i) target standard errors (confidence interval widths) for each estimate of interest, (ii) the anticipated outcome event risk in the validation population, (iii) the prediction model's anticipated (mis)calibration and distribution of linear predictor values in the validation population, and (iv) potential risk thresholds for clinical decision making. Software code is discussed. We show how to derive the anticipated linear predictor distribution based on the C statistic or D statistic reported in the model development study.
**Results:** We illustrate our proposal for external validation of a prediction model for mechanical heart valve failure with an expected outcome event risk of 0.018. Calculations suggest at least 9835 participants (177 events) are required to precisely estimate the calibration and discrimination measures, with this number driven by the calibration slope criterion, which we anticipate will often be the case. Also, 3554 participants (64 events) are required to precisely estimate net benefit at a risk threshold of 8%. Lastly, we discuss the importance of also plotting simulated calibration curves for the sample size identified, to visually assess whether confidence interval bands are suitable in key ranges of predicted risks relevant to clinical decision making.
**Conclusions:** Our approach allows researchers to gauge the sample size required when designing a study for validating a model's predictive performance in new data. The calculations can also inform whether the sample size of an existing (already collected) dataset is adequate for external validation.

**References:** *[1] Riley RD, Debray TP, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. 2021 (submitted)*

# SESSION OC7E

## OC7E-5 The nonnegative garrote as a flexible approach for model selection in low and high dimensional data

**Edwin Kipruto, Willi Sauerbrei**
Institute of Medical Biometry and Statistics, University of Freiburg, Germany

**Background:** Prediction and descriptive models play an important role in medical and methodological research. As a result, many approaches for model selection have been proposed including the nonnegative garrote (Breiman 1995, NNG). Like other penalized approaches, it combines variable selection with shrinkage. However, little attention has been given to the original nonnegative garrote despite some of its good conceptual properties. Its unpopularity is probably caused by dependence on least square estimates which does not have a unique solution in high dimensional data and performs poorly in high degree of multicollinearity. However, it has been shown that NNG is a flexible approach that can be used in combination with other estimators besides least-squares such as ridge or elastic net hence the aforementioned challenges can be circumvented (Yuan and Lin 2007). Despite this proposal, it is hardly used in practice. Here our main interest is on descriptive modeling and as a byproduct, we will present results of prediction.
**Objectives:** The goal of this analysis is to evaluate the performance of nonnegative garrote in low and high degree of multicollinearity as well as in high dimensional data and compare results with competing approaches using three real dataset. In addition, the preliminary results from a simulation study in low dimensional data will be presented.
**Methods:** We evaluated four penalized regression methods namely nonnegative garrote, lasso, adaptive lasso and relaxed lasso, and two classical variable selection methods: best subset and backward elimination with and without post-estimation shrinkage.
**Results:** Nonnegative garrote can be used with other initial estimators besides least squares in highly correlated data and in high dimensional datasets. Only small differences in predictions were observed in methods while considerable differences were observed in the number of variables selected.
**Conclusion:** The proposed initial estimates can be used as an alternative to least squares estimates to fit nonnegative garrote in highly correlated data and in high dimensional data. With this extension NNG is a promising approach for model selection in various situations.

# SESSION OC7F

## SESSION OC7F Flexible modelling and spatial data analysis

## OC7F-1 Conditional and unconditional logistic regression analysis of matched case-control studies of recurrent events

**Yin Bun Cheung[1], Xiangmei Ma[1], K.F. Lam[2], Jialiang Li[3], Chee Fu Yung[4], Grant Mackenzie[5], Paul Milligan[6]**
1 Center for Quantitative Medicine, Duke-NUS Medical School, Singapore
2 Department of Statistics and Actuarial Science, University of Hong Kong, China
3 Department of Statistics and Applied Probability, National University of Singapore, Singapore
4 Infectious Disease Service, KK Women's and Children's Hospital, Singapore
5 Medical Research Council Unit The Gambia at LSHTM, Fajara, Gambia
6 Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, UK

Case-control studies can estimate incidence rate ratio in fixed or dynamic populations by using incidence density sampling, which is also called concurrent sampling. This involves matching cases and controls according to the time of events. We review the arguments for and against the use of conditional and unconditional logistic regression models for the analysis of individually matched case-control studies. While unconditional logistic regression analysis that adjusts for case-control matched sets as indicator variables generates a bias, the method used in parsimonious formulations does not. While both conditional and properly formulated unconditional logistic regression models have similar properties in the studies of outcome events that are non-repeatable, currently there is no valid method for statistical inference based on conditional logistic regression in the studies of recurrent events such as episodes of pneumonia or hospital readmissions. In this context, we propose to apply unconditional logistic regression with adjustment for time in quintiles and residual times within each quintile (9 degrees of freedom), with a robust standard error to allow for clustering of observations within persons. The methods are evaluated in simulations in realistic scenarios that resemble the studies of pneumonia and malaria. The results show that in case-control studies of non-repeatable events or first events using incidence density sampling, the proposed method and conditional logistic regression analysis give highly comparable results in terms of relative bias, root mean square error and coverage probability. However, only the proposed method provides correct statistical inference in the studies of recurrent events. We use data from a study of pneumonia and a study of malaria to illustrate.

## OC7F-2 On the implications of influential points for the selection and reproducibility of MFP models

**Willi Sauerbrei[1], Edwin Kipruto[1], Patrick Royston[2]**
1 Institute of Medical Biometry and Statistics, University of Freiburg, Germany
2 MRC Clinical Trials Unit at UCL, London WC1V 6LJ, UK

The number of covariates is often too large and a more parsimonious model is needed. The multivariable fractional polynomial (MFP) combines variable selection using backward elimination and function selection for continuous variables using fractional polynomial (FP) functions, thus a nominal significance level is required for each part (https://mfp.imbi.uni-freiburg.de/). A function selection procedure (FSP) compares the best fitting FP2 function in three steps with exclusion of the variable, the linear function and the best FP1 function. The latter tests are only conducted if earlier tests are significant. Regression diagnostics, such as detection of influential points (IP) and plots of residuals may reveal lack of fit or other peculiarities of a selected model. Single observations may be responsible for the selection of a linear, a monotonic non-linear FP1 or even one of the non-monotonic FP2 functions. Based on real data, Royston and Sauerbrei (2008) designed the 'ART' study, an artificial data set with six continuous and four categorical variables. They identified IPs using leave-one-out and illustrated that if the sample size is too small MFP would have low power to detect non-linear effects and the selected model may differ substantially from the true model. Using ART, we extended the investigation of IPs and their effect on selected univariate functions and the multivariable model. We propose diagnostic plots for the combination of two or more points and approaches to identify influential points in multivariable analysis. Using non-overlapping parts from the dataset, we consider model reproducibility and investigate the effect of sample size on MFP models. IPs can have a severe effect on FP functions, specifically in small datasets. Regression diagnostics can be used to identify IPs and their effect can be downweighed. Too small sample sizes often result in wrongly postulating linear effects with further implications for the selected multivariable model. If the sample size is large and regression diagnostics are carefully done, MFP is a suitable approach to identify variables with a stronger effect and suitable functional forms for continuous variables. For better illustration we use a structured profile which provides an overview of all analyses conducted.

# SESSION OC7F

## OC7F-3 Worldwide age specific Human Papillomavirus prevalence patterns: a mixed effects binomial method to cluster trajectories

**Maxime Bonjour[1,2,3], Hadrien Charvat[4], Fabien Subtil[2,3,5], Iacopo Baussano[1]**

1 Early Detection, Prevention and Infections Branch, International Agency for Research on Cancer, Lyon, France
2 Service de Biostatistique-Bioinformatique, Pôle Santé Publique, Hospices Civils de Lyon, France
3 Université de Lyon, Université Lyon 1, Villeurbanne, France
4 Cancer Surveillance Branch, International Agency for Research on Cancer, Lyon, France
5 UMR CNRS 5558, Laboratoire de Biométrie et Biologie Évolutive, Villeurbanne, France

**Introduction:** To cluster genital HPV age-prevalence trajectories from locations around the world, a clustering methodology taking into account heterogeneity is developed, compared by simulations with a methodology without heterogeneity and then applied to 29 locations worldwide.
**Material and method:** A clustering algorithm is created based on the binomial likelihood, with random effects in the age-prevalence trajectories. A clustering version of the EM algorithm is proposed to maximize the loglikelihood of the model. Applying different number of groups, the best model is selected according to classification criteria and clinical relevance. Simulations were performed to compare fixed and mixed effect classification models in terms of proportion of well classified trajectories, with different scenarios in terms of number of measurements at each time and variance of random effects. The models with (mixed) or without (fixed) heterogeneity are applied on genital HPV age-prevalence trajectories data for 29 locations worldwide.
**Results:** In the simulations, the proportions of mean trajectories well classified were systematically higher with the mixed effect model, from 30 to 50% compared with the fixed effect model. However, the validity of the classification obtained with the mixed effect model depends on the number of measurements at each time, decreasing with a decrease in measurements. For the application part, the fixed effect model selected a model with 4 groups that corresponded to translations of the 2 groups obtained with the mixed effect model. Features of the populations (sexual behaviors and demographic data) by groups are represented for both selected models.
**Discussion:** There is always heterogeneity in biological data and taking it into account can modify the clusters obtained. One limit of the mixed effect classification model is the high number of measurements required at each time. Fixed effect classification model classifies trajectories according to their level and shape whereas mixed effect model leads to classifications based mainly on the shape. The two models have different and complementary information, one may help identifying sexual behaviors linked with the overall intensity of HPV prevalence, and the other one factors linked with the evolution of HPV over age, allowing to target health actions on specific population subgroups.



SESSION OC7F

## OC7F-4 Treating ordinal outcomes as continuous quantities: when, why and how

**Yilin Ning[1,2], Peh Joo Ho[3,4], Nathalie C Støer[5,6], Ka Keat Lim[7,8], Hwee-Lin Wee[3,9], Mikael Hartman[1,2,3,10], Marie Reilly[11], Chuen Seng Tan[3]**

1 NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore
2 Yong Loo Lin School of Medicine, Department of Surgery, National University of Singapore and National University Health System, Singapore
3 Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore
4 Genome Institute of Singapore, Singapore
5 Norwegian National Advisory Unit on Women's Health, Oslo University Hospital, Norway
6 Department of Research, Cancer Registry of Norway, Norway
7 Department of Population Health Sciences, School of Population Health & Environmental Sciences (SPHES), Faculty of Life Sciences & Medicine, King's College London, United Kingdom
8 Programme in Health Services & Systems Research, Duke-NUS Medical School, Singapore
9 Department of Pharmacy, Faculty of Science, National University of Singapore, Singapore
10 Department of Surgery, National University Hospital, Singapore
11 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden

Ordinal variables, such as quality of life scores and patient-reported outcomes, are common in studies of population health and patient care. Analysing such outcomes often utilizes the linear regression model to estimate the effect of an exposure or intervention of interest. The magnitude of the effect is quantified by the difference in mean ordinal scores of the two (or more) groups being compared, and this quantity is useful for the assessment of clinical significance. However, this approach may be inappropriate as it assumes the ordinal outcome is a proxy for the continuous scale but does not assess this assumption.
The cumulative link model, which is an appropriate model for assessing the effect of an exposure on an ordinal outcome, is less well known and not widely used. Here we propose a new procedure using this model to assess the proxy assumption and to estimate the difference in mean ordinal scores when appropriate. As an illustration, the procedure is applied to five subscales of fatigue measured using the Multidimensional Fatigue Inventory to investigate the effect of time since diagnosis on fatigue among breast cancer survivors in Singapore.
A statistically significant improvement over time since cancer diagnosis was found in the General Fatigue and Mental Fatigue scores, but only General Fatigue satisfied the proxy assumption. As such, we can only draw conclusions on the magnitude of change in General Fatigue score, which is expected to be 1-unit for every 6.5 additional years since diagnosis and clinical significance (i.e., a 2-unit difference) achieved at the 13-th year.
The procedure offers a seamless way to assess both statistical and clinical significance of an effect on ordinal outcomes when the proxy assumption is appropriate. Where the assumption is not appropriate, only the statistical significance should be reported.

# SESSION OC7F

## OC7F-5 Health map for HealthGap: Defining a geographical catchment to examine cardiovascular risk in Australia

**Karen E. Lamb[1], Ximena Camacho[1], Ping-Wen Lee[2], Javier Haurat[2], Lukar E. Thornton[3], Maureen Turner[2], Julie A. Simpson[1], Aneta Kotevski[4], Luke Burchill[4]**

1 Melbourne School of Population and Global Health, University of Melbourne, Australia
2 BioGrid, Australia
3 Institute for Physical Activity and Nutrition, Deakin University, Australia
4 Department of Medicine, University of Melbourne, Australia

Linked administrative data provide opportunities to understand how people navigate and receive care within the health care system. The HealthGap pilot is a population-based cohort study linking general practice data to three major tertiary hospitals in Melbourne, Australia. HealthGap aims to understand health inequities in cardiovascular risk and treatment between Indigenous and non-Indigenous Australians and to recalibrate cardiovascular risk models for Indigenous people. As available data covered a only subset of hospitals in which cardiovascular events could be observed, the first aim was to define a catchment around HealthGap hospitals to estimate the denominator population for risk modelling.

Geocoded hospital location data were overlayed on postcode shape files from the Australian Bureau of Statistics. The number of cardiovascular events by postcode was extracted for each HealthGap hospital overall and by Indigenous status. Catchments were defined using three approaches and compared to determine the optimal definition. The first two approaches utilised first- and second-order neighbours of hospital postcodes, while the third was based on the spatial distribution of observed events. In this latter approach, postcodes were grouped into deciles based on numbers of events, with those in the top decile classified as locations with high cardiovascular burden and thus the largest at-risk population. Each approach was applied separately to each study hospital.

The first-order neighbours, second-order neighbours and spatial event distribution approaches captured 27-58% (27-41% Indigenous), 55-79% (59-65% Indigenous) and 72-95% (76-88% Indigenous) of cardiovascular events presenting at HealthGap hospitals, respectively. Although none of the approaches was able to exclude proximal non-study hospitals at which patients may also have presented, the impact was minimised by restricting to major cardiovascular events.

Geographic catchments provide an alternative approach to defining at-risk populations when denominator data is not readily available. Of the three approaches trialled in the HealthGap study, the spatial event distribution catchment performed best and was deemed the optimal approach to defining the denominator population. These techniques should be applied with caution in order to minimise the impact of proximal non-study hospitals, and refined based on local knowledge of processes of care.

# SESSION OC7G

SESSION OC7G — Methods for clinical research

## OC7G-1 Bayesian extrapolaton from pre-clinical data to human

**Sandrine Boulet[1], Moreno Ursino[2], Robin Michelet[3], Charlote Klof[3], Emmanuelle Comets[4,5,6], Sarah Zohar[1]**

1 INSERM, Centre de Recherche des Cordeliers, Université de Paris, Sorbonne Université, France
2 F-CRIN PARTNERS Platorm, Assistance Publique-Hôpitaux de Paris, Université de Paris, France
3 Department of Clinical Pharmacy and Biochemistry, Insttute of Pharmacy, Freie Universitaet Berlin, Germany
4 Université de Paris, INSERM, IAME, F-75006, Paris, France
5 INSERM, CIC 1414, 35700 Rennes, France
6 Univ Rennes-1, France

**Background:** During the initial phases of drug development, the toxicity and mechanisms of action of a candidate drug are explored in preclinical studies, building candidate pharmacodynamic models. These models can be used to define a dose range that is likely to be effective and of low toxicity in human studies. Enriching these models with other sources of information (elicited expert data, data from previous experiments or data on similar molecules) can be particularly interesting.

**Aim:** We propose a Bayesian approach of extrapolation from pre-clinical data to human, taking into account all external sources of information (data, published articles, etc.).

**Method:** Our work is inspired by a real-life example in oncology, the inhibition of TGF-beta signaling to block tumor growth [1]. We consider several *in vivo* PK/PD experiments from literature. Bayesian models are used on the dose-outcomes (PD surrogates of efficacy or toxicity) relationship with mixture distributions constructed from elicited expert data, data from previous experiments and data from similar molecules, as prior distributions. Maximum tolerated doses determined during Phase I and II are assumed to be expert elicited data. Thereafter, each of the *posterior* distributions resulting from these analyses is used to transfer information via extrapolation to obtain a dose distribution in humans. The final recommended dose range for first-in-human study is then deduced from these distributions.

**Results:** We perform an extensive simulation study. Compared to the standard methods that use only the preclinical estimated dose in the extrapolation, our simulations suggest that using all the available external information leads to a better (in terms of efficacy and toxicity) dose-range choice in first in human clinical trials.

**Conclusion:** Using the proposed Bayesian approach, which incorporates all available external information, allows a better dose selection for human trial, possibly reducing the trials failure rate due to wrong dose panel selection.

This work is part of the European FAIR project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 847786.

**References:** *[1] Gueorguieva et al., British Journal of Clinical Pharmacology (May 2014); 77(5):796–807.*

# SESSION OC7G

## OC7G-2 Analysing ordinal endpoints in two-arm randomized clinical trials

**Anne-Laure Boulesteix, Christian Bihl, Eva Hoster**
Institute for Medical Information Processing, Biometry and Epidemiology, LMU Munich, Germany

We compare the type 1 error and power of several methods for the analysis of ordinal endpoints in two-arm randomized clinical trials. The considered methods are the two-sample Wilcoxon rank sum test (which ignores that the endpoint is categorical), the chi-square test of independence and the corresponding exact test (which ignore the ordinal scale of the endpoint), the chi-square test for trend, the likelihood-ratio test in the proportional odds ordinal logistic regression model as well as different (exact and asymptotic) versions of an original approach based on the distribution of the maximally selected chi-square statistic. In this context, "maximal selection" refers to the selection of a cutpoint considered to dichotomize the ordinal endpoint. The latter new approach yields an additional interpretable output (namely the optimal cutpoint) and is based on an intuitive principle that can be understood as a correction procedure for multiple testing. The results of our extensive simulation study suggest that all tests, except the exact version of the new approach, have an adequate type 1 error below but close to the nominal level. However, they show different performance patterns in terms of power depending on the sample size and the distribution of the ordinal endpoint in the two treatment arms. The considered methods are illustrated through an application to data from four randomized clinical trials investigating the effect of new therapies on the outcome of lymphoma patients.

## OC7G-3 Simulations to assess the impact of cross-over due to COVID-19 in a cluster-randomized non-inferiority trial

**Cattram Nguyen[1,2], Grant Mackenzie[2,3], John Carlin[1,2]**
1 Murdoch Children's Research Institute, Australia
2 Department of Paediatrics, University of Melbourne, Australia
3 Medical Research Council Unit The Gambia at London School of Hygiene & Tropical Medicine, UK

**Background:** The COVID-19 pandemic has led to disruptions to clinical research globally, including the suspension of thousands of clinical trials. In rural Gambia, a cluster-randomised, non-inferiority trial is underway to compare two schedules of the pneumococcal conjugate vaccine: 3 doses (standard) versus 2 doses (intervention). In 2020, this trial was temporarily suspended due to the COVID-19 pandemic. This led to treatment non-adherence whereby children residing in the 2-dose clusters received the standard 3-dose schedule. The aim of the current study is to use simulations to assess the impact of treatment non-adherence on study results.
**Methods:** The simulations replicated the trial design: 35 clusters in the 2-dose arm, 33 clusters in the 3-dose arm, 60 individuals per cluster. The primary outcome was a binary variable, the contrast between groups expressed as a risk ratio, and the non-inferiority margin was 1.38. The outcome data were simulated using a multilevel logistic model, with simulation settings based on observed data: prevalence of the outcome in the standard group (13%, 15%), odds ratio comparing the two groups (1, 1.1, 1.2) and percentage of non-adherence in the intervention arm (3%, 4%, 6%, 8%, 10%). Risk ratios were estimated using generalised estimating equations. We estimated power, bias and coverage using 2000 replications for each scenario.
**Results:** The simulation results indicated that with small amounts of non-adherence (i.e. 3-10%) in one arm only, there was minimal bias across the scenarios. Estimates of coverage were within nominal levels. There was little impact of non-adherence on study power, but as expected for this non-inferiority study design, power did decrease as the assumed odds ratio increased.
**Conclusion:** This study illustrated methods for assessing the potential impact of the COVID-19 pandemic on trial results, and indicated minimal impact of participant non-adherence on trial results. We are undertaking further simulations to assess bias with more non-adherence, and evaluating methods for handling the non-adherence (e.g. two-stage least squares regression using cluster-level summaries for estimating the local average treatment effect). This work will inform the analysis for the current trial and is relevant to other clinical studies that have been disrupted by the COVID-19 pandemic.



# SESSION OC7G

## OC7G-4 Analysis methods for personalized randomized controlled trial (PRACTical)

**Kim May Lee[1], Rebecca Turner[2], Ian White[2], A Sarah Walker[2]**
1 Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom
2 MRC Clinical Trials Unit, University College London, United Kingdom

For some clinical areas, multiple interventions are available, but it is not known which works best, and some interventions are not suitable for some patients. A personalized randomized controlled trial (PRACTical) design has been proposed for this setting. In this, each patient is randomised between the interventions for which they are eligible. The design can be considered as a platform that investigates a set of interventions on multiple subgroups of patients, where subgroups have "personalized" randomization lists that reflect their eligible interventions. The aim is to produce rankings of interventions that can guide the choice of intervention for each individual, rather than to estimate relative intervention effects.
We explore three analysis methods for this novel design. First, various analysis approaches at the subgroup level are possible but unlikely to provide inference of adequate precision. Second, we can utilise all indirect evidence by combining evidence from all subgroups as in a standard network meta-analysis, where the evaluation of interventions on a subgroup is analogous to a standard trial. This method adjusts for the randomization lists in the model, which would fail to converge if the number of subgroups is large. Third, a method resolves this issue by estimating each pairwise comparison separately from the appropriate subset of data and combining the estimates for different comparisons. Approaches 2 and 3 assume the indirect evidence on intervention comparisons is consistent with the direct evidence.
By performing simulation studies with a binary outcome, we evaluate the performance of the three analysis methods in terms of the properties of the intervention effect estimates and some new performance measures that summarise the benefit of treating according to the intervention rankings and the chances of picking best or near-best interventions. We find that approaches 2 and 3 provide estimates with good properties when there are no qualitative or quantitative subgroup-by-intervention interactions, and are reasonably robust to plausible levels of interaction. The overall rankings produced by these methods are likely to be suitable for determining intervention policies and guiding individual clinical decisions.

# SESSION OC7G

## OC7G-5 MultiNet: a computational algorithm to analyze high-dimensional epigenetic correlation structures

**Sara Prada Alonso**[1,2]
1 Mathematics, University Santiago de Compostela, Spain
2 Biostatistics, Clinipace, Spain

Derived from the lack of efficient algorithms of big data analysis, and motivated by the importance of finding a structure of correlations in (epi)genomics, we present a computational algorithm of topological data analysis (TDA) called MultiNet [1] with the main aim of developing a novel analytical tool to analyze the correlation structure of a high-dimensional epigenetic dataset. Additionally, MultiNet detects epigenetic patterns associated with data characteristics as a cancer disease status. This fact makes MultiNet a very powerful diagnostic tool to identify disease-associated biomarkers to be used in medical practice.

MultiNet is based on the idea of translating the data into the language of algebraic topology to extract their main characteristics (the "shape" of the data). MultiNet converts a big data cloud into a correlation network where each node is a cluster of variables and they are joined as per their Pearson correlation, extending the Mapper design [2] as it follows a "divide and conquer" strategy to analyze the correlation of the data by overlapping data windows and doing feature modeling. This network is a simpler object that allows to analyze the data cloud in a very complete and fast way.

MultiNet follows the main machine learning algorithm steps of data collection (DNA methylation arrays), exploration (remove sites with higher known variability), model training (creating the network and applying statistical models like random forest to detect significantly differentiated markers among sample groups), evaluation (test the significant sites on new datasets and detect disease-related biomarkers), and learning (apply the algorithm to explore deeply regional patterns on chromosomes). Moreover, MultiNet has more functionalities implemented as network differentiation models or plots for biological interpretation.

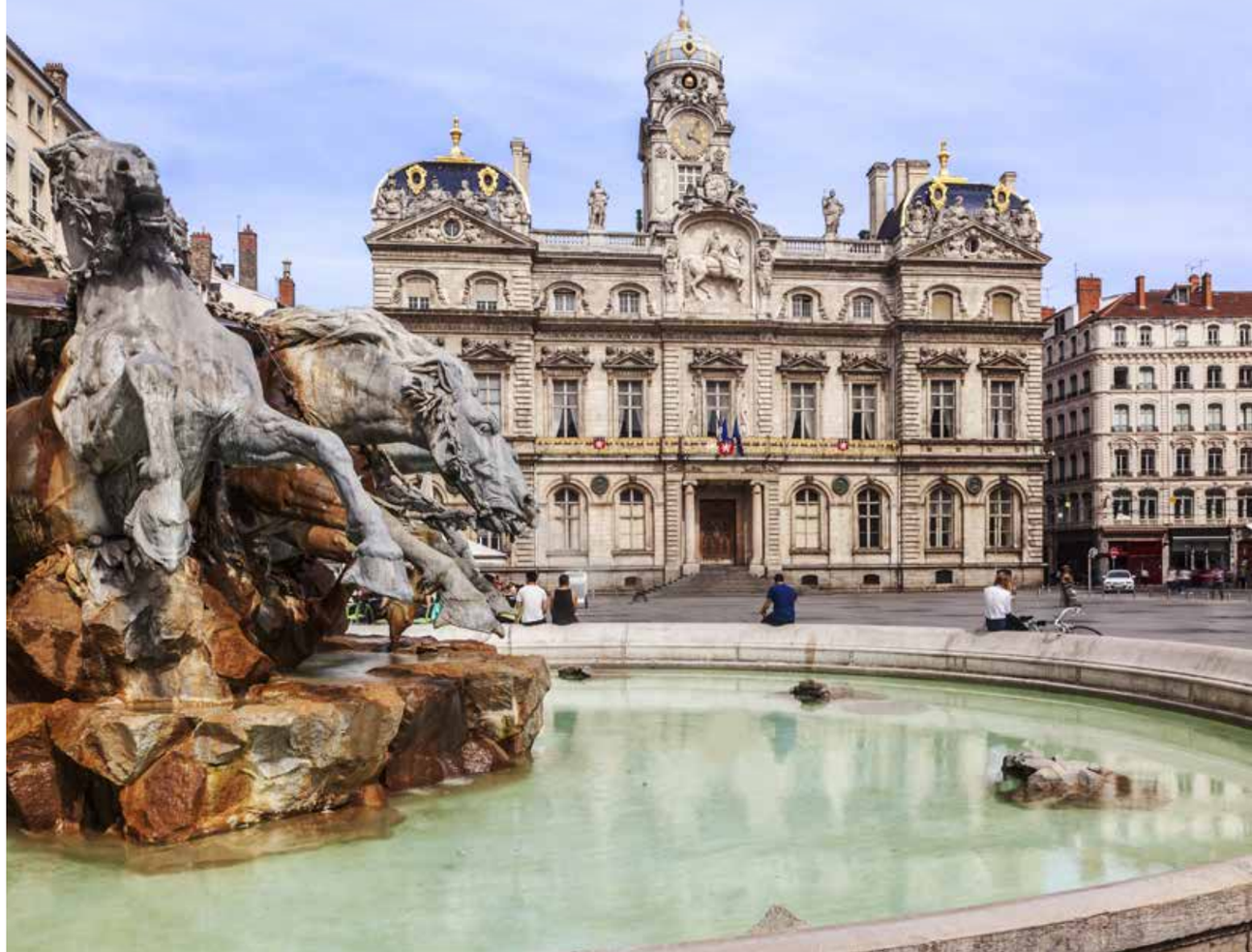
MultiNet identifies, for instance, several markers associated to prostate and colorectal cancer with a very good sample prediction rate. Also, it identifies important genes associated with the aging process.

Overall, this work establishes a novel perspective of analysis and modulation of hidden correlation structures, specifically those of great dimension and complexity, contributing to the understanding of the epigenetic processes, and that is designed to be useful for non-biological fields too.

**References:** *[1] https://github.com/SPRADA1/MultiNet. [2] Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition, Gurjeet Singh, F. Mémoli and G. Carlsson, PBG@Eurographics, 2007.*



Poster Sessions



# POSTER SESSION 01

## POSTER SESSION 01 — Methods for clinical research

### P01-01 An early-phase clinical trial design in oncology with generalization ability

**Shinjo Yada[1], Ryuji Uozumi[2]**
1 Biostatistics Department I, Data Science Division, A2 Healthcare Corporation, Japan
2 Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine, Japan

Recent years have seen progress in precision medicine that uses biomarkers to predict the actions and effects of pharmaceuticals. Finding investigational new drugs using biomarkers has become an important part of drug development. Research into using image data as a biomarker to indicate therapeutic effects has accelerated the progress of precision medicine. A diverse array of biomarkers is used in clinical trials, and strong correlations between biomarkers are believed to exist. Clinical trials are often conducted on a limited number of cases. Therefore, it becomes difficult to improve the generalization ability of conventional statistical models that express the relationship between dose and outcome if patients' biomarkers are included among the explanatory variables. In the presentation, we propose an approach that applies a Gaussian process, which is a form of machine learning, to an early-phase clinical trial in oncology. The treatment (dose), the status of the biomarkers, and the interaction between biomarker status and treatment (dose) are the input values; the presence or absence of efficacy and toxicity are the output values. Using previously accumulated patient data, efficacy and toxicity are predicted for each treatment (dose) based on subsequent patients' biomarkers status. From the predicted values, the treatment (dose) believed to be optimal for the patient are adaptively determined. We examined the operating characteristics of the proposed approach by simulating several scenarios. The results of a simulation study suggested the proposed approach obtains a desirable selection percentage of the optimal dose compared with conventional approach.

### P01-02 A flipped classroom approach for teaching medical statistics and statistical software training

**Andreas Allgöwer, Rainer Muche, Ulrike Braisch, Marianne Meule, Benjamin Mayer**
Institute for Epidemiology and Medical Biometry, Ulm University, Germany

Teaching biometry often relies on the use of statistical software courses. The aim is to impart hands-on experience for students beside the theoretical contents of medical biometry, enabling them to handle own research projects in the future autonomously to some extent. Our experience from such courses is that the technical realisation of the practical components is time-consuming, thus leaving only a small amount of time for exchange, questions and interpretation of statistical methods. The time until everyone feels sufficiently save in terms of the knowledge imparted is heterogeneous.
We will upend this scenario within the didactic method of a Flipped Classroom (https://facultyinnovative.utexas.edu/flipped-classroom). Before the course, students are ought to familiarize themselves with the statistical software and to work on exercises. Thus, they will have the possibility to adjust for both individual learning pace and experience. In the beginning of the course technical problems can be solved and – ideally – there will be more time left to discuss biometrical topics. The advantage is that the lecturers accompany and supervise this phase.
Presumptions for the implementation of such a concept in statistical software courses are, that
- the students have unlimited access to the software (anytime, anywhere) so that students have the opportunity to individually schedule the required time for preparation and
- a bundle of self-explanatory utilities (script, learning-videos, exercises, data, moodle-page) is available for autonomous training.

This approach will be implemented in our teaching of medical students at Ulm University. SAS Studio / SAS on Demand for Academics [1] offers a statistical software platform version that enables students to have unlimited access free of charge. The only presumption is that an internet connection and a browser are available, which may be assumed to be standard nowadays. A script [2] and other required learning materials are already available in moodle (in German, for our students for free). Altogether, this seems to be a reasonable basis for the execution of a Flipped Classroom. We will present the idea and first evaluation data. In a second step the program will be evaluated by a cluster randomised study.

**References:** *[1] SAS on Demand for Academics. Available from: https://www.sas.com/en_us/sofware/on-demand-for-academics.html [2] Büchele G, Rehm M, Muche R: Medical Statistics using SAS Studio in SODA (in German). Springer Verlag, Heidelberg, 2019*



POSTER SESSION 01

### P01-03 A comparison of dual biomarker threshold identification procedures within a confirmatory clinical trial

**Ben Lanza[1], Chris Harbron[2], Nigel Stallard[1], Deepak Parashar[1]**
1 Statistics and Epidemiology Unit, Warwick Medical School, United Kingdom
2 Roche Pharmaceuticals, United Kingdom

**Background:** Often, targeted therapies only show benefit in a 'sensitive' subgroup of the patient population. This can cause such treatments to be overlooked in broad 'all-comers' trials, due to dilution of the observed treatment effect. Patient subgroups can be defined by continuous biomarker values, in conjunction with a threshold value to dichotomise the population into sensitive and non-sensitive. Such biomarker-based subgroups are frequently identified retrospectively or in an exploratory manner within trials, which can lead to inefficiencies and delays in patient care. Identifying and validating biomarker-based subgroups within a single confirmatory trial is therefore key. Moreover, there is increasing evidence to suggest that multiple biomarkers are needed to sufficiently identify sensitive patients for some drugs or drug combinations. In this work, a variety of dual biomarker threshold identification procedures are applied in a phase III trial setting and their performance contrasted.
**Methods:** In this work, it was of interest to identify thresholds for two continuous biomarkers simultaneously, which dichotomise the respective biomarkers into sensitive and non-sensitive patients, thus defining a two-dimensional patient subgroup i.e. patients who are defined as sensitive for both biomarkers. Four methods were implemented within Freidlin and Simon's Adaptive Signature Design (ASD) framework, these being: a grid search, a modelling-based method, recursive partitioning and prognostic peeling. Methods were contrasted by their ability to accurately identify biomarker threshold locations and by the proportion of trials which achieved significant efficacy, both overall and subgroup specific. This work was carried out using a simulation study.
**Results:** In the simulation study, recursive partitioning methods showed the best overall performance, with respect to both the threshold identification accuracy and trial operating characteristics. All methods suffered when the expected proportion of sensitive patients was low and when the magnitude of treatment effect was modest.
**Conclusions:** Dual biomarker threshold identification can be successfully incorporated into a confirmatory phase III setting, without jeopardising the ability to detect an overall treatment effect. In such low dimensional settings, recursive partitioning methods should be taken into consideration.

# POSTER SESSION 01

## P01-04 Multifactor intervention efficacy on MACE and mortality in diabetic kidney disease: a cluster-randomized controlled trial

**Pia Clara Pafundi[1], Vittorio Simeon[2], Paolo Chiodini[2], Luca De Nicola[1], Raffaele Galiero[1], Roberto Minutolo[1], Ferdinando Carlo Sasso[1]; on behalf of NID-2 study group Investigators**

1 Department of Advanced Medical and Surgical Sciences, University of Campania "Luigi Vanvitelli", Naples, Italy

2 Department of Physical and Mental Health and Preventive Medicine, University of Campania "Luigi Vanvitelli", Naples, Italy

Diabetic kidney disease (DKD) associates with a very-high cardiovascular risk [1], advocating implementation of an intensive and multifactorial risk factors therapy. NID2 aimed to assess efficacy of a multifactorial intervention, versus Standard-of-Care (SoC), on major fatal/non-fatal cardiovascular events (MACEs) in DKD patients with albuminuria and diabetic retinopathy.

NID2 is a multicentre, cluster-randomized, open-label clinical trial on 395 DKD patients from 14 Italian diabetology clinics, aged >40 years, with negative CV events history. Centres were randomly assigned to a multifactorial intensive therapy (MT, n=207) of main cardiovascular risk factors and SoC (n=188). Primary endpoint was MACEs occurrence by end of follow-up phase. Secondary endpoints included single components and all-cause death.

Standardized differences (SDiff) cut-offs criteria by Leyrat et al. were used to establish baseline covariates' imbalance in cluster randomization [2]. To evaluate a global imbalance, c-statistic was further calculated. P-values to consider clustering were computed by generalized estimating equations (GEE) model with cluster as group variable. Distribution of dependent variable and link-function was used as appropriate (gaussian/identity for continuous variable, binomial/logit for dichotomous variable). Groups comparison at end of intervention was also performed by GEE, further adjusting for baseline values. Median follow-up was calculated by inverse Kaplan-Meier procedure and primary endpoint following intention-to-treat principle, with event curves based on Kaplan–Meier analysis. Due to cluster-randomization, a Cox shared-frailty model was fitted to calculate HR and 95% Confidence Interval. Across centres, frailties are assumed as gamma-distributed latent random effects affecting hazard multiplicatively.

Intervention lasted on median 3.84 and 3.40 years in MT and SoC, respectively. At end of intervention, targets achievement was significantly higher in MT. 74 MACEs were recorded (50 vs. 24 in MT), with an unadjusted HR 0.28 (95%CI 0.13-0.63; p=0.002). During global 13 years follow-up 262 MACEs were recorded (116 in MT vs. 146 in SoC). The adjusted Cox shared-frailty model demonstrated 52% lower risk of MACEs in MT arm (aHR 0.478, 95%CI 0.30-0.74, p=0.001). Similarly, all-cause death risk was 47% lower (aHR 0.53, 95%CI 0.29-0.93, p=0.027). In conclusion, MT reduces over the long-term MACEs and mortality risk in high-risk DKD patients. MT shows an early benefit on MACEs.

**References:** *[1] Rawshani A, et al. Risk Factors, Mortality and Cardiovascular Outcomes in Patients with Type 2 Diabetes. NEJM. 2018;379(7):633-44 [2] Leyrat C et al. Propensity score to detect baseline imbalance in cluster randomized trials: the role of the c-statistic. BMC Med Res Methodol. 2016;16:9*

# POSTER SESSION 01

## P01-05 Challenges in Factorial Design Randomized Control Trials

**Ioana R. Marian, Susan J. Dutton, Pradeep S. Virdee, Anita Mansouri, Nicholas Peckham, Mae Chester-Jones, Sally Hopewell**

Centre for Statistics in Medicine, University of Oxford, United Kingdom

**Description:** Randomised controlled trials (RCTs) using a factorial design enable the assessment of two or more interventions within a single trial. Compared to multi-arm trials, factorial RCTs are more efficient as they require fewer participants, with the assumption that the interventions act independently of each other (i.e. no interaction effect is present). This supposition creates specific challenges in the design, analysis and reporting of a factorial RCT, which if ignored can lead to biased results [1,2].

**Objective:** To evaluate current methodology and reporting of published reports of 2x2 factorial design RCTs. Additionally, to assess how frequently trial design methods differ in reporting of results compared to those pre-specified in the protocol/statistical analysis plan (SAP).

**Methods:** We searched PubMed to identify primary reports of 2x2 factorial design RCTs published between 01 January 2018 and 04 March 2020. The corresponding trial protocol and/or SAP were collected, where available. Data from both primary reports and protocol/SAP were extracted and compared on the trial characteristics (disease, sample size, funding, etc.) and approach to factorial design-specific methodology, such as design rationale or consideration for a treatment interaction in the sample size and analysis as indicators of potential challenges.

**(Preliminary) Results:** The review included a purposeful sample of 100 factorial RCTs. The majority (23%, n=23/100) were conducted in cardiology; the median sample size was 258 (interquartile range 120 to 693); 44% (n=44/100) were multicentre; 61% (n=61/100) were funded by non-industry. The rationale for a factorial design was often efficiency in assessing multiple treatments in one RCT (44%, n=44/100). 12% (n=12/100) explicitly assumed no treatment interaction in the outset and 4% (n=4/100) reported powered sample size to detect an interaction. The primary outcome analysis was conducted for the main effects in 43% (n=43/100), as a four arm comparison for 25% (n=25/100) and both in 32% (n=32/100). Of 60 articles reporting testing an interaction, 83% (n=50/60) reported non-significant interactions. Protocols/SAPs were available for 37% (n=37/100) of the published primary reports. 65% (n=24/37) intended to assess for an interaction in the analysis (as reported in the protocol/SAP) and 17% (n=4/24) did not report this in the final report.

**References:** *1 Montgomery AA, Astin MP, Peters TJ. Reporting of factorial trials of complex interventions in community settings: A systematic review. Trials. 2011;12. doi:10.1186/1745-6215-12-179, 2 Kahan BC. Bias in randomised factorial trials. Stat Med 2013;32:4540–9. doi:10.1002/sim.5869*

# POSTER SESSION 01

## P01-06 Non-inferiority trials with indirect evidence of assay sensitivity using network meta-analysis

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

**Background:** The choice of a non-inferiority (NI) margin and assurance of assay sensitivity are well-known issues in 2-arm NI trials. The conclusion that a NI trial has assay sensitivity is based on the following three considerations: (i) historical evidence on the efficacy of the treatment effect, (ii) the constancy assumption, and (iii) the quality of the NI trial (ICH-E10, FDA NI guidance). A 3-arm NI trial including both a placebo and a reference treatment, called the gold standard design, is strongly recommended to assess assay sensitivity. However, there are concerns about the ethics and feasibility of including a placebo; consequently, practical applications of the 3-arm NI trial have not progressed. Therefore, there is a need for a practical method to assess assay sensitivity in the 2-arm NI trials.
**Objective:** We propose a new practical approach to confirm assay sensitivity in a 2-arm NI trial. This method involves assessing the assay sensitivity using the indirect effects of a reference treatment and a placebo by integrating data from previous trials and the 2-arm NI trial using network meta-analysis.
**Method:** To assess assay sensitivity, it is necessary to demonstrate that the acceptable minimum effective value of a test treatment in the 2-arm NI trial is superior to placebo (Hida & Tango, 2018). Since the 2-arm NI trial does not include a placebo, we are forced to use historical trial results as external information. In other words, the proposed method uses network meta-analysis to assess and obtain indirect evidence on the substantial superiority of the reference treatment over placebo. The performance of this method is investigated in terms of the actual type I error rate, joint power, and calculated sample size using simulations of several scenarios based on the data from clinical trials.
**Results and Conclusions:** The level of evidence for the proposed method may be lower than that for the gold standard design, owing to the use of external information. However, the performance of this method as per the results of various simulations suggests that it will be useful as one of the methods to assess assay sensitivity of the 2-arm NI trial.
**Reference:** Hida E. & Tango T. Pharmaceutical Statistics. 2018. 17(5). 489-503

## P01-07 Biomarker-based Bayesian randomized clinical trial for population finding with efficient reduction of sample size

**Akiyoshi Nakakura[1], Satoshi Morita[1], Yasuo Sugitani[2], Hideharu Yamamoto[2]**
1 Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine, Kyoto, Japan
2 Biometrics Department, Chugai Pharmaceutical Co., Ltd., Tokyo, Japan

The benefits and challenges of incorporating biomarkers into the clinical development of new agents have been increasingly discussed. In many cases, more accurately identifying a sensitive subpopulation of patients, a larger sample size, thereby, a higher development cost and a longer study period may be required. Thus, we consider designing interim analyses with decision criteria to improve the efficiency of the study design that is aiming to reduce the expected number of patients enrolled onto a clinical trial. The clinical trial analyzes a time-to-event endpoint such as progression-free survival time and the decision criteria accounts for the amount of accumulated information of observed events. We discuss a Bayesian randomized clinical trial design incorporating a predictive biomarker measured on a graded scale for the development of a new molecular targeted treatment Extensive simulation studies evaluate the operating characteristics of the proposed method, including the correct identification probabilities of the desired subpopulation under a wide range of clinical scenarios.



# POSTER SESSION 01

## P01-08 A methodological review of phase I designs with late-onset toxicities and incomplete follow-up

**Zhulin Yin[1], Adrian Mander[2], Christina Yap[1]**
1 Clinical Trials and Statistics Unit, Institute of Cancer Research, United Kingdom
2 Centre for Trials Research, Cardiff University, United Kingdom

**Background:** Conventional phase I designs, such as 3+3 and CRM design, are conducted based on cytotoxic agents, where the acute toxicity is likely to occur within the first cycle of treatment. These designs require previous participants to be fully followed up before further dose assignment. However, some therapies such as immunotherapies and molecularly targeted agents may have late-onset toxicities. Suspending new recruitment and waiting for the full observation of DLT outcome may lead to a prolonged trial and increase trial costs, Similar difficulty also arises when accrual is fast as many may have incomplete follow-up [1, 2]. Several designs have been proposed to address this issue, but the uptake has been slow. We conduct a methodological review to provide a comprehensive overview of the designs and their characteristics.
**Methods:** We performed searches in PubMed in November 2020. Phase I designs that clearly stated late-onset/pending toxicity consideration were included. We also checked the references of these identified papers to ensure no designs are missed. Key characteristics such as the trial design, methodology, advantages and limitations, and how the designs have been implemented in published trials, are extracted.
**Results:** Our search yielded 23 designs, where 11 (47.8%) are parametric, e.g. TITE-CRM and 12 (52.2%) are non/semi-parametric designs, e.g. TITE-BOIN and Rolling 6. Only 5 (21.7%) designs have been implemented in published clinical trials. We analyzed the time from publication of a novel design to a first published trial application using the Kaplan-Meier estimates. The probability of implementation at 5 years is 0.11, 95% CI [0, 0.24]. Time-related weight function is one typical way to deal with late-onset toxicity: weight function is applied to the dose-toxicity model (e.g. TITE-CRM) or to the toxicity outcome (e.g. TITE-PIPE). TITE-CRM has been implemented most often – amongst those published trials, fifteen used the uniform weight function and five assigned different weights on different DLT follow-up period.
**Conclusions:** Intelligent trial designs that allow for more rapid trial completion and achieve high accuracy in determining the right dose are much needed in practice. This comprehensive review enhances knowledge and provides guidance for investigators to choose among such study designs.
**References:** *[1] Lee, S. M., Backenroth, D., Cheung, Y. K., Hershman, D. L., Vulih, D., Anderson, B., Ivy, P., & Minasian, L. (2016). Case Example of Dose Optimization Using Data From Bortezomib Dose-Finding Clinical Trials. Journal of clinical oncology: official journal of the American Society of Clinical Oncology, 34(12), 1395–1401. [2] Liu, S., Yin, G., & Yuan, Y. (2013). BAYESIAN DATA AUGMENTATION DOSE FINDING WITH CONTINUAL REASSESSMENT METHOD AND DELAYED TOXICITY. The annals of applied statistics, 7(4), 1837–2457.*

## P01-09 Patient-specific dose finding in seamless phase I/II clinical trials

**M. Iftakhar Alam, Shantonu Islam Shanto**
Institute of Statistical Research and Training, University of Dhaka, Bangladesh

This paper incorporates a covariate to determine the optimum dose in a seamless phase I/II clinical trial. A binary covariate and its interaction effect are assumed to keep the method simple. Each patient's outcome is assumed to be trinomial, and the continuation ratio model is utilized to model the dose-response data. The Bayesian approach estimates parameters of the dose-response model. At each stage of a trial, we allocate that dose to a patient for which the estimated probability of efficacy is maximum subject to the constraint that the estimated probability of toxicity is no more than a target value. Also, we allow the design to stop early for futility and/or toxicity. Eight plausible dose-response scenarios are investigated to check the proposed methodology. A simulation study shows that covariate consideration can enhance the identification of the optimum dose when it is appropriate to do.

# POSTER SESSION 01

## P01-10 The true power of clinical trials in pediatric cancers and other rare diseases

**Audrey Mauguen**
Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York NY, US

**Background:** Clinical trials are challenging in rare diseases like pediatric cancers, where the accrual is limited. In these trials, inference assumptions are the same as in common diseases, ie the sample comes from a quasi-infinite population. This leads to overestimating the variance of the treatment effect. The finite-population correction factor (FPCF) is often used in surveys, but not in clinical trials. With few assumptions, the use of the FPCF can improve trials efficiency, showing that the power of those trials is higher than it appears.
**Methods:** First, a simulation study assessed the standard-error of the mean (sem) treatment effect and coverage of the 95% confidence interval with and without the FPCF. Second, a corrected power of a z-test was derived. Finally, the impact on the sample size calculation was investigated. The impact of using the FPCF versus the naive approach is assessed for varying treatment effect, sample size and population size.
**Results:** The simulations results confirmed the overestimation of the sem with the naïve estimator. Depending on the scenario, the gain in power reached 10.0%, 14.1% and 12.9% to detect a difference in treatment effect of 10%, 15% and 20%, respectively. The gain increased with the sample size. It was negligible for n=30, and in scenarios with high power (>95%). This gain in power translated into a decrease in sample size: if the naive calculation leads to a sample size of 10% the population size, then the sample size can be divided by 1.1; if the naïve calculation leads to a sample size of 50% the population size, then the sample size can be divided by 3, in order to reach the planned type-I error and power. A Desmoplastic Small Round Cell Tumor trial is presented where the sample size is decreased from 32 to 27 patients.
**Conclusion:** When dealing with rare diseases like pediatric cancers, the power of clinical trials is higher than it appears. The gain in efficiency was seen with reasonable sample sizes and treatment differences, showing it can be useful in pediatric cancers clinical research, when the population size is approximately known.

## P01-11 Outcomes reported in randomized clinical trials of depression in geriatric patients: a methodological review

**Myanca Rodrigues[1], Alessia D'Elia[2], Stephanie Sanger[3], Sameer Parpia[1], Lehana Thabane[1], Zainab Samaan[2]**
1 Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton ON, Canada
2 Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton ON, Canada
3 Health Sciences Library, McMaster University, Hamilton ON, Canada

**Background and statistical challenges:** Major depressive disorder (MDD or depression) is prevalent among older adults aged 65 and older. The effectiveness and safety of interventions used to treat MDD is often assessed through randomized controlled trials (RCTs). However, heterogeneity in the selection, measurement, and reporting of outcomes in RCTs creates challenges for the comparison and interpretation of results, and limits their utility in clinical decision-making. Core outcome sets (COS), developed through systematic scans of the literature, have been proposed as a viable solution to address the heterogeneity of outcome selection in RCTs. A COS represents a minimum set of outcomes that must be measured and reported in trials pertaining to a particular illness. There is presently no COS for use in RCTs that evaluate interventions for geriatric populations with MDD.
**Objectives:** We will conduct a methodological review of the literature for outcomes reported in geriatric depression trials to assess the heterogeneity of outcome measures.
**Methods:** RCTs evaluating pharmacotherapy, psychotherapy, or any other intervention for older adults with depression that have been published in the last 10 years will be located using electronic database searches (MEDLINE, EMBASE, PsycINFO, and CINAHL). Reviewers will conduct title and abstract screening, full-text screening, and data extraction of trials eligible for inclusion independently and in duplicate.
**Analysis:** Outcomes will be synthesized and mapped to a core-outcome domain framework commonly used in biomedical research comprising five areas: physiological/clinical, life impact, resource use, adverse events, and death. We will also summarize characteristics associated with studies (e.g., the number of single-arm, parallel, multi-arm, and crossover trials) and outcomes (e.g., total number of outcomes per trial, number of trials with discernable primary outcomes).
**Expected results:** 'Depression severity' is expected to be the most commonly-used outcome measure in trials of elder adults. We anticipate inconsistency in the definition and measurement of outcomes across RCTs, and few trials which specify a single, discernable primary outcome.
**Conclusions:** The findings from our methodological review will inform the development of a COS for geriatric MDD, with the eventual aim of reducing variability in outcome selection, measurement, and reporting for this clinical population.



# POSTER SESSION 01

## P01-12 One small clinical trial design to provide additional evidence of treatment effects than single-arm trials

**Akimitsu Miyake[1], Tomoharu Sato[2], Tomomi Yamada[1], Eisuke Hida[2]**
1 Department of Medical Innovation, Osaka University Hospital, Japan
2 Graduate School of Medicine, Osaka University, Japan

**Background:** There are some cases where the traditional randomized controlled trial designs are difficult to conduct in small populations such as rare diseases and pediatric diseases area. In such small clinical trials, many single-arm trials to assess within-patient comparisons are conducted due to feasibility. The efficacy of a test drug is evaluated based on a pre-specified threshold from the evidence of natural history or external information. However, in even well-controlled single-arm trials, there are problems of bias in the observed treatment effects. Therefore, the development of new clinical trial design is needed which provides an adequate estimate of the therapeutic effect of the test drug as well as conventional efficacy evaluation based on the threshold.
**Objective:** We propose a new trial design that makes a level of evidence strengthen without increasing the required sample size compared with the single-arm trials.
**Method:** It is a method for estimating the effect size of a test drug in addition to a threshold-based efficacy assessment, using a design similar to the delayed start design (D'Agostino RB. 2009) where subjects are allocated to treatment at different time points (e.g., Group1 (G1): period1=placebo, period2= test drug, Group2 (G2): period1 and 2=test drug). In other words, by integrating the data of period2 in G1 and period1 in G2, we can assess therapeutic efficacy compared with the same threshold as single-arm trials. Additionally, the comparison of the data of period1 in both groups allows us to estimate the effect size of the test drug with certain accuracy. We also take into account period2 in G2 to estimate efficacy more accurately and assume a correlation between period1 and period2 to derive more unbiased and interpretable estimation. For practical application, we perform various simulation scenarios with parameter changed in actual clinical settings.
**Result and Conclusion:** We have confirmed that the assessment based on the estimation derived from integration of period2 in G1 and period1 in G2 achieves the similar power compared with single-arm trials. And the effect size of the test drug can be obtained with certain degree of accuracy through comparing period1 in both groups.
**Reference:** *D'Agostino RB Sr. The delayed-start study design. N Engl J Med. 2009 Sep 24; 361(13):1304-6.*

## P01-13 SteppedPower, an R Package for Power Calculation in Stepped Wedge Cluster Randomised Designs

**Philipp Mildenberger, Jochem König**
Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center Mainz, Germany

Stepped wedge cluster randomised trials (SWCRT) are a versatile alternative to parallel cluster randomised designs. They are increasingly popular in health services research for evaluation of complex interventions. Various approaches to power calculation have been presented. Hussey, Hughes [1] introduced an approach based on generalised least squares, which has since been refined; additionally, design formulae have been proposed [2]. Recently, a maximum-likelihood approach to power calculation for binary outcomes was presented. Power calculation for SWCRT is still a matter of ongoing research, and no ready software is at hand for various situations, designs, and statistical models. We have therefore set up an R package for power calculation that addresses current deficits and offers more flexibility concerning designs, underlying statistical models and methods of analysis. We extend the generalised least squares method first proposed in [1]. The design matrix and covariance matrix are constructed explicitly, which makes this approach very flexible. The use of sparse matrix algorithms and aggregation on cluster level makes computation efficient, thus calculation of (very) large designs becomes feasible. In particular, cluster level aggregation facilitates the calculation for closed cohort designs. For binary and count outcomes to be analysed with generalised mixed models with non-linear link functions, tools for translating between conditional and marginal means and effects are provided. It further offers out-of-the-box tools to visualise cluster- and period importance. Settings with few clusters can lead to anti-conservative analysis; we therefore offer some common design-driven degrees-of-freedom adjustments. We present the features implemented in the SteppedPower package and illustrate some of them with real world examples including cross-sectional and cohort type SWCRTs. The user-friendly software presented fills a gap in power calculation tools for cluster randomised trials. Work is in progress to extend visualisation tools to explore sensitivity to misspecification of the covariance structure.
**References:** *[1] M. A. Hussey and J. P. Hughes. Design and analysis of stepped wedge cluster randomized trials. Contemporary clinical trials, 28(2):182–191, 2007. [2] F. Li, J. P. Hughes, K. Hemming, M. Taljaard, E. R. Melnick, and P. J. Heagerty. Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. Statistical Methods in Medical Research, 2020*

# POSTER SESSION 01

## P01-14 Investigating the operating characteristics of clinical trials with borrowing from external data

**Annette Kopp-Schneider, Silvia Calderazzo, Manuel Wiesenfarth**
Division of Biostatistics, German Cancer Research Center, Germany

In the era of precision medicine, novel designs are developed to deal with flexible clinical trials that incorporate many treatment strategies for multiple diseases in one trial setting. This situation often leads to small sample sizes in disease-treatment combinations and has fostered the discussion about the benefits of borrowing of external or historical information for decision-making in these trials. Several methods have been proposed that dynamically discount the amount of information borrowed from historical data based on the conformity between historical and current data. It is of major importance for regulators and clinicians to investigate the operating characteristics of trial designs that include borrowing from external information.
The objective of the research is to identify correct simulation approaches to investigate the operating characteristics of trial designs that include borrowing from external information and to develop methods for informative display of results from the simulation studies.
We will consider Bayesian phase II clinical trials where efficacy is evaluated on the basis of posterior probabilities. Borrowing from external data will be achieved by various methods, including hierarchical models and robust dynamical priors. Clinical trial operating characteristics will be either simulated by Monte Carlo methods or calculated analytically if this is feasible.
We will illustrate appropriate and inappropriate simulation setups for investigating the operating characteristics of a clinical trial with borrowing from external data, compare the results of the setups and relate the findings to simulation studies that have been published recently. The results from the simulation studies can be used to characterize the properties of various Bayesian borrowing methods.

## P01-15 Introducing GINGER – A General simulation-INterpolation tool for designing multiGroup ExpeRiments

**Lena Jiricka, Georg Heinze**
Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria

In animal experiments it is often investigated whether there is a difference in means of an outcome variable between two different groups. If an experiment is conducted with more than two groups, usually ANOVA is used, in which first a global test of any difference between groups is performed. In a multifactor experiment, one can then test for interactions between effects and for main effects. If interactions exist (which is most often assumed), the ultimate question is if there are differences between the individual experimental groups defined by different combinations of factor levels. Therefore, the experiment should be planned in order to have sufficient power to detect the smallest biologically relevant effect between any two experimental groups, even if more than two groups were included in the experiment.
Existing software for sample size calculation either focuses on power of the global ANOVA test, or on contrasts between groups, but without considering multiplicity corrections (other than simple Bonferroni-type corrections). In analyses, however, typically corrections are needed for all-pairwise comparisons (Tukey HSD correction) or many-to-one comparisons (Dunnett correction), or for a number of specific contrasts between groups.
Therefore, we developed a tool that combines elements of simulation, interpolation and exact computation of sample sizes, and implemented it in GINGER, a web-based application hosted at https://clinicalbiometrics.shinyapps.io/GINGER/. Based on our experience with experimental design of animal trials, GINGER computes sample sizes for typical animal experiments with continuous outcome variables. In contrast to existing calculators, it also considers Tukey or Dunnett corrections. The tool is based on a simulation-interpolation approach. Results are obtained instantaneously as the tool builds on pre-simulated data. We exemplify the use of the tool by means of carrying out several sample size calculations for real experiments. We also compare results with existing software where this is possible. To maintain the tool, we are hosting it in a public repository allowing for full version control and bug reporting. Ideas for further developments include extensions of the range of scenarios covered, further multiplicity corrections and options to simulate repeated experiments which are often carried out to adjust for environmental confounding.



# POSTER SESSION 01

## P01-16 Substitution of study control group by historic controls: Effect on study results using the example pain therapy for endometriosis

**Marius Sieverding[1,2], Christoph Gerlinger[1,3], Christian Seitz[1]**
1 Bayer AG, Germany
2 Berlin School of Public Health, Germany
3 Department of Gynecology, Obstetrics and Reproductive Medicine, University Medical School of Saarland, Homburg/Saar, Germany

The experience of pain is modulated by physiological and psychological factors. Therefore, determining the effect of pain treatment usually requires a sufficiently large control population who receives placebo. Ethical consideration, in contrast, call for having as few patients as possible enduring the trial in the placebo group. Effort has been made for methods to reduce or substitute the control groups. The use of synthetic or historic control arm using data from previous trials or even Real World Data/Real World Evidence gained acceptance by regulation authorities. For pain medication regarding endometriosis, this method is promising but has not been tested so far. For this case study, study data from an already published clinical study [1] on the use of 2mg Dienogest daily to treat endometriosis-associated pelvic pain (EAPP) was used and efficacy was re-evaluated with a historic control arm based study data from an published study [2] using an CCR1 antagonist to treat EAPP. First, the full treatment [1] and historic control [2] groups were compared for several efficacy parameters. In addition, Propensity Score matching (PS) on all baseline variables was used to match between the treatment and historic control arm. To evaluate the effect of matching on PS, the same efficacy parameters were evaluated between matched treatment and control pairs as well. This case study has shown that even for studies which are very similar in design, heterogeneity and between-study variation is present. With the use of a historic control arm, it was possible to reproduce similar results than in the original study, while the PS matching improved the comparability considerably. For the main endpoint (pain measured on Visual Analog Scale), PS matching was able to reproduce the original study results. The method in general has proven to be useful while emphasis has to be given to the appropriate selection mechanism as well as the underlying assumptions.

**References:** *[1] Strowitzki T, Faustmann T, Gerlinger C, Seitz C. Dienogest in the treatment of endometriosis-associated pelvic pain: a 12-week, randomized, double-blind, placebo-controlled study. Eur J Obstet Gynecol Reprod Biol. 2010;151:193. [2] Trummer D, Walzer A, Groettrup-Wolfers E, Schmitz H. Efficacy, safety and tolerability of the CCR1 antagonist BAY 86-5047 for the treatment of endometriosis-associated pelvic pain: a randomized controlled trial. Acta Obstet Gynecol Scand. Juni 2017;96(6):694–701.*

## P01-17 Design optimization and intermediate safety reporting for a randomized controlled biomarker trial

**Andrej Schwabe, Jan Wiemer**
Thermo Fisher Scientific, B·R·A·H·M·S GmbH, Hennigsdorf, Germany

The specification of optimal estimands and estimators is a great biostatistical challenge for every randomized controlled trial (RCT), in particular concerning safety. We illustrate and discuss the specification of safety endpoints, analysis populations and reporting formats for an interventional rule-out biomarker trial when pilot study results are available. This work was performed for the RCT "IDEAL" set up to analyze the effect of a biomarker to assist decisions on hospitalization for a specific population of patients presenting to the emergency department (ED). The rate of hospital admissions was reduced in the treatment arm vs. standard care, 40% vs. 60%. Safety was characterized by mortality, later (possibly delayed) hospital admission and re-presentation to the ED within 28 days. Typically, safety analysis is conducted by quantifying adverse events per study arm and comparing between them. In the present study a biomarker was used to facilitate the stratification of patients into a low severity group with less medical surveillance (patients not admitted to hospital, the rule-out decision) and a higher severity group with high medical surveillance (patients admitted to hospital). Thereby, the study arms were split into two patient subgroups with different risks for adverse events. Accordingly, it seemed advisable to compute adverse event rates specifically for the study arm subgroups of non-hospitalized patients with potentially increased risk. Severe adverse events like death could be considered too rare for rule-out patients to construct a meaningful statistical safety criterion (single-case review, not observed in pilot). For the less severe adverse events ED re-presentation and later hospitalization a good balance had to be achieved concerning (a) larger follow-up times increasing the number of observed events (preferable for higher statistical power) and (b) stronger expected causal relationship between intervention and adverse events (attenuation of treatment effect with time). The gold standard of safety analysis is a confirmatory study proofing prespecified performance, typically non-inferiority. At the stage of a pilot study, we recommend reporting of (a) all analyzed safety endpoints with point estimates and uncertainties, (b) study-arm specifically and for the difference between study arms, (c) uncertainty by intuitively well-interpretable Bayesian credible intervals and (d) verbal statements.

# POSTER SESSION 01

## P01-18 Using routinely collected data to conduct a pragmatic randomized controlled trial: an example addressing antibiotic prescription and resistance monitoring in Swiss primary care

**Florian S. Halbeisen[1,2], Heiner C. Bucher[1,2], Soheila Aghlmandi[1,2]**
1 Basel Institute for Clinical Epidemiology & Biostatistics, University Hospital Basel and University of Basel, Basel, Switzerland
2 Department of Clinical Research, University Hospital Basel, Basel, Switzerland

**Background:** Antibiotic consumption is very high in primary care in Switzerland, and Nesting intervention trials into routinely collected registry data is an innovative approach to addressing clinical problems in need of system-wide interventions such as antibiotic overuse in primary care.
**Objective:** To reduce antibiotic use in primary care by providing personalized antibiotic prescription feedback to individual primary care physicians in Switzerland.
**Methods:** We conducted a nationwide pragmatic randomized intervention trial of routine antibiotic prescription feedback in general practitioners (GP). We used routinely collected individual patient claim data from the three largest health insurers to prepare interventional feedback and to assess endpoints. The target population consists of the top 75% antibiotics prescribers among all GPs who see at least 100 patients a year. The intention-to-treat analysis has been done. The prescription rates are calculated per 100 consultations for each year.
**Results:** We randomized 3,426 Swiss GPs in a 1:1 ratio to intervention and control arms. The 1,713 GPs on the intervention arm only once at the beginning of the trial received evidence-based guidelines for the management of acute respiratory and urinary tract infections. Then they received quarterly personalized antibiotic prescription feedback (see Figure 1). The 1713 GPs in the control group were not actively notified about the study and received no guidelines and no prescription feedback. The two-year intervention phase started in January 2018 and ended on December 31st, 2019. The prescription rates are decreasing gradually per each year of intervention comparing feedback vs. control 7%, and 18%, respectively (See Table 1).
**Discussion:** Our trial demonstrates that using the quarterly feedback the antibiotic prescription rate decreased by 18% in the intervention vs. control group in the second year of the intervention.

| | MEDIAN RATES (Q1 TO Q3) |
|---|---|
| **Baseline year (1st Jan – 31st Dec. 2017)** | |
| Control | 7.98 (7.39 to 8.49) |
| Feedback | 7.93 (7.39 to 8.42) |
| **First-year of intervention (1st Jan – 31st Dec. 2018)** | |
| Control | 7.71 (7.45 to 8.76) |
| Feedback | 7.64 (7.44 to 8.66) |
| **Second-year of intervention (1st Jan – 31st Dec. 2019)** | |
| Control | 7.69 (7.34 to 8.28) |
| Feedback | 7.51 (7.12 to 8.27) |

**Table 1.**
Median antibiotic prescription rates per 100 consultations



# POSTER SESSION 01

## P01-19 Statistical considerations in using a novel consensus building technique to estimate action thresholds in clinical decision making

**Isabel Brosius[1], Alfred Kipyegon Keter[1,2,3], Lutgarde Lynen[1], Bart K.M. Jacobs[1]**
1 Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium
2 Centre for Community Based Research, Human Sciences Research Council, Pretoria, South Africa
3 Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium

**Background:** Policy makers and health care workers are often confronted with uncertainty, both in the development of guidelines and patient management. In clinical decision making, such decisions are linked to an action threshold, e.g. the therapeutic threshold in case of treatment decisions. Because of common contention among stakeholders about the associated benefits and harms and the difficulty in weighing the harms, a predetermined agreed-upon threshold is important. Existing methods to estimate thresholds produce inconsistent results and often fail to account for harms that are hard to quantify and for the inherent variation among stakeholders.
**Methods:** We are piloting an adapted version of a formal consensus method in different clinical decision-making settings: switching to second-line for patients with presumptive HIV treatment failure, initiating treatment for tuberculosis in patients with presumptive tuberculosis, and enforcing self-isolation for presumptive SARS-CoV-2 infection. Experts and stakeholders are invited to formulate and reflect on the potential harms of wrong decisions: either taking unneeded action (false positive) or refraining from required action (false negative). The panel rates the extent to which each of these harms should be taken into account on a modified Likert scale, a process that is repeated after discussion. In the final step, each of the harms are weighed against each other. The action threshold is estimated as the probability of disease or infection at which the expected harms associated with false positives are equal to the expected harms of false negatives.
**Results and Discussion:** Respondents agreed more on the ratings after panel discussion and results were similar between different expert panels presented with the same statements. While estimated action thresholds were also similar between panels, it is uncertain if individual agreement is consolidated in the weighing phase given considerable variation remains between experts within the same group. Design choices, such as the number of harm-describing statements, the sequence of discussion thereof and the time spent on each statement may be sources of bias. Given the absence of a gold standard, validation of the estimated thresholds is impossible. Considerable challenges in the data-analysis and statistical inference process of this novel technique remain.

## P01-20 Flexible software framework to compare Bayesian hierarchical models across basket trial designs

**Stephan Wojciekowski**
Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach Riss, Germany

**Context:** Master protocols such as basket trials become increasingly important before moving to late phase pivotal trials. Commonly applied models for this purpose are Bayesian hierarchical models (BHM), which evaluate a trial's outcome factoring in the similarity of the strata's results by dynamically borrowing information across the strata. The decision to move to a pivotal trial can be formalized with a go / no-go decision-making framework, which enables the calculation of operating characteristics of a trial design. Often, the path to this decision is lined with one or several interim analyses.
**Objective:** The objective is a framework and its software implementation that allow to compare the operating characteristics of basket trial designs across several BHMs with binary endpoints.
**Methods:** An R package has been built that provides functions for the simulation, analysis and evaluation of basket trials with binary endpoints. The BHMs proposed by Berry et al. (2013) and Neuenschwander et al. (2016), as well as a modified BHM that combines both approaches, are implemented in JAGS. The runtime of the simulations has been optimized by applying the BHMs to unique trial realizations across scenarios, parallelization and storage of interim results.
**Results:** The implemented framework allows for an arbitrary number of strata, number of interim analyses, and staggered recruitment. The functions for trial evaluation enable highly customizable go / no-go decision rules for each decision point. This allows to assess the decision probabilities, biases and mean squared errors for very flexible trial designs, as well as the analysis of a such a trial's outcome. The implementation runs comparatively fast due to the performance optimizations and is available on CRAN.
**Conclusions:** The resulting R package "bhmbasket" provides a framework that facilitates the design selection for basket trials and enables the comparison of operating characteristics of different designs and BHMs.

# POSTER SESSION 01

## P01-21 Assessment of Clinical Trial Missing Data During a Pandemic: A Tipping Point Analysis Case Study

**Nicole C. Close**
Biometrics Department, EmpiriStat Inc., Kitty Hawk NC, United States

In December 2019 the COVID-19 outbreak emerged in China and quickly spread becoming a declared global pandemic by the World Health Organization (WHO) in March 2020. For over a year quarantines, travel restrictions, interruptions to supply chains, social distancing, face coverings, and site restrictions have led to many difficulties in adhering to and completing study protocols. While some clinical trials have been halted or suspended, others have had to implement different mitigation strategies to assure the safety of participants and continued to collect data. The effects of these difficulties, even with mitigation strategies, have led to an increase in the amount of missing clinical trial data.
Missing data may have an influential impact on the primary analysis. This case example explores a device study that was mid-way in data collection during the declaration of the pandemic, that implemented mitigation strategies for data collection at home by the participants without visiting the clinic, and other measures. Due to missing data, several sensitivity analyses needed to be considered. Ultimately a tipping point analysis was used to assess how severe a departure from the missing data assumptions by the main estimator must be to change the conclusions of the primary analysis was conducted. Tipping point adjustments were made between the pandemic and non-pandemic missingness, by type of data collection (in clinic versus by participant) and reason of missingness. Altering the assumed probability that the participant would have adhered to treatment through the end of the study versus that they would have had other non-pandemic related events were considered and their outcomes imputed accordingly. Each tipping point analysis with the data will be presented to illustrate the methodology for each set of assumptions.



# POSTER SESSION 02



## P02-01 A comparison of multiple imputation strategies for handling missing data in repeatedly measured multi-item scales

**Rheanna Mainzer[1], Jemishabye Apajee[1,2], Cattram D. Nguyen[1,3], John B. Carlin[1,4], Katherine J. Lee[1,3]**
1 Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Parkville VIC, Australia
2 Quality Use of Medicines and Pharmacy Research Centre, University of South Australia, Adelaide SA, Australia
3 Department of Paediatrics, The University of Melbourne, Parkville VIC, Australia
4 Centre for Epidemiology and Biostatistics, The University of Melbourne, Parkville VIC, Australia

Medical research often involves using multi-item scales to assess individual characteristics, disease severity and other health-related outcomes. It is common to observe missing data in the scale scores, due to missing data in one or more items that make up that score. Multiple imputation (MI) is a popular method for handling missing data. However, it is not clear how best to use MI in the context of scale scores, particularly when they are assessed at multiple waves of data collection resulting in large numbers of items. The aim of this research is to provide practical advice on how to impute missing values in a repeatedly measured multi-item scale using MI when inference on the scale score is of interest. We evaluated the performance of five MI strategies for imputing missing data at either the item or scale level using simulated data and a case study based on four waves of the Longitudinal Study of Australian Children. MI was implemented using both multivariate normal imputation and fully conditional specification, with two rules for calculating the scale score. A complete case analysis was also performed for comparison. Our results suggest that imputing at the item-level across all waves when you have many items measured at each wave can lead to biased inference even when computationally feasible.

## P02-02 A Hidden Markov model segmentation to identify trajectories in sleep apnoea patients

**Alphanie Midelet[1,2], Sébastien Bailly[1,3], Ronan Le Hy[2], Marie-Caroline Schaeffer[2], Jean-Louis Pépin[1,3]**
1 Grenoble Alpes University, Inserm, CHU Grenoble Alpes, HP2, France
2 Probayes, Montbonnot-Saint-Martin, France
3 EFCR Laboratory, Grenoble Alpes University Hospital, France
4 AGIR à dom. Homecare charity, Meylan, France

**Objective:** Obstructive sleep apnoea (OSA) is affecting nearly one billion people worldwide. Continuous positive airway pressure (CPAP), the reference treatment, is nightly used by million individuals globally with remote telemonitoring providing daily information on CPAP usage and efficacy. The knowledge gain by this avalanche of related data is hampered by the lack of relevant data mining. We aimed to implement state of the art data science methods for describing heterogeneity and diversity of CPAP telemonitoring time series of residual apnoea + hypopnoea index (rAHI).
**Methods:** We analysed a CPAP telemonitoring database of 2,860 CPAP-treated OSA patients to model and cluster rAHI trajectories. Our primary objective was to use Hidden Markov models (HMMs) as a probabilistic model-based approach to extract unrevealed features from rAHI time-series. Our secondary goals were to identify clusters of rAHI trajectories and their relation to CPAP treatment outcomes, namely adherence and leaks.
**Results:** From the telemonitoring records of 2,860 CPAP-treated patients (age: 66·31 ± 12·92 years, male gender 69·9%), HMM modelling revealed three distinct states differing in two complementary domains: variability inside a given state and probability for shifting from one state to another. Six clusters of rAHI telemonitoring trajectories were identified from very well controlled CPAP-treated patients (Cluster 0: 669 (23%); mean rAHI of 0·58 ± 0·59 events per hour) to the most unstable cluster (Cluster 5: 470 (16%); mean rAHI of 9·62 ± 5·62 events per hour). CPAP adherence was half an hour significantly higher in cluster 0 compared to clusters 4 and 5 (p-value<0·01). Leaks were also significantly higher in cluster 5.
**Interpretation:** We propose a new analysis based on Hidden Markov Models and supported by machine learning that might constitute a backbone for deployment and dissemination of digital health solutions for improving interpretation of telemonitoring of CPAP-treated patients. This method allows to visualise and reveal novel interesting features through the graph of the HMM states for guiding tailored CPAP follow-up management.

# POSTER SESSION 02

## P02-03 A comparison of statistical methods to compensate for missing data in longitudinal cluster-randomised trials

**Courtney McDermott[1]*, Barbara Dooley[2], Mary Codd[1], Ricardo Segurado[1]***
1 School of Public Health, Physiotherapy, and Sports Science, University College Dublin, Ireland
2 School of Psychology, University College Dublin, Ireland

Missing data in clinical trials can introduce bias and reduce power in analyses, potentially causing researchers to miss important intervention effects. This is a particular concern for longitudinal cluster-randomised controlled trials (LCRCT), as missing data may cumulate longitudinally in addition to occurring at the level of individual observations, individual trial participants, or clusters of participants. Both multiple imputation (MI) and full information maximum likelihood (FIML) estimations are expected to generate similar results with appropriately adjusted standard errors when compensating for missing data in LCRCTs. Despite the advantages and robustness of these methods, little research has directly compared the performance of these two methods in compensating for missing outcome data in a three-level structure. The aim of this research was to compare the performance of these two techniques in compensating for missing outcome values in both real and simulated LCRCT data.

Simulated datasets were modelled off the CopeSmart Trial, a LCRCT that aimed to increase emotional self-awareness of 560 Irish secondary school students from ten schools over an eight-week period. Missingness was introduced at varying proportions using three mechanisms: missing completely at random (MCAR); missing dependent on covariate x (MAR); and missing not at random (MNAR). The data were then analysed using the two aforementioned techniques and the bias, power, and coverage were averaged over 1000 simulates. A cluster-level covariate and individual-level covariate were also simulated to be correlated with the outcome variable. These were separately added to the MI and FIML models to determine if the addition of correlated covariates would reduce bias when data were MNAR. When the data were MCAR or MAR, both methods produced similar, unbiased estimates of the intervention effect. Both methods were biased when the data were MNAR, though MI had reduced power compared to FIML. When adding the covariates, both methods improved in terms of bias and coverage. However, the FIML method appeared to perform better consistently across all scenarios. When applying these methods to the CopeSmart data, the FIML was again superior. We recommend researchers employ FIML when compensating for missing outcome data in LCRCTs.

***Acknowledgements:** *This research is funded by the HRB-TRMN through a Study Within a Trial (SWAT) Grant (R17496).*

## P02-04 Effect of impaired vision on physical activity from childhood to adolescence

**Lisanne A. Horvat-Gitsels, Mario Cortina-Borja, Ameenat O. Solebo, Jugnoo S. Rahi**
Population, Policy and Practice Research and Teaching Department, UCL Great Ormond Street Child Health Institute, London, United Kingdom

**Background:** International physical activity (PA) guidelines are set irrespective of disabilities. Yet the levels of and changes in PA across transition from childhood into adolescence among those with impaired vision are not well understood due to the challenges of longitudinal population-based studies of rare conditions.

**Objective:** To determine whether children and adolescents with impaired vision can achieve PA levels equivalent to those without impaired vision.

**Methods:** Data from the Millennium Cohort Study of children born in the United Kingdom in 2000-01 and followed-up to age 14 years (n=11,571). Participants were grouped by eye conditions causing no, unilateral, or bilateral impaired vision based on parental self-report on vision and treatment coded by clinical reviewers. There were 16 types of PA reported by parents, teachers, and/or participants, which covered physical education (PE) at school, organised sports, self-organised sports, and hobbies. Age-related trends in reported PA types were modelled using ordinal and logistic regression. Objective accelerometer-based time spent in moderate-to-vigorous PA (MVPA) were modelled by quantile regression to assess differences by vision status and reported PA types.

**Results:** Bilateral impaired vision was associated with having difficulties with PE (aOR=4.67, 95% CI 2.31-9.41), finding oneself "to be bad at PE" (3.21, 1.44-7.15), and enjoying indoor PA (2.08, 1.14-3.85). Unilateral impaired vision was associated with having difficulties with PE (1.80, 1.26-2.59), below-average abilities in PE (2.27, 1.57-3.28), and high level of participation in organised sports (0.45, 0.56-0.98). Age-related trajectories in PA levels and time spent in MVPA did not differ by impaired vision. The internationally recommended level of ≥60 MVPA min/day was achieved by 50% of those aged seven and 41% of those aged 14. The main contributing factors were the levels of PE and organised sports.

**Conclusions:** Our findings show that children with impaired vision can achieve healthy levels of PA equivalent to those without impaired vision, although those who had suboptimal levels of PA continued on that level into adolescence. Population-wide public health programmes to increase PA levels are needed, as well as interventions specific to children and adolescents with impaired vision to encourage participation in PE and organised sports.



POSTER SESSION 02

## P02-05 Predictors of Multidrug Resistance in Nosocomial Pneumonia among Intensive Care Units' Patients of a Tertiary Hospital, Egypt

**Mina Yakoub[1], Fayek Elkhwsky[1], Iman El Sayed[1], Ayman El Tayar[2]**
1 Department of Biomedical Informatics and Medical Statistics, Medical Research Institute, Alexandria University, Egypt
2 Damanhour Medical National Institute, General Organization of Teaching Hospitals and Institutes, Egypt

**Background:** Ventilated-Acquired Pneumonia (VAP) and Nonventilated Hospital Acquired pneumonia (NV-HAP) remain critical public health problems. Increasing antimicrobial resistance has exaggerated the management and costs of ICU patients.

**Objectives:** To identify the predictors of Multidrug Resistance (MDR) in Nosocomial Pneumonia (NP) among ICU patients and report the microbiological profile of NP.

**Subjects and Methods:** A prospective longitudinal study was performed at a tertiary hospital's general ICUs from 2018-2019. We included adult patients admitted for at least 72 hours before signs appear. We studied the causative organisms, antibiotic susceptibility, and resistance patterns of these patients. We utilized the Relative Risk (RR) binomial model to determine the predictors of MDR. Statistical analyses were conducted using SPSS® version 23 and R® software.

**Results:** The incidence rate of MDR was 1.48 per 100 person-days (95% CI 1.21 – 1.78 per 100 person-days). Although Acinetobacter baumannii (21.05%), Klebsiella pneumoniae (40.60%), Pseudomonas aeruginosa (18.80%) were the most isolated bacteria among VAP patients, only Klebsiella pneumoniae (42.11%) and Pseudomonas aeruginosa (23.68%) were the most predominant organisms detached from NV-HAP. In gram-positive bacteria, overall antibiotic nonsusceptibility was high for cefuroxime, cefoxitin, as well as cefotaxime 100%. The highest antibiotic susceptibility of gram-positive bacteria was detected for vancomycin 93.75%. Among gram-negative bacteria, the antibiotic nonsusceptibility was maximum for cefuroxime 100%. Maximum antibiotic susceptibility of gram-negative bacteria was observed for colistin 47.46%. The relative risk model clarified that the most independent predictors for MDR were A. baumannii (RR= 3.71, P=0.014), K. pneumoniae (RR= 2.78, P<0.001), P. aeruginosa (RR=2.66, P=0.002), and duration in ICU (RR=0.99, P=0.001). Nevertheless, the relative risk of the ICU duration was not expected as the length of stay of patients in ICU did not affect the multidrug resistance.

**Conclusion:** Predictors of MDR were A. baumannii, K. pneumoniae, and P. aeruginosa, however duration in ICU did not alter MDR. Nonsusceptibility of gram-positive and gram-negative bacteria was high for commonly used antibiotics. Susceptibility in gram-positive bacteria was high for vancomycin, while high susceptibility of gram-negative bacteria was for colistin.

# POSTER SESSION 02

## P02-06 Modelling of longitudinal data to predict cardiovascular disease risk: a methodological review

**David Stevens[1,2], Ruwanthi Kolamunnage-Dona[3], Stephanie Harrison[1,2], Deirdre Lane[2], Gregory Lip[2]**

1 Liverpool Centre for Cardiovascular Science, University of Liverpool, UK
2 Cardiovascular and Metabolic Medicine, Institute of Life Course and Medical Sciences, University of Liverpool, UK
3 Department of Health Data Science, Institute of Population Health, University of Liverpool, UK

**Background:** Risk prediction informs understanding and management of cardiovascular disease (CVD). Many CVD risk prediction models only use one data point per patient, which does not account for temporal change in cardiovascular risk factors. Longitudinal data permits study of change in risk factors over time, within person-variance, and risk prediction based on risk factor trajectories. Analysis of longitudinal data adds complexity, such as dealing with dependence between individual observations and missing or incomplete data. The aim of this review was to identify methods used in modelling longitudinal data focusing on trajectories of risk factors to predict CVD risk.
**Methods:** The Medical Literature Analysis and Retrieval System Online (MEDLINE - Ovid) was searched from inception until 3rd June 2020 for studies meeting pre-specified inclusion and exclusion criteria: adults only; outcome including risk of CVD; and longitudinal design with ≥3 time-points. Reviews or those without individual patient data were excluded. Search terms related to "longitudinal" and "CVD" were used, including multilevel, change, slope, heart disease, myocardial infarction, and stroke. One author screened search outputs and extracted data; other authors discussed and resolved queries. Current statistical methods, assumptions, flexibility, and availability of the software were compared.
**Results:** Searches returned 2601 studies; 82 studies were included. Studies were divided into three approaches for modelling CVD risk using longitudinal data: single-stage models including basic summary measures, two-stage models using a summary measure or estimated longitudinal parameter as a covariate in a survival model, and joint models fitting longitudinal and survival data simultaneously. Single-stage models were used in 41 (50%), 30 (36%) used two-stage models, 8 (10%) used joint models, and 3 (4%) used simple statistical tests only. Over time, use of two-stage models and joint models were more prevalent, with an increase in CVD risk prediction models created using longitudinal data. Most studies used SAS, R or Stata. Joint and two-stage models allow for greater flexibility than single-stage models, although software for joint models is often more restricted.
**Conclusions:** Although the use of two-stage and joint models is becoming more prevalent, many studies still underutilise their longitudinal data when modelling cardiovascular risk.

## P02-07 Impact of model misspecification on model-based bioequivalence

**Mélanie Guhl[1], François Mercier[2], Satish Sharan[3], Kairui Feng[3], Guyoing Sun[4], Wanjie Sun[4], Stella Grosser[4], Liang Zhao[3], Lanyan Fang[3], France Mentré[1], Emmanuelle Comets[1], Julie Bertrand[1]**

1 University of Paris, INSERM, IAME, UMR 1137, France
2 Department of Biostatistics, Roche Innovation Center Basel, Basel, Switzerland
3 Division of Quantitative Methods and Modeling, Office of Research Standards, Office of Generic Drugs, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring MD, United States
4 Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring MD, United States

To assess the bioequivalence (BE) of a generic, we compare its pharmacokinetic (PK) to a reference drug, through two parameters of interest: the area under the curve of plasma concentration as a function of time, and the maximal concentration. When the conventional non-compartment analysis (NCA is not feasible due to sparse sampling, an alternative to compute these parameters is the model-based (MBBE) approach. Both methods can be similarly applied to compare two drug formulations. We compared NCA and MBBE approaches to determine possible impact on the BE determination using data from a biosimilarity study of a monoclonal antibody developed at Roche, as an example. For MBBE, the structural model was selected on real data on Bayesian Information Criterion, on both the original and sparsified data with less sampling points. This analysis inspired a simulation study with rich and sparse PK designs. On rich designs, we compared NCA and MBBE in terms of type I error, also exploring the impact of the treatment effect model used. On sparse designs, we investigated the impact of using a misspecified structural model and the relevance of a model selection step prior to the BE analysis. Both approaches were concordant on the real data. A two-compartment model with treatment effects on all PK parameters was found to best fit the data. BE of the formulations could not be shown. The design did not impact these results. On simulations, the type I error of the NCA and MBBE approaches were similar and close to the nominal level. The MBBE approach maintained a controlled type I error except when the structural or treatment effect model was misspecified. The step of selection of the structural model enabled to achieve a controlled type I error. MBBE was a robust alternative to NCA. For the first time, a simulation study shows how model selection is key to maintaining an appropriate type I error for BE testing.



# POSTER SESSION 02

## P02-08 Cluster randomised controlled trial of lifestyle intervention for adolescents' health using 'SPRAT' programme

**Kazue Yamaoka[1,2], Junko Watanabe[3], Mariko Watanabe[4,5], Misa Adachi[1,6], Asuka Nemoto[1], Toshiro Tango[1,7]**

1 Teikyo University Graduate School of Public Health, Tokyo, Japan
2 Tetsuyu Crinical Research Center, Tokyo, Japan
3 The Department of Nutrition Management, Minami Kyushu University, Miyazaki, Japan
4 Showa Women's University, Tokyo, Japan
5 Prefectural University of Kumamoto, Kumamoto, Japan
6 Nutrition Support Network LLC, Sagamihara, Japan
7 Center for Medical Statistics, Tokyo, Japan

**Background:** Lifestyle modifications to reduce subjective psychosomatic symptoms (SPS) is an important topic worldwide. We developed a school-based lifestyle education programme involving parents to reduce SPS in adolescents (SPRAT). That approach aimed to reinforce the role of parent participation in adolescents' healthy lifestyle modifications to reduce SPS and increase enjoyment of school life.
**Objectives:** This study aimed to evaluate the effectiveness of SPRAT in reducing SPS among adolescents.
**Design:** Cluster randomised clinical trial with two intervention arms.
**Setting:** Voluntary middle schools in Japan.
**Participants:** Middle school students with their informed consents.
**Interventions:** SPRAT with 6 months' intervention and usual school program (control).
Primary and secondary outcome measures: The SPS score was assessed at baseline and 2, 4, and 6 months thereafter. Proportions of lifestyle factors achieved such as enjoyment of school life were the secondary outcomes. Change from baseline (CFB) at 6 months was the primary endpoint.
**Results:** The participants used on the intention-to treat analysis were 951 (90.2%) and 1035 (84.6%) in the SPRAT and control groups, respectively. The CFB in the 6-month SPS score adjusted for baseline was lower in the SPRAT group compared to the control group but was not statistically significant -0.95 (p=0.081). Good effect was observed in CHB at 4-month (-1.60, 95%CI: -2.87 to -0.33). Improve of energy intakes in breakfast and lunch, and improved of lifestyle factors (enjoying school life, staple food consumed per breakfast, and main dishes consumed per lunch) were also observed.
**Conclusion:** Although the results for primary outcome were not significant, good effects were observed in some secondary outcomes. These findings will contribute to meeting the critical need to develop effective and practical measures to minimise SPS, and its potential influence among adolescents.
**Trial registration number** UMIN000026715.
**Reference:***[1] Watanabe J, et al. BMJ Open 2018;8:2:e018938.*

# POSTER SESSION 02

## P02-09 Assessing the role of hyperventilation in patients with traumatic brain injury: longitudinal data analysis from the CENTER-TBI

**Matteo Petrosino[1], Paola Rebora[1], Stefania Galimberti[1], Giuseppe Citerio[2], Maria Grazia Valsecchi[1]**

1 Bicocca Bioinformatics Biostatistics and Bioimaging Center B4, School of Medicine and Surgery, Monza, Italy

2 Neurointensive Care Unit, San Gerardo Hospital, School of Medicine and Surgery, Monza, Italy

In mechanical ventilated patients with traumatic brain injury (TBI) admitted in intensive care units (ICU), the management of bloody partial pressure of carbon dioxide ($paCO_2$) is controversial: hyperventilation leads to vasoconstriction, lowering ICP, but also might lead to an increased risk of cerebral ischemia. Guidelines suggest that the optimal $paCO_2$ value lies between 35-45 mmHg. Nonetheless, ICU operators manipulate these values in order to balance for intracranial pressure (ICP), as values exceeding 20 mmHg are potentially dangerous. We used data from the CENTER-TBI study, a worldwide longitudinal prospective collection of TBI patient data, 1) to describe the management across centers of paCO2 in ICU patients with TBI and 2) to assess the impact of hyperventilation to 6-months outcome (Extended Glasgow Outcome Scale, GOSE). To model the clustered longitudinal $paCO_2$ profiles, a linear mixed effects model was used, while to quantify heterogeneity in $paCO_2$ management through the median odds ratio (MOR), a logistic mixed effects model was performed on daily hyperventilation ($paCO_2$<30 mmHg). In both models, the fixed effects were deployed for baseline covariates and daily maximum ICP value, while a two-level hierarchical structure, i.e. patients nested within centers, was used for the random intercept effects. Finally, a logistic model on GOSE was adjusted by the standard trauma-related covariates at baseline and by two covariates representing a summary of the $paCO_2$ and ICP longitudinal profiles (i.e. the area under the ICP trajectory in time from 20 mmHg and the area between 30 mmHg and the $paCO_2$ trajectory in time). We evaluated data of 1100 mechanically ventilated TBI patients collected between December 2014 and December 2017 in 36 centers. A large variability in the ORs on hyperventilation usage was shown across centers (MOR=2.04) and further investigations should clarify the nature of this unexplained heterogeneity. Moreover, $PaCO_2$ showed an important impact on the 6-months neurological outcome. We will show the clinical findings of the study and we will discuss the methodological challenges we encountered that are mainly due to the hierarchical structure of data and to the relationship between $paCO_2$ and ICP.

## P02-10 A new approach to measure frailty in the context of COVID-19 population

**Francesca Graziano[1], Stefania Galimberti[1], Paola Rebora[1], Paolo Mazzola[2,3], Giuseppe Bellelli[2,3], Maria Grazia Valsecchi[1]**

1 Bicocca Bioinformatics Biostatistics and Bioimaging B4 Center, University of Milano - Bicocca, Monza, Italy

2 School of Medicine and Surgery, University of Milano - Bicocca, Milan, Italy

3 Acute Geriatric Unit, San Gerardo Hospital, Monza, Italy

Frailty is a clinical syndrome, resulting from interaction of the age-related decline in physiologic systems with chronic diseases. It is a multidimensional biological and physiological phenomenon able to capture the "chronological age-independent" health status. During the coronavirus (COVID-19) pandemic, it became clear that frail subjects are those at highest risk of mortality. Some tools to measure individual's frailty have been proposed and the most widespread is the cumulative frailty index (FI) defined as proportion of deficits in a large number of domains.

The aim of this study was to develop a new frailty score through the latent variable approach in a sample of patients diagnosed with COVID-19. We explored the quantification of the frailty status through structural equation modeling based on the data of consecutive COVID-19 patients admitted at San Gerardo hospital in Monza, Italy (n=448) from February until December 2020. A total of 41-items assessing the health status (e.g., comorbidities, functional abilities, habits, and laboratory tests) at admission were considered to create the underlying latent frailty phenotype. Confirmatory Factor Analysis (CFA) models with different structures were explored and the diagonally weighted least squares (WLSMV) estimator was used to estimate the parameters. Models were evaluated with several goodness-of-fit and summary statistics. We carried out a correlated CFA model with one latent factor and 41-items as indicators and a nested CFA model with one general factor (frailty) that accounts for the communality of the items, and other accounting for the influence of the specific items (functional abilities). Both models showed good fit and adequate summary indices (Comparative Fit Index, CFI, and Tucker Lewis index, TIL, more than 0.95, Root Mean Square Error of Approximation, RMSEA, less than 0.05) and suggested that frailty score was strongly represented by a cluster of functional abilities. Results confirmed the presence of a latent variable underlying different organ involvements. Our approach allowed to identify different clusters of items and variables strongly associated with the frailty status by a weighted approach. The prognostic ability of the proposed score on hospital mortality will be compared with standard tools, such as the cumulative FI and Clinical Frailty Scale.



# POSTER SESSION 02

## P02-11 Pain Management in Immediate Life Support Ambulances

**Mauro Mota[1], Carla Henriques[2], Ana Cristina Matos[3]**

1 Institute of Biomedical Sciences Abel Salazar, University of Porto, Portugal

2 CMUC, Coimbra University, and Polytechnic of Viseu, Portugal

3 Research Centre in Digital Services, and Polytechnic of Viseu, Portugal

This study evaluates different approaches in the pre-hospital treatment of pain in trauma patients. Data were collected in immediate life support ambulances (ASIV), in mainland Portugal and the Azores, from March 1, 2019 to April 30, 2020. Pain management may include pharmacological measures, which have the disadvantage of possible negative effects, and non-pharmacological ones, which have also been shown to be effective in reducing pain (e.g. Pierik et al., 2015). Several non-pharmacological measures were studied: relationship-based measures (therapeutic touch, active listening, hand holding and therapeutic presence without the use of touch); cryotherapy; heat application; distraction; immobilization; extremity elevation; presence of family and friends; comfort measures (comfortable position). Pain was assessed in three moments, before (T1), during (T2) and after (T3) nurses' interventions, using an 11-point Numeric Rating Scale (NRS) validated by Bijur et al. (2003). The effect of pain management measures was first studied through the change in the patients' level of pain in the first and third moments of pain assessment (Δ_PAIN). Furthermore, linear mixed-effects models with random intercepts were used to account for the repeated measurements in pane level of the same individual. Models were adjusted for patient-related variables including age, gender, anatomical location of trauma and type of injury. Of the pharmacological measures, only morphine was shown to have a significant effect on the decrease of pain intensity between the first moment, T1, and the third, T3. Two non-pharmacological measures, relationship measures and cryotherapy, were also significantly associated with pain reduction, even after adjusting for morphine, which supports the effectiveness of these non-pharmacological measures as reliable alternatives to analgesics.

**References:** *Bijur, P. E., Latimer, C. T., & Gallagher, E. J. (2003). Validation of a verbally administered numerical rating scale of acute pain for use in the emergency department. Academic emergency medicine, 10(4), 390-392., Pierik, J. G., IJzerman, M. J., Gaakeer, M. I., Berben, S. A., van Eenennaam, F. L., van Vugt, A. B., & Doggen, C. J. (2015). Pain management in the emergency chain: the use and effectiveness of pain management in patients with acute musculoskeletal pain. Pain medicine, 16(5), 970-984.*

## P02-12 Statistical methods for estimating sources of variability in count-based biomarkers

**Kostas Tryposkiadis[1,2], Alice Sitch[1,2], Malcolm Price[1,2], Jon Deeks[1,2]**

1 Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham, United Kingdom

2 NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, United Kingdom

**Background:** Analysis using random effects linear models is the established method used in biological variability studies to attribute the observed variability arising from between-patient differences, within-patient differences, and measurement error. However, these models assume underlying normality, and thus may not be applicable for biomarkers based on counts.

**Aim:** To present methods for estimating sources of variability in count-based biomarkers and apply and compare the approaches in a case study of patients with Sjogren's syndrome.

**Methods:** Both Poisson and negative binomial models are appropriate for analysis of count data, and methods for obtaining between and within-patient variance estimates are described in Leckie et al [1]. We analysed the biomarker data using random effects Poisson and negative binomial models, and for comparison, using a random effects linear regression model. The intra class correlation (ICC) was calculated as a ratio of the between-patient variance over the total variance, and was compared across the different models. The AIC and BIC criteria were used to assess each model's performance. Data from 32 patients with Sjogren's syndrome was used as a case study, considering the focus score, calculated for each salivary gland observed in each biopsy as the number of foci over the glandular area, multiplied by 4. Between-patient and between-gland within-patient sources of variability were estimated.

**Results:** The ICC estimates obtained from Poisson (0.323) and negative binomial models (0.310) were similar, and higher than the linear regression model (0.222). AIC and BIC values were similar for Poisson (AIC=463.63, BIC=469.84) and negative binomial models (AIC=465.55, BIC=474.87) and indicated both were a better fit than the linear regression model (AIC=632.69, BIC=642.01).

**Conclusion:** It is important to properly model the distribution of biomarkers based on count data to correctly estimate sources of variability and measurement error.

**References:** *[1] Leckie et al, 2019. arXiv: 1911.06888 [stat ME].*

# POSTER SESSION 02

## P02-13 Longitudinal progression of frailty in older population and risk of adverse events: An application of joint models

**Leena Elhussein, Danielle E. Robinson, Antonella Delmestri, Daniel Prieto-Alhambra, Alan Silman*, Victoria Y. Strauss***

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Windmill Road, University of Oxford, UK

* Joint last authorship

**Context:** A validated electronic frailty index (eFI) based on cumulative deficits in the electronic primary care record has been implemented in UK primary care to identify older people living with frailty, to better target clinical care [1]. Assessing the change of eFI over time could enhance understanding of eFI progression and variation, and the nature of its association with the risk of adverse events.

**Objective:** To evaluate the application of joint models to frailty progression and risk of death or unplanned hospital admission in elderly people using routinely collected data.

**Methods:** Patients identified from the UK Clinical Practice Research Datalink (CPRD GOLD) linked to Hospital Episode Statistics and Office for National Statistics mortality register aged over 65 at January 1st 2009 were included. eFI scores were calculated annually to assess progression of frailty for up to 5 years. Outcomes of mortality and unplanned hospital admission were assessed during the same time window. Joint modelling combines time and event into one model. Joint models were fit to assess the risk of all-cause mortality and unplanned hospital admission, separately. Both models were adjusted for baseline age and gender in their longitudinal and survival parts.

**Results:** A total 475,698 were included in this study. Mean baseline age was 75.1 (SD 7.4) and 211,849 (44.5%) were male. 244,240 had unplanned hospital admissions and 54,107 died during follow-up. eFI score showed a deterioration in 155,732 (63.8%) of those hospitalised and 43,484 (80.4%) of those who died. eFI slope over time was 0.41 (95%CI 0.41, 0.41) for hospitalisation and 0.42 (0.42, 0.42) for death. eFI progression was associated with a higher risk of death and hospitalisation; hazard ratio of the association term (eFI) was 1.12 (1.12, 1.12) for hospitalisation and 1.13 (95%CI 1.13, 1.14) for death.

**Conclusion:** This example demonstrates the applicability of joint models in this area and the potential value of using frailty trajectories over a cross-sectional assessment. It can be used to inform health practitioners on the risk of adverse outcomes for frail patients based on their eFI progression. Thus, they will be able to target the limited resources towards the most vulnerable patients.

**References:** *[1] Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age and Ageing 2016;45(3):353-60. doi: 10.1093/ageing/afw039*

# POSTER SESSION 02

## P02-14 Semi-variogram approach to estimate within-subject variability in repeated measurements

**Simon Baldwin[1,2], Alice Sitch[1,2], Yemisi Takwoingi[1,2], Jonathan J. Deeks[1,2]**

1 Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham, UK

2 National Institute for Health Research Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, UK

**Context:** Components of variability in biomarker measurements are best estimated in Biological Variability Studies (BVS). Longitudinal BVS aim to estimate components of variability in repeated biomarker measurements, ideally taken on all subjects at the same time-points. However, conducting prospective longitudinal BVS is not always feasible. We are investigating how measurement error and true change components of within-subject variability can be estimated retrospectively using pre-existing longitudinal data, such as in clinical trial and routine datasets. Understanding these components of variability is critical in evaluating properties of biomarkers and monitoring programmes.

**Objective:** To demonstrate a semi-variogram approach based on the work of Peter J Diggle et al [1], to estimate components of within-subject variability in repeated systolic blood pressure measurements on children in the UK monitored every 6 months over 42 months. Subjects were treated for an initial episode of nephrotic syndrome as part of the PREDNOS RCT [2].

**Methods:** Variability in measurements on subjects over time was attributed to three components: differences in mean measurements between-subjects; differences in change-from-mean measurements between-time-points within-subjects, referred to as the 'signal'; and differences in measurement errors within-time-points within-subjects, referred to as the 'noise'. The semi-variogram approach considers differences in measurements between-time-points within-subjects; the semi-variances of these differences increase towards a ceiling comprised of the combination of 'signal' plus 'noise'. The semi-variogram is therefore a plot of semi-variance of differences on the y-axis, against time separation on the x-axis. Extrapolation of the intercept is by a regression model, which estimates the 'noise'. The 'signal' can be estimated from the modelled estimate at the ceiling, minus the 'noise'.

**Results:** In 159 subjects, the 'noise' estimate accounted for 81.0% (95% CI: 53.6%, 94.0%) of the total within-subject variability. Such results could be used to indicate when a biomarker monitoring programme may be more likely to detect measurement error ('noise') over true change ('signal').

**Conclusion:** The semi-variogram approach has potential in retrospectively estimating components of within-subject variability using pre-existing datasets. However, further research is due to investigate the precision of estimates, and risks of bias from missing measurements and missing documentation on the context in which measurements were taken.

**References:** *[1] Diggle PJ, Heagerty P, Liang K, Zeger SL. Analysis of Longitudinal Data. 2nd ed. New York: Oxford University Press; 2002. [2] Webb NJA, Woolley RL, Lambe T, Frew E, Brettell EA, Barsoum EN, et al. Sixteen-week versus standard eight-week prednisolone therapy for childhood nephrotic syndrome: The PREDNOS RCT. Health Technol Assess. 2019;23(26):1-109.*

# POSTER SESSION 02

## P02-15 Non-linear dynamic indices summarize densely sampled longitudinal data

**Mireya Diaz**
Division of Epidemiology and Biostatistics, Homer Stryker M.D. School of Medicine, Western Michigan University, Kalamazoo MI, United States

**Background:** Densely sampled longitudinal data have become ubiquitous with wearable device measurement technology presenting a challenge for traditional statistical analysis. However, these data also present an advantage over traditional longitudinal sampling. These allow learning about the dynamics of the underlying process, which generates them. Traditional and modern time-series and dynamic models can be applied to these data. The aim of this project is to assess how well-established indices from non-linear dynamics represent these data and how to best use these indices within other statistical modeling frameworks.
**Methods:** Monthly glycated hemoglobin measurements were captured for up to nine years. These measurements come from patients with Type I diabetes without microvascular complications from the Diabetes Control and Complications Trial. Indices from Poincare Plots were estimated from these monthly series for each individual. Overall (global) and yearly indices (local) were calculated to identify the most suitable formulation for both estimation and prediction purposes. We assessed the fitting via mean squared root error of three different formulations of the indices in case we would like to include these densely sampled longitudinal data within a joint model for example: (1) model based on global parameters, (2) model based on global average and local dynamics, (3) model based on both local average and dynamics.
**Results:** For series represented by larger Poincare indices (those above 1), the formulation that used at least one global parameter resulted in smaller errors. When indices were below 1, the formulation that used both local parameters for signal reconstruction resulted in smaller errors.
**Discussion:** Non-linear dynamic indices can be used to summarize densely sampled longitudinal data. Results are somewhat counterintuitive in the sense that for large variation across yearly windows, global parameters provide a good fit when you would think that local parameters would give a better description of the observations encompassed in the corresponding analytic window. Future work includes: (1) examining further details of the information contained in the analytic window to identify the underlying reasons for the apparently paradoxical findings, and (2) examining patients with other disease profiles to assess the robustness of the methods for different disease conditions.

## P02-16 Kernel density estimation for circular data about COVID-19 in the Czech Republic

**Stanislav Katina[1,2], Stanislav Zámečník[1], Ivana Hórová[1]**
1 Institute of Mathematics and Statistics, Masaryk University, Brno, Czech Republic
2 Institute of Computer Science of the Czech Academy of Sciences, Prague, Czech Republic

The term circular statistics describes a set of techniques used to model distributions of random variables that are cyclic in nature and these approaches can be easily adapted to temporal data recorded, e.g., daily, weekly or monthly. One of the nonparametric possibilities how to analyze these data is through kernel estimations of circular densities where the problem of how much to smooth, i.e., how to choose the bandwidth, is crucial.
In this presentation we describe the existing methods: cross-validation method, smoothed cross-validation, adaptive method and propose their modifications. We apply these methods on real data from the Institute of health information and statistics of the Czech Republic about total (cumulative) number of persons with a proven COVID-19 infection according to regional hygienic stations, number of cured persons, number of deaths and tests performed for whole country and regions coded based on nomenclature of territorial units for Statistics (NUTS). The results are visualized as circular histograms (rose diagrams) and calculated standardized characteristics are superimposed with choropleth map, where NUTS are shaded in diverging color scheme. All statistical analyses are performed in the R software.

**Acknowledgment:** *The work was supported (partly) by the long-term strategic development financing of the Institute of Computer Science (RVO:67985807) and specific research of Masaryk University as support for student projects (MUNI/A/1615/2020).*

**References:** *Horová, I., Koláček, J. and Zelinka, J. (2012), Kernel Smoothing in MATLAB: theory and practice of kernel smoothing, World scientific, Singapore. - Taylor, C. C. (2008). Automatic bandwidth selection for circular density estimation. Computational Statistics & Data Analysis, 52(7), 3493–3500. - Ley Ch., Verdebout T. (2019), Applied directional statistics: Modern methods and case studies, Chapman and Hall, London.*



# POSTER SESSION 02

## P02-17 Functional analysis of temporal data about patient's health condition after total knee replacement

**Markéta Janošová[1], Stanislav Katina[1,2], Libor Nečas[3,4]**
1 Institute of Mathematics and Statistics, Masaryk University, Brno, Czech Republic
2 Institute of Computer Science of the Czech Academy of Sciences, Prague, Czech Republic
3 Orthopedic Clinic, Martin University Hospital, Martin, Slovakia
4 Jessenius Faculty of Medicine, Comenius University, Martin, Slovakia

Persistent knee pain while walking or at rest, often caused by osteoarthritis (destruction of cartilage and changes of its mechanical properties), leads to total knee replacement (TKR) using arthroplasty implants. Patient Reported Outcome Measure questionnaires (PROMs) are commonly used to evaluate patient's condition. Some of the most commonly used PROMs for patients after TKR are Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and Knee Society Score (KSS). WOMAC was designed to measure patient's degree of pain, stiffness and functional limitations of the affected joint through patient's self-evaluation. KSS combines patient's objective and functional characteristics and is filled by attending physician. Both of these questionnaires are used at Martin University Hospital during examinations of patients after TKR, where the data about TKR are recorded in the Slovakian Arthroplasty Register (SAR). This study includes 2295 patients, who underwent primary TKR between January 1st, 2006 and December 31st, 2020. Patients were monitored before the surgery and then approximately 3, 6, 12, 24, and 36 months after the surgery. If a revision of primary TKR recorded in SAR was performed during the course of this follow-up, the patient is excluded from the subsequent statistical analyses. Our aim is to retrospectively evaluate patient's condition using their WOMAC scores and KSS and compare the condition and its changes through time among three structurally different types of knee implants, such as cruciate retaining, posterior stabilized condylar constrained and hinge knee implant, using functional data analysis and cubic splines. Statistical analyses were carried out using R software environment.

**Acknowledgment:** *The work was supported (partly) by the long-term strategic development financing of the Institute of Computer Science (RVO:67985807) and specific research of Masaryk University as support for student projects (MUNI/A/1615/2020).*

**References:** *Ramsay, J.O., Silverman, B.W., 2005: Functional Data Analysis. New York: Springer-Verlag - Nečas, L., Katina, S., Uhlárová J., 2013: Survival analysis of total hip and knee replacement in Slovakia 2003–2011. Acta Chirurgiae Orthopaedicae et Traumatologiae Cechoslovaca 80,1, 1–85. - Kahn T.L., Soheili A., Schwarzkopf R., 2013: Outcomes of total knee arthroplasty in relation to preoperative patient-reported and radiographic measures: data from theosteoarthritis initiative. Geriatric Orthopaedic Surgery & Rehabilitation 4,4: 117–126. - Jiang, C., Liu, Z., Wang, Y., Bian, Y., Feng, B., Weng, X., 2016: Posterior cruciate ligament retention versus posterior stabilization for total knee arthroplasty: A meta-analysis. PLOS ONE 11,1: 1–15.*

## P02-18 Multivariate Analysis of Blood Biomarkers in Amyotrophic Lateral Sclerosis

**Solmaz Yazdani[1], Christina Seitz[1], Anikó Lovik[1], John Andersson[1], Caroline Ingre[2], Fang Fang[1]**
1 Institute of Environmental Medicine, Karolinska Institute, Solna, Sweden
2 Department of Clinical Neuroscience, Karolinska Institute, Solna, Sweden

**Context:** Amyotrophic lateral sclerosis (ALS) is a disease characterised by the selective degeneration of upper and lower motor neurons, resulting in paralysis of the respiratory muscles and death, typically within 3-5 years following symptom onset. It is a rare disease, and the pathogenesis is not well understood. The current study aims to expand on our current knowledge about the role of inflammatory biomarkers in the disease progression.
**Problem and Aim:** There are few patients, the patients are very heterogeneous (both with regards to clinical characteristics and blood biomarkers), and the biological analysis of the samples is resource intensive, resulting in a small sample size and a large number of variables with missing data. Due to the small sample sizes, in this type of data usually each biomarker is analysed separately. Our goal is to predict disease progression (employing the ALSFRS-R scale) and survival based on the inflammatory biomarkers measured at the time of diagnosis (and used all together in one analysis) and other clinical characteristics.
**Methods:** We used multiple imputation to deal with the incompleteness, which allowed us to use data from all 85 patients, then an exploratory factor analysis identified five latent constructs, which we used in a cluster analysis to classify patients into groups that differ in functional decline based on a linear mixed model excluding clinical characteristics. We compared the differences in survival for the obtained clusters with Cox proportional hazards model.
**Results:** Five factors, collecting variables of specific cell types, were obtained. K-means clustering of the factor scores resulted in four clusters (each with at least ten patients). Two clusters had slow and two had faster decline in function, which was also reflected in the survival analysis. The differences between the clusters were attributed to a specific factor containing all variables of the same cell type (resting regulatory T cells).
**Conclusions:** Despite the sample size issues, we found clinically relevant latent constructs which allowed to classify the patients according to functional decline.

# POSTER SESSION 02

## P02-19 Possibilities and challenges when analysing large longitudinal data from a population based Norwegian registry (MoBa study)

**Milada Cvancarova Småstuen, Lisbeth Valla, Nina Misvær, Randi Andenæs, Sølvi Helseth**
Department of Public Health, Oslo Metropolitan University, Norway

**Background:** Large registry-based data are a valuable source of information and are used extensively in medical research. All individuals born in Norway are given a unique ID number which makes it possible to link information from several registries. However, analyzing such data requires a high level of expertise and collaboration both from a data analyst and health professionals. The Norwegian Mother and Child cohort study (MoBA) established in 1999, collects longitudinal data on demographic and background characteristics of infants born in Norway, their mothers, and fathers. Child development and personality traits are assessed using questionnaires. However, different versions of the same questionnaires have been used over time with different numbers of included items thus making it challenging to use these data in longitudinal studies.
**Objective:** To estimate associations between disruptive sleep patterns and colic at 6 months and child development later in life assessed at 18 months, 3 and 5 years using ASQ (Ages and Stages Questionnaire) and CBCL (Child Behavior Check-list) questionnaires.
**Methods:** We used z-scores to evaluate differences between groups of children both at given time points and across the whole follow up as the number of included items varied for both questionnaires. The z-scores were constructed as follows: at each assessment point the dataset was divided into two groups based on the exposure variable (e.g. having a sleep problem or not). Those who did not report having sleeping problems served as the reference population. Changes at given time points and across the follow up trajectory were estimated with linear mixed models for repeated measures with unstructured covariance matrix to account for dependencies within included individuals.
**Results:** We have analysed 75188 children using data collected at four time points. Our data revealed very small but statistically significant differences between children with disruptive sleep patterns early in life and the reference population.
**Conclusion:** Our analyses revealed that disruptive sleep early in life does not have a clinically relevant negative effect on child development later in life. Correct and clinically relevant interpretation of results from large registry based studies require a close collaboration between medical professionals and statisticians.



# POSTER SESSION 02

## P02-20 Estimating time to confirmed disease progression in observational data sources with irregular visit schedules

**Thomas P.A. Debray[1,2], Massimiliano Copetti[3], Robert W. Platt[4], Gabrielle Simoneau[5], Changyu Shen[6], Fabio Pellegrini[7], Carl de Moor[6]**
1 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Netherlands
2 Smart Data Analysis and Statistics, Utrecht, Netherlands
3 Unit of Biostatistics, Fondazione IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy
4 Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal QC, Canada
5 Biogen Canada, Mississauga ON, Canada
6 Biogen Inc., Cambridge MA, United States
7 Biogen International GmbH, Zug, Switzerland

**Background:** Multiple sclerosis (MS) is a chronic, progressive neurological disease. While long term follow-up of patients in observational data sources offer countless opportunities for comparative effectiveness research in MS, they also entail specific challenges. For example, time to confirmed disease progression (CDP), an important outcome in MS, is defined from treatment initiation until an increase on a severity score, provided the increase can be confirmed 6 months later. This definition is challenging to apply in observational data sources where patients follow irregular visit schedules because a progression and its confirmation may not be recorded during a visit.
**Objective:** To compare two single imputation approaches (last observation carried forward and rounding) to a multiple imputation approach (linear mixed model accounting for autocorrelation) for analyzing time to CDP in observational data sources.
**Methods:** Data were generated under various scenarios such that the informative visit mechanism was dependent on the observed covariates, on the treatment received or on the underlying outcome process. We compared the three imputation approaches to recover the missing EDSS scores, time to CDP and the hazard ratio. In a real data example, we compared dimethyl fumarate to fingolimod on time to CDP using data from the Multiple Sclerosis Partners Advancing Technology and Health Solutions database.
**Results:** The linear mixed model led to the lowest mean squared error for recovering the time to CDP across all simulation scenarios. Hazard ratio estimators were generally less biased and more efficient with the linear mixed model, especially when the visit process depended on the treatment or outcome. The linear mixed model also yielded much better coverage in these situations. In the real data example, preliminary results found diverging hazard ratio estimates across the three imputation methods, although confidence intervals were wide and overlapping.
**Conclusions:** Last observation carried forward and rounding allow to recover individual trajectories of disease progression, but the corresponding comparative effectiveness analysis does not account for uncertainty of the imputations and may be inaccurate. More advanced methods based on linear mixed models can improve imputations of time to CDP, and thus lead to more reliable inference on treatment effect.

# POSTER SESSION 03

POSTER SESSION 03 — Causal inference

## P03-01 Statistical methods to analyse ordinal categorical data arising from the clinical trial of drugs from the pharmaceutical industry

**Nazneen Shariff**
Edinburgh University, Medical Statistics, Medical School., Edinburgh, United Kingdom

The pharmaceutical industry gives rise to square tables of ordinal categorical data from the clinical trial of newly manufactured drugs. The Data arising from such trial have three dimensions: Drug X pre-treatment score X post-treatment score. The Proportional Odds Model is assumed to fit the data. The model is initiated by an underlying grouped continuous unobservable random variable. The ordinal categories on the post-treatment score is taken as a grouped continuous random variable. The data is stratified according to the pre-treatment score for purpose of homogeneity. Statistical tests of the null hypothesis: are formulated through the parameters of the proportional odds model. Three methods for testing the null hypothesis are examined: Regular Wilcoxon test, The classical Analysis of Covariance (ANACOVA) and a novel procedure existing in literature: Wilcoxon Van Elteren test for stratified data. A simulation study was carried out to examine the existing methods with the new method of application. The three methods are evaluated on a real data set of square ordinal responses from the clinical trial of a new drug.
**Conclusions:** The following conclusions can be drawn from the study of the statistical methods on square tables of ordinal data: The Wilcoxon Van Elteren test is proving to be most efficient in the presence of a pre- and post-treatment relationship when compared with the regular Wilcoxon test and the ANACOVA. The ANACOVA appears to be robust in that it gives the correct nominal significance level. but it loses power relative to the Wilcoxon Van Elteren test when there is a strong relationship between the pre-and post- treatment score. In analyzing ordinal data from a drug induced symptom change study, the post-treatment score can be identified as a sensitive variable in detecting treatment differences. Theoretical calculations on the parameter 'q' give consistency to the application of the Wicoxon Van Elteren test.

## P03-02 Causal inference with skewed outcome data: Moving beyond the "ignore or transform" approach

**Daisy A. Shepherd[1,2], Margarita Moreno-Betancur[1,2]**
1 Clinical Epidemiology & Biostatistics Unit, Murdoch Children's Research Institute, Australia
2 Department of Paediatrics, The University of Melbourne, Australia

With continuous outcomes, the average causal effect is typically defined using a contrast of mean potential outcomes. However, in the presence of skewed outcome data, the mean may no longer be a meaningful summary statistic and the definition of the causal effect should be considered more closely. When faced with this challenge in practice, the typical approach is to either "ignore or transform" – ignore the skewness in the data entirely, or transform the outcome to obtain a more symmetric distribution for which the mean is interpretable. In many practical settings, neither approach is entirely satisfactory. An appealing alternative is to define the causal effect using a contrast of median potential outcomes. Despite being a widely acknowledged concept, there is currently limited discussion or availability of confounder-adjustment methods to generate estimates of this parameter.
Within this study, we identified and evaluated potential confounding-adjustment methods for the difference in medians to address this gap. The methods identified are multivariable quantile regression, adaptations of the g-computation approach, weighted quantile regression and an IPW estimator [1]. The performance of these methods was assessed within a simulation study, and applied in the context of an empirical study based on the Longitudinal Study of Australian Children. Results indicated that the performance of the proposed methods varied considerably depending on the simulation scenario, including the severity of skewness of the outcome variable. Nonetheless, the proposed methods provide appealing alternatives to the common "ignore or transform" approach, enhancing our capability to obtain meaningful causal effect estimates with skewed outcome data.
**References:** *[1] Zhang Z, Chen Z, Troendle JF, Zhang J. Causal Inference on Quantiles with an Obstetric Application. Biometrics. 2012;68(3):697-706. doi:10.1111/j.1541-0420.2011.01712.x*



POSTER SESSION 03

## P03-03 Marginal Structural Models with Latent Class Growth Modeling of Time-varying Treatment

**Awa Diop[1], Caroline Sirois[2], Denis Talbot[1]**
1 Département de médecine sociale et préventive, Université Laval, Canada
2 Faculté de pharmacie, Université Laval, Québec QC, Canada

A well-known treatment to prevent Cardiovascular disease (CVD) are statins. Several randomized controlled trials (RCTs) have shown the efficacy of statins to prevent a first CVD event, that is, for primary prevention. Despite current evidence, it is still not clear if these conclusions can be applied to older adults as they are often excluded from RCTs. Moreover, there is little evidence concerning the effectiveness of statins for primary prevention among older adults in a real-life setting. Analysis of observational data could add crucial information on the benefits of actual statin's patterns of use. However, the number of unique treatment trajectories increases exponentially with the length of follow-up in longitudinal studies. Latent class growth models (LCGM) are increasingly proposed as a solution to summarize the observed longitudinal treatment in a few distinct groups. It is known that LCGM cannot be effectively combined with standard approaches, such as covariate adjustment in an outcome regression model, as they fail to control confounding bias. Marginal structural models (MSMs) are popular for their ability to deal correctly with time-varying treatment and covariates. We propose to use LCGM to classify individuals into a few latent classes based on their medication adherence pattern, then choose a working MSM that relates the outcome to these groups. We showed that the data-driven estimation of the trajectory groups can be ignored. As such, parameters can be estimated using inverse probability of treatment weights and conservative inferences can be obtained using a standard robust variance estimator. Simulation studies are used to illustrate our approach and compare it with unadjusted, baseline covariates adjusted, time-varying covariates adjusted and inverse probability of classes weighting adjusted models. We found that our proposed approach yielded estimators with little or no bias. We will apply our LCGM-MSM approach to a database composed by 572 822 Quebecers aged 66 or more and who are statins initiators to estimate the effect of statin-usage trajectories on a first CVD event. Our proposal is relatively simple to implement and we expect it to yield results that are clinically meaningful, easy to interpret and statistically valid.

## P03-04 Multidimensional mediators: Comparison between statistical methods using Data simulation

**Nadia Dardenne, Anne-Françoise Donneau**
Public Health Department, Biostatistics, University of Liege, Belgium

**Background:** Various statistical methods exist to evaluate mediation effect of a multidimensional score. Based on our experiences with quality of life and health literacy data, few differences in results of mediation effect are highlighted between methods. Data simulations could help to confirm these results using broader examples.
**Objectives:** This study compares, through simulations, natural effect models (NEM) and structural equation modelling (SEM).
**Methods:** Data were simulated using a Normal distribution, considering different correlation values between the independent variable (X), the dependent variable (Y) and the multidimensional scores $M^{1,2,3}$. A 0.4 correlation between $M^{1,2,3}$ were considered. Mediation effect was assessed with NEM using all scores (model 1) and a sum of the scores (model 2), and with SEM (model 3). The number of simulation was fixed to 2000 and sample size to 30, 50, 150 and 500. The percentage of direct and indirect effect rejection were calculated and agreement between models was assessed using the Bowker test and Cohen's Kappa coefficient.
**Results:** First, we simulated data without correlation between X and respectively $M^{1,2,3}$ and Y. As expected, all models concluded to a non-significant direct and indirect effect. Kappa values were higher when comparing model 1 and 2 and increased with sample size (from 0.4 to 0.9). When adding moderate correlation between only X and Y, less than 1% (up to 0%) of the simulated data leaded to significant indirect effect while more than 85% (up to 100%) concluded to significant direct effect (Kappa > 0.7). Secondly, data were simulated with a zero to moderate correlation between X and Y and moderate correlations between $M^{1,2,3}$ and Y. The larger the sample size, the more the mediation effect was found to be partial rather than total with better agreement between models 1 and 2 (Kappa > 0.5). Distribution of mediation effect was however different according the methods ($p < 0.0001$). For larger sample size, all concluded to partial mediation.
**Conclusion:** Differences between methods appear especially for small samples and when mediation effect was assumed. Further simulations should be carried out by varying the number of mediators, adding covariates and for other distributions.
**References:** *[1] Steen, J, Loeys, T, Moerkerke, B & Vansteelandt, S. Medflex : An R package for flexible mediation analysis using naturel effect models. Journal of statistical sofware 2017, 76 (11) : 1-46 https://www.Jstatsoft.Org/article/view/v076i11. [2] Rossel, Y. Lavaan: An R package for Structural Equation Modeling. Journal of Statistical Sofware 2002, 48 (2) : 1-36 http://www.jstatsoft.org/v48/i02*

# POSTER SESSION 03

## P03-05 The role of the matching algorithm in an analysis of the effect of hemoadsorption in patients with sepsis

**Priska Heinz[1], Pedro David Wendel Garcia[2], Ulrike Held[1]**

1 Department of Biostatistics, at Epidemiology, Biostatistics & Prevention Institute, University of Zurich, Switzerland
2 Institute of Intensive Care Medicine, University Hospital of Zurich, Switzerland

In an observational study of 208 patients with therapy refractory septic shock, collected at the intensive care unit of the university hospital Zurich, the objective was to estimate the treatment effect of a hemoadsorption device (Cytosorb®) on in-hospital mortality. Septic shock is a life-threatening condition, incepted by a dysregulated hyperinflammatory immune response, the so-called cytokine storm, which is associated with a highly elevated mortality. In this context, hemoadsorption is discussed as effective treatment to reduce cytokine levels and inflammatory mediators in the blood. Lacking international validated protocols or guidelines, the decision to employ hemoadsorption remains fully at the discretion of the treating clinician. A set of predefined confounder variables was used for matching on treatment decision.

The research question led us to set up a simulation study, following Morris et al. [1], to evaluate five different matching algorithms (nearest, optimal, caliper, full and genetic matching), and to compare their results regarding covariate balance as measured with the standardized mean difference, coverage, and precision of estimated odds ratios [2]. The odds ratios were estimated by unadjusted and adjusted logistic regression, employing the same covariates for adjustment than used for matching.

Results of the simulation study showed that good balance could be achieved with full, genetic, and caliper matching, and these matching algorithms also yielded estimated treatment effects closest to the true values. Full matching was the only matching algorithm that could not achieve a coverage rate of more than 90%. We could show that additional adjustment for covariates provided better estimates. A considerable proportion of patients was lost during the matching process. In our simulation study, 18% of the treatment cohort was discarded after caliper matching leading to a substantial loss of power.

In conclusion, matching is a powerful tool for the analysis of observational studies, nonetheless the matching algorithms, potential loss of power and unmeasurable confounding should be considered when assessing the results for generalizability.

**References:** *[1] Morris, T. P., White, I. R., and Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. Statistics in medicine, 38, 2074–2102. [2] Simulation study protocol: https://osf.io/unbka/*

# POSTER SESSION 03

## P03-06 Estimating the causal effect of direct-acting antiviral agents on kidney function in a clinical cohort of chronic Hepatitis C patients

**Adrienne O'Donnell[1], Nathan Pham[2], Leandra Battisti[3], Sara Lodi[1]**

1 Department of Biostatistics, Boston University School of Public Health, United States
2 Department of Internal Medicine, Boston Medical Center, United States
3 Department of Pharmacy, Boston Medical Center, United States

**Introduction:** Treatment with direct-acting antiviral (DAA) agents is highly effective at clearing Hepatitis C virus (HCV) among individuals with chronic HCV-infection: the cure rate of DAAs in clinical trials is >95%. However, the effect of DAAs on kidney function remains uncertain. Previous studies describe differential changes in glomerular filtration rate (GFR), a measure of kidney function, before and after receipt of DAA [1]. Interpretation of these results is limited by lack of a contemporary comparator group not receiving DAA. We aim to use causal inference methods to estimate 2-year change in GFR under two counterfactual interventions: had all versus no patients received DAA.

**Methods:** We used electronic health records (EHR) from Boston Medical Center, an urban safety-net hospital in the US. We included DAA-naive patients with chronic HCV-infection engaged in care, a Fibrosis-4 Score and GFR measurements between 2014 and 2018. We used the g-computation algorithm [2], which works by first fitting parametric models for the densities in the observed data, and then simulating a dataset under each counterfactual intervention by fixing the treatment variable, DAA. All models were adjusted for baseline demographic and clinical time-varying confounders. We estimated and compared the mean 2-year change in GFR in each simulated dataset, using 500 bootstrap samples to obtain 95% confidence intervals. Sensitivity analyses were run to ensure the parametric assumptions were appropriate given the complex nature of EHRs and the fact that implementing this method with a continuous outcome is fairly novel.

**Results:** The study included 2307 individuals with median age [IQR] 54 [42-60], of whom 67% were male, 18% had advanced fibrosis or cirrhosis, and 41% had chronic kidney disease (GFR≤90) at baseline. Estimated 2-year mean change in GFR was -5.8 ml/min/1.73m2 (95%CI:-7.7;-4.0) under DAA receipt and -4.9 (-7.3;-1.9) under no DAA receipt, with a mean difference of -0.9 (-4.4;2.3).

**Conclusions:** We found no causal effect of DAA on kidney function in this sample. This finding suggests continued clinical monitoring of patients' kidney function is important, even after receipt of DAA. In future work, we intend to optimize methods for handling the challenges that come with utilizing EHRs to answer causal questions.

**References:** *[1] Sise ME, Chute DF, Oppong Y, et al. Direct-acting antiviral therapy slows kidney function decline in patients with Hepatitis C virus infection and chronic kidney disease. Kidney Int. 2020 Jan;97(1):193-201. doi: 10.1016/j.kint.2019.04.030. [2] Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. Math Model. 1986;7(9-12):1393-1512. doi: 10.1016/0270-0255(86)90088-6.*

# POSTER SESSION 03

## P03-07 Should multiple imputation be stratified by exposure when estimating causal effects via outcome regression?

**Jiaxin Zhang[1,2], S. Ghazaleh Dashti[2], John B. Carlin[1,2], Katherine J. Lee[1,2], Margarita Moreno-Betancur[1,2]**
1 Clinical Epidemiology and Biosatistics Unit, Department of Paediatrics, University of Melbourne, Australia
2 Clinical Epidemiology and Biosatistics Unit, Murdoch Children's Research Institute, Australia

Despite recent advances in methods for causal inference in epidemiology, outcome regression remains the most widely used method for estimating causal effects in the simple time-fixed exposure and outcome setting. Missing data are common in epidemiologic studies and complete-case analysis (CCA) and multiple imputation (MI) are two commonly used methods for handling them. In randomized controlled trials (RCTs), it has been shown that MI should be conducted separately by treatment group, but the question of whether to impute by exposure group has not been addressed for observational studies, in which causal inference is understood as the task of emulating an RCT. We designed a simulation study to evaluate and compare the performance of five missingness methods: CCA, MI on the whole cohort, MI including an interaction between exposure and outcome in the imputation model, MI including interactions between exposure and all incomplete variables in imputation models, and MI conducted separately by exposure group. Bias, precision and confidence interval coverage were investigated. We generated data based on an example from the Victorian Adolescent Health Cohort Study, where interest was in the causal effect of adolescent cannabis use on young adulthood depression and anxiety in females. Three exposure prevalence scenarios and seven outcome generation models were considered, the latter ranging from no interaction to a strong positive or negative interaction between exposure and a strong confounder. Two missingness scenarios were examined: one with the incomplete outcome, the other with incomplete outcome and confounders, each with three levels of complexity in terms of the variables or interaction on which missingness depended. The simulation results show the relative bias of analysis approaches across all scenarios considered. MI by exposure group usually led to the least bias. Considering the overall performance, MI by exposure group is recommended if MI is adopted in settings such as these.

## P03-08 Pathways to inequalities in child mortality: a population level study in Wales

**Daniela K. Schlüter[1], Eric T.C. Lai[1,2], Hoda Abbasizanjani[3], Rowena Griffiths[3], Ashley Akbari[3], David Taylor-Robinson[1]**
1 Department of Public Health, Policy and Systems, University of Liverpool, United Kingdom
2 Jockey Club Institute of Ageing, The Chinese University of Hong Kong, Hong Kong
3 Swansea University Medical School, Swansea University, United Kingdom

**Context:** There has been an unprecedented rise in infant mortality rates in the UK since 2014, especially in disadvantaged areas. This trend is concerning since infant mortality is a sensitive indicator of the prevailing social conditions affecting health across the life course. Identifying potentially modifiable factors on the pathway linking childhood socio-economic conditions (SECs) to child mortality is important to inform public health policies to reduce health inequalities.
**Objectives:** To assess the extent to which intervening on maternal health, perinatal factors and/or birth outcomes might reduce inequalities in child mortality.
**Methods:** We conducted a causal analysis of linked population level data from the SAIL Databank on all singletons born in Wales between 2000 and 2019 and their mothers. The exposure of interest was mother's quintile of small area deprivation 3-years prior to pregnancy; the outcome was child mortality between birth and age 15-years. The data included gestational age, birthweight, parity, maternal age, maternal health conditions before and during pregnancy, pregnancy complications, congenital anomalies, smoking during pregnancy and perinatal maternal mental health. Using the framework of interventional disparity measures, we estimated the contribution of a block of factors relating to maternal health and perinatal factors, and that of a block of birth outcomes to inequalities in child mortality. Confidence intervals will be calculated by non-parametric bootstrap.
**Results:** Initial results are based on a complete-case analysis of data on 463,200 live births out of which 1,719 died by age 15. The probability of having died by age 15 was 1.37 times as high in the most deprived quintile compared to the least. After shifting the distribution of maternal health, perinatal factors and birth outcomes in the most deprived population quintile to that in the least deprived quintile, the survival probability ratio between the most and least deprived children was reduced to 1.09.
**Conclusions:** Child mortality is a rare event but with clear socio-economic patterning. Initial results indicate that maternal health, perinatal factors and birth outcomes may explain most of the observed inequalities. Further analyses will aim to disentangle the contribution of these mediating blocks to identify potential public health policy entry points.



# POSTER SESSION 03

## P03-09 Target Trial Emulation and Missing Eligibility Data: A study of Palivizumab for child respiratory illness

**Daniel Tompsett, Bianca de Stavola, Ania Zylbersztejn, Pia Hardelid**
UCL Great Ormond Street Institute of Child Health, London, United Kingdom

Target trial emulation (TTE) seeks to apply the principles of a Randomised Controlled Trial (RCT) onto analyses of observational data in order to improve the quality of analysis. However, selection into the target trial necessitates that the data on eligibility is observed, and thus in practice, a TTE typically only includes those with complete case eligibility data.This presents a problem when eligibility data is Missing at Random (MAR), or Missing not at Random (MNAR). When the estimand of interest is the Average Treatment Effect (ATE), an analysis of complete case TTEs will often suffer from bias and issues of internal validity if the treatment has a heterogeneous effect on the outcome. The objective of this work is to investigate the source and size of any bias caused by taking complete case eligible individuals when eligibility is either MAR, or in particular, MNAR. We propose a solution whereby Multiple Imputation (MI) is used to predict missing eligibility data, which is then used to recruit individuals into a Target Trial, a method which, to our knowledge, is rarely proposed. This will be investigated using both simulation, and practical clinical data. Notable aspects of the work include imputation of eligibility criteria for inclusion into a target trial, and the Multiple Imputation of eligibility data using MNAR imputation models.
We apply this work to an analysis of the effect of the effect of Palivizumab on hospitalisation in premature infants due to Respiratory Syncytial Virus (RSV) in England, by constructing a target trial using MNAR Multiple Imputation of gestational age. Results are yet to be finalised.

## P03-10 Targeted causal quantile estimation for measurement of postdischarge opioid use in a text-to-web survey

**Chris J. Kennedy[1,2], P. Nina Scalise[2], Kortney Robinson[2], Brandon Booth[2,3], Jayson Marwaha[1,2], Aaron Fleishman[2], Gabriel Brat[1,2]**
1 Department of Biomedical Informatics, Harvard Medical School, Boston MA, United States
2 Department of Surgery, Beth Israel Deaconess Medical Center, Boston MA, United States
3 College of Medicine, Howard University, Washington DC, United States

**Background:** We developed an automated system for recruiting discharged surgical patients via text message to report opioid consumption in an online survey. We planned to assess its effectiveness for estimating opioid use in surgical cohorts while addressing the potential for nonresponse bias.
**Methods:** Patients who underwent a surgical procedure at our institution between 2019 and 2020 were surveyed to quantify opioids consumed after discharge. Opioid consumption was measured through either survey response or electronic health record (EHR) documentation when no opioids were prescribed. Factors sourced from the EHR were tested for association with survey response and with opioid use, and were used to estimate a propensity score (PS) for measurement of opioid consumption among those prescribed opioids. The PS was estimated using logistic regression (LR) and ensemble machine learning (stratification, linear regression, LR, lasso, random forest, BART), and was evaluated by its area under the curve (AUC). The median and 75th percentile of opioid consumption were estimated for the 10 most common surgical procedure groups in morphine milligram equivalents (MMEs), including adjustment for observed nonresponse patterns using inverse probability weighting (IPW) and doubly robust targeted learning (TL) using quantile regression. SMS-recruited survey results were benchmarked against estimates calculated from a previous phone-based survey.
**Results:** We evaluated 6,553 surgical patients, of which 71% were prescribed opioids, 24% completed the survey, and 44% had opioid consumption measured. Characteristics significantly associated with opioid measurement include age, length of stay, and current tobacco use. The ensemble was able to predict opioid measurement with AUC (95% C.I.) = 0.593 (0.577 - 0.608), compared to 0.575 (0.560 - 0.591) for logistic regression (p = 0.017). Unadjusted opioid consumption was substantially underestimated when compared to earlier phone-based results. IPW and TL adjustment reduced underestimation bias of median MMEs for 60% of surgical procedures, and 90% of procedures when estimating the 75th percentile of MME consumption.
**Conclusion:** Adjustment for nonresponse using TL or IPW resulted in increased ranges for estimated opioid consumption. Without appropriate statistical adjustment, nonresponse can strongly bias estimates of typical opioid consumption as collected in patient surveys.

# POSTER SESSION 03

## P03-11 Does early weight gain mediate the causal effects of severity and duration of depression on the onset of metabolic syndrome?

**Khalil El Asmar[1,2], Sarah El Omari[2], Bruno Falissard[3], Romain Colle[1,4], Laurent Becquemont[1,5], Emmanuelle Corruble[1,4]**

1 CESP, MOODS Team, INSERM, Faculté de Médecine, Univ Paris-Saclay, France
2 Department of Epidemiology and Population Health, Faculty of Health Sciences, American University of Beirut, Lebanon
3 CESP/INSERM, U1018 Centre de Recherche en Epidemiologie et Santé des Populations, Paris, France
4 Service Hospitalo-Universitaire de Psychiatrie de Bicêtre, Hôpitaux Universitaires Paris-Saclay, Assistance Publique-Hôpitaux de Paris, Hôpital de Bicêtre, France
5 Centre de recherche clinique, Hôpitaux Universitaires Paris-Saclay, Assistance Publique-Hôpitaux de Paris, Hôpital de Bicêtre, France

**Objectives:** In clinical research, the metabolic syndrome (MetS) has thusfar been used as an observed variable. We conducted a confirmatory factor analysis of the MetS in order to identify whether there's a single latent MetS factor suggestive of a common pathophysiology, and to use SEM to analyze whether early weight gain mediates the causal effects of severity and duration of depression on the onset of MetS.
**Methods:** In the prospective METADAP cohort, 260 non-overweight patients with a Major Depressive Disorder (MDD) were assessed for early weight gain (EWG) (>5%) after one month of antidepressant treatment, and for the later incidence of MetS after three and six months of treatment. Measurement models of MetS at 3 and 6 months were tested. The variables were chosen based on a modified version of the International Diabetes Federation definition. An exploratory factor analysis (EFA) using principal components analysis was performed to assess the dimensionality of the constructs. Confirmatory factor analysis (CFA) was subsequently used to test the structure of the extracted factor models. Two SEMs were fitted to test the mediating role of EWG on the effects of the severity and the duration of depression on the onset of MetS at 3 and 6 months.
**Results:** EFA of each of MetS at 3 and 6 months revealed a unidimensional structure. CFA showed that the two one-factor measurement models significantly reflected the constituent observed variables. An error term correlation path between triglycerides and HDL-C was added producing measurement models of moderate to good fit (CFI of 1 and 0.889, RMSEA of 0 and 0.089, and SRMR of 0.034 and 0.064, respectively). SEMs incorporating either model (CFI of 0.738 and 0.661, RMSEA of 0.086 and 0.096, and SRMR of 0.069 and 0.079, respectively) revealed that the relationship between duration and severity of depression on one hand, and MetS on the other is totally mediated by EWG. EWG was more positively and significantly correlated with MetS6 than with MetS3 (standardized coefficients of 0.23 versus 0.3; $p<0.05$).
**Conclusion:** Clinicians should monitor their patients for EWG in addition to the severity of their conditions, in order to minimize risk of onset of MetS.



POSTER SESSION 03

## P03-12 Pathway specific population attributable fractions

**Maurice O'Connell, John Ferguson**
HRB Clinical Research Facility, School of Medicine, National University of Ireland Galway, Ireland

Population attributable fractions (PAF) represent the relative change in disease prevalence expected if an exposure was absent from the population. What percentage of this effect acts through particular pathways may be of interest. E.G. the effect of sedentary lifestyle on stroke may be mediated by blood pressure, BMI and several other mediators. Path-specific PAFs (PS-PAFs) represent the relative change in disease prevalence from an intervention that, conditional on observed covariates, shifts the distribution of the mediator to its expected distribution in a hypothetical population where the risk factor was eliminated. A related (more mechanistic) definition examines disease prevalence expected from an individual-level intervention assigning each individual the mediator they would have received if the risk factor had been eliminated.
Our aim here is not to decompose the total PAF for a risk factor into an additive sum over mediating pathways, but to instead fairly compare disease burden attributable to differing mediating pathways and as a result gain insights into the dominant mechanisms by which the risk factor affects disease on a population level. While PS-PAFs corresponding to differing pathways (mediating the same risk factor-outcome relationship) will each usually be less than the total PAF, they will often sum to more than the total PAF. In this manuscript, we present definitions, identifiability conditions and estimation approaches for PS-PAFs under various study designs. We illustrate results using INTERSTROKE [1], an international case-control study designed to quantify disease burden attributable to a number of known causal risk factors. A R package will be available.

**References:** *[1] M. J. O'Donnell, S. L. Chin, S. Rangarajan, D. Xavier, L. Liu, H. Zhang, P. Rao-Melacini, X. Zhang, P. Pais, S. Agapay, et al., "Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (interstroke): a case-control study," The lancet, vol. 388, no. 10046, pp. 761-775, 2016.*

# POSTER SESSION 04

## POSTER SESSION 04 — High Dimensional Data

### P04-01 A methodological approach to assess data quality from the Clinical Practice Research Datalink

**Pradeep S. Virdee[1], Alice Fuller[2], Michael Jacobs[3], Tim Holt[2], Jacqueline Birks[1]**
1 Centre for Statistics in Medicine, NDORMS, University of Oxford, Oxford, United Kingdom
2 Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom
3 BMS Haematology, John Radcliffe Hospital, Oxford University Hospitals, Oxford, United Kingdom

**Introduction:** Electronic health records (EHRs) are databases that store routinely-collected anonymised patient data. EHR data is increasingly used for research over time. However, data validation and quality assessments are often not performed. Previous systematic reviews have identified barriers and highlighted the need for generalised approaches to assess EHR data quality. We derived a methodological approach to assess data quality from the UK Clinical Practice Research Datalink (CPRD), demonstrated with application to the full blood count (FBC) blood test, which consists of up to 20 parameters. We provide recommendations for researchers who wish to access and analyse EHR data.
**Methods:** Laboratory data from CPRD was accessed for primary care patients aged at least 40 years at study entry with at least one FBC blood test. Medical codes and entity codes, two coding systems used within CPRD, were used to identify FBC blood test records (step 1) and cross-checked (step 2), with mismatches explored (step 3) and, where possible, rectified (step 4). The reliability of units of measurement are described (step 5) and reasons for missing data discussed (step 6).
**Results:** There were 138 medical codes and 14 entity codes for the FBC in the data (step 1). Medical and entity codes consistently corresponded to the same FBC parameter in 95.2% (n=217,752,448) of parameters (step 2). In the 4.8% (n=10,955,006) mismatches, the suggested FBC data was often already present elsewhere in the patient's record (step 3). The most common parameter rectified was mean platelet volume (n=2,041,360), which we identified as not having an existing entity code, and 1,191,540 could not be rectified (step 4). Units of measurement were often missing, partially entered, or did not correspond to the blood value (step 5). One FBC parameter, red blood cell distribution width, had the most missing data (98% of 16,537,017 FBC tests), which we identified to be a result of haematology laboratories supressing output before delivering test results to primary care practices (step 6).
**Conclusion:** Performing data quality checks can help to understand the extent of any issues in the dataset. We emphasise balancing large sample sizes with reliability of the data.



POSTER SESSION 04

### P04-02 Modeling Child Mortality in the presence of Clustering

**Beryl Ang'iro[1], Roel Braekers[1], Nelson Owuor[2]**
1 Center for Statistics, Hasselt University, Belgium
2 School of Mathematics, University of Nairobi, Kenya

**Introduction:** Though Under Five Child Mortality (U5CM) rate has significantly reduced globally, it is still a major public health problem in Low and Middle Income Countries (LMIC) compared to High Income Countries. In 2019 for instance, 1 in 13 children died before reaching their fifth birthday in Sub-Saharan Africa. Kenya is among the countries that have recorded high U5CM rate in the world. In 2020, U5CM in Kenya stands at 43.2 deaths per 1000 live births. This rate is almost twice the newly set third Sustainable Development Goals (SDG) target that seeks to reduce U5CM to about 25 per 1000 live births in all countries by 2030. The study analyses data from the Kenya Demographic and Health Survey (KDHS) conducted in 2014 which is associated with challenges such as clustering, variable selection, high level of missingness among others. We model U5CM in the presence of clustering. When observations are clustered within groups or multiple event times are clustered within individuals, dependence between event times in a cluster is of interest.
**Objective:** To establish the determinants of U5CM accounting for clustering at household level.
**Methods:** Two Random Survival Forest (RSF) models are fitted; the log-rank split-rule and the log-rank score split-rule. This is an attempt to handle the problem of non-proportional hazards that is often associated with survey data. We focus attention on the gamma frailty model with households as the random effects (frailty).
**Results and Conclusion:** We apply the two RSF models to the data. The two models identified the number of children under five years in the household as the most important predictor of U5CM. Both models are similar in identifying the same covariates affecting mortality of children under five, that is Number of children under five in the household, duration of breastfeeding, births in the past five years, total children ever born, total living children, mother's education and toilet facility. The gamma frailty model shows an increase in the effect of the same covariates on the outcome (U5CM). If dependence within the household is ignored, the effects of these covariates are underestimated.
**References:** *[1] W. H. Organization, et al., Global health observatory data, under-five mortality, 2016 (2018). [2] L. Duchateau, P. Janssen, The frailty model, Springer Science & Business Media, 2007.*

### P04-03 Nguyen's Information Criteria (NIC)

**Jean-Michel Nguyen**
Techniques de l'Ingénierie Médicale et de la Complexité - Informatique, Mathématiques, Applications (TIMC – IMAG) -UMR 5525, Université Grenoble Alpes – CNRS, France

**Objective:** To develop a new strategy for the selection of variables for Big Data.
**Methods:** ROP classification trees have the particularity of including neurons in the nodes [1], leading to obtain perfect tree (PT), with no classification errors. By assembling PT in large quantities, a new family of ten statistical information criteria, including three dimensions, to describe and understand how a feature contributes to a PT classification was developped [2]. The first category represents the informative and predictive quality of a feature and included 6 criteria. Among them, the first criterion, NIC1, is defined as the probability that a feature will obtain a PT. The second category represents the information measuring the proximity of a feature in the pathogenic sequence to a state Y and concerned 2 criteria. Among them, NIC8 is defined as the one-node model probability. The third category represents the information measuring the complexity of the relationship between a feature and state Y and concerned 2 criteria. Among them, NIC10, is defined as the probability to get a unique solution. With these 10 criteria, the variables can be ranked to an order of importance, according to a hierarchical strategy or a score. Two public databases were used (Wisconsin Breast Cancer, 569 observations, 30 features and GSE22513, placitaxel resistance genomic expression, 54675 probes and 14 duplicate observations) to compare the performance of this new statistical information in the selection of variables, against Gini information or against the SVM-RFE (support vector machines - recursive feature elimination) cross-validation procedure.
**Results:** The first NIC allowed the Akaike information criterion to be minimized more quickly than data obtained with the Gini index when the features were introduced in a logistic regression model. The selected features based on the NICScore showed a slight advantage compared to the SVM-RFE procedure.
**Conclusion:** A field of research for the development of criteria for evaluating the pathophysiological proximity of features to a biological event of interest and the complexity of their relationships was open. With this information, a mapping of the cascade of features leading to the event of interest can be expected.
**References:** *[1] Nguyen,J.M. and Antonioli,D. A New neuron, a 3D-propagation: the Regression OPtimized (ROP) concept, performances, applications, perspectives, submitted. [2] Jean-Michel Nguyen, Pascal Jézéquel, Pierre Gillois, Luisa Silva, Faouda Ben Azzouz, Sophie Lambert-Lacroix, Philippe Juin, Mario Campone, Aurélie Gaultier, Alexandre Moreau-Gaudry and Daniel Antonioli. Random Forest of Perfect Trees: Concept, Performance, Applications, and Perspectives. Bioinformatics, in press.*

# POSTER SESSION 04

## P04-04 Long-term oral prednisolone exposure for bullous pemphigoid: a population-based study using 'big data' and missing data algorithms

**Monica S.M. Persson[1], Karen E. Harman[2], Kim S. Thomas[2], Jo R. Chalmers[2], Yana Vinogradova[2], Sinead M. Langan[3], Julia Hippisley-Cox[4], Sonia Gran[2]**

1 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
2 Centre of Evidence Based Dermatology, School of Medicine, University of Nottingham, UK
3 Department of Non-communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, UK
4 Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Oral prednisolone is the mainstay of treatment for bullous pemphigoid, a rare auto-immune blistering skin disorder of older people. Moderate to high dose treatment is often initiated in secondary care, but then continued in primary care. The aim is to use 'big data' and missing data algorithms to describe long-term oral prednisolone prescribing in UK primary care for adults with bullous pemphigoid, for the first time ever.

A cohort study using routinely collected data from the Clinical Practice Research Datalink was undertaken to identify people (≥18 years) with an incident diagnosis of bullous pemphigoid in the UK between 1998-2017. Oral prednisolone exposure was characterised in terms of the proportion of individuals with bullous pemphigoid prescribed oral prednisolone following diagnosis. Prednisolone dose and duration were extracted when available and imputed when missing. These data were often missing when prescriptions were issued with information restricted to the free text field, such as "Take as indicated by your dermatologist". Implausible and missing values were handled using the DrugPrep algorithm with the decisions described and validated by Joseph et al. (2019). Following the diagnosis of bullous pemphigoid, 2,312 (69.6%) of 3,322 people with incident bullous pemphigoid received a prescription for oral prednisolone in primary care. Of the users, only 321 (13.9%) people had complete data for all prescriptions. For the remaining people, the dose, start date, or treatment duration were imputed for at least one prescription. The median duration of exposure was 10.6 months (IQR 3.4 to 24.0). Of those prescribed prednisolone, 71.5% were continuously exposed to prednisolone for >3 months, 39.8% for >1 year, 14.7% for >3 years, 5.0% for >5 years, and 1.7% for >10 years. The median cumulative dose was 2,974mg (IQR 1,059 to 6,456).

A high proportion of people with incident bullous pemphigoid are treated with oral prednisolone in UK primary care. The potential iatrogenic risks posed to this population of predominately older people are high. Clear communication between primary and secondary care and consideration of steroid-sparing alternatives may be appropriate, and, where prednisolone is deemed the safest options, appropriate monitoring and prophylaxis for potential side effects are important measures.

**References:** *[1] Joseph RM, van Staa TP, Lunt M, Abrahamowicz M, Dixon WG. Exposure measurement error when assessing current glucocorticoid use using UK primary care electronic prescription data. Pharmacoepidemiology and drug safety. 2019;28(2):179-86.*

# POSTER SESSION 04

## P04-05 Exploring risk stratification in cardiomyopathies using a deep learning approach for survival prediction

**Ilaria Gandin[1], Alessia Paldino[2], A. Pio d'Adamo[3], Matteo Dal Ferro[4], Marco Merlo[5], Giulia Barbati[1]**

1 Biostatistics Unit, Department of Medical Sciences, University of Trieste, Italy
2 Department of Life Sciences, University of Trieste, Italy
3 Department of Medical Sciences, University of Trieste; Institute for Maternal and Child Health IRCCS "Burlo Garofolo", Trieste, Italy
4 Cardiothoracic Department, Azienda Sanitaria Universitaria Giuliano Isontina, Trieste, Italy
5 Department of Medical Sciences, University of Trieste; Cardiothoracic Department, Azienda Sanitaria Universitaria Giuliano Isontina, Trieste, Italy

Cardiomyopathies are primary myocardial disorders characterized by an important genetic background. Most severe outcomes are cardiovascular (CV) death and life-threatening arrhythmias. An appropriate risk stratification, which can be relevant for clinical management of patients, is still lacking. This is a preliminary study aimed to provide a survival prediction model for a better risk stratification.

We consider a cohort of 1453 patients enrolled in the Heart Muscle Disease Registry of Trieste (Italy), one of the largest and best characterized cardiomyopathy cohorts. Cause-specific survival models will be estimated for two endpoints: 1) CV death; 2) the occurrence of sudden cardiac death (SCD) or major ventricular arrhythmias (MVA). Baseline predictor factors will include 18 phenotypic variables spanning clinical, echocardiographic, ECG, Holter-monitoring domains. A subgroup of 467 patients have been screened for genetic rare variants in 35 candidate genes with Next Generation Sequencing (NGS) technology. Resulting variations will be annotated using multiple pathogenicity scores and then aggregated at gene-level.

Preliminary analysis includes comparison of cumulative incidence curves with the appropriate test for competing risk. Given the relevant number of candidate predictors, the possible presence of non-linear effects and interaction-effects (especially for genetic factors), in the next stage of the study we will apply the recently published multilayer deep neural network (DNN) for survival prediction by Sun et al. (2020). The algorithm consists in a feed-forward neural network with partial likelihood estimation with Efron's approach and L1 penalty for the resulting loss function. Prediction importance measures can be obtained using the LIME method.

The cohort appears appropriate for investigating long-term outcomes: the median follow-up is 110 months (IQR 26-208). Number of events are 180 and 241 for the CV death and SCD/MVA respectively. At the current stage of the analysis, we observe an unexpected trend in the reduction of arrhythmic events for patients carrying variant of uncertain significance (p=0.087) that is stronger in males (p=0.049).

Our first results support the need for a more complex multi-parametric prognostic model, accounting for the impact of distinct genes. The application of interpretability strategies will enhance the identification of clinically meaningful subgroups of patients.

# POSTER SESSION 04

## P04-06 Can animal studies on rodents help better understand Alzheimer's disease in humans?

**Tereza Dračková[1], Stanislav Katina[1,2], Petra Majerová[2], Andrej Kováč[2]**

1 Institute of Mathematics and Statistics, Masaryk University, Brno, Czech Republic

2 Institute of Neuroimmunology, Slovak Academy of Science, Bratislava, Slovakia

Alzheimer's disease (AD), the major cause of dementia, is a widespread neurodegenerative disease induced by misfolded protein tau. The neurofibrillary degeneration significantly correlates with disease progression, thus better understanding of the tau cascade represents an important target for potential therapeutic strategies for the clinical course of AD.

In this study we investigate and compare the protein expression in choroid plexus tissue from transgenic rodent models expressing human truncated tau with three microtubule-binding repeats (SHR24) and spontaneously hypertensive rats (SHR) used as controls. SHR24 line satisfied several key histological criteria typical for neurofibrillary degeneration in human AD, the rats developed tau pathology in the spinal cord and partially in the brainstem, and in the motor cortex. Choroid plexus tissue was isolated from 4 to 14 months old rats (with 2 months step). Our data contain normalized abundances of 11 245 proteins in 52 animal samples. Also raw abundance, spectral count and other variables are included. 2209 proteins were identified by 3 or more peptides. In average a protein was identified by 9 peptides.

Data were assessed using multidimenstional statistical methods as PCA, hierarchical clustering and PLS-DA in order to find differences in protein expression between transgenic and contol rats. The R software was used to carry out the statistical analysis.

**Acknowledgment:** *The work was supported (partly) by specific research of Masaryk University as support for student projects (MUNI/A/1615/2020).*

**References:** *[1] Zilka N, Z Kazmerova, S Jadhav, P Neradil, A Madari, D Obetkova, O Bugos, M Novak. Who fans the flames of Alzheimer's disease brains? Misfolded tau on the crossroad of neurodegenerative and inflammatory pathways. Journal of Neuroinflammation 9,1 (2012): 1–9. [2] Levarska, L, N Zilka, S Jadhav, P Neradil, M Novak. Of rodents and men: the mysterious interneuronal pilgrimage of misfolded protein tau in Alzheimer's disease. Journal of Alzheimer's Disease 37,3 (2013): 569–577. [3] Novak P, S Katina, et al. Safety and immunogenicity of the tau vaccine AADvac1 in patients with Alzheimer's disease: a randomised, double-blind, placebo-controlled, phase 1 trial. The Lancet Neurology. 2017 Feb 1;16(2):123–34.*



POSTER SESSION 04

## P04-07 Non-linear and non-additive associations between the pregnancy exposome and birthweight

**Elena Colicino[1], Federico Ferrari[2], Whitney Cowell[1], Megan M. Niedzwiecki[1], Nicolo Foppa Pedretti[1], Anu Joshi[1], Robert O. Wright[1,3], Rosalind J. Wright[1,3,4]**

1 Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York NY, United States

2 Department of Statistical Science, Duke University, Durham NC, United States

3 Department of Pediatrics, Kravis Children's Hospital, Icahn School of Medicine at Mount Sinai, New York NY, United States

4 Institute for Exposomic Research, Icahn School of Medicine at Mount Sinai, New York NY, United States

Birthweight is an indicator of fetal growth and environmental-related alterations of birthweight have been linked with multiple disorders and conditions progressing into adulthood. Although a few studies have assessed the association between birthweight and the totality of exposure, herein 'exposome', in maternal urine and cord blood; no prior research has considered a) the maternal serum prenatal exposome, which is enriched for hormones, and b) non-linear and synergistic associations among exposures.

We measured the maternal serum exposome during pregnancy using an untargeted metabolomics approach and birthweight for gestational age (BWGA) z-score in 410 mother-child dyads enrolled in the PRogramming of Intergenerational Stress Mechanisms (PRISM) cohort. We leveraged a Bayesian factor analysis for interaction to select the most important metabolites associated with BWGA z-score and to evaluate their linear, non-linear and non-additive associations. We also assessed the primary biological functions of the identified proteins using the MetaboAnalyst, a centralized repository of curated functional information. We compared our findings with those of a traditional metabolite-wide association study (MWAS) in which metabolites are individually associated with BWGA z-score.

Among 1110 metabolites, 46 showed evidence of U-shape associations with BWGA z-score. Most of the identified metabolites (85%) were lipids primarily enriched for pathways central to energy production, immune function, and androgen and estrogen metabolism, which are essential for pregnancy and parturition processes. Metabolites within the same class, i.e. steroids and phospholipids, showed synergistic relationships with each other.

Our results support that the aspects of the metabolic exposome during pregnancy contribute linearly, non-linearly and synergistically to variation in newborn birthweight.

**References:**

*[1] Ferrari, F., Dunson, D. B., 2020. Bayesian Factor Analysis for Inference on Interactions. Journal of the American Statistical Association.*

*[2] Pang, Z., et al., 2020. MetaboAnalystR 3.0: Toward an Optimized Workflow for Global Metabolomics. Metabolites. 10, 186*

# POSTER SESSION 05

POSTER SESSION 05 | Machine Learning

## P05-01 Machine learning to support Reinke's edema diagnosis from voice recordings

**Mari Carmen Robustillo Carmona, Carlos Javier Pérez Sánchez, María Isabel Parra Arévalo, Mario Madruga Escalona**

Department of Mathematics, University of Extremadura, Badajoz, Spain

Voice diseases mean an important disorder that affects people's communication. One of the most prevalent voice disorders is Reinke's edema (RE), an inflammatory disease of the vocal folds, which makes the voice to become deep and hoarse. Machine learning techniques fed by acoustic features extracted from voice recordings can be used as a non-invasive and low-cost tool to diagnose RE. Specifically, several feature selection and classification models have been implemented and evaluated to determine how accurate they are to distinguish between healthy and pathological voices, as well as to discover the most relevant features for the detection of RE. Two databases have been considered: i) the commercial database MEEI [1] (53 healthy subjects and 25 RE patients), and ii) our own database collected at the San Pedro de Alcántara Hospital in Cáceres, Spain (30 healthy subjects and 30 RE patients). Voices were recorded following a research protocol based on the phonation of the vowel /a/ in a sustained way. Different acoustic features were extracted from the voice recordings. After that, a feature selection employing the recursive feature elimination technique and several classifiers were applied to the different selected predictors. Ten models were analyzed, being classified classified into ensemble and non-ensemble ones. They include a broad variety of methods, such as decision trees, close neighbors, neural networks, support vector machines, Bayesian classification, regression analysis and linear discriminant analysis. All classification models were validated using a 10-fold stratified cross-validation, this process is replicated a total of 500 times, taking a new partition of the data in each repetition, so the mean and standard deviation of each evaluation metric is calculated.

The best performance has been achieved with an ensemble model based on neural networks, obtaining an accuracy of 100% at MEEI database and 95.49% at our own database. The obtained results are competitive with those found in the scientific literature, so that these techniques could be successfully used within a computer aided system to support the diagnosis of Reinke's edema disease.

**References:** *[1] Massachusetts Eye and Ear Infirmary (MEEI), Voice disorders database, Version 1.03 (cdrom), Lincoln Park, NJ: Kay Elemetrics Corporation, 1994.*

# POSTER SESSION 05

## P05-02 The Need for Expanded Standardized Reporting for Machine Learning Methods in Clinical Prediction

**Kelly Reeve[1] *, Begum Irmak On Seker[2] *, Heidi Seibold[3], Ulrich Mansmann[2], Ulrike Held[1]**

1 Epidemiology, Biostatistics & Prevention Institute and CRPP Precision MS, University of Zurich, Switzerland

2 Institute for Medical Information Processing, Biometry & Epidemiology, Ludwig-Maximilians-Universität, Germany

3 Institute of Computational Biology, HelmholtzZentrum München, Germany

* Equal contribution

Reporting of clinical prediction modeling is known to be poor, which hampers risk of bias assessment and validation and ultimately hinders translation of this research into clinical practice. Tools such as TRIPOD (Collins et al., 2015) and PROBAST (Moons et al., 2019) aim to improve reporting and enhance understanding of such studies, however, the current surge in popularity of machine learning (ML) may compound these reporting problems. Although Collins and Moons announced the development of an ML-specific TRIPOD in The Lancet in 2019, researchers in this field need basic guidance today. In this work we aim to: (1) highlight the major reporting deficiencies according to current guidelines, (2) describe the ML-specific analysis details currently being reported, and (3) make suggestions on which current items require communication in greater detail and which additional items should be introduced.

This work is based on an ongoing Cochrane Review of clinical prediction models in multiple sclerosis (MS). Of the 58 studies included after screening the initial database search results, 24 utilize ML methods to predict future clinical outcomes in MS. Even though a majority of these ML studies have been published after the introduction of TRIPOD in 2015, unclear reporting of data sources, participants, missing data/loss to follow up, and final models persists. The details reported regarding model development and performance evaluation vary greatly, suggesting that authors may not be aware of the details necessary for study assessment or of best practice regarding data pre-processing, hyperparameter tuning, nested resampling, and clinically relevant evaluation. While current tools provide a strong foundation, ML-based clinical prediction studies require reporting of additional details which are not yet addressed by these tools. Several PROBAST items require re-interpretation for fair assessment of the ML studies. Specific guidance with respect to sample size requirements, data leakage, hyperparameter tuning, optimism-adjusted model performance, and model sharing is necessary. A major challenge in creating such guidance will be to capture all algorithms in use and yet-to-be-created, while also helping the user to assess studies within the context of each specific algorithm.

**References:** *[1] Gary S. Collins, Johannes B. Reitsma, Douglas G. Altman, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. Ann Intern Med. 2015; 162:55-63. doi:10.7326/M14-0697. [2] Karel G.M. Moons, Robert F. Wolff, Richard D. Riley, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med. 2019; 170:W1-W33. doi:10.7326/M18-1377*

# POSTER SESSION 05

## P05-03 Virtual biopsy in action: a radiomic-based model for CALI prediction

**Francesca Ieva[1,2], Giulia Baroni[1], Lara Cavinato[1], Chiara Masci[1], Guido Costa[3,4], Francesco Fiz[5], Arturo Chiti[4,5], Luca Viganò[3,4]**

1 MOX laboratory, Department of Mathematics, Politecnico di Milano, Italy
2 CADS - Center for Analysis, Decisions and Society, Human Technopole, Milan, Italy
3 Division of Hepatobiliary & General Surgery, Department of Surgery, Humanitas Clinical & Research Center – IRCCS, Milan, Italy
4 Department of Biomedical Sciences, Humanitas University, Milan, Italy
5 Department of Nuclear Medicine, Humanitas Clinical and Research Center – IRCCS, Milan, Italy

**Background & Aims:** Chemotherapy-associated liver injuries (CALI) have a major clinical impact, but their non-invasive diagnosis is still an unmet need. The present work aims at elucidating the contribution of radiomic analysis to diagnosis of sinusoidal dilatation, nodular regenerative hyperplasia (NRH) and non-alcoholic steatohepatitis (NASH).
**Methods:** Patients undergoing liver resection for colorectal metastases after oxaliplatin- or 67 irinotecan-based chemotherapy between January 2018 and February 2020 were retrospectively analyzed. Radiomic features were extracted from a standardized volume of non-tumoral liver parenchyma outlined in the portal phase of preoperative post-chemotherapy computed tomography (CT). Multivariate logistic regression models and CART were applied to identify predictors of CALI and internally validated.
**Results:** Overall, 78 patients were analyzed. Three fingerprints derived from radiomic features were considered as independent predictors of grade 2-3 sinusoidal dilatation: GLRLM_f3 (OR=12.25), NGLDM_f1 75 (OR=7.77), and GLZLM_f2 (OR=0.53). The combined clinical/radiomic predictive model had 82% accuracy, 64% sensitivity, and 91% specificity (AUC=0.87 vs AUC=0.77 of the model without radiomics). Three radiomic parameters were independent predictors of NRH: conventional_HUQ2 (OR=0.76), GLZLM_f2 (OR=0.05), and GLZLM_f3 (OR=7.97). The combined clinical/radiomic model had 85% accuracy, 81% sensitivity, and 86% specificity (AUC=0.91 vs AUC=0.85 without 80 radiomic features). One radiomic feature was associated with NASH: conventional_HUQ2 (OR=0.79). Steatohepatitis was predicted with 91% accuracy, 86% sensitivity, and 92% specificity (AUC=0.93 vs AUC=0.83 without radiomic features). In the validation set, accuracy was 72%, 83 71%, and 91% for sinusoidal dilatation, NRH, and NASH, respectively.
**Conclusions:** Radiomic analysis of liver parenchyma may provide a signature that, in combination with clinical and laboratory data, improves diagnosis of CALI.

## P05-04 Creation of adverse drug reactions assessment tool

**Roma Puronaitė[1,2,3], Greta Burneikaitė[2,3], Dovilė Ramanauskaitė[2,3], Justas Trinkūnas[2,4], Monika Grigentytė[2,3], Audronė Jakaitienė[1,3], Laimis Dambrauskas[2,3], Edita Kazėnaitė[2,3]**

1 Faculty of Mathematics & Informatics, Institute of Data Science & Digital Technologies, Vilnius University, Lithuania
2 Vilnius University Hospital Santaros Klinikos, Lithuania
3 Faculty of Medicine, Vilnius University, Lithuania
4 Department of Information Systems, Vilnius Gediminas Technical University, Lithuania

**Clinical research problem (context):** International health organizations around the world are actively pursuing pharmacovigilance by collecting data on adverse drug reactions (ADRs). ADRs profile and clinical manifestation become most important when choosing best treatment options and considering patient related drug safety decisions. Due to different levels of drug use, peculiarities of medical practice, population age and comorbidities, the epidemiological incidence of ADRs varies. Various scales or classifiers are used for the analysis of ADRs, but there is no universally agreed tool in clinical practice. A detailed analysis of ADRs may contribute to the country's pharmacovigilance and ease the burden on the health care system.
**Statistical challenges (objective):** We aimed to develop a new ADR's assessment tool to be used in clinical practice by adapting most common ADR's assessment scales. The purpose of this tool is to answer to the clinician whether ADRs are clinically relevant. Thus, there is a need to minimize the number of items for convenience in everyday use with minimal loss of accuracy.
**Statistical methods:** The study analyzes the documentation of patients for whom ADR was confirmed. Epidemiological data, drug classes and patient risk factors were analyzed. The most commonly described ADR assessment scales or questions of ADR type, causality, predictability, severity and preventability in the literature were selected. Data was evaluated on selected scales by two independent investigators and statistical analysis was performed to assess the inter-rater agreement. Machine learning was used to reduce a set of items and create a combined reduced questionnaire as a new ADRs assessment tool. Data were analyzed using the R software.
**Results and Conclusions:** The significant variables from the different scales were determined using a machine learning algorithm. This method was used for the creation of a new ADR assessment tool which after validation will be implemented into the health information system and used in clinical practice.



# POSTER SESSION 05

## P05-05 Use of machine learning models combined with innovative interpretation methods to identify prognostic factors

**Romane Péan, Adrien Darbier, Rim Ghorbal, David Pau, Cyril Esnault, Mélina Gilberg, Julien Dupin, Alexandre Civet**
Roche, France

**Objectives:** The objective of this exploratory analysis is to identify the prognostic factors of the complete pathological response (pCR) post neoadjuvant treatment using Machine Learning methods. This research complements a retrospective national observational study including HER2+ early breast cancer patients receiving trastuzumab-based neoadjuvant and adjuvant therapy in France.
**Methods:** Prognostic factors were identified using a 2-step method. The first step consists in using a Support Vector Machine (SVM) model, trained and optimized through grid search with a cross-validation on a learning dataset (75% of patients), then tested on a validation dataset (25% of patients) including the clinical characteristics of the patients as well as the characteristics of the centers. Then the model was applied to the analysable population (n = 301) to calculate the global accuracy, sensitivity and specificity. The second step consists in identifying the variables influencing the prediction of pCR, using three agnostic interpretation methods: SHAP (with SHAPLEY values) to identify the most important variables and the magnitude of their impact on the model, Partial Dependence Plot (PDP) to visualize the overall impact (negative or positive) of each variable on the model, and Local Interpretable Model-agnostic Explanations (LIME) to further explain the prediction of pCR for each patient.
**Results:** The final SVM model presented an accuracy of 0.65, a sensitivity and a specificity of 0.47 and 0.80 respectively. The SHAP results for this model showed that the variables with the most impact on pCR status are the stage at diagnosis, the T classification, the presence of invaded lymph nodes and the estrogen receptors. Individual results from LIME and PDP approaches tended to reinforce global results from the SHAP method. Results are consistent with knowledge from literature.
**Conclusion:** We showed the interest of new interpretation tools for identifying potential prognostic factors from more complex Machine Learning models. These models allow a consideration of all the variables without a priori which is a strength for a more in-depth analysis of medical data. Other Machine Learning models (K-nearest neighbors, DecisionTree, RandomForest, XGBoost, AdaBoost) are being tested and will be compared to SVM in order to improve predictive quality of the model.

# POSTER SESSION 06

POSTER SESSION 06 | Epidemiology

## P06-01 The impact of left-truncation of exposure in environmental case-control studies: evidence from breast cancer risk associated with airborne dioxin

**Yue Zhai[1,2,3,4], Amina Amadou[1,5], Catherine Mercier[2,3,4], Delphine Praud[1,5], Elodie Faure[6], Jean Iwaz[2,3,4], Gianluca Severi[7], Francesca Romana Mancini[7], Thomas Coudon[1,5], Béatrice Fervers[1,5], Pascal Roy[2,3,4]**

1 Département Prévention Cancer Environnement, Centre Léon Bérard, Lyon, France
2 Service de Biostatistique -Bioinformatique, Pôle Santé Publique, Hospices Civils de Lyon, Lyon, France
3 Laboratoire de Biométrie et Biologie Évolutive, CNRS UMR 5558, Villeurbanne, France
4 Université Claude Bernard Lyon 1, Lyon, France
5 Inserm UA 08 Radiations: Défense, Santé, Environnement, Lyon, France
6 Centre de Recherche en Epidémiologie et Santé des Populations (CESP, Inserm U1018), Facultés de Médecine, Université Paris-Saclay, UPS UVSQ, Gustave Roussy, Villejuif, France
7 Department of Statistics, Computer Science and Applications (DISIA), University of Florence, Italy

**Background:** In epidemiology, left-truncated data may bias exposure effect estimates. We analyzed the bias induced by left truncation in estimating breast cancer risk associated with exposure to airborne dioxins.
**Methods:** Simulations were run with exposure estimates from a geographic information system-based metric and considered two hypotheses for historical exposure, three scenarios for intra-individual correlation of annual exposures, and three exposure-effect models. For each correlation/model combination, 500 nested matched case-control studies were simulated and data fitted using a conditional logistic regression model. Bias magnitude was assessed by estimated odds-ratios (ORs) vs. theoretical relative risks (TRRs) comparisons.
**Results:** With strong intra-individual correlation and continuous exposure, left truncation overestimated the Beta parameter associated with cumulative dioxin exposure. Versus a theoretical Beta of 4.17, the estimated mean Beta (5%; 95%) was 73.2 (67.7; 78.8) with left-truncated exposure and 4.37 (4.05; 4.66) with lifetime exposure. With exposure categorized in quintiles, the TRR was 2.0, the estimated $OR_{Q5}$ vs. $_{Q1}$ 2.19 (2.04; 2.33) with truncated exposure vs. 2.17 (2.02; 2.32) with lifetime exposure. However, the difference in exposure between Q5 and Q1 was 18x smaller with truncated data, indicating an important overestimation of the dose effect. No intra-individual correlation resulted in effect dilution and statistical power loss.
**Conclusions:** Left truncation induced substantial bias in estimating breast cancer risk associated with exposure with continuous and categorical models. With strong intra-individual exposure correlation, both models detected associations, but categorical models provided better estimates of effect trends. This calls for careful consideration of left truncation-induced bias in interpreting environmental epidemiological data.

# POSTER SESSION 06

## P06-02 Lower Limit of Quantification in various distributed data: examining confidence interval variations

**Tanja Bülow[1], Ralf-Dieter Hilgers[1,2], Nicole Heussen[1,2]**

1 Department of Medical Statistics, RWTH Aachen University, Germany
2 Center for Biostatistics and Epidemiology, Sigmund Freud Private University Vienna, Austria

Single or multiple lower limits of quantification (LLOQ) appear in concentration measurement data due to one or more laboratories involved in the quantification of observations, in which some observations are too low to be quantified with required precision. As the missing data mechanism is not random, most statistical methodology to handle missing data is not applicable. In clinical practice, simple imputation methods are often used to receive substitution values for the missing observations. Nevertheless, they lead to severe bias in estimating parameters such as the mean and variance. Even procedures relying on the assumption of normally distributed concentration data are little robust against distributional model misspecification [1]. Interpretation of confidence intervals (CI) rather than only point estimates would lead to an assessment of precision of the respective point estimates. The objective is to investigate robustness and precision of different types of CI's applied to newly developed parametrical point estimation methods for different distributional assumptions. With this we aim to show the advantage of interpreting CI's rather than point estimates in this missing data situation.
We transfer existing maximum-likelihood based approaches relying on the normal distribution assumption for LLOQ's to other distributional assumptions. With suitable approaches at hand for specific distributions, we not only investigate the robustness of the point estimates for mean and variance, but also compare bootstrap CI's with parametrical CI's to evaluate the characteristics of the CI types with respect to coverage probability and precision. The performance of the most popular simple imputation method is compared to our approach. The proposed procedure will be demonstrated using data from a cohort study [2], in which the underlying distribution varies depending on the chosen clinical parameter.
The variety of distributional assumptions for which the methods are applicable gives the applicant a broadly usable tool to handle LLOQ affected data with appropriate approaches. Before choosing the appropriate method, the distribution of the data at hand should be examined. In case of sureness about the underlying distribution, interpretation of CI's will broaden the possible conclusion. Under uncertainty, CI's prove to deliver more robust interpretation possibilities for the intriguing parameters than point estimates.

**References:** *[1] Berger, T, Hilgers, R-D, Heussen, N. Observations below multiple lower limits of quantification: How to estimate the mean and variance. Biometrical Journal. 2019; 61: 1258– 1269. [2] Kempf, K, et al.. The epidemiological Boehringer Ingelheim employee study–Part I: Impact of overweight and obesity on cardiometabolic risk. Journal of Obesity. 2013; 159123.*

# POSTER SESSION 06

## P06-03 Diversity indices and statistical methods used in studies addressing dysbiosis applied to compositional data of the gut microbiota

**Asuka Nemoto**

Teikyo University Graduate School of Public Health, Japan

**Background:** Technological innovations, such as high-throughput sequencing, have promoted the direct determination of the taxonomic composition of the gut microbiota in samples, boosting research on the role of the gut microbiota in human health. We focused that some diseases result from dysbiosis (imbalances in the microbial community).
**Objectives:** To illustrate the diversity measures and statistical methods used in articles addressing dysbiosis through a literature review.
**Methods:** We searched PubMed using the search terms "microbiome, gut, 16s, taxonom*, diversity, richness" on December 1, 2020. All abstracts identified were reviewed to include studies in humans or animal models addressing the research question on the association between dysbiosis and disease. The diversity measures and statistical methods reported were identified.
**Results:** The initial searches yielded 144 articles. After screening titles/abstracts and selecting studies, 33 articles were analyzed. The measures used by the authors of those articles for within-group diversity (alpha diversity) were Chao1 or ACE for richness and Shannon or Simpson for diversity (or evenness). the measures used for between-group diversity were "distance" using UniFrac, Bray-Curtis method, etc. Parametric or non-parametric methods for comparisons between groups of continuous variables were used in those articles for the alpha diversity measure. The methods used for "distances" between the groups were visual illustration by principal coordinates analysis and test by PERMANOVA.
**Discussion:** We found the following methods used in most studies: comparing alpha diversity between groups; comparing groups based on the "distance" between samples. There seems to be a need for other approaches adequate the research questions from the following aspects. Microbiome compositional data are essentially relative abundance of each microbe in the sample and are multivariate data structured in a multilevel hierarchy. Each microbe influences the host in multiple ways that are partially similar and partially different from each other.
**Conclusion:** A proposal and adoption of new methods taking the following points into account are awaited: all microbes are not needed considering; microbes may be mutually exchangeable when focusing on the specific biochemical reaction; utilization of other information than the relative abundance of each microbe; a multilevel hierarchical inter-variable structure.

## P06-04 Subsequent primary neoplasms in bladder cancer patients

**Lucie Pehalova[1,2], Denisa Krejci[1,2], Ladislav Dusek[1,2]**

1 Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic
2 Institute of Biostatistics and Analyses, Medical Faculty, Masaryk University, Brno, Czech Republic

**Background:** Numbers of patients who develop subsequent primary neoplasms have markedly increased recently. The development of subsequent primary neoplasms nowadays presents a major clinical problem because subsequent primary tumours are the main cause of morbidity in a large proportion of long-term cancer survivors [1]. This makes subsequent primary neoplasms a challenge and an opportunity for future oncology research.
**Objective:** This study aimed to perform a comprehensive analysis documenting the risk of subsequent primary neoplasms in patients with bladder cancer.
**Methods:** The Czech National Cancer Registry was the main data source, containing records of all cancers diagnosed in the Czech Republic since 1977. The risk of development of subsequent primary neoplasm after bladder cancer was evaluated by the standardised incidence ratio (SIR) with corresponding 95% confidence interval (CI) [2].
**Results:** A total of 71,982 patients with bladder cancer were diagnosed in 1977–2018, out of which 12 375 (17.2%) developed subsequent primary neoplasm. Bladder cancer patients of younger age, early clinical stage, and male sex were shown to be at higher risk of development of subsequent primary neoplasm. The risk of development of any malignant neoplasm (C00–C97) was approximately 1.7 higher in persons with bladder cancer than in the general Czech population (SIR:1.66; CI:1.64–1.69). The highest-risk diagnoses that occurred after bladder cancer were lung cancer (SIR: 2.79; CI: 2.67–2.91) and laryngeal cancer (SIR: 2.52; CI: 2.08–3.03). The median time to the development of a subsequent neoplasms (for all malignant neoplasm combined) was 5.6 years. The shortest times were recorded for oesophageal cancer (4.5 years) and laryngeal cancer (4.7 years). In contrast, the longest times were reported for thyroid cancer (7.1 years) and chronic lymphocytic leukaemia (6.6 years).
**Conclusion:** To our knowledge, this is the first population-based study documenting the risk of incidence of subsequent primary neoplasms in bladder cancer patients that takes into account such a long time period. Conclusions from the performed analysis might be useful to set up correctly follow-up procedures for bladder cancer patients in specialised centres and in GP surgeries. A correctly adjusted follow-up might improve the prognosis of patients.
**References:** *[1] Zaorsky NG, Churilla TM, Egleston BL, et al. Causes of death among cancer patients. Ann Oncol. 2017;28(2):400-407.*
*[2] Schoenberg BS, Myers MH. Statistical methods for studying multiple primary malignant neoplasms. Cancer. 1977;40(4 Suppl):1892-1898.*



# POSTER SESSION 06

## P06-05 On heuristic detection of maternal-age-related increase of birth defect risk: Experience, issues, alternatives

**Jan Klaschka[1], Marek Malý[1], Antonín Šípek[2]**

1 Department of Statistical Modelling, Institute of Computer Science of the Czech Academy of Sciences, Czech Republic
2 Department of Medical Genetics, Thomayer University Hospital, Czech Republic

One of the focuses of our research is detection of increased congenital anomaly (birth defect) risk related to high or low maternal age. The size and onset of the increase depend on the anomaly type. When events (anomalies) are frequent and the risk increase is big and takes place far from the age scale ends, joinpoint Poisson regression, for instance, may be a right choice. It fits well, for example, Czech 2013 – 2017 Down syndrome data (on both born children and terminated pregnancies), where it shows a consistent risk increase along the entire age scale, first slow, and accelerated since the age of 32. Such methods may, nevertheless, fail for rare anomalies with a moderate risk increase at age close to extremes. For such situations, we have designed, and presented at ISCB 2020 [1] a heuristic method. Each year on the age scale splits the scale into two opposite tails. Risks in the two tails are compared by relative risk (RR) and Fisher test. The attribute of suspect risk increase belongs to a tail if it yields RR over 2 and significant unadjusted Fisher test, or is nested within such tail. A stronger attribute of verified risk increase is given to tails with the former attribute that are nested within a tail with significant Bonferroni-adjusted Fisher test. A new alternative method variant compares all tails with a common reference age interval from the lower to the upper quartile. (Considered are tails disjoint with the interval.) Otherwise, the definitions of suspect and verified risk increase remain the same. Numerical differences between the two method variants on real data are minor. For example, the former variant finds a verified risk increase at 18 or less years, and the new one at 19 or less in the 1992 – 2016 anencephaly incidence. Both variants assess equally the risk increase from 42 years as suspect. The new variant is, however, more logically consistent, as it always transforms, unlike the former one, monotone risk by age curves into monotone RR by age curves.
**Acknowledgements:** *Supported by Czech Health Research Council grant No. 17-19622A.*
**References:** *[1] Klaschka, J, Malý, M., Šípek, A. (2020) A simple heuristics for detection of age-related increase of risk of rare congenital anomalies. Poster, ISCB 2020.*

## P06-06 Linkage of national clinical datasets without patient identifiers using probabilistic methods

**Helen A. Blake[1,2], Linda D. Sharples[3], Katie Harron[4], Jan H. van der Meulen[1,2], Kate Walker[1,2]**

1 Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, UK
2 Clinical Effectiveness Unit, Royal College of Surgeons of England, UK
3 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, UK
4 University College London (UCL) Great Ormond Street Institute of Child Health, UK

Linkage of electronic health records from different sources is increasingly used to address important clinical and public health questions. However, obtaining linked data can be a complex and lengthy process, often involving transfer of sensitive patient information to a trusted third party. Alternative methods that do not require the use of patient identifiers would accelerate the use of linked datasets. We developed a step-by-step process for probabilistic linkage of national clinical and administrative datasets without patient identifiers and validated it against deterministic linkage using patient identifiers. Probabilistic linkage was carried out using seven indirect identifiers: small area of residence (Lower Super Output Area); hospital trust; date of surgery; responsible surgeon; age; sex; and surgical procedure. We used electronic health records from the National Bowel Cancer Audit (NBOCA) and Hospital Episode Statistics (HES) databases for 10,566 bowel cancer patients undergoing emergency surgery in the English National Health Service. Probabilistic linkage without patient identifiers linked 81.4% of NBOCA records to HES, compared to 82.8% using deterministic linkage. The approach had over 96% sensitivity and 90% specificity compared to deterministic linkage using patient identifiers. Of the 176 records that linked probabilistically but not deterministically, 143 (81%) agreed on small area of residence (Lower Super Output Area) and ≥4 other indirect identifiers, suggesting that most are true links and the specificity of the probabilistic linkage is therefore likely underestimated. No systematic differences were seen between patients that were and were not linked, and regression models for mortality and length of hospital stay according to patient and tumour characteristics were not sensitive to the linkage approach. Probabilistic linkage without patient identifiers was successful in linking national clinical and administrative datasets for patients undergoing a major surgical procedure. The approach can be used as an alternative to deterministic linkage using patient identifiers, or as a method for enhancing deterministic linkage. It has important implications as it allows analysts outside highly secure data environments to carry out the linkage process while protecting data security and maintaining – and potentially improving – linkage quality.

# POSTER SESSION 06

## P06-07 Chronic exposure to multiple air pollutant and risk of breast cancer: A nested case-control within the E3N cohort

**Bibi. F. Ngaleu[1], Béatrice Fervers[1,2], Thomas Coudon[1,2,3], Delphine Praud[1,2], Leny Grassot[1,2], Floriane Deygasa[2], Elodie Faure[4], Florian Couvidat[5], Gianluca Severi[4,6], Francesca Romana Mancini[4], Pascal Roy[7], Amina Amadou[1,2]**

1 Department of Prevention Cancer Environment, Centre Léon Bérard, Lyon, France
2 Inserm UA 08 Radiations: Défense, Santé, Environnement, Lyon, France
3 Ecole Centrale de Lyon, Ecully, France
4 Centre de Recherche en Epidémiologie et Santé des Populations (CESP, Inserm U1018), Facultés de Médecine, Université Paris-Saclay, UPS UVSQ, Gustave Roussy, Villejuif, France
5 National Institute for industrial Environment and Risks (INERIS), Verneuil-en-Halatte, France
6 Department of Statistics, Computer Science and Applications (DISIA), University of Florence, Italy
7 Service de Biostatistique -Bioinformatique, Pôle Santé Publique, Hospices Civils de Lyon, France

**Background:** Studies have suggested that exposure to environmental pollutants, particularly those with endocrine disrupting properties, have a role in breast cancer (BC) development. Exposure to pollutant mixtures is recognized as the real-life scenario of populations. However, the effects of mixtures are seldom analyzed.

**Objectives:** In this study, we applied a new statistical method to assess the complex effect of exposure to a mixture of four xenoestrogens air pollutants (Benzo-[a]-pyrene (BaP), cadmium, dioxins, polychlorinated biphenyls 153 (PCB153)) on the risk of BC.

**Methods:** The study was conducted on 5,222 cases and 5,222 matched controls nested within the French E3N cohort from 1990 to 2011. Annual air concentrations of pollutants were simulated with the chemistry-transport model (CHIMERE) and were assigned to subjects using their geocoded residential history. Mean exposures were calculated for each subject from their cohort inclusion to their index date. We employed a new statistical approach, Bayesian Kernel Machine Regression (BKMR), to investigate the relative risk of the joint effect of co-exposure of four xenoestrogens pollutants on the risk of BC. Due to the high correlation between pollutants, a hierarchical selection of variables with 5000 iterations by a MCMC algorithm is performed. This quantifies the relative importance of each pollutant within the group using the probability of posterior inclusion. To account for the matching design, we used the fitted probit model with a vector (saturated on the matching variables), that was close to the values of the conditional logistic regression models. The estimation of the function will be visualized by examining the dose-response relationship between each exposure, the statistical interactions between pollutants, and the joint association between the mixture and the BC risk.

**Expected Results/ Conclusions:** The analyses are ongoing. This study is the largest to date to evaluate the impact of multi-pollutant mixtures on the risk of BC, the most common cancer in women. The results will allow our understanding of the mixture effect in the context of a highly nonlinear, dose-response relationship, and to estimate overall, single-exposure, as well as interactive health effects.

# POSTER SESSION 06

## P06-08 Use of innovative methods to estimate a reliable French pathological complete response rate on real world data

**Adrien Darbier, Romane Péan, Rim Ghorbal, David Pau, Cyril Esnault, Mélina Gilberg, Alexandre Civet, Julien Dupin**

Roche, France

**Objectives:** We aim to obtain the most reliable estimate of pathological Complete Response (pCR) rate post neo-adjuvant treatment by exploring a new approach and using several known methods, from a retrospective national observational study including HER2+ early breast cancer patients receiving trastuzumab-based neoadjuvant and adjuvant therapy in France.

**Methods:** As patient data (n=301) were only available for the sample of centers included in the registry (n=48), the exhaustive hospitalization database of center characteristics (PMSI, n=460) was used to estimate a representative French pCR rate. The estimation was done using a 5-step approach based on the generation of a modified sample of centers: Firstly, clinical relevance and correlation with pCR allowed selecting center characteristics. Secondly, a propensity score (PS), using either logistic regression or the Covariate Balancing Propensity Score (CBPS) [1], grouped the selected characteristics. Thirdly, the PS was used to create the modified sample, exploring two methods: Inverse Probability of Treatment Weighting (IPTW) and an innovative oversampling-based matching method [2]. Next, the balance was assessed by the standardized mean differences (SMD) of each center characteristic between the modified sample and the PMSI. Finally, the modified samples were used to estimate the pCR rate at patient level.

**Results:** Two characteristics, center types and regions, were retained to estimate PS. A satisfying overlap between estimated PS distributions was observed (AUC logistic regression: 0.63 / AUC CBPS: 0.62). Each combination between PS estimation methods and sample correction methods improved the balance with the PMSI compared to the original cohort. A negligible imbalance (SMD < 10%) on all characteristics was only observed for the logistic regression + IPTW and CBPS + matching combinations. The pCR rates estimated on these samples (resp. 42.7% and 42.3%) were close to that observed in the initial cohort (42.9%).

**Conclusions:** Data from a dedicated registry were joined with the PMSI to reliably estimate the pCR rate in the French population, by combining propensity matching and innovative methods. Such methodology could be useful to extrapolate results from a single study to a more general population.

**References:** *[1] Imai K., Ratkovic M., (2014). Covariate Balancing Propensity Score, Journal of the Royal Statistical Society, Series B (Statistical Methodology). [2] Esnault C., Génin M., Ekhtiari S., Civet A., (2019). Projeter l'efficacité d'un traitement en vie réelle (VR) à partir d'une étude clinique randomisée (ECR) : une approche utilisant le suréchantillonnage, Poster présenté au 11ème Colloque Données de Santé en vie réelle.*

## P06-09 Sensitivity analyses for measurement error using regression calibration or simulation-extrapolation

**Linda Nab[1], Rolf H.H. Groenwold[1,2]**

1 Department of Clinical Epidemiology, Leiden University Medical Center, Netherlands
2 Department of Biomedical Data Sciences, Leiden University Medical Center, Netherlands

Measurement error is common in various domains of epidemiology. In case of random measurement error, measurements randomly fluctuate around their true value. Random measurement error in an exposure introduces bias in the exposure outcome association. Methods to correct for this bias are regression calibration and simulation-extrapolation, which can also be applied in the absence of validation data about the true measurements. Both methods require assumptions about the variance of the random measurement error. A simulation study was conducted comparing the performance of simulation-extrapolation and regression calibration for correction of measurement error of the exposure variable. The performance of the two methods was evaluated assuming the absence of validation data, yet with correct assumptions about the measurement error variance. Studied scenarios differed regarding sample size, reliability of the measurements, and precision of the estimated measurement error variance. Simulation-extrapolation and regression calibration were evaluated in terms of bias, mean squared error, and 95% confidence interval coverage. Across the evaluated scenarios, regression calibration generally resulted in less bias than simulation-extrapolation (median percentage bias of 2% for regression calibration (interquartile range 2%) vs -9% for simulation-extrapolation (interquartile range 21%)). Simulation-extrapolation was however generally more efficient in terms of mean squared error (median percentage decrease in mean squared error of 16% for simulation-extrapolation vs regression calibration (interquartile range 14%)). The quantification of the performance of the two methods in a broad range of settings was used as the input for a framework guiding sensitivity analyses for random exposure measurement error in the absence of validation data.

# POSTER SESSION 07

## POSTER SESSION 07 — Survival analysis

### P07-01 Modeling Non-Proportional Hazards for Overall Survival Time for Cancer Treatments

**Shoichi Irie**

Department of Data Analysis, Topoi, Inc., Japan

**Background:** The Cox's proportional hazards regression is a quite common and useful model in many medical researches. What it essentially means is that the ratio of the hazards for any two individuals is constant over time. However, the proportional hazards (PH) assumption is potentially a major constraint. When the PH assumption is violated, the hazard ratio (HR) is expressed as a function of time.

**Objective:** In this study, we confirmed that the overall survival (OS) time for cancer treatments does not satisfy the PH assumption in many cases, and then we proposed a way of modeling such cases where the PH assumption is violated.

**Method:** Among the cancer drugs used in Japan, we chose 50 drugs which include the Kaplan-Meier curve in its attached document, and we confirmed whether the proportionality is satisfied. We also developed a non-proportional hazards model in which the HR is considered to be a time-dependent function.

**Results and Conclusion:** Among 50 OS time figures, 35 (70%) cases did not satisfy the PH assumption significantly. It was confirmed that the HR as a time-dependent function can be successfully modeled in all 50 cases by expressing it with some parameters such as convergent value and measure.

### P07-02 Sensitivity of results to missing data for clinical trials with discrete, longitudinal outcome measurements

**Isabelle L. Smith[1], Linda Sharples[2], Jane Nixon[1]**

1 Clinical Trials Research Unit, University of Leeds, United Kingdom

2 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, United Kingdom

**Background:** Multi-state models (MSM) are structures that can represent transition of patients through different disease categories (states). Analysis of discrete, interval censored longitudinal data using MSM is established and can lead to increased power for clinical trials, compared to analysis of aggregated data using binary or time to event methods. However, longitudinal data often suffers from intermittent missing measurements, which may depend on the true disease state. If the state is derived from composite data, consideration of the quantity and reason for missing components, and the potential association with the latent (missing) state is required.

**Objectives:** To investigate methods for handling non-ignorable missing data in a MSM framework.

**Methods:** We investigate joint (selection) models for the multi-state process and the probability of missing data. Such selection models are equivalent to hidden Markov models, where an additional 'state' is used to represent a misclassification of the underlying latent state, whilst the observed data are assumed to be accurate. For interval-censored data, misclassification probabilities and transition intensities for the MSM may be estimated simultaneously using the 'msm' program in r-CRAN library. The model was applied to a dataset from a pressure ulcer prevention trial, where different assumptions for the missing state were considered.

**Results:** Exploration of the motivating trial dataset identified 'key' components in the definition of the disease that would be associated with the true disease state and would be informative if missing. Based on these key components, three candidate definitions for missing state mechanism were identified for our motivating example. Applying misclassification models demonstrated that there was important variation in the point estimate and precision (and therefore hypothesis tests) of treatments effects between different missing data assumptions. Further, the probability of missing state was higher for non-healthy true (latent) states compared to the healthy state for all definitions.

**Conclusions:** Sensitivity to non-ignorable missing data can be accommodated in MSM using hidden Markov models. The definition of missing data for an endpoint derived from composite data needs careful consideration in the context of the research setting.



POSTER SESSION 07

### P07-03 New Application of competing risks model in IgA nephropathy to explore the severity-dependent urinary remission

**Masako Nishikawa[1], Kentaro Koike[2], Keita Hirano[2,3], Tetsuya Kawamura[2]**

1 Clinical Research Support Center, The Jikei University School of Medicine, Japan

2 Department of Internal Medicine, The Jikei University School of Medicine, Japan

3 Department of Internal Medicine, Ashikaga Red Cross Hospital, Japan

**Outline of Clinical research and its problem:** Nationwide prospective cohort study of immunoglobulin A nephropathy (IgAN) has been conducted throughout Japan to confirm risk classification ability of severity grading systems (HG,CG,RG). Patients with IgAN were registered between April 1, 2005 and August 31, 2015. The primary outcome (PO) was a 50% increase in serum creatinine from baseline or dialysis induction, whichever is earlier. The secondary outcomes (SOs) were proteinuria remission and hematuria remission. SOs were preferred events. The follow-up data were collected every 6 months. The final observation date was January 31, 2018. The time-to-event data were treated as censored at the latest respective examination date, if the respective events were not confirmed. With conventional statistical method (log-rank test), the association of severity grading systems and PO was clear, however, their association with hematuria remission was unclear.

**The objective:** To make the association between the severity grading systems and hematuria remission clearer.

**Statistical methods:** Proposed statistical method is a new application of competing risks model with elaborate data handling. In the database, there were such data as the latest examination dates between PO and SOs were different in the same patients. And some patients experienced urinary remission after 50% increase in serum creatinine. While all patients transferred to dialysis never achieve urinary remissions, some patients can experience PO after urinary remission in real world. Since 50% increase in serum creatinine was surrogate for dialysis, we treated PO and respective SO in the framework of (semi-)competing risks. If the lengths of follow up records were different between PO and a SO in the same patient, we truncated the longer time-to-event data at the shorter one and used the event status at the truncated time in the competing risks analyses.

**Results:** The association of HG and the SOs was confirmed by Fine-Gray model [1] with contrasts or Gray's test, and the cumulative incidence function was shown by severity level. The influence of the SOs to the PO was confirmed by multivariate Cox regression with time-dependent remissions status, and cumulative joint incidence functions [2] after respective SO will be plotted.

**References:** *[1] Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association 94(446): 496-509, 1999. (2) Nishikawa M, Tango T, Ogawa M. Non-Parametric inference of adverse events under informative censoring. Statistics in Medicine 25: 3981-4003, 2006*

# POSTER SESSION 07

## P07-04 Frailty Multi-state Model with Time-dependent Covariate for Prediction of Colorectal Cancer Progression

**Robabeh Ghodssi-Ghassemabadi[1], Ebrahim Hajizadeh[1], Mahmoud Mahmoodi[2], Shaghayegh Kamyan[3]**

1 Department of Biostatistics, School of Medical Sciences, Tarbiat Modares University, Tehran, Iran

2 Department of Epidemiology & Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

3 Department of Clinical Oncology, Imam Hossein Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran

**Background:** Colorectal cancer (CRC) patients usually have a complex medical history. They often experience recurrence (loco-regional or metastases) as an intermediate event which increases the risk of mortality. Although the disease progression is highly dependent on clinical or pathological stages, but the disease course of the same stage patients can differ considerably. Multi-state models are used to describe the progression of a complex disease which occupies several states over time. Predictions in survival analysis become adjustable when considering intermediate events, so the result of multi-state survival models are more accurate than considering just one event [1]. Usually homogenous Markov modeling are applied to analyze multi-state models, but the homogeneity assumption is unrealistic when considering time-dependent covariate in modeling a disease course.

**Objective:** We evaluated the disease course of colorectal cancer by utilizing a parametric multi-state model with time-dependent covariates with and without frailty term in the model.

**Methods:** We obtained the data of newly-diagnosed CRC patients who had undergone curative surgery and admitted to the Clinical Oncology Department at Imam-Hossein Hospital, Tehran, Iran, between 2002 and 2013. The last date of follow-up was May 2018. Demographic characteristics and clinical data of all patients were obtained through their medical records and follow-ups. A non-recursive illness-death model was considered for modeling CRC evolution in which the initial state (1) was alive without recurrence, the transient state (2) was alive with recurrence, and the only absorbing state (3) was death of any causes. We used piecewise-constant approximation for Weibull transition-specific model and compared its result when incorporating log-normal frailty into the model [2].

**Result:** A total of 339 CRC patients with mean age 53.32 ± 11.44 years of were included. The median follow-up of the patients was 6.2 years. Of the whole, 40.12% of patients experienced recurrence of whom 80.1% died and 10.9% of patients died without recurrence. The AICs were calculated to be 1702.893 for model I and 1527.75 for model II.

**Conclusion:** The incorporation of frailty into the parametric multi-state model resulted in a better fit of model for prediction of CRC progression.

**References:** *[1] Andersen PK, Keiding N. Multi-state models for event history analysis. Statistical Methods in Medical Research2002; 11:91–115. [2] Hout A. Multi-State Survival Models for Interval Censored Data. Boca Raton, FL: CRC Press; 2017.*



# POSTER SESSION 07

## P07-05 Time to readmission among newborns: time for a reappraisal?

**Carly E. Milliren[1], Patrice Melvin[2], Al Ozonoff[3]**

1 Institutional Centers for Clinical and Translational Research, Boston Children's Hospital, United States

2 Center for Applied Pediatric Quality Analytics, Boston Children's Hospital, United States

3 Division of Infectious Diseases, Boston Children's Hospital, United States

Traditional analyses of hospital readmissions calculate time to readmission relative to index visit discharge. In the context of newborns, the classic readmission definition can be problematic particularly when comparing groups with disparate birth lengths of stay, as is often the case when studying neonates with conditions requiring longer post-natal hospitalization. For this study population, age from birth versus age at discharge may differ by weeks or months. We compare two methods of examining readmissions within the first year for infants diagnosed with neonatal opioid withdrawal syndrome (NOWS) compared to normal newborns (average LOS: 17 days vs. 2 days). First we applied the traditional definition to examine readmission timing from *birth discharge* using crude estimates of proportions and a Cox regression model. Second, we defined readmission timing by *day of life* and compare the corresponding proportions, then fit a Cox model with left truncation to allow delayed entry of hospitalized neonates into the at-risk period at time of discharge. Results using the traditional definition indicated normal newborns were at highest risk of readmission within the first few days since discharge while infants with NOWS were at higher risk later into infancy resulting in violation of proportional odds, an assumption which the Cox model requires for validity. We examined the hazard function and constructed a piecewise model predicting early readmissions (<25 days) and late readmissions (≥25 days). In adjusted models, NOWS infants had 1.5 times the hazard of late readmissions with no difference in early readmissions. Models predicting readmission by day of life indicated no violations of proportional odds, and overall estimates of one-year hazard ratios were similar [1.76 (95% CI: 4.40-2.22) vs. 1.55 (1.09-2.22)]. Crude estimates differed substantially between methods particularly within the first 30-days but converged at later time points through one-year. These methods indicate similar overall findings between the two approaches though readmissions indexed to day of life offers a more intuitive interpretation with no issues of non-proportionality. Advances in time-to-event modeling available in most statistical packages allow for easy incorporation of left truncation, which is particularly useful in the context of readmissions for newborns.

## P07-06 Studying the longitudinal trajectory of potassium in heart failure patients through dynamic survival models

**Caterina Gregorio[1,2], Francesca Ieva[2], Giulia Barbati[1]**

1 Biostatistics Unit - Department of Medical Science, University of Trieste, Italy

2 MOX – Department of Mathematics, Politecnico di Milano, Italy

**Background:** Potassium plays a fundamental role in the heart functioning. In patients affected by Heart Failure (HF), the disease itself together with the pharmacological treatment can alter potassium values. In clinical practice, dangerous changes are identified according to a single measurement and a cutoff which has been questioned by recent studies. This would be a trivial thing if it didn't lead cardiologists to decide for the discontinuation of life-saving treatments. Clinical research highly needs new methods to better explore the dynamic impact of potassium on survival for personalized optimization of the treatment in HF.

**Objectives:** The aim of this study is to propose a dynamic survival model to study the association between individual potassium trajectories and survival, which could provide an alternative to identify patterns associated with a lower survival probability.

**Methods:** The data comes from the administrative regional data of the Friuli Venezia Giulia Region, integrated with the Outpatient and Inpatient Clinic E-Chart. We exploited the continuous, longitudinal nature of potassium representing it as a functional datum in order to go beyond the cut-off paradigm. The two main approaches to dynamic survival modelling that have been considered to study the association between potassium and the outcome are: joint modelling and landmarking.

**Results:** The study included 3678 patients affected by HF who were observed for a median time of 45 months (IQR: 25-68). Over this period, the median number of potassium measurements per subject was 16 (IQR: 7-31). Moreover, the survival probability after 4 years of follow-up was 0.65 (95%CI: 0.64-0.67). The analyses highlighted some novel insights of the relationship between potassium and survival. They confirmed the need of using the longitudinal trend of potassium to identify when a patient shows a potassium trajectory which increase the risk of events.

**Conclusions:** This work leads to promising new directions for the treatment of HF patients and the developing of personalized treatment tools. Future research should further investigate the estimation of personalized treatment schedules based on the potassium trajectory and the risk of adverse outcomes to avoid premature discontinuation of life-saving treatments in patients affected by HF.

# POSTER SESSION 07

## P07-07 Use of electronic health records to enhance data from a single clinical trial evaluating maintenance therapy in non-small cell lung cancer patients

**Ilse Cuevas Andrade[1]*, Devleena Ray[1]*, Keith Abrams[2], Laura Gray[1], Nuala Sheehan[1], Sylwia Bujkiewicz[1]**

1 Department of Health Sciences, University of Leicester, United Kingdom

2 Centre for Health Economics, University of York, United Kingdom

* Joint first authors

Randomised control trials (RCTs) are considered as gold standard for evidencing treatment efficacy and subsequent decision making in health care research. However, there has been a shift in focus towards the prospects of real-world evidence (RWE) in complementing RCTs to support decision making and enhance estimation of treatment effects. We aim to develop and compare methods to combine registry data with existing trial evidence to improve inference by using simulated Systemic Anti-Cancer Therapy (SACT) data available from the Simulacrum database. We explore the potential of simulated registry data in emulating the control arm of the completed PARAMOUNT trial investigating the effects of pemetrexed maintenance therapy on overall survival in non-small cell lung cancer (NSCLC) patients. We intend to evaluate the effects of combining RWE from SACT and PARAMOUNT trial data on the results that were presented at NICE. Methods to adjust for selection bias between both populations will be used to then enhance the comparison and analysis of the survival estimation curves.

Synthetic patient data for non-squamous NSCLC was obtained from the Simulacrum database. Since there were no patients in the Simulacrum database that received pemetrexed as a maintenance therapy after cisplatin-pemetrexed induction therapy it was not possible to estimate comparative effectiveness. Therefore, we selected patients who received either a standard treatment or no treatment following initial cisplatin-pemetrexed therapy for a duration approximately equivalent to that in the PARAMOUNT trial to create a comparable synthetic control arm for a single-arm approach. We further evaluated adjusting the synthetic control arm data to reflect the control arm of the PARAMOUNT trial using a variety of methods, including; matching, re-weighting and regression-based adjustment. There were 973 synthetic patients in Simulacrum and 939 patients in the trial who received cisplatin-pemetrexed induction therapy. We present the results of time-to-event analysis evaluating overall survival by amalgamating the synthetic data with reconstructed data from the PARAMOUNT trial.

The results presented demonstrate the potential effects of combining RWE with existing trial data and highlight the need to develop more sophisticated methods which may guide decision making in settings with scarce experimental data in the future.

## P07-08 Modelling the length of stay of COVID-19 patients using a multistate approach

**Michael Lauseker, Ludwig Christian Hinske, Ulrich Mansmann**

Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig-Maximilians-Universität München, Germany

The aim was to predict the length of stay of COVID-19 patients in different units of the hospital: general ward, intermediate care (IMC) and intensive care unit (ICU).

A semi-Markov multistate model with the states “general ward”, “IMC”, “ICU”, “dead” and “recovered” was estimated. Patients could move repeatedly between the initial states “general ward”, “IMC” and “ICU”, while recovered and dead were seen as absorbing states. The patients’ sex, age, type of hospital admission and time period (wave) were considered as covariates. Transition hazards for the twelve different possible transitions were estimated using a parametric approach with different distributions for the transitions. The approach is opposed to a Cox model approach with regard to the prediction of deaths and sojourn times.

The study population consisted of a sample of all patients that received inpatient COVID-19 treatment at the university hospital. Estimated sojourn times in different states were obtained by simulating the model 10000 times and averaging over the simulations. With increasing age, patients stayed longer both in regular ward and ICU up to an age of about 80 years. For older patients, the sojourn times decreased due to increased mortality. Males had slightly longer sojourn times in all states (except IMC) than females. At the moment, the parametric model and the Cox model yielded similar results.

Combined with a model for hospital admissions, this model can be used to estimate hospital occupancy. A possible limitation of the model is that general circumstances could hardly be modelled, the capacity limit of the hospital could e.g. influence the length of stay as well. A future extension of the model should be the inclusion of the SARS-CoV-2 mutation.



# POSTER SESSION 07

## P07-09 Landmarking: An R package for analysis using landmark models

**Isobel Barrott, Jessica Barrett**

MRC Biostatistics Unit, University of Cambridge, United Kingdom

The landmarking approach allows survival predictions to be updated dynamically as new measurements from an individual are recorded. It was first described by Van Houwelingen (2007). The idea is to set predefined time points, known as ‘landmark times’, and form a model at each landmark time using only the individuals in the risk set. Here I present ‘landmarking’ an R package which allows the user to perform analysis using the landmarking approach, offering benefits over the existing package ‘dynpred’ (Van Houwelingen and Putter, 2011). The main benefit of the ‘landmarking’ package is that it allows for mixed effect modelling of the repeat measurements, in addition to the option of using the last observation carried forward (LOCF). Mixed effects modelling has the following advantages over LOCF: it allows for missing data in the repeat measurements, it provides improved precision when there are infrequent measurements, and it reduces measurement error. Moreover, this package allows the user to model competing risks in the survival data using either cause-specific Cox regression or Fine-Gray regression.

**References:** *[1] Ellie Paige, Jessica Barrett, David Stevens, Ruth H Keogh, Michael J Sweeting, Irwin Nazareth, Irene Petersen, and Angela M Wood. Landmark models for optimizing the use of repeated measurements of risk factors in electronic health records to predict future disease risk. American journal of epidemiology, 187(7):1530–1538, 2018. [2] Hans van Houwelingen and Hein Putter. Dynamic prediction in clinical survival analysis, 2011. [3] Hans C Van Houwelingen. Dynamic prediction by landmarking in event history analysis. Scandinavian Journal of Statistics, 34(1):70–85, 2007.*

## P07-10 The risk of valvular heart disease after childhood cancer: contribution of dose-volume histogram parameters

**Stefania Chounta[1,2,3,4], Sarah Lemler[4], Nadia Haddy[1,2,3], Imene Mansouri[1,2,3], Brice Fresneau[5], Neige Journy[1,2,3], Charlotte Demoor-Goldschmidt[1,2,3,6], Hubert C. Hounsossou[7], François Pein[8], Ibrahima Diallo[1,2,3], Carole Rubino[1,2,3], Florent de Vathaire[1,2,3], Véronique Letort[4], Rodrigue S. Allodji[1,2,3,7]**

1 Université Paris-Saclay, UVSQ, Univ. Paris-Sud, INSERM, CESP, Villejuif, France

2 INSERM, CESP, Cancer and Radiation Team, Villejuif, France

3 Gustave Roussy, Department of Clinical Research, Cancer and Radiation Team, Villejuif, France

4 Laboratory of Mathematics in Interaction with Computer Science (MICS), Centrale Supélec, France

5 Gustave Roussy, Department of Pediatric Oncology, Université Paris-Saclay, Villejuif, France

6 CHU de Nantes, Pediatric Oncology, Nantes, France

7 Polytechnic School of Abomey-Calavi (EPAC), University of Abomey-Calavi, Cotonou, Benin

8 Département de Recherche, Institut de Cancérologie de l’Ouest, site René Gauducheau CLCC Nantes-Atlantique, Saint-Herblain, France

**Background:** Childhood cancer survivors are at increased risk of developing Valvular Heart Disease (VHD). Despite the large size of voxelized dosimetric data currently provided by individual dose estimates in radiotherapy, most studies are limited to single variable approaches such as mean dose to the heart (MHD) to study its relationship with the risk of VHD. Therefore, we used the French Childhood Cancer Survivor Study (FCCSS) cohort including 7670 five-year survivors, to investigate the potential predictive capability of dose-volume histogram in the risk of VHD after childhood cancer.

**Methods:** Individual dose volume histograms for the whole heart were obtained, and MHD was calculated, as well as the doses (Gy) delivered to the v% of the heart volume ($D_v$ in Gy), and the volume percentages of heart receiving ≥d Gy ($V_{\geq d}$). Their role in the occurrence of VHD was investigated using the Cox proportional hazard regression model and penalized Cox regression (LASSO, Ridge, Elastic Net) when faced with multicollinearity. Models were compared with each other via classic information criteria and their efficacy was evaluated through performance indices.

**Results:** 82 patients had developed a severe VHD (grade ≥ 3). Overall, patients treated with radiotherapy had an approximately 2-fold ($CI_{95\%}$: 1.16, 3.42) risk increase after adjustment for chemotherapy exposure. MDH was 23.7 Gy for patients who developed a VHD while almost 7 Gy for the entire cohort. The risk for of VHD increased 12-fold ($CI_{95\%}$: 7.02, 21.83) when MHD was over 20 Gy. The risk increased 30-fold ($CI_{95\%}$: 16.07, 58.8) as the volume having received ≥ 30 Gy increased. Exposure to chemotherapy seems to increase by almost 2-fold the risk of VHD in most of the alternative adjustments. Multivariable approaches seem to provide better predictions than the binomial model, but overall the model that studies irradiation dose-effect relationship adjusted on the MDV appears to be the closest to the true model according to its AIC and a combination of decorrelated volume indicators seems to provide the best prediction (C-index: 0.783).

**Conclusions:** Findings may be useful for patients and doctors both before treatment and during long-term follow-up for VHD in survivors of childhood cancer.

**References:** *[1] Haddy, N. et al. Cardiac Diseases Following Childhood Cancer Treatment: Cohort Study. Circulation 133, 31–38 (2016). [2] Cutter, D. J. et al. Risk for Valvular Heart Disease After Treatment for Hodgkin Lymphoma. JNCI J. Natl. Cancer Inst. 107, (2015).*

# POSTER SESSION 07

## P07-11 The estimation of adjustment factors for expected mortality rates with application in comorbidity adjusted lifetables

**James C.F. Schmidt[1], Paul C. Lambert[1,2], Michael Sweeting[1]**

1 Biostatistics Research Group, Department of Health Sciences, University of Leicester, United Kingdom

2 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Published life tables can be used with Relative Survival (RS) techniques in the study of excess deaths in a disease-specific population. Using population-based mortality rates as expected rates circumvents obtaining large control samples as comparators when case-only data are available (e.g. from disease registries). For RS methods to be unbiased, expected mortality rates should represent rates the exposed population would experience if unexposed. However, published life tables are usually stratified by a small number of factors, such as age, sex and calendar year. Cardiovascular disease (CVD), the leading cause of mortality worldwide, has an increased burden of comorbidities (such as diabetes), compared to those without [1]. Hence RS methods applied to CVD require further adjustment of the published rates for comorbidity. We extend on previously developed methods for estimating adjustment factors for expected mortality rates [2] to incorporate time-varying adjustment factors. We describe how analyses are undertaken using both a Poisson and Flexible Parametric Survival Model (FPSM) approach. Poisson models, using a generalized linear model framework, require the splitting of data by relevant timescales (age and calendar year). Attained age and year are included in the model as restricted cubic splines to allow flexibility in the hazard functions. Person years and expected mortality are both included as offsets. FPSMs have no need for data splitting, representing an appealing method for large datasets, incorporating smoothed expected rates and by constraining coefficients, we can estimate deviations from the smoothed expected rates in the cohort. Using a cohort of 1.8 million patients from primary care data, a baseline Charlson Comorbidity Index (CCI) is derived from linked secondary care data, with patients tracked for mortality. CCI score-group adjustment of published rates is performed with initial baseline CCI score group analysis revealing that those with no comorbidity have lower absolute rates than the general population, while those with CCI scores greater than 0 have adjustment factors greater than one. Comorbidity adjustment factors vary by age, and calendar year. Extensions to the research will investigate the effect of accounting for lagged CCI measures and the inclusion of updated CCI scores over time.

**References:** *[1] Kendir, C. et al. (2017) 'Cardiovascular disease patients have increased risk for comorbidity: A cross-sectional study in the Netherlands', The European journal of general practice, pp. 1-6. [2] Bower, H. et al. (2018) 'Adjusting Expected Mortality Rates Using Information From a Control Population: An Example Using Socioeconomic Status', American Journal of Epidemiology, 187(4), pp. 828-836.*

# POSTER SESSION 07

## P07-12 Multiple Cox regression analysis to investigate a biomarker in IgA nephropathy disease: different approaches

**Ana Cristina Matos[1], Carla Henriques[2], Bernardo Faria[3]**

1 Research Centre in Digital Services, and Polytechnic of Viseu, Portugal

2 CMUC, Coimbra University, and Polytechnic of Viseu, Portugal

3 Nephrology and Infectious Disease R&D Group, INEB, Institute of Investigation and Innovation in Health (i3S), University of Porto, Portugal

IgA nephropathy (IgAN) is a common worldwide glomerulonephritis, and the use of biomarkers is an important vehicle in identifying subgroups of patients with IgAN. Glomerular C4d (C4dG) is a robust marker of IgAN patients with poor outcome prognosis. In a retrospective cohort study, our aim was to investigate the significance of arteriolar C4d (C4dA) in an IgAN group of 126 patients and to compare it with clinical and histological markers of disease progression, particularly with C4dG.

The effect of C4dA on survival was evaluated using two approaches: first, predictors were selected with a stepwise forward procedure from a set of the previously identified variables (through simple Cox regression analysis) as being related to disease progression. The second approach was motivated by the reduced sample size and the number of events that limits the number of predictor variables to be included in the multivariate model. So, two models were first obtained: one selecting variables from the set of significant clinical ones and another choosing between the histological significant variables. Finally, a multivariable model was constructed with the histological and clinical variables of the two previous models. C4dA and C4dG were added separately to these models to evaluate and compare the impact of these two markers on survival. Both C4dA and C4dG showed to be associated with renal survival, but in the final model, C4dA remained significant, while C4dG did not (p=0.054) and the Akaike information criterion was slightly lower for the C4dA model.

Harrell's C indexes were calculated for both final models and their values were validated using bootstrapping. Slightly higher values were obtained for both the C4dA model than for the C4dG model, but without reaching a statistical significance. Likelihood ratio tests were used to compare the crude models with the model having both C4dA and C4dG as predictors: the inclusion of C4dA affords a statistically significant improvement in the prediction of the survival model with C4dG alone (p=0.012), but the reverse is not true (p=0.280). These findings show that C4dA is a robust biomarker predicting the progression of kidney disease in IgAN and compared to be superior to C4dG.

**References:** *[1] Bernardo Faria, Pedro Canão, Qingqing Cai, Carla Henriques, Ana Cristina Matos, Felix Poppelaars, Mariana Gaya da Costa, Mohamed R. Daha, Roberto Silva, Manuel Pestana, Marc A. Seelen, (2020). Arteriolar C4d in IgA Nephropathy: A Cohort Study, American Journal of Kidney Diseases, Volume 76, Issue 5, Pages 669–678, https://doi.org/10.1053/j.ajkd.2020.03.017.*

# POSTER SESSION 07

## P07-13 Analysis of Results of Total Knee Replacement Failure Using Cox Proportional Hazard Model with Time-Dependent Covariates

**Veronika Bendová[1], Stanislav Katina[1,2], Libor Nečas[3,4]**

1 Institute of Mathematics and Statistics, Masaryk University, Brno, Czech Republic
2 Institute of Computer Science of the Czech Academy of Sciences, Prague, Czech Republic
3 Orthopedic Clinic, Martin University Hospital, Martin, Slovakia
4 Jessenius Faculty of Medicine, Comenius University, Martin, Slovakia

Total knee replacement (TKR) surgery is the most common treatment of osteoarthritis of the knee. Good health, as well as other factors, influence successful and prompt recovery of patients that underwent this surgery. Physicians are interested in quantifying the effect of patient's well-being on the failure of the TKR that might come during the follow-up. Therefore, their aim is to monitor health of patient before and after the TKR surgery. Two suitable tools for the assessment of patient's health state have been proposed: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) along with the Knee Society Scoring System (KSS). Both have the ability to assess the subjective health state of a patient, while the KSS aspires to evaluate the objective health state as well.

Our study includes data about 2295 patients, who have undergone primary TKR surgery between January 1st, 2006 and December 31st, 2019, from Orthopaedic Clinic of Martin University Hospital. The data were recorded in the Slovak Arthroplasty Register (SAR). WOMAC and KSS related questionnaires have been recorded for each patient in the study in four time points during planned examinations: before TKR surgery, three months, six months, and 12 months after TKR surgery. The aim of this registry-based study is to show the relationship of primary TKR failure on WOMAC and KSS scores, age, sex, and diagnosis by means of Cox proportional hazards model with time-dependent covariates stratified based on type of implants, such as cruciate retaining, posterior stabilized condylar constrained and hinge knee implant. Statistical analyses were carried out using R software environment.

**Acknowledgment:** *The work was supported (partly) by the long-term strategic development financing of the Institute of Computer Science (RVO:67985807) and specific research of Masaryk University as support for student projects (MUNI/A/1615/2020).*

**References:** *[1] Klein, J.P., Moeschberger, M.L., 2003: Survival Analysis: Techniques for Censored and Truncated Data. Springer-Verlag, New York. [2] Nečas, L., Katina, S., Uhlárová J., 2013: Survival analysis of total hip and knee replacement in Slovakia 2003–2011. Acta Chirurgiae Orthopaedicae et Traumatologiae Cechoslovaca 80,1, 1–85.*

## P07-14 Joint contribution of positive and total lymph nodes number in predicting overall survival of esophageal cancer

**Yuqin Cao[1,2,3,4], Mathieu Fauvernier[1,2,3], Yajie Zhang[4], Hecheng Li[4], Pascal Roy[1,2,3,5]**

1 Université Claude Bernard - Lyon 1, Villeurbanne, France
2 Service de Biostatistique-Bioinformatique, Pôle Santé Publique, Hospices Civils de Lyon, France
3 Laboratoire de Biométrie et Biologie Évolutive, CNRS UMR 5558, Villeurbanne, France
4 Department of Thoracic Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
5 Université de Lyon, France

The number of positive lymph nodes (PLN) serves as the current criteria in the TNM staging of esophageal cancer, but whether to involve the prognostic influence of total lymph nodes resected (TLN) is still controversial. This study compared the AIC and Harrell's C-index of various parametric models using three different modeling strategies: (1) categorization of PLN and TLN; (2) penalized natural spline of lymph node ratio (LNR, the ratio of PLN to TLN); (3) penalized natural splines of PLN and TLN. The interaction between covariates and the hypothesis of proportionality were tested. According to the original cohort of population-based data in the USA and external validation of hospital-based data in China, the better modeling strategy to analyze the joint contribution of PLN and TLN was to construct a proportional hazard model with penalized natural splines of both covariates without interaction. A more aggressive adjuvant therapy could be proposed to patients with low TLN in the context of a randomized clinical trial.



POSTER SESSION 07

## P07-15 Modelling the duration of recurrent events from interval-censored data

**Mamadou Saliou Kalifa Diallo[1,2], Abdoulaye Toure[1,2], Eric Delaporte[1], Jean-Francois Etard[1], René Ecochard[3]**

1 IRD/INSERM/Montpellier University, Montpellier, France
2 Centre de Recherche et de Formation en Infectiologie de Guinée, Université Gamal Abdel Nasser de Conakry, Conakry, Guinea
3 Hospices Civils de Lyon, Service de Biostatistique, Université de Lyon, France

**Background:** Recurrent event data arise frequently in longitudinal studies when subjects are monitored discretely. Symptomatic episodes that occur in survivors of epidemics such as Ebola are a good illustration of the recurrent events termed "recurrent episode". In the literature, several statistical methods have been proposed to analyze the time to first event or the risk of recurrent events but the analysis of event durations especially from interval censored data remain scarce.

**Objective:** We aim to present a new approach to estimate the event-duration of recurrent events from interval censored data and assess predictors of this duration.

**Methods:** We divided the patient's visit history into segments composed of two consecutive visit dates. Four situations were observed: a) the symptom was present all along the time segment; b) the symptom was absent; c) the symptom occurred during the time segment; and d) the symptom stopped during the time segment. The missing start/end dates were imputed either deterministically (mid-point) or stochastically using either uniform on the interval or according to maximum likelihood (Turnbull). In the latter case, we created five imputed datasets. A simulation was then performed to assess the properties of the estimators. We calculated all duration and their 95% CI using Rubin's rules incorporating within and between imputation variability. The predictive value of several factors on the duration of symptoms was then estimated by mixed-effect regression. We have applied this method to a prospective cohort study who followed 802 Ebola virus disease survivors over 48 months in Guinea. Patients were assessed at inclusion and every 6 months up to 48 months. Clinical symptoms were recorded at each visit.

**Results:** Our simulation study demonstrates that this approach is a good strategy in estimating the duration of events and assess impact of predictors on it. We reconstructed our database and estimate the duration of symptoms day by day to get the number of days each symptom was present or absent for each patient.

**Conclusion:** We conclude that the method could be a promising and useful tool to be used for estimating the event-duration of recurrent events from interval censored data in longitudinal study.

## P07-16 Relation between women empowerment and birth interval: A survival analysis approach

➜ *CFCD AWARD*

**Rubaiya Rubaiya, Samia Sultana Sanzi**

Institute of Statistical Research and Training, University of Dhaka, Bangladesh

Birth intervals and birth spacing patterns provide important information about women's reproductive behavior and dynamics of the fertility process. Birth intervals can be considered as time-to-event or survival data, where the events are the childbirth i.e first birth, second birth, and so on. Women empowerment may have important influence on birth intervals and birth spacing patterns. This study attempts to investigate the effects of women empowerment on the birth interval in Bangladesh. Four indicators have been considered to measure four dimensions of women empowerment: the level of education, participation in household (HH) decisions, freedom in movements, and employment status. Several socioeconomic and demographic variables are used as covariates in this study. The Cox proportional hazard model is used for analyzing the Bangladesh Demographic and Health Survey (BDHS, 2014) data. From Kaplan-Meier and Log-rank test it is observed that women empowerment is one of the key factors for increasing the birth interval, which subsequently contribute to the improvement of the health status and the level of development. Socioeconomic and demographic variables are affecting the birth intervals. Increases in mother's age at birth make the birth interval shorter. On the contrary women from rich socio-economic groups have longer birth interval. Birth order play a vital role in determination of birth interval. From our model we concluded that higher birth order makes the length of birth interval shorter. It is also observed that birth interval is long if previous birth is male. Urbanization can make the birth interval longer. Efforts should be made to increase the women empowerment, particularly girls' education, participation in HH decisions, freedom in movement and employment of women in economic activity.

**References:** *[1] Bates, L. M., Maselko, J., and Schuler, S. R. (2007). Women's education and the timing of marriage and childbearing in the next generation: evidence from rural Bangladesh. Studies in Family Planning, 38(2):101-112. [2] Chakraborty, N., Sharmin, S., and Islam, M. A. (1996). Differential pattern of birth intervals in Bangladesh. Asia-Pacific population journal, 11(4):73-86.*

# POSTER SESSION 07

## P07-17 Impact of model choice when studying the relationship between blood pressure variability and risk of stroke

**Hugues de Courson, Antoine Barbieri, Christophe Tzourio, Karen Leffondré**

Univ Bordeaux, INSERM, BPH, U1219, Bordeaux, France

Long-term blood pressure variability (BPV) is an increasingly recognized vascular risk factor. However, quantifying its impact on the risk is challenging. Most previous epidemiological studies used a Cox model with BPV derived as a fixed-in-time covariate conditioning on the future. The objective was to compare the results of commonly used models with time-dependent Cox model and joint shared random effect models, using data from a large randomized clinical trial.

We used data from a secondary stroke prevention trial, PROGRESS, which included 6105 subjects. A total of 727 patients experienced a first stroke recurrence. The mean follow-up was 4.3 years and the median number of blood pressure (BP) measurements per patient was 12. Hazard ratio (HR) of BPV were estimated from six models. Commonly used Models 1 and 2 first derived standard deviation of BP (SDBP) measures observed over the whole follow-up (including or excluding BP measures observed after the first stroke recurrence, respectively), and then included SDBP as a fixed-in-time covariate in a Cox model estimated on the whole follow-up. Model 3 derived SDBP using data from the first year of follow-up, and used it as the baseline value in a Cox model estimated on the remaining follow-up. In Model 4, SDBP was included as a time-dependent covariate in a Cox model. Models 5 and 6 were shared random-effect models. In Model 5, the longitudinal marker was time-dependent SDBP, and its current true value was the main covariate in the survival part [1]. In Model 6, the longitudinal marker was time-dependent BP and was modeled with a subject-specific error variance which was the main covariate in the survival part [2].

While Models 1-3 produced opposite results (for a 5mmHg increase in BPV, HR=0.75, 95% confidence interval (CI) [0.68, 0.82], HR=0.99 [0.91, 1.08], HR=1.19 [1.10, 1.30], respectively), Models 4-6 resulted in a similar moderate positive association (e.g. HR=1.08 [0.99-1.17] for model 5).

The modelling of BP variability strongly affects its estimated effect on the risk of stroke. Further methodological developments are needed to account for the dynamics of both BP and BPV over time, to clarify the specific role of BPV.

**References:** *[1] Rizopoulos D. JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data. J Stat Softw. 2010;35(9):1-33. [2] Barrett JK, Huille R, Parker R, Yano Y, Griswold M. Estimating the association between blood pressure variability and cardiovascular disease: An application using the ARIC Study. Stat Med. 2019;38(10):1855-1868. doi:10.1002/sim.8074*

# POSTER SESSION 07

## P07-18 A jack-knifed version of the log-rank test in small samples: when bias meets variance to increase test power

**Lubomír Štěpánek[1,2], Filip Habarta[1], Ivana Malá[1], Luboš Marek[1]**

1 Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business, Czech Republic

2 Department of Biomedical Statistics, Institute of Biophysics and Informatics, First Faculty of Medicine, Charles University, Czech Republic

The problem of comparing two time-to-event survival curves is common in biostatistics and is often solved by the log-rank test's application. However, it could be complicated when two groups of individuals, which the survival curves are based on, are generally of small size, pulling the log-rank test power down. There are some workarounds by which the small sample "curse" could be overcome but are usually of low statistical power, too.

This preliminary study addresses the small sample size issue and related survival curves compared by various jack-knifing samples. We assume the paradigm that jack-knifed tuples are still valid for statistical inference purposes since, compared to bootstrapped or else resampled samples, it does not contain any new piece of data. Supposing we build all possible jack-knifed samples by leaving each individual out, by leaving all combinations of each two individuals out, every three individuals out, etc., we always get a superset as a union of all the jack-knifed samples. The higher the jack-knife degree is, i. e. the higher level of individuals' left-out combinations, the larger the union of jack-knifed samples (and lower estimates' variance). However, the higher the degree of observations' left-out combinations, the more distorted the original data could be (and higher estimates' bias is). The bias-variance trade-off originated from the jack-knifing's growing degree, i. e. the level of individuals' left-out combinations, was measured by the simple summation of the variance and bias deviations compared to the original data. The first type error rate of such jack-knifed log-rank test was also considered.

Using multiple simulations and applying real-world COVID-19 data, we researched optimal degrees of the left-out combinations of individuals to minimize the variance and bias deviations' summations. Interesting consequences also resulted in the first type error rate of the jack-knifed log-rank test and its statistical power.

The jack-knife version of the log-rank test seems to be an alternative for small sample-based survival curves comparison. Analytical derivations are required to investigate jack-knifing's first type error rate and start a new R package development, implementing the proposed methods.

# POSTER SESSION 08

POSTER SESSION 08 — Epidemic modeling

## P08-01 Exploring the Sensitivity of Extended SIR Models Through Randomized Simulations and Multiple Factor Analysis

**Jiayu He[1,2,3], Linh Tang[2,4], Senay Gokcebel[1,2,5], Bowen Mince[2]**

1 Department of Biology, Grinnell College, Grinnell IA, United States
2 Department of Mathematics and Statistics, Grinnell College, Grinnell IA, United States
3 Department of Classics, Grinnell College, Grinnell IA, United States
4 Department of Computer Science, Grinnell College, Grinnell IA, United States
5 Department of Economics, Grinnell College, Grinnell IA, United States
These authors contributed equally to this work.

COVID-19 has been modelled since its emergence in several ways, most prominently using(Susceptible-Infected-Removed) SIR models. This paper aims to evaluate the robustness of extended SIR models by using multivariate factorial analysis on 13 parametric assumptions to determine which methods are most influential in overcoming the COVID-19 pandemic. Overall, we show that the COVID-19 epidemic projections are very sensitive to minor changes in assumptions, even when using parametric assumptions within ranges given by the CDC. The spread and disease burden depended upon very distinct parameters based on the primary response that was measured, and key parameters impacted different facets of disease burden.
Specifically, we find that testing, as well as isolation and quarantine measures are most effective in containing the spread and alleviating disease burden. Similarly, the accuracy of diagnostic tests carry great importance. Since each of the parameters used in all COVID-19 model projections are estimated values, better care should be used to understand the variability of these parameter estimates when models are shared with the public.

## P08-02 Monte Carlo simulation of the COVID-19 spread using an agent-based modelling in Russian regions

**Mikhail Kirillin[1], Ekaterina Sergeeva[1], Aleksandr Khilov[1], Valeriya Perekatova[1], Daria Kurakina[1], Nikolai Saperkin[2]**

1 Institute of Applied Physics, Russian Academy of Sciences, Russia
2 Department of Epidemiology, Microbioloogy and EBM, Privolzhsky Research Medical University, Russia

**Introduction:** Prediction of COVID-19 outbreak and timely introduction of preventive measures require reliable tools for epidemic spread simulations. The aim of this study was simulation of the COVID-19 spread in early and peak stages in Russian regions using a Monte Carlo agent-based model.
**Methods:** We fitted a general pooled model, in which all individuals (agents) may interact with each other. Conditional on a simulated scenario, each agent has its appropriate binary states, which are governed by the Monte Carlo based random values. The model accounts for the population age structure. Infection transmission coefficient was directly related to the average number of individuals, to which an infected agent may transmit the infection within one week given no restriction measures to be applied. The model also accounts for the efficiency of isolation measured by the 'self-isolation index'. For prognosis, we considered positive and negative scenarios. The testing model also includes the number of daily PCR tests for COVID-19 and the increase in its accuracy.
**Results:** To provide a best-fit scenario, we manipulated with several key parameters of the model. They were the number of initial infected agents, percentage of deaths among agents in the critical state, and virus transmission coefficient. Our agent-based model of COVID-19 epidemic spread widely employed the rigorous methodology of Monte Carlo simulation principles. The model was validated on the statistical data for daily new cases and deaths in representative regions of Russia, such as Moscow and Novosibirskaya oblast'.
**Conclusions:** We suggest that agent-based modelling can be succefully used for elucidating the COVID-19 trends. Traditional simulation approaches based on derivatives of a SIR model, although being quite efficient, suffer from not accounting for random factors. Agent-based models provide a convenient solution which allows accurately accounting for age distribution, variations in self-isolation strategies and testing protocols, super-spreaders etc. Despite local features of territories, epidemic curves can be predicted correctly for different regions. In this situation we use the same model parameters, except the initial number of infected, which serve as a tuning parameter of the model.
**Acknowledgments:** *The study is supported by RFBR, CNPq, and NSFC (project no. 20-51-80004).*



# POSTER SESSION 08

## P08-03 Similarities between the COVID-19 spread in Romanian counties identified through data clustering

**Cristina Gena Dascalu[1], Ramona Diana Feier[2], Magda Ecaterina Antohe[3]**

1 Medical Informatics and Biostatistics Dept., "Grigore T. Popa" University of Medicine and Pharmacy, Romania
2 Faculty of Medicine, "Dimitrie Cantemir" University, Romania
3 Removable Prostheses, Oral Implantology, Dental Technology Dept., "Grigore T. Popa" University of Medicine and Pharmacy, Romania

The purpose of our study was to analyze the dynamics of COVID-19 spread in Romanian counties over a period of 299 days (April 2nd 2020 - January 25nd 2021), in order to identify possible similarities that may contribute to a better understanding of the mechanisms of disease spread. The data used in our study are the numbers of active cases for each county in Romania, as well as Bucharest and the whole country, reported daily by the Romanian Ministry of Health (https://datelazi.ro); based on these values, we calculated the daily number of new cases and we reported them at the total number of inhabitants per counties, as it was enlisted by INS (Romanian National Institute of Statistics). We expressed these values as ratios at 100.000 habitants, in order to gain consistency and relevancy (otherwise, the raw values were small in magnitude and difficult to interpret).
In order to identify similarities between counties, we decided to use data clustering techniques. We performed the hierarchical clustering of counties, recorded as variables, using the between-groups linkage method and testing different distances: Euclidean, squared Euclidean, Chebyshev, Block and Minkowski. The result of this approach was that we identified rather the "outliers" among counties, because, no matter the distance we used, we found a big cluster, which includes most counties in Romania, a small cluster including 5 counties, situated approximatively in the same geographic area (Transylvania – Alba, Brasov, Cluj, Sibiu and Timis) and a few counties with specific evolution: Bucuresti (which is the biggest city from Romania), Constanta (located in the SE of Romania, along the seaside), Ilfov (the rural area around Bucuresti) and Salaj (located also in Transylvania, in the neighborhood of Cluj, which belong to the first small cluster). Since we are dealing with time series, this method for data analysis isn't actually the first choice (the calculation of cross-correlations is the classical approach), but it lead us to an interesting conclusion – there are no significant differences between most counties in Romania in which concerns the disease's spread pattern – with only a few exceptions which were clearly identified.

## P08-04 A new epidemic model for the Covid-19 pandemic

**Ettore Rocchi[1], Sara Peluso[1], Davide Sisti[1], Margherita Carletti[2]**

1 Department of Biomolecular Sciences, Unit of Biostatistics and Biomathematics, Urbino University, Italy
2 Department of Pure and Applied Sciences, Urbino University, Italy

We present a new model to describe the Covid-19 pandemic, that takes in account both the possibility of a re-infection of the recovered subjects and the differentiation between symptomatic and asymptomatic infected subject.
The model, denoted as Θ-S(I)RD, is a 6-compartment model, described by as many ordinary differential equations. The six compartments are represented by Susceptible (S), Symptomatic Infected (Is), Asymptomatic Infected (Ia), Recovered from Asymptomatic fraction (Ra), Recovered from Symptomatic fraction (Rs), Deceased (D).
The biological assumptions are as follows: (i) no entry or exit from the territory (closed territory); (ii) the contagiousness of the infected is immediate (therefore the compartment of the Exposed is not considered); (iii) a loss of immunity is considered (in the first version of the model it is considered at a constant rate); (iv) mortality and birth rate affect, as a first approximation, only the Susceptible compartment; (v) the Asymptomatic Infected compartment includes both the fraction identified by diagnostic evaluation and the unidentified one; (vi) there is no lethality in the Asymptomatic Infected fraction.
Some numerical simulations have been performed, but at present, the validity of the model has not been verified by fitting with real data, also due to poor data quality. However, we will perform the fitting on the basis of the data referring to the compartment of the Deceased (the only compartment characterized by hard data, at the moment).

# POSTER SESSION 08

## P08-05 Spatial analyses of the first wave of COVID-19 cases in Hong-Kong using Poisson intrinsic and Besag York Mollié conditional autoregressive models under a Bayesian paradigm

**Ornella Wafo Noubissie, Demy Dam, Zharmaine Ante**
McGill University, Montreal QC, Canada

**Background:** As the COVID-19 pandemic continues to evolve, identifying high risk areas and populations becomes increasingly important. Spatial analyses are limited, yet such insight can inform targeted responses.
**Methods:** We conducted exploratory analyses of the spatial clustering of COVID-19 cases by Hong Kong district and investigated dependence on socioeconomic and demographic factors. These include population density, age and sex composition, working population, healthcare capacity, and post-secondary education attainment, among others. Spatial trends were tested using Moran's I statistic with contingency- and distance-based neighbourhood structures. Cases were modeled using Poisson intrinsic and Besag York Mollié (BYM) conditional autoregressive (CAR) Bayesian models.
**Results:** Moran's I test showed evidence of positive spatial autocorrelation. The WAIC revealed the distance-based binary-weighted BYM CAR Ridge model with five-nearest neighbours performed the best. The analyses showed positive spatial autocorrelation of cases, as districts with the highest standardized incidence ratio were clustered in southern Hong Kong, which is evidence of positive spatial clustering.
**Conclusion:** These results can inform facilitate the identification of COVID-19 clusters and their determinants to better anticipate the course of the pandemic and design focused interventions to effectively control the spread of COVID-19. Nonetheless, as the pandemic continues to evolve, trends may change.

## P08-06 Indices of inequality to monitor temporal and geographic trends in COVID-19 incidence and death data

**Kirsi Manz, Ulrich Mansmann**
Institute for Medical Information Processing, Biometry & Epidemiology (IBE), Ludwig-Maximilians-University Munich, Germany

There is a clear need to monitor pandemic activity at different levels. Useful metrics and their follow-up over time have moral as well as policy implications. For example, differences in observed incidence and death are explained by differences in the infection fatality rate (IFR). It is of interest to explore inequality in IFR across geographic areas and over time. Inequalities in IFR can be argued to be of moral significance (unfair), and such inequalities may also have policy implications as countries/regions should aim for the lowest possible IFR.
Among the most common metrics for measuring inequality are the Gini index, the Theil index, and the Hoover index [1]. They share four relevant properties: Anonymity or symmetry, scale independence or homogeneity, population independence, and the transfer principle. Inequality is different from variability: The Gini index of two populations may be different while the variability within each population is the same. We also study the more general concept of entropy [1].
The approach is also inspired by the discussion of heavy tail properties for distributions related to global incidence or mortality data [2].
We observe that classic inequality measures regarding infection and mortality require a new interpretation within the analysis of global infection data. Due to the equality of income or wealth within a society, low values of the Gini index (≤ 0.4) are associated with positive connotations such as fairness and justice. In the epidemiological setting, on the other hand, high values of the Gini index (>0.8) reflect the goal striven for: the epidemic is confined to one (or few) region(s), while the surrounding area is not affected - the epidemic has not spread. However, the analysis of inequality for the IFR rather follows the classical interpretation: single very high values must be avoided. We explore these concepts using 2020 COVID-19 pandemic data on infection incidence, mortality, and IFR at multiple scales: global, continental, national, and federal regions.

**References:** *[1] Atkinson A (1970). "On the Measurement of Inequality" Journal of Economic Theory. 2 (3):244–63.*
*[2] Cirillo P, Taleb NN (2020) Tail risk of contagious diseases. Nat. Phys. 16, 606–613. https://doi.org/10.1038/s41567-020-0921-x*



# POSTER SESSION 08

## P08-07 Bayesian disease mapping of standardized infection fatality rate using the example of COVID-19 in Bavaria

**Kirsi Manz, Ulrich Mansmann**
Institute for Medical Information Processing, Biometry & Epidemiology (IBE), Ludwig-Maximilians-University Munich, Germany

During the ongoing COVID-19 pandemic, a major effort has been made to provide up-to-date (daily/weekly) snapshots of current infection incidence and mortality measures. This information is provided as listings or on maps. However, such maps are limited to showing either infection incidence or, in rare cases, mortality, and lack simultaneous visualization of both quantities. Further, such combined presentation of pandemic data across regions needs to be standardized for comparison purposes. We introduce the regional standardized infection fatality rate sIFR as the ratio of observed to expected infection mortality. It turns out that the sIFR can be calculated as the ratio of two standardized measures - the standardized mortality rate SMR and the standardized incidence rate SIR. The sIFR describes the ratio of the regional deviation in the mortality process to the regional deviation in the infection process. Providing sIFR on maps is of interest for comparing different regions and identifying regional hotspots or areas of concern. Both the SMR and the SIR are relative risks (RR). Thus, the sIFR can be understood as the ratio of relative risks (RRR). Bayesian disease mapping is suited for analysis of relative disease risk measures. We simultaneously estimate SMR and SIR within the Bayesian framework of autoregressive convolution models [1]. This also yields the sIFR, taking into account the neighborhood structure and correcting for statistical artefacts introduced by regions with a small number of observed data. To demonstrate the application of the method, we estimate the sIRF using the COVID-19 pandemic data from a large German federal state (Bavaria). We use aggregated data from four periods of three months each between February 2020 and January 2021.
The sIFR and its components SMR and SIR are shown on maps. The naive IFR decreases during the first three periods and then increases again. Regional sIFRs change over time. Identifying major deviations in sIFR can help inform decision making between emphasizing measures for infection control and mortality reduction.

**References:** *[1] Besag, J., York, J. and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics, 43, 1-59.*

## P08-08 Mathematical Modelling of COVID-19 Epidemics in Tokyo Metropolitan and New York City

**Kazumi Omata[1,2], Hiroaki Mitsuya[1,2,3]**
1 National Center for Global Health and Medicine, Japan
2 Departments of Hematology, Rheumatology, and Infectious Diseases, Kumamoto University, Japan
3 Experimental Retrovirology Section, Center for Cancer Research, National Cancer Institute, National Institutes of Health, United States

The comparative study of the COVID-19 epidemics in Tokyo Metropolitan and New York City (NYC) is an intriguing research subject in two points. First, the large contrast of the epidemic sizes in the two cities attracts our great attention. Second, the two cities have been publishing high-quality datasets for the COVID-19 epidemics, so that epidemiological studies should definitely make use of them to obtain useful information. Datasets used are from the daily case reports by Tokyo and NYC Governments. The timings of infections and the effective reproductive ($R_e$) numbers were determined from February 2020 through January 2021, using the back-calculation method and exponential growth model. The determined $R_e$ numbers were substituted into a modified SEIR compartment model, and theoretical epidemic curves were derived. The change-with-time of $R_e$ numbers showed irregular evolution in Tokyo, while that in NYC shows a large and one-time peak in $R_e$ numbers at the early stage of the epidemics followed by small fluctuation. The greatest $R_e$ numbers since March 2020 were 4.19 (95%CI: 3.63-4.75) and 11.7 (95%CI: 10.5-12.8), and the duration periods of $R_e$ numbers exceeding 2.0 in March were less than 9 days and more than 10 days in Tokyo and NYC, respectively. These results suggest that the difference in the amplitude and duration period of large $R_e$ numbers at the early stage decided the issue in the two cities. With respect to our analyses using the compartment model, we showed the possibility that the medical treatments have been improved and the aged individuals have been getting cautious to COVID-19 in Tokyo since July 2020, whereas such changes arose in NYC after November. While various factors can account for the epidemic contrast in the two cities, we proposed a difference in medical intervention as an additional factor, i.e., whether that is individual-oriented countermeasure or not. Although the time lags between infection and detection impose us to accept a few days for the duration of large $R_e$ numbers, we must avoid the duration period of more than 10 days. Careful medical (individual-oriented) interventions must be also implemented irrespective of the $R_e$ numbers.

# POSTER SESSION 08

## P08-09 Application of a spatio-temporal SVEIRD model to COVID-19 epidemic in the Czech Republic

**Jake Doody[1], Ashok Krishnamurthy[2], Bedřich Sousedík[1]**

1 Department of Mathematics and Statistics, University of Maryland, Baltimore County, United States

2 Department of Mathematics and Computing, Mount Royal University, Alberta, Canada

Data assimilation is a general Bayesian technique for repeatedly and optimally updating an estimate of the current state of a dynamic model. We present advanced methods of Bayesian data assimilation to epidemiology, specifically the Optimal Statistical Interpolation (Cobb et al., (2014), in this case to capture the transmission dynamics and the spatial spread of the ongoing COVID-19 epidemic in the Czech Republic. The machinery of data assimilation acts to integrate daily incidence, recovery and death data (Komenda et al., 2020) as made available by the Ministry of Health of the Czech Republic into a fully spatial Susceptible-Vaccinated-Exposed-Infectious-Recovered-Dead (SVEIRD) compartmental model for the tracking process. Rather than representing the population as a linked set of regions or districts, we represent the population as a gridded map. Each grid cell has a population count, which is divided into disease compartments. Each grid cell can transmit disease to its neighbors, with probabilities that decline exponentially with the Euclidean distance. We use the proposed spatial SVEIRD model to estimate and project the number of newly infected and death cases up to August 1, 2021. We use mathematical modeling in order to provide insights that would support public health agencies towards informed, data-driven decision making.

**References:** *[1] Cobb L., Krishnamurthy A., Mandel J., Beezley J., Bayesian tracking of emerging epidemics using ensemble optimal statistical interpolation (EnOSI) . Spatial and Spatio-temporal Epidemiology, 2014;10:39-48. [2] Komenda M., et al., Complex reporting of the COVID-19 Epidemic in the Czech Republic: Use of an Interactive Web-Based App in Practice, J Med Internet Res, 2020;22(5):e19367*

## P08-10 Impact of STI screening intensity on antibiotic exposure: A modelling study among men who have sex with men in Belgium

**Achilleas Tsoumanis[1,2], Niel Hens[3,4], Chris Richard Kenyon[5,6]**

1 Clinical Trials Unit, Institute of Tropical Medicine, Antwerp, Belgium

2 Centre for Health Economic Research and Modeling Infectious Diseases, Vaccine, and Infectious Disease Institute (VAXINFECTIO), University of Antwerp, Belgium

3 Centre for Health Economics Research and Modelling Infectious Diseases (CHERMID), Vaccine and Infectious Disease Institute (VAXINFECTIO), University of Antwerp, Belgium

4 Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Data Science Institute, Hasselt University, Belgium

5 HIV/STU Unit, Institute of Tropical Medicine, Antwerp, Belgium

6 Department of Medicine, Division of Infectious Diseases and HIV Medicine, University of Cape Town, South Africa

**Background:** Neisseria gonorrhoeae (NG) could become untreatable in the near future, as it has developed resistance to all class antibiotics it has been exposed to. While treatment of symptomatic NG in core groups, such as men who have sex with men (MSM) is crucial, screening programs that target asymptomatic NG cases may contribute to an excessive exposure of the population to antibiotics and thus contribute to emergence of antibiotic resistance in NG. It is important to ensure that screening has benefits that outweigh the risks of increased antibiotics resistance.

**Methods:** We used a network-based mathematical model of NG transmission dynamics among MSM in Belgium to estimate the prevalence of NG in the population and the amount of antibiotic uptake. The model simulates the daily transmission of NG among three anatomical sites (pharynx, urethra, rectum). Low- and high risk behaviours, are modelled for a more realistic approach. The effects of different screening intensities on NG prevalence and antibiotic exposure were explored.

**Results:** The model was simulated in a population of 10.000 Belgian MSM over a period of 10 years. Different combinations of screening intensity (annual, biannual or every 3 months) and coverage (5% - 50% of the population) were compared to no screening. Annual screening of 50% of the population resulted in a prevalence of 7.6% in the pharynx, 3.8% in the urethra and 9.4% in the rectum, compared to 9.5% (pharynx), 4.8% (urethra) and 11.9% (rectum) in the no-screening scenario. In the most intensive scenario (3-monthly, 20% coverage) the prevalence was reduced to 3.8% (pharynx), 2.1% (urethra) and 5.9% (rectum). If 50% of the population were screened on an annual basis, the proportion of the asymptomatic population exposed to antibiotic treatment increased from 9.6% to 26.4% with 3-monthly screening of 30% of the population.

**Discussion:** All scenarios reduced the prevalence of NG in all anatomical sites, compared to no screening. However, the most screening-intensive scenarios, exposed a large part of the population to antibiotics, which could result in the emergence of antibiotic resistant NG and other organisms.



POSTER SESSION 08

## P08-11 Epilocal: A real-time tool for local epidemic monitoring

**Marco Bonetti[1], Ugofilippo Basellini[2]**

1 Carlo F. Dondena Research Center, Bocconi Institute for Data Science and Analytics, and Covid Crisis Lab, Bocconi University, Milan, Italy

2 The Max Planck Institute for Demographic Research (MPIDR), Rostock, and the Institut national d'études démographiques (INED), Aubervilliers, France

**Background:** The novel coronavirus (SARS-CoV-2) emerged as a global threat at the beginning of 2020, spreading around the globe at different times and rates. Within a country, such differences provide the opportunity for strategic allocations of health care resources.

**Objective:** We aim to provide a tool to estimate and visualize differences in the spread of the pandemic at the subnational level. Specifically, we focus on the case of Italy, a country that has been harshly hit by the virus.

**Methods:** We model the number of SARS-CoV-2 reported cases and deaths as well as the number of hospital admissions at the Italian subnational level with Poisson regression. We employ parametric and nonparametric functional forms for the hazard function. In the parametric approach, model selection is performed using an automatic criterion based on the statistical significance of the estimated parameters and on goodness-of-fit assessment. In the nonparametric approach, we employ out-of-sample forecasting error minimization.

**Results:** For each province and region, fitted models are plotted against observed data, demonstrating the appropriateness of the modeling approach. Moreover, estimated counts and rates of change for each outcome variable are plotted on maps of the country. This provides a direct visual assessment of the geographic distribution of risk areas as well as insights on the evolution of the pandemic over time.

**Contribution:** The proposed Epilocal software provides researchers and policymakers with an open-access real-time tool to monitor the most recent trends of the COVID-19 pandemic in Italian regions and provinces with informative graphical outputs. The software is freely available and can be easily modified to fit other countries as well as future pandemics.

**References:** *[1] Bonetti, M. and Basellini U. (2021). Epilocal: a real-time tool for local epidemic monitoring. Demographic Research 44:307-332.*

# POSTER SESSION 09

## POSTER SESSION 09 — Individual prediction and precision medicine

### P09-01 A Bayesian nonparametric approach for modeling SF-6D health state utility scores

**Samer A. Kharroubi**
Department of Nutrition and Food Sciences, Faculty of Agricultural and Food Sciences, American University of Beirut, Lebanon

**Background:** Typically, models that were used for health state valuation data have been parametric. Recently, many researchers have explored the use of non-parametric Bayesian methods in this field.
**Objectives:** In the present paper we report on the results from using a nonparametric model to predict a Bayesian SF-6D health state valuation algorithm along with estimating the effect of the individual characteristics on health state valuations.
**Methods:** A sample of 126 Lebanese members from the American University of Beirut valued 49 SF-6D health states using the standard gamble technique. Results from applying the nonparametric model were reported and compared to those obtained using a standard parametric model. The covariates' effect on health state valuations was also reported.
**Results:** The nonparametric Bayesian model was found to perform better than the parametric model 1) at predicting health state values within the full estimation data and in an out-of-sample validation in terms of mean predictions, root mean squared error and the patterns of standardized residuals, and 2) at allowing for the covariates' effect to vary by health state. The findings also suggest a potential age effect with some gender effect.
**Conclusion:** The nonparametric model is theoretically more flexible and produces better utility predictions from the SF-6D than previously used classical parametric model. In addition, the Bayesian model is more appropriate to account the covariates' effect. Further research is encouraged.

### P09-02 Small-sample accuracy of approximations of individual polynomial growth curves' prediction error in linear mixed regression

**Math J.J.M. Candel**
Department Methodology and Statistics, CAPHRI, Maastricht University, Netherlands

Empirical Bayes estimates of random coefficients in mixed model regression allow for estimating individual growth curves. The expected average squared prediction error (ASPE) quantifies the prediction error of estimated growth curves. From the mean squared error of Bayes estimates based on the method of moments (Rao, 1975), two asymptotic expressions for the expected ASPE can be derived for empirical Bayes estimates based on maximum likelihood estimation. In an extensive Monte Carlo study, the small-sample accuracy of these expressions for the expected ASPE, when employing restricted maximum likelihood estimation, is examined for linear and quadratic trends across time. Varied are the number of subjects, the number of time points, relevant model parameters, as well as the time span of prediction. For the best approximation formula underestimation of the expected ASPE was smaller than 2% for a linear trend, and smaller than 5% for a quadratic trend. These results also held under various violations of normality for the random coefficients in mixed model regression. To obtain a safe upper bound for the prediction error when planning a study on individual growth curves, that is, when deciding upon the study length and the number of measurement occasions, these results suggest to multiply the prediction error calculated by the best approximation formula by a factor 1/0.98 for a linear trend and by a factor 1/0.95 for a quadratic trend.

**References:** *[1] Rao, CR. Simultaneous estimation of parameters in different linear models and applications to biometric problems. Biometrics. 1975; 31: 545-554.¨*



POSTER SESSION 09

### P09-03 Assessment of performance measures for external validation of multivariable prediction models: Simulations

**Harald Heinzl[1], Harbajan Chadha-Boreham[2]**
1 Section for Clinical Biometrics, CeMSIIS, Medical University of Vienna, Austria
2 Clinical Biostatistics Consultancy, Dijon, France

Multivariable prediction models (MPMs) are developed for the diagnosis of diseases or the prediction of their progress. Before an MPM can be used in clinical practice, it has to be internally and externally validated.
Commonly used performance measure for external validation are the C-statistic, the Brier score and the R-squared measure (Collins et al. 2014). We consider requirements and conditions for the assessment of these performance measures in a simulation framework, and we will present the prototype of such a simulation framework.
A key feature of external validation is generalisability, an umbrella term for transportability and reproducibility of an MPM. If the development population and the external validation population are closely related, then we can assess the reproducibility of the model; if they are not closely related, then transportability will be the issue (Debray et al. 2015). When developing our simulation framework to assess the impact of relatedness between development and external validation population on performance measures, a crucial point was to establish a simple, reasonable and easy to communicate approach.
Our simulation results show that the C-statistic, the Brier score and the R-squared performance measures work well in transportability settings. On the other hand, the results revealed unexpected behaviour of the Brier score in reproducibility settings. Furthermore, our simulation framework is simple, reasonable and easy to communicate.

**References:** *[1] Collins et al.: External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Medical Research Methodology 2014 14:40. doi:10.1186/1471-2288-14-40. [2] Debray et al.: A new framework to enhance the interpretation of external validation studies of clinical prediction models. Journal of Clinical Epidemiology 2015; 68:279-289. doi: 10.1016/j.jclinepi.2014.06.018*

### P09-04 Mathematical proof of the equivalence between Post-test Predictive Probabilities and Predictive Values

**Ettore Rocchi, Sara Peluso, Davide Sisti**
Department of Biomolecular Sciences, Unit of Biostatistics and Biomathematics, University of Urbino "Carlo Bo", Italy

In the last years, in order to evaluate of diagnostic performance, many Authors (see for example Crouser et al, 2019) use the so-called Post-Test Predictive Probability (Positive and Negative) in addiction to the usuals indices, such as Sensitivity, Specificity, Positive Predictive Value and Negative Predictive Value, Positive Likelihood Ratio and Negative Likelihood Ratio. Positive Post-Test Predictive Probability is formulated as a function of Pre-test Predictive Probability and Positive Likelihood Ratio, while Negative Post-Test Predictive Probability is formulated as a function of Pre-test Predictive Probability and Negative Likelihood Ratio. The Pre-test Predictive Probability corresponds to the Prevalence. Moreover, simply by starting from the definition of Positive Predictive Value and Negative Predictive Value based on the Bayes' theorem, through a few mathematical passages (in detail, making explicit the Likelihood Ratios as functions of Sensitivity and Specificity), we demonstrate that they are equivalent to Positive Post-Test Predictive Probability and to the complement of Negative Post-Test Predictive Probability, respectively.
In conclusion, we think that it is unnecessary, if not detrimental and confusing, to introduce other parameters in addition to Sensitivity, Specificity, Positive Likelihood Ratio, Negative Likelihood Ratio, Positive Predictive Value, and Negative Predictive Value in order to evaluate a diagnostic test performance, especially if they don't get any additional information.

**References:** *[1] Crouser ED, Parrillo JE, Seymour CW, Angus DC et al. Monocyte Distribution Width: A Novel Indicator of Sepsis-2 and Sepsis-3 in High-Risk Emergency Department Patients. Crit Care Med 2019; 47:1018-1025*

# POSTER SESSION 09

## P09-05 Development and validation of a clinical risk score to predict the risk of SARS-CoV-2 infection from administrative data

**Laura Savaré[1,2], Valentina Orlando[3], Federico Rea[4]**

1 MOX - Modelling and Scientific Computing Laboratory, Department of Mathematics, Politecnico di Milano, Italy
2 Center for Analysis, Decisions and Society, Human Technopole, Milan, Italy
3 Center of Drug Utilization and Pharmacoeconomics, University of Naples Federico II, Italy
4 National Centre for Healthcare Research & Pharmacoepidemiology, University of Milano-Bicocca, Italy

**Background:** The novel coronavirus (SARS-CoV-2) pandemic spread rapidly worldwide, early increasing exponentially in Italy. To date, there is a lack of studies describing the clinical characteristics of the population most at risk of infection.
**Objectives:** Our aim was to identify clinical predictors of SARS-CoV-2 infection risk and to develop and validate a score predicting SARS-CoV-2 infection risk comparing it with unspecific surrogates.
**Methods:** A retrospective case/control study using administrative health-related database was carried out in Southern Italy (Campania region) among beneficiaries of Regional Health Service aged over 30 years. For each subject with Covid-19 infection confirmed diagnosis (case), up to five controls were randomly matched for gender, age and municipality of residence. Odds ratios and 90% confidence intervals for associations between candidate predictors and risk of infection were estimated by means of conditional logistic regression. SARS-CoV-2 Infection Score (SIS) was developed by generating a total aggregate score obtained from the assignment of a weight at each selected covariate, according to the least absolute shrinkage and selection operator (LASSO) method, using the coefficients estimated from the model. Finally, the score was categorized by assigning increasing values from 1 to 4. To evaluate the clinical utility of SIS for predicting infection and to compare the discriminatory ability of specific and unspecific predictors of SARS-CoV-2 infection, ROC curves and corresponding AUCs were used.
**Results:** Subjects suffering from diabetes, anaemias, Parkinson's disease, mental disorders, cardiovascular and inflammatory bowel and kidney diseases showed an increased risk of SARS-CoV-2 infection. Similar estimates were recorded for men and women and younger and older than 65 years. Fifteen conditions significantly contributed to the SIS. As SIS value increases, risk progressively increases, being the odds of SARS-CoV-2 infection among people with the highest SIS value (SIS= 4), 1.74 times higher than those unaffected by any SIS contributing conditions (SIS=1). However, there was no evidence that specific scores had different discriminatory ability.
**Conclusions:** This study identified conditions and diseases making individuals more vulnerable to SARS-CoV-2 infection. Our results can be a tool supporting decision-makers for the identification of the population most vulnerable to Covid-19, allowing them to adopt preventive measures.

# POSTER SESSION 09

## P09-06 Subject-specific networks as features for predictive modelling - A scoping review of methods

**Mariella Gregorich[1], Federico Melograna[2], Martina Sunqvist[3], Stefan Michiels[3], Kristel Van Steen[2,4], Georg Heinze[1]**

1 Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria
2 BIO3 - Laboratory for Systems Medicine, KU Leuven, Belgium
3 Service de Biostatistique et d'Epidémiologie, Gustave Roussy, Oncostat U1018, Inserm, University Paris-Saclay, labeled Ligue Contre le Cancer, Villejuif, France
4 BIO3 – Laboratory for Systems Genetics, GIGA-R Medical Genomics, University of Liège, Belgium

**Background:** Recent advances in biotechnology (e.g. imaging modalities, microarrays and sequencing) enable the acquisition of high-dimensional data on individuals, posing challenges for prediction models which traditionally use covariates such as patient characteristics. Alternative forms of covariate representations for the features derived from these modern data modalities should be considered that can utilize their intrinsic interconnection. The connectivity information between these features can be represented as a network defined by a set of nodes and edges and specified for each individual in the dataset. Then, global or local graph-theoretical features may yield potential prognostic biomarkers instead of or in addition to traditional covariates and may replace the often unsuccessful search for individual biomarkers in a high-dimensional predictor space.
**Methods:** We conducted a scoping review to identify, summarize and critically appraise the state-of-art in the use of subject-specific networks for predictive modelling in biomedicine published during the time period 2000-2020 in the electronic databases PubMed, Scopus, Embase and arXiv.
**Results:** Our scoping review revealed that network approaches have been predominantly applied in neurological and pathopsychological studies, followed by genomics, cardiology and pathology (N=143). Data-driven network construction was mainly based on Pearson correlation coefficients of time series, but also alternative approaches (e.g. Gaussian graphical modelling, measures of distributional similarity) could be found. For independent variables measured only once for a subject, individual network construction was primarily based on leave-one-out resampling quantifying each individual's contribution to the overall group-level structure. Graph-theoretical features were mainly assessed locally (nodes) and used for classification between outcome groups by supervised learning techniques (support vector machines, random forests). While the value of the subject-specific network framework was highlighted in most of the identified studies, we found methodological concerns regarding network sparsification approaches and graph-theoretical feature extraction and selection.
**Conclusion:** This study demonstrates the potential of subject-specific networks as a new powerful approach to predictive modelling in medicine, in particular for modern data modalities spanning a high-dimensional predictor space. Although the intersection of network science and predictive modelling is a promising area of research, more work is needed on adequate network sparsification and outcome-guided graph-theoretical feature extraction and selection.

# POSTER SESSION 09

## P09-07 Individual dynamic predictions in joint analysis of non-linear longitudinal model and parametric competing risk model: application to sepsis patients

**Alexandra Lavalley-Morelle[1], Jean-François Timsit[1,2], Jimmy Mullaert[1], Outcomerea network**

1 Université de Paris, IAME, Inserm, France

2 AP-HP, Bichat University Hospital Medical and Infectious Diseases ICU, France

**Introduction:** Systemic inflammatory reaction to an infection, called sepsis, potentially leads to fatal organ dysfunctions [1]. The SOFA score [2], widely used in Intensive Care Unit (ICU) context, allows to quantify these organ dysfunctions. The objective of the work is to propose an original joint modelling approach based on a longitudinal model for daily SOFA measurement, coupled with a competing risk survival model to provide dynamic predictions of the risk of death during ICU stay.

**Methods:** Training data contains 4050 patients admitted in ICU for sepsis randomly taken from the OUTCOME REA database. The joint model is composed of a non-linear mixed effects sub-model to model individual SOFA evolution, and a full parametric sub-distribution hazard model, adjusted on age, to model the risk of death in the presence of ICU discharge as a competing event. Parameters are estimated with SAEM algorithm implemented on Monolix. This modelling approach is original and allows more flexibility than previous reported joint models based on linear mixed effects models [3]. Performances are assessed on a validation set composed of 1996 other patients randomly taken from the same database. Predictive performances are based on a landmark analysis with four landmark times at 5, 7, 10 and 15 days after admission and a horizon time of 30 days after admission. For each landmark time we report time-dependent AUC and Brier score.

**Results:** Median age was 65.1 years (IQR : 52.7 – 76.2) and median admission SOFA was 6 (IQR : 4 – 9). The 30-day mortality was 19%. A 15-day follow-up allows to well predict the 30-day post admission mortality with an AUC of 0.90 and a Brier score of 0.10. Early landmark times also showed good properties to predict 30-day mortality (for landmark=5d: AUC = 0.83, Brier score = 0.11).

**Discussion:** The joint modelling approach allows to quantify the future risk of death associated with any individual SOFA history, while genuinely overcoming issues with time-dependant covariates in sub-distribution hazard models. Despite better predictive performances than previous studies, further investigations on the properties of such models (e.g. validity of inference), based on simulation, are needed.

**References:** *[1] Singer et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3), JAMA (2016)*
*[2] Jones et al. The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation, Critical Care Medicine (2009)*
*[3] Musoro et al. Dynamic prediction of mortality among patients in intensive care using the sequential organ failure assessment (SOFA) score: a joint competing risk survival and longitudinal modeling approach, Statistica Neerlandica (2018)*



# POSTER SESSION 09

## P09-08 Estimating the treatment selection ability of a marker: review, comparison, and improvement of existing approaches

**Yoann Blangero[1,2,3,4], Muriel Rabilloud[1,2,3], Fabien Subtil[1,2,3]**

1 Service de Biostatistique-Bioinformatique, Pôle Santé Publique, Hospices Civils de Lyon, France

2 Université de Lyon, Université Lyon 1, Villeurbanne, France

3 UMR CNRS 5558, Laboratoire de Biométrie et Biologie Évolutive, Villeurbanne, France

4 Soladis, Lyon, France

Treatment-selection markers are essential for precision medicine; they help clinicians choosing between two (or more) treatments. Most of these markers were initially continuous; hence, the early phases of treatment-selection marker assessment required methods to measure the overall ability of the marker for treatment selection without threshold defined. Historically, marker assessments have been carried out through analyses of treatment-marker interactions. However, the conclusions of such analyses depend on the model chosen (additive or multiplicative), and these assessments do not reflect the effect of a marker-based treatment choice on patients' health. More recently, several indexes have been proposed, some purely statistical and some grounded in the medical decision-making theory. Graphical approaches have also been suggested but without sum-up indexes. This work presents and compare some of the existing approaches. First, the context of treatment-selection marker assessment is presented with associated questions. Second, some existing methods are described, insisting on the underlying links between them and on possible improvements. More precisely, the work focuses on the ability of comparing the indexes stemming from various studies. Then, a simulation study is performed to compare the statistical properties of the methods considered (i.e., mean relative bias, RMSE, power, coverage, and mean width of the confidence intervals) by varying the ability of the marker for treatment selection and the risk of event (e.g, treatment failure, disease progression, etc.) in each treatment arm. Within the context of treatment selection, approaches to marker evaluation should reflect the consequences of the marker-based treatment choice on patients' health. They should also provide indexes able to yield an interpretation that is independent of the risk of event in each treatment arm to enable comparisons between several studies.

## P09-09 Double cut-point identification of continuous diagnostic test variables for clinical use

**Laura Antolini[1], Laura Cristoferi[1,2,3,4], Davide Paolo Bernasconi[1], Marco Carbone[1,2,3], Maria Grazia Valsecchi[1]**

1 School of Medicine, Università Milano-Bicocca, Italy

2 Division of Gastroenterology, Center for Autoimmune Liver Diseases, Department of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy

3 European Reference Network on Hepatological Diseases (ERN RARE-LIVER), San Gerardo Hospital, Monza, Italy

4 Bicocca Bioinformatics Biostatistics and Bioimaging Centre - B4, School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy

The identification of classification rules on a continuous diagnostic test is often of interest in laboratory/clinical data applications such as the case ofvibration-controlled transient elastography (VCTE) (Fibroscan, Echosens, Paris), a simple and non-invasive realiable biomarker of liver fibrosis. In order to confirm or exclude the presence of liver fibrosis using VCTE results (avoding performing more invasive tests i.e. liver biopsy) two cut-points could be identified with two different approaches. The first consists in giving a predicted risk equal to a sufficiently low/high fixed values, and the second in using fixed negative and positive predictive values. Both approaches include an intermediate risk classification. However, the commonly used approaches identify double or single cut-points on the ground of "reversed probabilities" of the distribution of the diagnostic result in diseased/non diseased populations by: fixed values of sensitivity and specificity, maximum sensitivity*specificity, maximum Youden function, nearest to (0,1) point on the ROC curve, etc. In our opinion, the use of these methods is due to three reasons:

1. the presence of case-control data where the predicted risk cannot be calculated;
2. the non monotonicity of the nonparametric estimate of predictive values as function of the diagnostic test result;
3. the advocated advantage of cut-points identified on measures that do not depend on the prevalence.

Of note, the latter aspect is in contrast with the clinical use of the diagnostic test in a population with a given prevalence and a classification rule that should depend on the numerical predicted risk [1]. In this work we show how in case of cross sectional/cohort data, such as the motivating application [2], the identification of two cut-points can be carried out to achieve the desidered sufficiently low/high fixed values for the predicted risk, or predictive values, instead of relying on sensitivity and specificity. We prove that in the latter case the interval between the two cut-points is slightly reduced with respect to the former. We discuss how cut-points could be reasonably modified to be used in populations with difference prevalence.

**References:** *[1] Trevethan R (2017) Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. Front Public Heal 5:307. [2] Cristoferi L, Nardi A, Viganò M, Rigamonti C, Degasperi E, Cardinale V et al. (2020) Accuracy of liver stiffness measurement in assessing liver fibrosis in naive patients with primary biliary cholangitis. J Hepatol 73:s401–s652*

# POSTER SESSION 09

## P09-10 Patient heterogeneity assessments via network-based ANOVA

**Diane Duroux[1], Federico Melograna[2], Kristel Van Steen[1,2]**

1 BIO3-GIGA R Medical Genomics, University of Liège, Belgium

2 BIO3-Department of Human Genetics, KU Leuven, Belgium

Studying heterogeneity between individuals lies at the core of precision medicine. Often, information about individuals can be represented as networks characterized by individual-specific edge values [1]. In this project, we develop a procedure to identify those networks that can be aggregated in a cluster where the determination of the final clustering is based on notions of statistical significant differences between clusters. In particular, we first use an unsupervised hierarchical algorithm to identify latent classes of similar networks. Similarity between networks is computed via appropriate distance measures between graphs (e.g. shortest-path kernel [2], DeltaCon [3], GTOM [4]). To determine the optimal number of clusters, we recursively test for distances between two groups of networks, progressing from the root node to the end nodes of the clustering dendrogram. The test itself is based on computing pairwise distances between graphs and is inspired by distance-wise ANOVA algorithms, commonly used in ecology [5]. Permutations are used to assess significance. We show the merits and pitfalls of our approach via extensive simulations and an application to inflammatory diseases. The results of our strategy pave the way towards deciding which networks can sensibly be aggregated. This is not only relevant in stratified medicine or for molecular reclassification of disease, but also in areas such as genetic epistasis detection in which conclusions need to be drawn from multiple statistical epistasis networks across different analysis protocols.

**References:** *[1] Kuijjer, M. L., Tung, M. G., Yuan, G., Quackenbush, J., & Glass, K. (2019). Estimating sample-specific regulatory networks. Iscience, 14, 226-240. [2] Borgwardt, K. M., & Kriegel, H. P. (2005, November). Shortest-path kernels on graphs. In Fifth IEEE international conference on data mining (ICDM'05) (pp. 8-pp). IEEE. [3] Koutra, D., Shah, N., Vogelstein, J. T., Gallagher, B., & Faloutsos, C. (2016). Deltacon: Principled massive-graph similarity function with attribution. ACM Transactions on Knowledge Discovery from Data (TKDD), 10(3), 1-43. [4] Yip, A. M., & Horvath, S. (2006, June). The Generalized Topological Overlap Matrix for Detecting Modules in Gene Networks. In BIOCOMP (pp. 451-457). [5] Anderson, M. J. (2005). Permutational multivariate analysis of variance. Department of Statistics, University of Auckland, Auckland, 26, 32-46.*

## P09-11 A comparison of prioritisation tools for CVD risk assessment in UK Biobank: Primary care records versus polygenic risk scores

**Ryan Chung[1], Angela Wood[2], Juliet Usher-Smith[3]**

1 Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, UK

2 Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, UK

3 Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, UK

**Background:** Public health guidelines in England recommend individuals at higher estimated risk of cardiovascular disease (CVD) be prioritised for formal CVD risk assessment. However, no prioritisation tool has been proposed. We propose and compare two strategies involving a prioritisation tool leveraging routinely collected data within primary care records or polygenic risk scores (PRS) followed by the recommended *QRISK2* CVD risk assessment model.

**Data:** 108,685 participants in UK Biobank aged between 40-69 years, with primary care records, measured biomarkers and genetic data, and without prevalent CVD or statin use. Incident CVD events were defined using Hospital Episode Statistics and death registry records.

**Methods:** Strategy 1 uses a prioritisation tool using multivariate-linear-mixed models to leverage sporadically recorded longitudinal primary care records followed by the *QRISK2* model. The prioritisation tool was derived using sex-specific Cox models with *QRISK2* risk factors, using last observed values and mixed model estimates. Strategy 2 uses PRS for the prioritisation tool followed by the *QRISK2* model. The prioritisation tool was derived using sex-specific Cox models with the risk factors: age, a PRS for coronary artery disease and a PRS for stroke. Prioritisation tools were applied to 40-54-year olds. Those in the top 50% of predicted risk, as well as all 55-69 year olds, underwent a formal risk assessment using the *QRISK2* model based on UK Biobank baseline data. For each strategy we estimated the % reduction in (i) the number needed to screen (NNS) to prevent one CVD event and (ii) identified CVD events, in comparison to carrying out formal CVD assessment on all 40-69-year olds. We applied a recalibration approach to enable performance to reflect that in the general English population.

**Results/Conclusions:** A prioritisation tool based on routinely available data outperforms a tool incorporating PRS. Strategy 1 led to reductions in NNS of 27.3% for women and 26.5% for men, whilst only reducing the percentage of identified CVD events by 0.05% for women and 1.5% for men. Comparatively, strategy 2 reduced the NNS by 26.9% for women and 25.1% for men, and reduced the percentage of identified CVD events by 0.67% for women and 3.3% for men.



# POSTER SESSION 09

## P09-12 Operational characteristics of generalized pairwise comparisons for hierarchically ordered endpoints

**Vaiva Deltuvaite-Thomas[1], Tomasz Burzykowski[1,2]**

1 International Drug Development Institute, Belgium

2 Hasselt University, Data Science Institute, I-BioStat, Belgium

The method of generalized pairwise comparisons (GPC) is a multivariate extension of the well-known non-parametric Wilcoxon-Mann-Whitney test. It allows comparing two groups of observations based on multiple hierarchically ordered endpoints, regardless of the number or type of the latter. The summary measure, ``net benefit'', quantifies the difference between the probabilities that a random observation from one group is doing better than an observation from the opposite group. The method takes into account the correlations between the endpoints. We have performed a simulation study for the case of two hierarchical endpoints to evaluate the impact of their correlation on the type-I error probability and power of the test based on GPC. The simulations show that the power of the GPC test for the primary endpoint is modified if the secondary endpoint is included in the hierarchical GPC analysis. The change in power depends on the correlation between the endpoints. Interestingly, a decrease in power can occur, regardless of whether there is any marginal treatment effect on the secondary endpoint. It appears that the overall power of the hierarchical GPC procedure depends, in a complex manner, on the entire variance-covariance structure of the set of outcomes. Any additional factors (such as thresholds of clinical relevance, drop out, or censoring scheme) will also affect the power and will have to be taken into account when designing a trial based on the hierarchical GPC procedure.

## P09-13 Landmark Prediction of Survival for HIV-infected Patients by Considering AIDS as an Intermediate Event

**Behnaz Alafchi[1], Hemin Shanazy[1], Jalal Poorolajal[2]**

1 Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Iran

2 Research Center for Health Sciences and Department of Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Iran

**Background:** The overall survival of HIV-infected patients may be shortened by rapid progression to acquired immunodeficiency syndrome (AIDS). So, instead of the occurrence of death as a single event, the progression of disease should be modeled. Here, progression-to-AIDS is considered as an intermediate event. Landmarking prediction method focused on dynamic prediction of survival time given the information of the intermediate events at the moment of prediction. In this study landmark method is used to dynamic prediction of survival of HIV-infected patients, given the history of AIDS status.

**Methods:** The information of 1530 HIV-infected patients, which was related to a registry-based retrospective cohort study were analyzed in this study. A method of landmarking was utilized for dynamic prediction of survival of the patients. Dynamic C-index and time-dependent area under the ROC curve (AUC) were used to evaluate the performance of the model.

**Results:** About 34.4% of patients were diagnosed at the advanced stage of HIV and their CD4 cell count were less than 200 cells/mm$^3$. The majority of HIV infected patients were male (76.5%) and about 47.7% of patients have received ART. The probability of dying within 5-years was lower in females ($p=0.027$) and patients who their CD4 cell count were lower at the time of diagnosis ($p<0.001$). The results also revealed that progression to AIDS can increase the hazard of death and shorten the survival time of HIV infected patients ($p=0.040$). Moreover, the probability of death within the next 5-years was lower in patients who received ART either their disease progressed to AIDS or not ($p<0.001$). The effect of ART on the survival time were related to the level of CD4 cell count, such that the impact of receiving ART on the hazard of death were lower in patients with lower level of CD4 cell count. The value of the dynamic C-index=0.79 and AUC=0.80 indicate a promising predictive accuracy.

**Conclusions:** As the disease progression could be considered in the landmark model and its performance was promising in analyzing the survival processes of HIV-infected patients, it could be an adequate choice for analyzing such survival data in the future studies.

**References** *[1] van Houwelingen H, Putter H. Dynamic prediction in clinical survival analysis: CRC Press; 2011.*
*[2] World Health Organization. HIV/AIDS. 2020 [Available from: http://www.who.int/news-room/fact-sheets/detail/hiv-aids.*

# POSTER SESSION 10

POSTER SESSION 10 | Meta analysis

## P10-01 Video-Assisted Non-Intubated Lobectomies for Lung Cancer: A Systematic Review and Meta-Analysis

**Luca Bertolaccini[1], Elena Prisciandaro[1], Niccolò Filippi[1], Lorenzo Spaggiari[1,2]**

1 Department of Thoracic Surgery, IEO, European Institute of Oncology IRCCS, Milan, Italy

2 Department of Oncology and Haemato-oncology, University of Milan, Italy

**Introduction:** To assess safety, feasibility and oncological outcomes of video-assisted non-intubated lobectomies for non-small cell lung cancer (NSCLC).

**Methods:** A comprehensive search performed in EMBASE (via Ovid), MEDLINE (via PubMed) and Cochrane CENTRAL from 2004 to 2020. Studies comparing non-intubated anaesthesia with intubated anaesthesia for video-assisted lobectomy for NSCLC were included. A systematic review and meta-analysis were performed by combining the individual studies' reported outcomes using a random effect model. For dichotomous outcomes, risk ratios (RR) was calculated, and for continuous outcomes, the mean difference (MD) was used.

**Results:** Three retrospective cohort studies were included. The comparison between non-intubated and intubated patients undergoing video-assisted lobectomy showed no statistically significant differences in postoperative complication rate (RR = 0.65; 95% confidence interval (CI) = 0.36 – 1.16; p = 0.30; $I^2$ = 17%), operating times (MD= -12.40; 95% CI = -28.57 – 3.77; p = 0.15; $I^2$ = 48%), length of hospital stay (MD = -1.13; 95% CI = -2.32 – 0.05; p = 0.90; $I^2$ = 0%) and number of dissected lymph nodes (RR = 0.92; 95% CI = 0.78 – 1.25; p = 0.46; $I^2$ = 0%).

**Conclusion:** Awake and intubated video-assisted lobectomies for resectable NSCLC have comparable perioperative and postoperative outcomes. The oncological implications of the non-intubated approach should be considered. Radical pulmonary resection and systematic lymph node dissection might prove more challenging and harder to achieve due to spontaneous ventilation maintenance. Further research should be focused on the safety and feasibility of anatomical lung resections under spontaneous breathing. The long-term benefits for patients with lung cancer still need to be carefully assessed.

## P10-02 Bias in evaluation of discrete surrogate outcomes, due to separation: a penalized likelihood solution

**Hannah Ensor, Christopher J. Weir**

Edinburgh Clinical Trials Unit, Usher Institute, University of Edinburgh, United Kingdom

**Background:** The use of a surrogate in place of the outcome of true clinical interest is one strategy intended to improve the efficiency of clinical trials and hence drug development programmes. Rigorous evaluation of a potential surrogate outcome is required to ensure the validity of inference about the effect of treatment on the true clinical outcome that has been replaced by a surrogate. Various techniques are available to assess surrogates: a leading method is use of a meta-analytical information theory-based approach, which enables a trial-level measure of surrogacy, $R^2_{ht}$, to be assessed across a range of outcome types including binary and ordinal discrete outcomes.

**Objectives:** In the context of surrogacy evaluation for discrete outcomes, separation (a zero cell count in the cross-tabulation of the surrogate and true clinical outcome within a randomised treatment group) causes bias in estimating the strength of surrogacy. We investigated the penalized likelihood technique of Firth (1993) as a possible solution to this.

**Methods:** Penalized likelihood adds a systematic bias to the score function to prevent bias in the resulting parameter estimate. We simulated multiple clinical trials for various scenarios (varying the strength of surrogacy; and the number and size of trials) to assess the value of penalized likelihood when investigating a potential binary surrogate outcome for an ordinal true clinical outcome. Outcomes and simulation parameters were based on data from a large clinical trial of compression stockings for prevention of deep vein thrombosis in immobile stroke patients.

**Results:** Compared to a default approach of omitting from the meta-analysis trials in which separation is present, we found that the penalized likelihood approach reduces the bias when estimating the strength of surrogacy using $R^2_{ht}$. It also estimates $R^2_{ht}$ more precisely, utilising all of the available data even in the presence of separation.

**Conclusions:** The adoption of the penalized likelihood approach into information theoretic surrogacy evaluation is a useful addition to address the issue of separation which arises frequently in the context of categorical outcomes.

**References:** *[1] Firth D. (1993) Bias reduction of maximum likelihood estimates. Biometrika 80:27-38.*



# POSTER SESSION 10

## P10-03 Developing a Novel Interactive Multifaceted Graphic for Treatment Ranking within Network Meta-Analysis

**Clareece R. Nevill, Nicola J. Cooper, Alex J. Sutton**

Department of Health Sciences, University of Leicester, United Kingdom

**Background:** For many diseases there are multiple interventions or treatments available, with varying advantages and disadvantages. Network meta-analysis (NMA) is an advantageous statistical methodology for synthesising study results looking at treatment efficacy. In contrast to standard pairwise meta-analysis, NMA has the ability to synthesis data on multiple treatments/interventions simultaneously, allowing comparison between all treatment pairs even where there is no direct evidence. Subsequently, a powerful feature of NMA is the ability to rank interventions.

**Objective:** This project had two aims: (1) ascertain current methods, challenges, and visualisations regarding treatment ranking within an NMA framework; (2) develop a novel graphic within MetaInsight (an interactive free NMA web application) to aid clinicians and stake-holders when making decisions regarding the 'best' intervention(s) for their patient(s).

**Methods:** Current literature regarding methodology and/or visualisation of treatment ranking over the last ten years was collated and studied. Based on the literature, a novel graphical visualisation was developed using RShiny and integrated within MetaInsight, which is currently hosted on shinyapps.io [crsu.shinyapps.io/MetaInsight].

**Results:** Bayesian analyses produce rank probabilities from which mean/median rank and surface under the cumulative ranking curve can be calculated. For frequentist analyses the P-score is available. The simpler methods may be easier to interpret but are often more unstable and don't encompass the whole analysis (and vice versa). To aid interpretation, and facilitate sensitivity analysis, an interactive graphic was developed that presented rankings alongside treatment effect and study quality results.

**Conclusions:** Treatment ranking is a beneficial tool for clinicians and stake-holders, but results should be interpreted cautiously and visualisation transparent and all-encompassing. A 'living' version of MetaInsight, with treatment ranking, would allow interested parties to follow the evidence base as it grows, empowering users to continuously know which treatment(s) is the 'best'.

## P10-04 Propensity Score-Integrated Meta-Analytic-Predictive Priors

**Carissa Reid, Oliver Sailer**

Biostatistics & Data Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany

In clinical trials there may be ethical, financial or feasibility constraints that make it difficult to randomize patients to a control group. Historical control groups may address this problem by borrowing information from external patients and thus, reducing the size of the concurrent control group. A Bayesian approach can be used to incorporate this information in the form of a prior distribution for the parameter of interest. Two popular approaches for defining this prior are the power prior and the Meta-Analytic-Predictive (MAP) prior (Neuenschwander, Capkun-Niggli, Branson & Spiegelhalter, 2010) approaches. Additionally, propensity scores can be used to select a subset of the historical patients that is comparable to the patients in the current study. Bayesian inference and propensity score methods have already been integrated for the purpose of forming historical controls using a propensity score-integrated power prior approach (Wang et. al, 2019). In this research, several new methods were implemented for integrating propensity score stratification or weighting with the MAP prior approach. In the first method, the patients are stratified using propensity scores and then the MAP prior is estimated separately in each stratum. The final estimate is a weighted sum of the estimates in each stratum using either equal weights or weights based on the overlap coefficient of the propensity score distributions. In the second method, one MAP prior is derived by replacing the study level of the standard hierarchical model with the strata level. In other words, the MAP prior is derived from a hierarchical model for strata which ignores the study attribute. Finally, in the third method, the MAP prior is derived from a hierarchical model for studies. However, the weight of each study in the estimation of the posterior mean is adjusted based on a summary statistic of the propensity scores of the patients in each study. Initial simulation results suggest that the second propensity-score integrated MAP method using a random effect per strata outperforms the other methods in terms of bias and mean-squared error. This simulation study will be expanded to consider further scenarios. Additionally, the methods will be applied to data on Alzheimer's disease.

**References:** *[1] Neuenschwander, B., Capkun-Niggli, G., Branson, M., & Spiegelhalter, D. (2010). Summarizing historical information on controls in clinical trials. Clinical Trials: Journal of The Society for Clinical Trials, 7(1), 5-18. doi: 10.1177/1740774509356002. [2] Wang, C., Li, H., Chen, W., Lu, N., Tiwari, R., Xu, Y., & Yue, L. (2019). Propensity score-integrated power prior approach for incorporating real-world evidence in single-arm clinical studies. Journal of Biopharmaceutical Statistics, 29(5), 731-748. doi: 10.1080/10543406.2019.1657133*

# POSTER SESSION 10

## P10-05 Using threshold analysis to assess the robustness of public health intervention recommendations from network meta-analyses

**Molly Wells, Sylwia Bujkiewicz, Stephanie J. Hubbard**

Biostatistics Research Group, University of Leicester, United Kingdom

In the appraisal of new clinical interventions, network meta-analysis (NMA) is commonly used to investigate the effectiveness of multiple interventions in a single analysis. The results from a NMA can be fed into a decision analytic model to assess the cost-effectiveness of the interventions [1]. However, in public health intervention appraisals, NMAs are not widely used. Public health interventions are often complex in nature, with multiple outcomes and multiple interventions containing multiple components. Most evidence in public health appraisals is summarised with a narrative review or pairwise meta-analysis [2].

The hesitancy to use NMA methods in public health intervention appraisals is usually stated to be due to high levels of heterogeneity between studies [2]. Heterogeneity can arise due to several reasons such as differing study populations, outcomes, and study designs. In public health, the studies are more likely to be prone to bias and often are of poor quality due to the nature of the interventions and the outcomes [2]. Threshold analysis is used to assess the robustness of intervention recommendations to bias adjustments in the effect estimates from NMA [1]. Threshold analysis quantifies how large a change in the effect estimates, from individual studies or intervention contrasts, would be needed to alter the intervention recommendations from the NMA and resulting conclusions. The threshold analysis [1] was applied to published network meta-analyses in various areas of public health. We were able to assess and quantify the robustness of intervention recommendations from several networks by assessing the credibility of results from individual studies in the networks as well as the intervention contrasts. We illustrate that threshold analysis allows researchers and policy makers to assess and quantify the credibility of their results to evidence that could be biased [1]. This highlights that the use of such methods should ease any hesitancy to use NMAs in public health intervention appraisals and increase the use of such methods.

**References:** *[1] Phillippo D, Dias S, Ades A, Didelez V, Welton N. Sensitivity of treatment recommendations to bias in a network meta-analysis. Journal of the Royal Statistical Society 2018;181(3):843-67. [2] Smith EA, Cooper NJ, Sutton AJ, Abrams KR, Hubbard SJ. A review of the quantitative effectiveness evidence synthesis methods used in public health intervention guidelines. BMC Public Health. 2021;21(1):278.*



POSTER SESSION 10

## P10-06 Meta-analysis methods used in systematic reviews of interrupted time series studies

**Elizabeth Korevaar[1], Amalia Karahalios[2], Andrew B. Forbes[3], Simon L. Turner[3], Steve McDonald[3], Monica Taljaard[4,5], Jeremy M. Grimshaw[4,5,6], Allen C. Cheng[3,7], Lisa Bero[8], Joanne E. McKenzie[3]**

1 School of Public Health and Preventive Medicine, Monash University, Australia
2 Centre for Epidemiology and Biostatistics, The University of Melbourne, Australia
3 School of Public Health and Preventive Medicine, Monash University, Australia
4 Clinical Epidemiology Program, Ottawa Hospital Research Institute, Canada
5 School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Canada
6 Department of Medicine, University of Ottawa, Canada
7 Infection Prevention and Healthcare Epidemiology Unit, Alfred Health, Australia
8 Faculty of Medicine and Health and Charles Perkins Centre, University of Sydney, Australia

**Background:** Systematic reviews are used to inform decision making, evaluating the effects of organisational, policy change, public health interventions or exposures. Such reviews may include evidence from interrupted time series (ITS) studies. A core component of many systematic reviews is meta-analysis, which is the statistical synthesis of results across studies. To date there have been no reviews examining the statistical approaches used to meta-analyse effect estimates from ITS designs.

**Objectives:** We aimed to examine: 1) whether reviewers reanalyse primary ITS studies included in reviews, and what reanalysis methods are used; 2) the meta-analysis methods used; 3) the effect measures reported; and 4) the tools used to assess the risks of bias or methodological quality of ITS studies.

**Methods:** We searched eight electronic databases across several disciplines between 2000 and 2019 to identify reviews that meta-analyse at least two ITS studies. We describe, at the review level, the type of interruption and methods used to assess ITS study risk of bias; and at the meta-analytic level, the effect measures, meta-analytic methods, and any methods used to reanalyse the primary ITS studies.

**Results:** Of 4213 identified records, 54 reviews were included. The common interruption evaluated was a public health policy intervention (32/54). The majority of reviews (34/54) meta-analysed results from separate studies (with 23/34 reanalysing the included ITS studies), and the remainder meta-analysed results from sites (e.g. states) within a study (20/54). Among reviews that analyse the ITS studies, the most common models were Poisson (8/41) and ARIMA (7/41), with 21 including both level and slope parameters. Most reviews used a two-stage meta-analysis (51/54), fitting a random effects model (35/51) with the DerSimonian and Laird variance estimator and Wald type confidence intervals (18/35). Several effect measures were meta-analysed, with the most common being an immediate level change (45/54). Of the reviews of studies, nearly all assessed the risk of bias of included ITS studies (33/34).

**Conclusion:** This review informs work that aims to examine the performance of different meta-analysis approaches to combining results from ITS studies, and demonstrates the need for improved statistical analysis reporting at the ITS analysis and meta-analysis levels.

# POSTER SESSION 10

## P10-07 Effects of probiotics on mortality and morbidity in preterm infants: a Bayesian network meta-analysis of randomized and non-randomized studies

**Jhalok Ronjan Talukdar[1], Sameer Parpia[1], Holly Crandon[1], Xiaoqin Wang[2], Lehana Thabane[1,3], Behnam Sadeghirad[1]**

1 Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton ON, Canada
2 Michael G. DeGroote Institute for Pain Research and Care, McMaster University, Hamilton ON, Canada
3 Biostatistics Unit, St Joseph's Healthcare-Hamilton, Hamilton ON, Canada

**Background and statistical challenges:** Necrotizing enterocolitis and late-onset sepsis are major causes of mortality and morbidity in preterm infants. It is argued that only randomized controlled trials (RCTs) should be included in the meta-analysis. The experimental setting, stringent inclusion and exclusion criteria, small sample size, high cost, short follow-up time and ethical restrictions limit the abilities of RCTs to apply the findings in real-world settings. The integration of findings from non-randomized studies (NRSs) will complement the findings of RCTs and address some of the limitations of both study designs. Bayesian network meta-analysis is suitable to incorporate data from multiple sources. There is a lack of network meta-analysis that assesses the effects of probiotics on mortality and morbidity in preterm infants combing evidence from both RCTs and NRSs. The study will fill this gap. This study will also guide combining evidence from RCTs and NRSs in network meta-analysis in the future.
**Objective:** The study aims to assess the preventive effects of probiotics on all-cause mortality, severe necrotizing enterocolitis, and late-onset sepsis in infants of low birth weight.
**Methods and analysis:** We searched MEDLINE, Embase, CINAHL, CENTRAL and Scopus databases on August 4, 2020, without any date or language restriction. We assessed the risk of bias using the Cochrane risk of bias instrument for randomized trials and Newcastle-Ottawa Scale for non-randomized studies. We will use Bayesian network meta-analysis to compare the effects of single vs multi-strain probiotics combining data from randomized trials and non-randomized studies (e.g., cohort, case-control, cross-sectional). We will use the GRADE approach to assess the certainty in of evidence of each outcome.
**Expected results:** We included 104 articles (71 RCTs and 33 NRS) for our network meta-analysis. We hope to finish the data analysis for this project by March 2021. Our preliminary analysis indicates that probiotics have beneficial effects on mortality and morbidity in preterm infants.
**Conclusions:** Probiotics have positive effects on mortality and morbidity in preterm infants. The integration of real-world evidence from NRS with RCTs has the potential to increase the precision of the evidence and help in the decision-making process.

# STRATOS MINI-SYMPOSIUM

## MS-1 Understanding and accounting for measurement error when prediction equations are used in observational studies

**Paul Gustafson**
University of British Columbia, Canada for TG4

When an important variable in an observational study is hard to measure, an appealing strategy is to predict its values from other variables. In essence, this induces Berkson measurement error, the implications of which may not be widely understood. We discuss three scenarios. In the first, the marginal distribution of the variable being predicted is of inferential interest. In the second and third, this variable is respectively an exposure variable and an outcome variable, with the exposure-outcome association being of interest. We show that both the implications of the measurement error, and appropriate mitigation strategies, can vary across these scenarios. As part of this discussion, we necessarily grapple with an assumption of non-differential measurement error. We consider both the general plausibility of this assumption, and how it facilitates statistical adjustment for the measurement error. The ideas presented will be illustrated via an example from nutritional epidemiology, with data from the Hispanic Community Health Study / Study of Latinos (HCHS/SOL).

## MS-2 Guidance for performance assessment in prediction models for survival outcomes

**David J. McLernon[1], Terry Therneau[2], Daniele Giardiello[3,4], Ben Van Calster[4,5], Laure Wynants[6], Maarten van Smeden[7], Ewout W. Steyerberg[4], on behalf of STRATOS TG6 and TG8**

1 Institute of Applied Health Sciences, University of Aberdeen, United Kingdom
2 Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester MN, United States
3 Netherlands Cancer Institute, Amsterdam, Netherlands
4 Department of Biomedical Data Sciences, Leiden University Medical Center, Netherlands
5 Department of Development and Regeneration, KU Leuven, Netherlands
6 School for Public Health and Primary Care, Maastricht University, Netherlands
7 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Netherlands

Risk prediction models need careful validation to understand their performance. Validation of models for survival outcomes poses challenges due to the censoring of observations and the varying time horizon at which predictions can be made. We aim to give a hands-on description of methods to evaluate predictions and decisions from survival models based on Cox proportional hazards regression. We discuss statistical measures, that evaluate model performance in terms of discrimination, calibration, and overall performance, and decision analytic measures. As a motivating case study, we consider the prediction of event free survival in breast cancer patients following surgery.

# STRATOS MINI-SYMPOSIUM

## MS-3 Guiding the path from Patient Reported Outcomes to treatment registration based on randomised and single arm studies: STRATOS engaged in the European IMI-SISAQOL project. (https://qol.eortc.org/projectqol/sisaqol/)

**Saskia le Cessie[1,2], Els Goetghebeur[2]**
1 LUMC, Clinical Epidemiology & Biomedical Data Sciences, Leiden, Netherlands
2 Ghent University, Belgium

Whether cancer presents itself as a late stage terminal disease or in almost chronic form, its impact on quality of life and other patient reported outcomes is typically overwhelming. This justifies a far greater role for patient reported outcome measures (PROMs) besides survival in primary and secondary parameters targeted for inference in cancer clinical trials. In addition, recent years saw many new cancer treatments entering the market based on evidence drawn from non-randomized – often single arm - studies. The challenge of reaching evidence on treatment effects without concurrent randomized control is daunting. In IMI-SISAQOL statisticians and clinicians from academia and industry team up with regulators and patient advocates, to reveal current practice in published and unpublished supporting material. We critically examine data, methods and results uncovering opportunities and threats to reliable, relevant and actionable evidence. Our aim is to give guidance on (un)acceptable methodological approaches in specific contexts. We discuss estimands that are or should be targeted with due attention to reference values, minimal acceptable difference and sensitivity analysis starting from available background when concurrent controls are missing while repeated measures inadvertently vary in timing.

## MS-4 Statistical analysis of high-dimensional biomedical data: Analytical goals, common approaches and challenges

**Jörg Rahnenführer**
Technische Universität Dortmund, Germany for TG9

In high-dimensional data (HDD) settings, the number of variables associated with each observed individual or experimental unit is very large. In biomedical research, prominent examples are omics data and electronic health records data. Advances in statistical methodology and machine learning methods offer new opportunities for innovative analyses of HDD, but at the same time require a deeper understanding of some fundamental statistical concepts. We discuss key aspects of HDD analysis to provide a gentle introduction both for non-statisticians and for classically trained statisticians with little experience specific to HDD analysis. Main analytical goals are outlined, and situations are identified where traditional statistical methods cannot, or should not, be used in the HDD setting, or where adequate analytic tools are still lacking.

## MS-5 Current education and practical guidance in statistical (non-linear) modeling for researchers with limited statistical background (level-1)

**Christine Wallisch[1], Geraldine Rauch[2], on behalf of TG2**
1 Vienna Medical University, Austria
2 Charité – Universitätsmedizin Berlin, Germany

In a recent systematic review about issues related to selection of variables and functional forms, we investigated the transferred knowledge of statistical modeling to medical researchers through series of statistical notes and tutorials in medical journals. We found that some areas of statistical modeling were underrepresented and code reproducing the presented results was mostly missing. In particular, Poisson regression, variable selection methods and methods to deal with non-linear relations in multivariable models were rarely found. The latter topic was addressed by us in an educational shiny app, with which one interactively learns how fractional polynomials, b-splines and natural splines handle non-linear relations between an outcome and an independent variable.



STRATOS MINI-SYMPOSIUM

## MS-6 In defense of correct use of statistical significance

**Michal Abrahamowicz[1], Marie-Eve Beauchamp[1], James Carpenter[2,3], Victor Kipnis[4]**
1 Department of Epidemiology & Biostatistics, McGill University, Montreal, Canada
2 Department of Medical Statistics, London School of Hygiene & Tropical Medicine, United Kingdom
3 MRC clinical trials unit at UCL, Holborn, London, United Kingdom
4 Biometry Research group, National Cancer Institute, United States

Recently, Amrhein et al, in a highly cited Comment in Nature (2019), recommended banning statistical significance, based on a pre-specified dichotomization of p-values, as a criterion for assessing the strength of the evidence provided by empirical studies. About 850 researchers, representing a wide range of empirical sciences, including several statisticians, endorsed this position, by signing the Amrhein et al's comment, which got > 100 citations in the 6 first months after its publication. However, other experienced statisticians raised serious concerns about the potential impact of such a black-and-white recommendation.

To stimulate further discussions about this controversial yet very impactful issue, we present some empirical evidence in 'defense' of an accurate use of statistical significance. We start by taking a 2nd look at the very flagship empirical example used by Amrhein et al to illustrate the absurd consequences of incorrect interpretation of the results of significance testing, applied separately to two independent studies and then informally compared. We demonstrate that a proper use of a simple statistical test of the significance of the difference between the results of the two studies eliminates the risk of incorrect inference and illogical conclusions. On the other hand, to illustrate the potentially risky implications of the research paradigm advocated by Amrhein et al, we briefly review some recent empirical studies that were cited to back up their decision to report 'effects' regardless of the lack of their statistical significance. Finally, we will clarify an important conceptual and formal difference between (i) p-values as used by Fisher in significance testing versus (ii) the role of an a priori chosen significance level (⍰) in hypothesis testing as suggested by Neyman and Pearson. Based on this empirical evidence and relevant statistical arguments, we conclude that the statistical research community should focus its efforts on better educating our collaborators to improve the understanding of the pros and cons of statistical significance testing and its accurate use in applications rather than on "political" interventions aimed at banning it from the applied research. We believe that the STRATOS initiative can play an important role on this front.

# EARLY CAREER BIOSTATISTICIANS' DAY

## ECB-1 *Invited* A replication crisis in methodological statistical research?

**Anne-Laure Boulesteix**
Institute for Medical Information Processing, Biometry and Epidemiology, Faculty of Medicine, Ludwig Maximilian University Munich, Germany

Statisticians are often keen to analyze the statistical aspects of the so-called "replication crisis". They condemn fishing expeditions and publication bias across empirical scientific fields applying statistical methods. But what about good practice issues in their own - methodological - research, i.e. research considering statistical methods as research objects? When developing and evaluating new statistical methods and data analysis tools, do statisticians adhere to the good practice principles they promote in fields which apply statistics? I argue that statisticians should make substantial efforts to address what may be called the replication crisis in the context of methodological research in statistics and data science, in particular by trying to avoid bias in their comparison studies based on simulated or real data. I will discuss topics such as publication bias, the design and necessity of neutral comparison studies and the importance of appropriate reporting and research synthesis in the context of methodological (bio)statistical research by drawing an analogy with clinical research.

## ECB-2 *(OC)* From Consulting to Research

**Max Behrens**
University Freiburg, Germany

Last year in October, I started my work at the University of Freiburg as a young researcher in the field of Biometrics. Before that I worked at KPMG, a big consulting firm, for almost two years. Thereby, I took a career path which is more frequently seen the other way around or never.
In my presentation, I would like to share the differences and similarities between the two different occupations. Especially, how knowledge and information is generated in consulting and how this differs from a career in research. I would like to exemplify this difference by talking about discussions with different people about a statistical analysis I was responsible for during my time as a consultant. As a comparison, I will talk about frequent discussions I had during my first research projects. Some of the major differences I found so far are about statistical accuracy, choice of the correct statistical method and reporting of results.
Freedom is certainly one of the similarities I experienced so far. Both career paths offer a wide extend of freedom on how you want to approach the tasks at hand and how you want to organise yourself. Hence, work routines are not focussed on daily reports but encourage defining your own milestones.
Lastly, I would like to share the reasons why I decided to start a career in research. The things I missed when I worked as a consultant and thus the things I hoped to find as a researcher. Among others, these things concern the exchange with peers about what is possible and not what is profitable.

## ECB-3 *(OC)* It's all about collaboration: PhD as an academia-industry partnership in an interdisciplinary environment

**Jinran Zhan**
University of Warwick, Coventry, United Kingdom

Industry-funded and interdisciplinary PhDs are both becoming more and more popular. The industrial aspect provides a valuable opportunity to gain insight into the pharmaceutical industry and develop a wide range of transferable skills. Interdisciplinarity encourages you to think critically from different perspectives and train your ability to communicate with diverse audiences.
Stepping out of the comfort zone, gaining practical experiences, collaborating with people from different disciplines and industry are all important aspects of this type of PhD. In this talk, I will share my personal experiences from being a mathematics and statistics undergraduate student to embarking on a programme in interdisciplinary biomedical research and finally to where I am today, doing a PhD in collaboration with a pharmaceutical company. I will highlight what I have learnt, the challenges, opportunities and benefits of working in a multidisciplinary environment as well as collaborating with an industry partner.



EARLY CAREER BIOSTATISTICIANS' DAY

## ECB-4 *(OC)* Two-piecewise model to assess the impact of a treatment on a quantitative repeated endpoint

**Constant Josse[1], Silvia Oghina[2], Mélanie Bezard[2], Thibaud Damy[2], François Montestruc[1]**
1 eXYSTAT, Malakoff, France
2 Henri Mondor Teaching Hospital, Créteil, France

**Introduction:** In clinical studies where the objective is to compare the evolution before and after treatment, the choice of the appropriate model for repeated values is critical for the study conclusions. This choice is emphasized when the number of repeated values per subject or the delays between these values are heterogeneous.
**Aim:** To choose an adequate two-piecewise regression model to assess the N-terminal pro-brain natriuretic peptide (NT-proBNP) slopes change before and after treatment.
**Methods:** In this monocentric, longitudinal, observational registry on the natural history of transthyretin amyloid cardiomyopathy, patients were considered if having at least 2 NT-proBNP values available (before and under treatment). Slopes before and after initiating treatment were evaluated and compared using a two-piecewise regression model for longitudinal data using a fixed change point that corresponded to the initiation of the treatment. To account for repetitions over time and heterogeneity of the different timepoints across patients, model using time as random with an unstructured (UN) variance-covariance matrix was used. Another model to analyze longitudinal data without random effect using repeated measures was performed. Choice of the most appropriate covariance matrix structure and change point value were explored.
**Results:** Data from 248 patients treated were included. Mean age was 76 years with 82% males. Median follow up was 17.5 months. In the two-piecewise regression random model with a fixed change point using UN matrix, the slope before was 0.0099 (95% CI: [0.0042; 0.0156]) compared to 0.0083 [0.0036; 0.0131] after treatment (p=0.70 for the slopes comparison). In the fixed model using repeated measures, the slope before was 0.0115 (95% CI: [0.0070; 0.0160]) compared to -0.0034 [-0.0093; 0.0024] after treatment (p=<0.01 for the slopes comparison).
**Conclusion:** When treatment follow-up, number of values and delays between these values are heterogeneous, the two-piecewise regression model using time as random effect is strongly recommended. This approach avoids false conclusions for a study where the objective is to compare the evolution before and after treatment.

# EARLY CAREER BIOSTATISTICIANS' DAY

## ECB-5 *Invited* Studying and working in developing nations

**Mavuto Mukaka**

University of Oxford, Malawi University of Science and Technology, Mahidol-Oxford Tropical Medicine Research Unit (MORU), Thailand

The basic education curriculum is generally similar across continents. This is also true for developed and developing countries. The main reason is that education systems were developed in developed countries and then spread across the globe. Despite this similarity, the amount of effort that students put into their education to achieve the same level of knowledge and capabilities varies widely between developed and developing countries.

In developing countries, the student to computer ratio is very low for science students. This makes life tough for statistics and computing students for whom having a computer is a necessity. Even where computers may be available due to donations from the developed world, the institutions struggle to get commercial software such as Stata or SAS. Of course, the alternatives are open platform software such as R, but then there are internet challenges for accessing updated packages.

One of the main challenges of working in developing countries is that it is not easy to access relevant short courses for one's career and relevant books. Students often fail to obtain the recommended books and subsequently rely mostly on lecture notes. This results in students being well developed in theoretical work but having limited application skills. Even students who may have their own computer and good internet face obstacles; it is difficult to find real data sets that students can use for practical work or dissertations. Thus, most graduates struggle with practical applications during their first few years of their first job.

In addition, though not specific to developing countries, communication is very disappointing. In many cases, collaborators take a long time to respond to emails. Student supervisors often take too long to give feedback to a student dissertation and surprisingly some students take too long to update their dissertations- in some cases up to the point of a supervisor chasing a student.

Working and studying in developing countries also has benefits. For example, people who work or study in developing countries often operate in real time disease scenarios; research questions are of direct public health interest. The impact of your research may be seen immediately as improvements in day-to-day health, updates in policy and guidelines, or enhancements to clinical practice, to name a few.

In summary, studying or working in developing countries is attractive in that disease scenarios and problems are real, making it more straightforward to find a research question of interest. However, one encounters many obstacles compared to someone working in developed countries. Those working in developing countries often take longer times to achieve the same career goals as those working in the developed world. The advantaged groups working in developing countries are those whose institutions are from developed countries performing a mission in developing countries. They have good access to the necessary study or work resources from the partner institutions in the developed world.

## ECB-6 *(OC)* A biostatistician's journey navigating ethical issues

**Charlotte Bolch**

Midwestern University, Glendale AZ, United States

The foundation of good statistical practice is based on transparent assumptions, reproducible results, and valid interpretations. Students and early career biostatisticians need to understand critical concepts of responsible data analysis, interpretation, and reporting. However, they may not have received formal training in how to navigate ethical issues in their professional lives. In this talk, I will share my experiences as a biostatistician working in the areas of industry and academic statistical consulting about strategies to prepare for potential ethical issues. The strategies will address reproducibility of results, data integrity methods, privacy and confidentiality of research subjects, and honest and meaningful communication of results. In addition, tools that biostatisticians can use to support ethical decision making will be discussed based on the American Statistical Association's Ethical Guidelines for Statistics Practice.



EARLY CAREER BIOSTATISTICIANS' DAY

## ECB-7 *(OC)* Developing a prediction model: Challenges and Lessons Learned

**Zheng Jing Hu[1], Salhab El-Helou[2], Sarah Khan[2], Kara Tsang[3], Sameer Parpia[1]**

1 Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton ON, Canada

2 Department of Pediatrics, McMaster Children's Hospital, Hamilton ON, Canada

3 Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton ON, Canada

**Introduction:** Late-onset neonatal sepsis (LONS) is a serious driver of morbidity and mortality among infants hospitalized in the neonatal intensive care unit (NICU). Many factors may predispose infants to a greater risk of LONS, and to date, numerous prediction models have been developed to predict this risk. This prediction model utilizes up-to-date statistical techniques and practice recommendations for model development and validation. We described the challenges, decision process and statistical considerations used to execute our modelling approach.

**Objective:** To develop a clinical prediction model to predict the risk of acquiring LONS in a NICU using demographic and clinical risk factors.

**Methods:** The study was conducted at the McMaster Children's Hospital NICU. For variable selection, we determined potential variables for inclusion a-priori based on the medical literature and consultation with experts. We examined variable transformations, interactions, and linearity. We used logistic regression with penalized maximum likelihood shrinkage to estimate our regression parameters. Internal validation was done using bootstrapping.

**Results:** We identified 8 variables based on the initial literature search; gestational age, central venous line (CVL), mechanical ventilation, total parenteral nutrition, acid inhibition, antibiotics at birth, Apgar score at 5-minutes, and the duration of rupture of membranes (ROM). After model exploration using graphs and descriptive statistics, two variables, acid inhibition and the duration of ROM were excluded from further consideration. Interaction terms between gestational age, and CVL and mechanical ventilation were included in the model. Apgar score was modelled using restricted cubic splines. Our model demonstrated a concordance statistic of 0.86, calibration slope of 0.93. The calibration curve showed minor overestimation when the observed risk of LONS was below 0.40, and consistent underestimation when the actual probability was 0.50 or higher.

**Discussion:** Model performance was limited by the absence of the duration of risk factors as predictor variables.

# Authors Index

# Authors Index



## D.



## E.



## F.



## G.

# Authors Index



## H.



## I.



## J.



## K.

# Authors Index



## L.



## M.

# Authors Index



## N.



## O.



## P.



## Q.



## R.

# Authors Index

# Authors Index

ISCB 42
LYON 2021