The 8th International Symposium on Biopharmaceutical Statistics (ISBS2026)

Evidence Insights – Statistical Innovation, Regulatory Science, and the Transformative Impact of AI in Biopharma

Organized by

The International Society for Biopharmaceutical Statistics (ISBS)

October 13 – 16, 2026

Novotel Madrid Center

Madrid, Spain

Table of Contents

Organizing Committees ..... - 3 -
Conference Sponsors ..... - 4 -
General Information ..... - 5 -
Keynote speeches and the 2026 Lai and Heyse Memorial Lecture Award Speeches ..... - 5 -
Special issue for ISBS2026 ..... - 5 -
Overview of Symposium Program ..... - 5 -
Short Courses ..... - 7 -
Scientific Program – Day 1 ..... - 9 -
Scientific Program – Day 2 ..... - 16 -
Scientific Program – Day 3 ..... - 23 -
Short Course Description ..... - 28 -
Abstracts for Keynote Speeches ..... - 41 -
Abstracts for presentations in parallel sessions ..... - 46 -

Organizing Committees

Executive Committee

María del Carmen Pardo, Complutense University of Madrid

Frank Fan, Bristol Myers Squibb

Toshi Hamasaki, The George Washington University

Ying Lu, Stanford University

Cindy Lu, AstraZeneca

Martin Posch, Medical University of Vienna

Wenjin Wang, Pfizer Inc

Jie Chen, Taimei Intelligence Biopharma

Scientific Program Committee (alphabetic order of last name)

Carl-Fredrik Burman, AstraZeneca

Claudia Coscia Requena, AstraZeneca

Guoqing Diao, Washington University

Xiangyi Duan, AstraZeneca

Alba Franco-Pereira, Universidad Complutense Madrid

Pepa Garcia de Polavieja, Novo Nordisk

Toshi Hamasaki, The George Washington University

Cindy Lu, AstraZeneca

Ariel Alonso Abad, L-BioStat, KU Leuven

Keaven Anderson, MSD

Zoran Antonijevic, Bioforum Group

Robert Beckman, Georgetown University

Marta Bofill Roig, University in Barcelona

David Calle, Fortrea

Hongyuan Cao, University of Wisconsin-Madison

Jie Chen, Taimei Intelligence

Haitao Chu, Pfizer

Olivier Collignon, GSK

María del Carmen Pardo, Complutense University of Madrid

Frank Fan, Bristol Myers Squibb

Bo Fu, Astellas

Danyu Lin, University of North Carolina at Chapel Hill

Ying Lu, Stanford University

Olga Marchenko, Bayer

Cyrus Mehta, Cytel

Elias Laurin Meyer, Berry Consulting

Satoshi Morita, University of Kyoto

Haitao Pan, St. Jude hospital

Fabio Pellegrini, Daiichi-Sankyo España

Teresa Pérez, Universidad Complutense Madrid

Martin Posch, Medical University of Vienna

Antonio Remiro-Azócar, Novo Nordisk

Kentaro Takeda, Astellas

Ming Tan, Georgetown University

Wim Vander Elst, Janssen Pharmaceutica

Wenquan Wang, Pfizer

William Wang, MSD

Chenguang Wang, Regeneron

Cuiyi Wang, Johnson & Johnson

Wenjin Wang, Pfizer

Kyle Wathen, Cytel

Samuel Wu, University of South Florida

Dong Xi, Gilead

Jin Xu, East China Normal University

Margaret Gamalo, Pfizer
Borja García, AEMPS
Alberto García Hernández, Argenx
Philip He, Daiichi-Sankyo España
Guanyu Hu, Michigan State University
Yi Huang, UMBC
Yuan Ji, University of Chicago
Franz König, Medical University of Vienna
Hana Lee, FDA
Li-an Xu, Pfizer
Shu Yang, NCSU
Jingjing Ye, DataDoctor
Lilly Yue, FDA
Xiang Zhang, CSL Behring
Heng Zhou, MSD
Jeni Zhou, BeOne
Yong Zhou, Indiana University
Sarah Zohar, Inserm

Local Organizing Committee

María del Carmen Pardo, Chair, Universidad Complutense Madrid
Rosa Alonso, Universidad Complutense Madrid
Mar Fenoy, Universidad Complutense Madrid
Alba Franco-Pereira, Universidad Complutense Madrid
David Lora, Universidad Complutense Madrid
Teresa Perez, Universidad Complutense Madrid

Conference Sponsors

The primary sponsor, **Complutense University of Madrid**, has provided tremendous support—including, but not limited to, securing and setting up the conference venue, planning the budget, and offering administrative and management assistance—to help make this conference possible.

Endorsement from other professional organizations:

- Biopharmaceutical Section of the American Statistical Association
- Center for Innovative Study Design, Stanford University
- International Chinese Statistical Association
- Pharmaceuticals and Medical Devices Agency (PMDA)

Additional sponsors:

- AMS (Advanced Medical Services) (Platinum partner)
- Bristol Myers Squibb (Gold partner)
- Cytel (Silver partner)
- MSD (Silver partner)

We would like to express our deepest gratitude to the above sponsors for their invaluable support of ISBS2026. The success of the Symposium wouldn't be possible without their generosity and commitment.

General Information

Conference Venue
Novotel Madrid Center
O Donnell 53
28009 MADRID, Spain

The ISBS2026 program includes:

- Short courses (Oct. 13, 2026)
- Three-day conference (Oct. 14 – 16, 2026)
  - Four keynote speeches
  - Forty parallel sessions

Keynote speeches and the 2026 Lai and Heyse Memorial Lecture Award Speeches

The Symposium will feature four keynote speeches, delivered by an esteemed group of speakers: two from the European Medicines Agency (EMA)—**Dr. Armin Koch** and **Dr. Peter Arlett**; one from the biopharmaceutical industry—**Dr. Weili He** of AbbVie, recipient of the **2026 Heyse Memorial Lecture Award**; and one from academia—**Professor Mark van der Laan** of University of California at Berkeley, recipient of the **2026 Lai Memorial Lecture Award**.

Special issue for ISBS2026

The Journal of Biopharmaceutical Statistics will publish a **special issue** highlighting topics presented at the **8th International Symposium on Biopharmaceutical Statistics (ISBS2026)**.

The special issue will follow the journal's **standard peer-review guidelines**, overseen by the **guest editors**. Submissions must be made via the **journal's online submission system**. Detailed submission instructions and timelines will be shared with **presenters** shortly after the conclusion of the Symposium.

Overview of Symposium Program

Short Courses

	Room				
Date/Time	Bahia	La Paz	Caracas	Londres	Paris
Tue. Oct. 13, 2026					
9:00 – 10:30	SC1	SC2	SC3	SC4	SC5 (C*)
10:30 – 11:00	*Coffee Break*				
11:00 – 12:30	SC1	SC2	SC3	SC4	SC5 (C*)
12:30 – 13:00	*Pre-Lunch Transition Break*				
13:00 – 14:30	*Lunch Break (Lunch on your own)*				
14:30 – 16:00	SC1	SC2	SC6	SC7	SC8 (C*)
16:00 – 16:30	*Coffee Break*				
16:30 – 18:00	SC1	SC2	SC6	SC7	SC8 (C*)

* C—Cancelled.

Scientific Program—Keynote and parallel presentations

Date/Time	Room				
	Convención A	Convención B	Bahia	La Paz	Caracas
Wed. Oct. 14, 2026					
8:30 – 10:00	Plenary session WK – Keynote Speech (Convención)				
10:00 – 11:00	*Coffee Break*				
11:00 – 12:30	WP1	WP2	WP3	WP4	WP5
12:30 – 14:00	*Lunch Break (Lunch provided)*				
14:00 – 15:30	WP6	WP7	WP8	WP9	WP10
15:30 – 16:30	Coffee Break				
16:30 – 18:00	WP11	WP12	WP13	WP14	WP15
Thur. Oct. 15, 2026					
8:30 – 10:00	Plenary session TK – Keynote Speech (Convención)				
10:00 – 11:00	*Coffee Break*				
11:00 – 12:30	TP1	TP2	TP3	TP4	TP5
12:30 – 14:00	*Lunch Break (Lunch provided)*				
14:00 – 15:30	TP6	TP7	TP8	TP9	TP10
15:30 – 16:30	*Coffee Break*				
16:30 – 18:00	TP11	TP12	TP13	TP14	TP15
Fri. Oct. 16, 2026					
8:30 – 10:00	Plenary session FK1 – Keynote Speech (Convención)				
10:00 – 11:00	*Coffee Break*				
11:00 – 12:30	FP1	FP2	FP3	FP4	FP5
12:30 – 14:00	*Lunch Break (Lunch provided)*				
14:00 – 15:30	FP6	FP7	FP8	FP9	FP10
15:30 – 16:30	*Coffee Break*				
16:30 – 18:00	Plenary session FK2 – Keynote Speech (Convención)				

WK (TK, FK) – Wednesday (Thursday, Friday) Keynote; WP (TP, FP) – Wednesday (Thursday, Friday) parallel session.

Short Courses

Tuesday October 13, 2026

9:00 – 18:00: *Registration* — Hotel Entrance Floor 0 Hospitality

SC1: **Improving Treatment Evaluation by Integrating Randomized Clinical Trials and Real-World Data: Causal Inference and Machine Learning Approaches (Full-Day)**

Room: Bahia

Time: 9:00 – 18:00

Instructors: **Shu Yang**, Professor of Statistics, North Carolina State University

SC2: **Applied Longitudinal Data Analysis with AI for Pharmaceutical Science (Full-Day)**

Room: La Paz

Time: 9:00 – 18:00

Instructors: **Depeng Jiang**, Professor of Biostatistics, University of Manitoba

SC3: **Implementing Bayesian Methodology in Clinical Trials: Navigating New FDA and EMA Guidance (Half-Day)**

Room: Caracas

Time: 9:00 – 12:30

Instructors: **Yong Zang**, Associate Professor of Biostatistics, Indiana University

Ying Yuan, Bettyann Asche Murray Distinguished Professor, Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas

SC4: **Graphical multiple comparison procedures: Combining flexibility with optimality (Half-Day)**

Room: Londres

Time: 9:00 – 12:30

Instructors: **Frank Bretz**, Distinguished Quantitative Research Scientist, Novartis

Yao Chen, Statistical Consultant, Novartis

Dong Xi, Gilead Sciences

SC5: **Data Visualization with Applications to the Life Sciences (Half-Day)**

Room: Paris

Time: 9:00 – 12:30 **(Cancelled)**

Instructors: **Richard Zink**, Principal Research Fellow, JMP Statistical Discovery

SC6: **Causal AI with Targeted Learning (Half-Day)**

Room: Caracas

Time: 14:30 – 18:00

Instructors: **Susan Gruber**, Co-Founder, Targeted ML Solutions

Alan Hubbard, Professor of Biostatistics and co-director of the Center for Targeted Machine Learning, University of California at Berkeley

SC7: **Baseline Covariate Adjustment in Randomized Clinical Trials (Half-Day)**

Room: Londres

Time: 14:30 – 18:00

Instructors: **Kelly Van Lancker**, Assistant Professor of biostatistics, Ghent University and Vrije Universiteit

Bohdana Ratitch, Bayer

SC8: **Hands-on Introduction to Trial Design with East Horizon(TM) – Methodologies and Practical Applications (Half-Day)**

Room: Paris

Time: 14:30 – 18:00 **(Cancelled)**

Instructors: Martin Kappler, Director of Customer Success, Cytel

Pantelis Vlachos, VP of Customer Success at Cytel

Scientific Program – Day 1

Wednesday October 14, 2026

8:30 – 18:00: *Registration* — Hotel Entrance Floor 0 Hospitality

8:30 – 10:00: *Welcome Remarks & Keynote Session* (WK) — Convención

Welcome Remarks: María del Carmen Pardo, Complutense University of Madrid

Chair: Ying Lu, Stanford University

Keynote Speech: **A Sad Farewell to the Study-Wise Type I Error? Considerations about Innovation and Conservativism in Regulatory Statistics**
Armin Koch, Hannover Medical School

10:00– 11:00: *Coffee Break*

11:00 – 12:30: *Parallel Sessions*

WP1: **Chasing the Estimand: Statistical and Regulatory Challenges in Platform Trials**

Room: Convención A

Organizer(s): Martin Posch, Medical University of Vienna

Chair: Martin Posch, Medical University of Vienna

Title & Speaker: **Non-concurrent controls in platform trials: Estimands, identification, and efficiency across endpoint types**
Michele Santacatterina, New York University

Estimands and estimation in platform trials with time trends
Marta Bofill Roig, Universitat Politecnica de Catalunya -BarcelonaTech

Regulatory update on platform trials
Benjamin Hofner, German regulatory authority PEI

Discussant
Ying Yuan, MD Anderson Cancer Center

WP2: **Innovative Statistical Frontiers in Clinical Research: Integrating AI-Driven Evidence Synthesis and Advanced Survival Methodologies**

Room: Convención B

Organizer(s): Danyu Lin, University of North Carolina, Chapel Hill

Ying Lu, Stanford University

Chair: Ying Lu, Stanford University

Title & Speaker: **Learning from literature: Integrating LLMs and Bayesian hierarchical modeling for oncology trial design**
Wei Wei, Yale University

Beyond RMST: shape-weighted estimands based on survival curves
Hajime Uno, Harvard University

MetaFemina: A large language-model–driven meta-analysis platform for studying associations between dietary supplements and breast, ovarian and uterine cancer risk
Yushu Shi, Cornell University

Semiparametric regression analysis of interval-censored data
Danyu Lin, University of North Carolina at Chapel Hill

WP3: **The Evolution of Multi-Component Strategies and Their Practical Value in Clinical Trials**

Room: Bahia

Organizer(s): Patrick Schloemer, Bayer AG

Chair: Silke Janitza, Bayer AG

Title & Speaker: **Not all Hierarchical Multi-Component Strategies are Created Equal: GPC, DOOR, and MOST**
Mickaël De Backer, UCB

Development and validation of a hierarchical composite endpoint for chronic kidney disease
Niels Jongs, University Medical Center Groningen

Increasing statistical efficiency in clinical trials using ordinal longitudinal data
Maximilian Rohde, Bristol Myers Squibb

Principal Component Analysis to identify multivariate patterns of patient-reported symptom change during adjuvant chemotherapy in early breast cancer: data from GEICAM/2003 02 study
Andrea Blasco, GEICAM Spanish Breast Cancer Group

WP4: **Multiple Testing, Subgroup Analysis & Exact Inference**

Room: La Paz

Chair: Jie Chen, Taimei Intelligence

Title & Speaker: **Assessment of global evidence against homogeneity for exhaustive subgroup treatment effect plots**
Frank Bretz, Novartis

A decision-theoretic interpretation of frequentist error rates in rare disease trials
Nico Bruder, Institute of Medical Statistics, Medical University of Vienna

Error-controlled effect screening in saturated factorial designs using APC analysis
Abu Zar Shafiullah, Thuenen Institute of Agricultural Technology

Simultaneous testing of hypotheses and alternatives: theory and applications to problems of network analysis
Petr Koldanov, NRU Higher School of Economics

WP5: **Transdisciplinary Decision Making for Biomarkers and Drugs at Multiple Levels**

Room: Caracas

Organizer(s): Robert Beckman, Georgetown University Medical Center

Chair: Cristiana Mayer, Johnson and Johnson Vision

Title & Speaker: **Holistic decision making at multiple levels**
Zoran Antonijevic, Bioforum Group

Applications to proof-of-concept trials and biomarker-drug co-development at the portfolio level
Robert Beckman, Georgetown University Medical Center

Exploring the effects of context on biomarker-drug co-development at the portfolio level using an R-shiny app
Patrick Roney, Georgetown University Medical Center

12:30 – 14:00: *Lunch Break (Lunch provided)*

14:00 – 15:30: *Parallel Sessions*

WP6: **Aligning Clinical Development with Regulatory Approval: Designing and Analyzing Studies that Deliver Decision-Relevant Evidence**

Room: Convención A

Organizer(s): Kathrin Bogner, AMS Advanced Medical Services GmbH

Chair: Alexandra Carls, AMS Advanced Medical Services GmbH

Title & Speaker: **How early design and analysis choices of pivotal trials shape HTA usability**
Stefanie Wüstner, AMS Advanced Medical Services GmbH

Non-linear power for Cox models: Design-by-shape sample size planning for spline effects
Stefania Lando, University of Padua

One-Class Support Vector Machines Integrated Bayesian Approaches for External Control Borrowing in Clinical Trials
Yi Huang, University of Maryland Baltimore County

Beyond go/no-go: An adaptive recruitment strategy for basket trials
Libby Daniells, MRC Biostatistics Unit, University of Cambridge

WP7: **Bayesian Borrowing of External Data**

Room: Convención B

Organizer(s): Carl-Fredrik Burman, AstraZeneca

Chair: Darren Scott, AstraZeneca R&D Cambridge

Title & Speaker: **Leveraging external-trial information to optimise patient allocation in group sequential designs**
Haiyan Zheng, University of Bath

Type 1 error and other metrics for RCTs borrowing external data
Annette Kopp-Schneider, German Cancer Research Center

From individuals to effects: Control-group borrowing via individual exchangeability and disjunctive subgroup treatment-effect sharing
Darren Scott, AstraZeneca R&D Cambridge

Bayesian meta-analytic methods for leveraging external data in surrogate endpoint evaluation
Sylwia Bujkiewicz, University of Leicester

WP8: **Regulatory Communication for Complex Trial Designs: Aligning Expectations Across Stakeholders**

Room: Bahia

Organizer(s): Florian Klinglmüller, Austrian Agency for Health and Food Safety

Chair: Yuki Ando, Pharmaceuticals and Medical Devices Agency

Title & Speaker: **Industry perspective**
Frank Bretz, Novartis

Panel discussion—Panelists

- Miya Saigusa, Pharmaceuticals and Medical Devices Agency
- Florian Klinglmüller, Austrian Agency for Health and Food Safety
- Frank Bretz, Novartis
- Hiromi Sugano, Pharmaceuticals and Medical Devices Agency

WP9: **Biostatnet: Connecting Biostatistics Research I**

Room: La Paz

Organizer(s): Teresa Pérez Pérez, Universidad Complutense de Madrid

Chair: Teresa Pérez Pérez, Universidad Complutense de Madrid

Title & Speaker: **Assessing surrogacy for a binary true endpoint and a continuous surrogate in longitudinal data using joint modeling**
Gökçe Deliorman, Universidad Complutense de Madrid

Evaluating causal inference methods for dynamic deprescribing strategies: a simulation study
Teresa Pérez, Universidad Complutense de Madrid

Causal inference in non-randomized research: The role of G-computation
Cristóbal Manuel Rodríguez Leal, Universidad Complutense de Madrid

Evaluating causal inference methods for continuous outcomes: power and performance analysis in small samples
Adrian Aurensanz Crespo, Universidad de Zaragoza

WP10: **Recent Advances in Dose Optimization and Related Issues for Clinical Trials**

Room: Caracas

Organizer(s): Heng Zhou, MSD

Chair: Yeonhee Park, Sungkyunkwan University

Title & Speaker: **Integrating the patient voice into early phase dose-finding oncology trials: From OPTIMISE-ROR objective setting to OPTIMISE-AR analytical insights**
Christina Yap, The Institute of Cancer Research in London

Optimizing Allocation-Probability Tests for Clinical Trials with Normal Outcomes and Unknown Variances
Philipp Besendorfer, University of Regensburg

Untangling the comparator web: A blending method for individualised treatment comparisons in EU joint clinical assessments
Alberto Vidal Diez, Cytel Inc.

Quantifying the impact of interim decisions in group sequential trials
Gianmarco Caruso, MRC Biostatistics Unit, University of Cambridge

15:30 – 16:30: *Coffee Break*

16:30 – 18:00: *Parallel Sessions*

WP11: **Agentic AI for Transformative Clinical Research and Drug Development**

Room: Convención A

Organizer(s): Haitao Pan, St. Jude Children's Research Hospital

Chair: Li-an Xu, Pfizer

Title & Speaker: **An AI companion for autonomous research with application to statistical methods for clinical trials**
Yuan Ji, University of Chicago

Closing the loop: AI-powered IPD reconstruction and robust evidence synthesis
Yanxun Xu, Johns Hopkins University

AI-Powered Digital Ecosystem and Value Revitalization in Pharma Industry
Jingjun (Jeannie) Qiu, FosunPharma

Discussant
Jie Chen, Taimei Intelligence Pharma R&D

WP12: **Evidence to Insights: Examples of Statistical Innovation from Platform Trials to AI-Enabled Real-World Evidence in Medical Devices**

Room: Convención B

Organizer(s): Cristiana Mayer, Johnson & Johnson Vision

Chair: Patrick Roney, Georgetown University

Title & Speaker: **A platform trial under a master protocol for iterative medical device innovation: The case study in myopia-correcting soft contact lenses**
Cristiana Mayer, Johnson & Johnson Vision

Optimizing pragmatic clinical trial design for medical devices: An AI-enabled target trial emulation framework using large-scale real-world data
Xiaofeng Wang, Cleveland Clinic

Traditional validation pathways of a biomarker and examples of AI created efficiencies
Hope Knuckles, Abbott

WP13: **The Expanding Role of Simulation Studies in the Design, Communication and Regulation of Modern Clinical Trials**

Room: Bahia

Organizer(s): Elias Laurin Meyer, Berry Consultants

Chair: Elias Laurin Meyer, Berry Consultants

Title & Speaker: **Using simulations to guide the process of trial design**
Dario Zocholl, National Center for Tumor Diseases (NCT), NCT Berlin, Universitätsmedizin Berlin, Berlin Institute of Health at Charité (BIH), and Max Delbrück Center

Bridging the gap: Utilising clinical trial simulations as a tool for regulatory dialogue and funding of platform trials
Michaela Maria Freitag, Institute of Biometry and Clinical Epidemiology, Charité - Universitätsmedizin

Simulation-driven collaborative trial design with regulators
Matteo Vestrucci, Berry Consultants

Simulation studies in regulatory science
Tobias Fellinger, AGES - Austrian Agency for Health and Food Safety (Österreichische Agentur für Gesundheit und Ernährungssicherheit)

WP14: **Causal inference methods for treatment effect estimation in randomised trials with intercurrent events – Part 1**

Room: La Paz

Organizer(s): Robin Ristl, Medical University of Vienna

Chair: Niels Hendrickx, Uppsala University

Title & Speaker: **Treatment switching in randomised trials with time-to-event outcomes – hypothetical estimand strategy via causal inference**
Robin Ristl, Medical University of Vienna

Causal inference methods for estimating treatment effects in the presence of rescue medication and treatment
Louise Jespersen, Medical University of Vienna

Causal inference methods for specific estimands in the context of randomised trials – scoping review
Paula Starke, University Medical Center Göttingen

TMLE for regulatory purposes: A landscape overview
Hana Lee, The US Food and Drug Administration

WP15: **Dose Optimization Incorporating Biomarker in Oncology Clinical Trials**

Room: Caracas

Organizer(s): Kentaro Takeda, Astellas Pharma

Chair: Ying Yuan, University of Texas MD Anderson Cancer Center

Title & Speaker: **Precision generalized phase I-II designs**
Yong Zang, Indiana University

Dose Optimization Design with Biomarker Identification for Cancer Clinical Trials
Shing Lee, Columbia University

A Bayesian Enrichment Design for Dose Optimization in Dose-Ranging Basket Trials
Yeonhee Park, Sungkyunkwan University

COMIC: A Bayesian dose optimization design for drug combination in multiple indications with application to CAR-T therapies
Kentaro Takeda, Astellas Pharma

Scientific Program – Day 2

Thursday October 15, 2026

8:30 – 18:00: *Registration* — Hotel Entrance Floor 0 Hospitality

8:30 – 10:00: *Keynote Session* (TK) — Convención

Chair: Wenjin Wang, Pfizer

Keynote Speech: **From Evidence to Decisions: How RWE, Causal Inference, and AI Are Reshaping Biopharmaceutical Statistics**
Weili He, AbbVie
Recipient of the 2026 **Heyse Memorial Lecture Award**

10:00 – 11:00: *Coffee Break*

11:00 – 12:30: *Parallel Sessions*

TP1: **Borrowing Strength or Inflating Risk? Bayesian Evidence in Rare Diseases**

Room: Convención A

Organizer(s): Martin Posch, Medical University of Vienna

Chair: Martin Posch, Medical University of Vienna

Title & Speaker: **Flexible and robust Bayesian information borrowing for clinical trials with longitudinal outcomes**
David Jesse, F. Hoffmann-La Roche AG, University Medical Center Göttingen

Robust Bayesian borrowing with local control of type I error and power
Lingjiao Wang, University of Warwick

Comparing Type I Error Measures in Hybrid Clinical Trials
Teresa Engelbrecht, Medical University of Vienna

Discussant
Sarah Zohar, Inserm

TP2: **Developing a Principled Framework for Exploratory Subgroup Analyses and Related Issues in Clinical Trials**

Room: Convención B

Organizer(s): Larry Leon, Statistical Methodology Research, MSD

Chair: Larry Leon, Statistical Methodology Research, MSD

Title & Speaker: **Statistical approaches and tools for identification of interpretable predictive subgroups with applications to consistency evaluation in multi-regional clinical trials (MRCTs)**
Larry Leon, Statistical Methodology Research, MSD

Overview of modern approaches for evaluating treatment effect heterogeneity from clinical and observational data

Ilya Lipkovich, Retiree

Application of spatio temporal modeling in long term phenotypic screening of patient derived organoids
Maryam Kazemi Naeini, Discovery Sciences, R&D, AstraZeneca

TP3: **Causal Inference Methods for Treatment Effect Estimation in Randomised Trials with Intercurrent Events – Part 2**

Room: Bahia

Organizer(s): Robin Ristl, Medical University of Vienna

Chair: Fei Yang, The University of Manchester

Title & Speaker: **Review of causal inference methods for randomised trials in European Public Assessment Reports**
Susanne Urach, Austrian Agency for Health and Food Safety

Estimating principal-stratum vaccine efficacy under incomplete regimen completion
Florian Klinglmüller, Austrian Agency for Health and Food Safety

Treatment discontinuation and while-on-treatment estimation
Niels Hendrickx, Uppsala University

Bayesian expert elicitation for estimand construction in rare disease: A cystic fibrosis case study
Woosub Shin, Université Paris Cité, Inserm / KU Leuven

TP4: **Advances in Clinical Endpoints for ALS and Other Rare Disease Trials: Joint Assessment of Function and Survival, Patient Preferences, and Estimands**

Room: La Paz

Organizer(s): Ruben van Eijk, University Medical Center Utrecht

Chair: Ying Lu, Stanford University

Title & Speaker: **Expressing the joint effect on function and survival using win probability**
Ying Lu, Stanford University

Integrating patient-reported preferences in the joint assessment of function and survival
Ruben van Eijk, University Medical Center Utrecht

Different estimands for function-related treatment effects and their relation to survival
Afroditi Tsatsi, University of Ioannina

A regulatory perspective on innovative endpoints for ALS trials
Stavros Nikolakopoulos, College ter beoordeling van geneesmiddelen & University of Ioannina

TP5: **Applications of AI to Clinical Trial Design, SAP Writing, and Evidence-Driven Strategy in Clinical Development**

Room: Caracas

Organizer(s): Li-An Xu, Pfizer Inc.

Chair: Cheng Zheng, Daiichi Sankyo Inc.

Title & Speaker: **Accelerating clinical trial design with agentic AI: From fragmented data to evidence-driven development strategies**
Xiaoyan Wang, Tulane University

AI for statistical analysis plan authoring: A retrieval-augmented, agent-orchestrated automation framework
Zhaohua Lu, Daiichi Sankyo Inc.

Digital twins: Opportunities and challenges in virtual clinical trials
Fangrong Yan, China Pharmaceutical University

Evidence-based clinical trial design: When the power of AI meets statistical rigor
Will Ma, HopeAI

12:30 – 14:00: *Lunch Break (Lunch provided)*

14:00 – 15:30: *Parallel Sessions*

TP6: **FDA Guidance on Bayesian Methodology in Clinical Trials: Challenges and Opportunities**

Room: Convención A

Organizer(s): Ying Yuan, University of Texas MD Anderson Cancer Center

Chair: Kentaro Takeda, Astellas Pharma

Title & Speaker: **Implementing FDA's Bayesian methodology guidance: Strategies for calibrating trial success criteria**
Ying Yuan, University of Texas MD Anderson Cancer Center

Use of Bayesian mechanistic digital twin in early phase clinical trials based on totality of data
Ruitao Lin, University of Texas MD Anderson Cancer Center

Quantifying effective sample size in Bayesian clinical trials
Chenguang Wang, Regeneron

Global sensitivity analysis of operating characteristics in type I error-calibrated quantitative decision-making clinical trial simulations
Harry Saxton, Astellas Pharma

TP7: **Adaptive and Integrated Designs for Modern Clinical Trials**

Room: Convención B

Organizer(s): Heng Zhou, MSD

Chair: Rajenki Das, MRC Biostatistics Unit, University of Cambridge

Title & Speaker: **Impact of missing data handling on response adaptive randomisation in rare disease trials**
Rajenki Das, MRC Biostatistics Unit, University of Cambridge

Clinical trial evaluation of digital medical devices for clinical decision support: a simulation study
Baptiste Archambaud, Inserm, Université Paris Cité

Beyond treatment effect drift in Bayesian dynamic borrowing: variance misspecification and type I error
Julien Tanniou, Saryga

Bayesian sample size determination for analysis of covariance in randomised controlled trials incorporating historical data
David Jones, University of Bath

TP8: **Why Statisticians Should Not Leave Safety Physicians Alone with Complex and Diverse** Data

Room: Bahia

Organizer(s): Richard Zink, JMP Statistical Discovery LLC

Chair: Richard Zink, JMP Statistical Discovery LLC

Title & Speaker: **The aggregate safety assessment planning (ASAP) process for multidisciplinary safety assessment**
Greg Ball, ASAPprocess

A coherent assessment of safety through a contemporary application of the Bradford-Hill criteria
Jürgen Kübler, QSciCon

Recurrence: An analysis of adverse events whose time has come in the evaluation of patient safety in clinical trials
Richard Zink, JMP Statistical Discovery LLC

Binomial sum variance inequality correction of 95% CIs of percentages in multicentre studies ensures approximately 95% coverage with minimal width
Paul Talsma, Phastar

TP9: **Machine Learning for Treatment Effect Heterogeneity Assessment in Randomized Trials: Methods, Inference, and Interpretation**

Room: La Paz

Organizer(s): Olga Marchenko, Bayer, Whippany, US

Chair: Silke Janitza, Bayer AG, Berlin, Germany

Title & Speaker: **Assessing Treatment Effect Heterogeneity using Machine Learning: Questions, Methods, Challenges**
Bohdana Ratitch, Bayer Inc.

Inference for ML-Driven Heterogeneous Treatment Effects in Randomized Experiments
Michael Lingzhi Li, Harvard Business School

Beyond the Average: Leveraging Treatment Heterogeneity for More Powerful Testing in RCTs
Stijn Vansteelandt, Ghent University

Understanding Drivers of Treatment-Effect Heterogeneity via Interpretable ML: Lessons from Recent Papers
David Svensson, AstraZeneca R&D Biopharmaceuticals

TP10: **New Collaboration Opportunities between Academia and Pharmaceutical Industry**

Room: Caracas

Organizer(s): Ran Dai, University of Nebraska Medical Center

Hongyuan Cao, University of Wisconsin-Madison

Chair: Ivan Chan, Bristol Myers Squibb

Title & Speaker: **Statistical leadership and innovation in accelerating drug development**
Ivan Chan, Bristol Myers Squibb

Implications for industry of the ICH E20 guideline on adaptive designs for clinical trials
Christopher Jennison, University of Bath

From genomic discovery to precision therapeutics in Alzheimer's disease
Zhongming Zhao, UTHealth Houston

15:30 – 16:30: *Coffee Break*

16:30 – 18:00: *Parallel Sessions*

TP11: **Bayesian Methods & Decision-Theoretic Frameworks**

Room: Convención A

Chair: Gianmarco Caruso, MRC Biostatistics Unit, University of Cambridge

Discussion Topics & Panelists: **On the calibration of Bayesian success criteria and operating characteristics for clinical trials**
Ying Yuan, UT MD Anderson Cancer Center

When should the next clinical trial be conducted? A Bayesian borrowing approach for AI-based digital medical devices
Caroline Lawless, Inserm, Université Paris Cité,

Bayesian simultaneous credible bands for polynomial regression
Fei Yang, The University of Manchester

An extended Bayesian semi-mechanistic dose-finding design for phase I oncology trials using pharmacokinetic and pharmacodynamic information
Yisheng Li, MD Anderson Cancer Center

TP12: **Recent Advances in Targeted and Causal Machine Learning in Randomized Trials and Beyond**

Room: Convención B

Organizer(s): Silke Janitza, Bayer AG

Chair: Bohdana Ratitch, Bayer Inc.

Title & Speaker: **Automated, efficient and model-free inference for randomized clinical trials via data-driven covariate adjustment**
Kelly Van Lancker, Ghent University and Vrije Universiteit Brussel

Causal inference via propensity scores for case-control and case-cohort studies
Mireille Schnitzer, Faculté de pharmacie, l'Université de Montréal

Coherent and efficient subgroup effect inference using the estimands framework and causal machine learning
Philippe Boileau, McGill University

The crucial role of Targeted Machine Learning to increase power in RCTs with small sample size
Susan Gruber, Targeted ML Solutions

TP13: **Multiplicity in Model Clinical Trials: Graphical, Adaptive, and AI-Augmented Designs**

Room: Bahia

Organizer(s): Dong Xi, Gilead Sciences

Chair: Yao Chen, Novartis

Title & Speaker: **Practical implementation of graphical multiple testing procedures in group-sequential designs**
Michael Grayling, Johnson & Johnson

AI for group sequential oncology design with multiple hypotheses
Keaven Anderson, MSD

Revisiting adaptive multi-arm two-stage designs
Daniel Sabanés Bové, RCONIS

Graph-based adaptive designs for multi-arm trials with multiple endpoints
Martin Posch, Medical University of Vienna

TP14: **Applications of AI for Drug Development**

Room: La Paz

Organizer(s): Samuel Wu, University of South Florida

Chair: Dong Xu, University of South Florida

Title & Speaker: **Open Source Drug Discovery in the AI Era**
Jake Chen, University of Alabama at Birmingham

Application of natural language processing and machine learning for drug target selection
Chao Fang, AstraZeneca

Bayesian optimization and protein structure modeling for AI-guided peptide therapeutic development
Dong Xu, University of South Florida

AI-driven optimization of patient-specific immune checkpoint therapy through tumor immune microenvironment modeling
Qingpeng Zhang, University of Hong Kong

TP15: **Confirmatory Evidence Generation: Master Protocols and Related Trial Design and Analytics Issues**

Room: Caracas

Organizer(s): Jingjing Ye, DataDoctor LLC & Platinum Life Excellence BioTech

Chair: Annette Kopp-Schneider, German Cancer Research

Title & Speaker: **Generalized Randomized Basket Trials**
Robert Beckman, Georgetown University Medical Center

A look at safety and benefit-risk assessment in master protocols
Alessandro Vagheggini, MSD Switzerland

A unified information–time paradigm for single-arm time-to-event trials: The window-cohort with calibrated follow-up requirement
Haitao Pan, St Jude Children's Research Hospital

Bayesian meta-analysis of continuous outcomes for postoperative pain: Should treatment effects be modelled as additive or multiplicative?
Suzanne Freeman, University of Leicester

Scientific Program – Day 3

Friday October 16, 2026

8:30 – 12:30: *Registration* — Hotel Entrance Floor 0 Hospitality

8:30 – 10:00: *Keynote Session* (FK1) — Convención

Chair: Florian Klinglmüller, AGES — Austrian Agency for Health and Food Safety

Keynote Speech: **Europe in Action: Regulatory Considerations on Clinical Evidence and Data Science**
Peter Arlett, Data Analytics and Methods Taskforce, European Medicines Agency

Panel Discussion: Chair: Florian Klinglmüller, AGES — Austrian Agency for Health and Food Safety

Panelists:

- Annette Kopp-Schneider, German Cancer Research Center
- Sarah Zohar, Inserm
- Frank Bretz, Novartis

10:00– 11:00: *Coffee Break*

11:00 – 12:30: *Parallel Sessions*

FP1: **Innovative Solutions to Address Significant Problems in Drug Development via Academia and Industry Partnership**

Room: Convención A

Organizer(s): Ming Tan, Georgetown University Medical Center

Chair: Chenguang Wang, Regeneron Pharmaceuticals

Title & Speaker: **Bias corrections and functional models for a variety of estimands in clinical trials**
Ming Tan, Georgetown University Medical Center

Innovative solutions to address significant problems in drug development via academia and industry partnership
Jeen Liu, Regeneron Pharmaceuticals

Panel discussion on innovative solutions to address significant problems in drug development via academia and industry partnership
Sejong Bae, Augusta University

Panel discussion on innovative solutions to address significant problems in drug development via academia and industry partnership
KentaroTakeda, Astellas

FP2: **Advancing Patient-Reported Outcomes in Oncology Trials: Estimand, Design Rigor, Results Interpretation, and Implications for Labeling & HTA**

Room: Convención B

Organizer(s): Leslie Meng, Boehringer Ingelheim

Chair: Alexandra Lauer, Boehringer Ingelheim

Title & Speaker: **Raising standards for patient-centered evaluation of benefit and tolerability in oncology trials**
Antoine Regnault, formerly Modus Outcome, A THREAD Company

Patient Reported Outcomes MythBusters and other considerations
Yovanna Castro, Patient-Centered Outcomes in Novartis

Patient-reported outcomes in EU-HTA: From considerations to implementation
Anton Schoenstein, Boehringer Ingelheim

The patient voice in focus: Advancing early phase oncology trials through PRO-driven innovation
Christina Yap, Institute of Cancer Research in London

FP3: **Biostatnet: Connecting Biostatistics Research II**

Room: Bahia

Organizer(s): Teresa Pérez Pérez, Universidad Complutense de Madrid

Chair: Teresa Pérez Pérez, Universidad Complutense de Madrid

Title & Speaker: **Sample size reestimation in two-arm trials using doubly robust estimators for marginal treatment effects**
Maria Lee Alcober, Universitat Politècnica de Catalunya - BarcelonaTech

Modeling approaches to incorporate non-concurrent controls in platform trials with time-to-event endpoints
Pavla Krotka, Universidad Politécnica de Cataluña

Expert-centred evaluation of CURIE-AI in biosciences: from gene ontology narratives to biomedical benchmarking across psoriasis, Crohn's disease, and adverse food reactions
Yovani Torres, BIOST3, Section of Statistics, Faculty of Biology, University of Barcelona

A framework for sample size calculation in Kolmogorov–Arnold network prediction models
Ester Rosa, University of Padua

FP4: **Causal Inference & Pharmacoepidemiology**

Room: La Paz

Chair: Alessandro Vagheggini, MSD Switzerland

Title & Speaker: **Propensity score methods in non-randomized clinical studies: A simulation framework for preterm infant growth**
Atiye Nazari, Nestlé Research

Efficiency and interpretation of marginal versus conditional covariate adjustment for the Cox model
Laetitia Fritz, Medical University of Vienna, Center for Medical Data Science, Institute of Medical Statistics

Weighted Cumulative Exposure modelling of the cumulative effects of time-varying drug exposures
Michal Abrahamowicz, McGill University

Novel methods to address measurement error in safety and effectiveness studies of prescription-based time-varying exposures
Steve Ferreira Guerra, Université de Montréal

FP5: **Adaptive Designs & Dose-Optimization**

Room: Caracas

Chair: Hui Zhang, Northwestern University

Title & Speaker: **Adaptive population selection designs for clinical trials with multiple endpoints**
Koko Asakura, National Cerebral and Cardiovascular Center

Great Wall: A generalized dose optimization design for drug combination trials maximizing survival benefit
Yan Han, Indiana University School of Medicine

Incorporating backfills in model-free dose-escalation designs
Nan Wang, AstraZeneca

Operationalizing allocation probability tests for RAR designs
Stina Zetterstrom, MRC Biostatistics Unit, University of Cambridge

12:30 – 14:00: *Lunch Break (Lunch provided)*

14:00 – 15:30: *Parallel Sessions*

FP6: **Panel Discussion: Statistical Leadership in a Transforming Biopharma Landscape**

Room: Convención A

Organizer(s): Jie Chen, Taimei Intelligence Pharma R&D

Chair: Jie Chen, Taimei Intelligence Pharma R&D

Discussion Topics & Panelists: Panelists:

- Keaven Anderson, MSD
- Ivan Chan, Bristol Myers Squibb
- Margaret Gamalo, Pfizer, Inc.
- Elaine Hoffman, CSL

Discussion topics:

- The evolving regulatory landscape and its implications for statistical practice and submissions;
- The expanding role of real-world evidence (RWE) in evidence generation and regulatory decision-making;
- The strategic integration of AI and automation into biostatistics workflows; and
- The changing expectations of statistical leaders as scientific partners, strategic advisors, and organizational influencers.

FP7: **AI, Digital Health & Model-Informed Drug Development**

Room: Convención B

Chair: Yi Huang, University of Maryland Baltimore County

Title & Speaker: **Computational prediction of druggable targets using temporally validated machine learning**
Mohammed Quazi, West Virginia University

Building confidence in generative AI: A multi-method framework for evaluating AI-driven document automation in drug development
Camilo Rojas Cifuentes, Johnson and Johnson Innovative Medicine

What is a virtual patient? Toward unifying definitions in model-informed drug development
Laurens Sluijterman, Radboud University Medical Center

TBEP-v2: Integrating knowledge graphs and explainable AI for traceable biomedical hypothesis generation
Shivansh Gupta, Indian Institute of Information Technology Allahabad, UP

FP8: **Biomarker Discovery, Validation & Translational Genomics**

Room: Bahia

Chair: Hui Zhang, Northwestern University

Title & Speaker: **Oral microbial signatures identified in head and neck cancers can guide targeted therapies to mitigate oral mucositis**
Kim-Anh Do, University of Texas M.D. Anderson Cancer Center

Powering up: Enhancing success rates in biomarker discovery, sample size calculation for feature selection
Fetene Tekle, Johnson & Johnson, Innovative Medicine R&D, Beerse, Belgium

An innovative statistical method for co-localization in super-resolution microscope images
Hui Zhang, Northwestern University

Evaluating the robustness of firth-type logistic regression methods for rare events under model misspecification and complex predictor structures
MD Fakrul Islam, Department of Statistics, George Mason University

FP9: **Leveraging External Information in the Design and Analysis of Clinical Trials**

Room: La Paz

Organizer(s): Samuel Wu, University of South Florida

Chair: Chenguang Wang, Regeneron

Title & Speaker: **A few cautionary notes for leveraging external data in clinical trials**
Bo Lu, Ohio State University

Leveraging external controls in clinical trials: estimands, estimation, assumptions
Fan Li, Duke University

External control data augmentation for randomized controlled recent-onset type 1 diabetes immunotherapy trials

Aidong Ding, Northeastern University

Leveraging external information in adaptive platform trial designs: two case studies
Samuel Wu, University of South Florida

FP10: **Advanced Statistical Estimation, Mixed Modeling & Software Tools**

Room: Caracas

Chair: Daniel Sabanes Bove, RPACT

Title & Speaker: **Beyond means and dispersions: Mixed-effects location-scale modeling for cluster-randomized trials**
Depeng Jiang, University of Manitoba

Conditional unbiased estimation of the selected treatment effects in a three-stage drop-the-losers trials with early futility
Yogesh Katariya, Indian Institute of Technology Kanpur

The RobinCar family: R tools for robust covariate adjustment in randomized clinical trials
Daniel Sabanes Bove, RPACT

Reframing evidence: Biostatistical innovations to close the women's health gap in trials and real-world data
Grammati Sarri, Cytel, Inc.

15:30 – 16:30: *Coffee Break*

16:30 – 18:00: *Closing Keynote* (FK2) — Convención

Chair: Ying Lu, Stanford University

Keynote Speech: **Targeted Learning for Generating Real World Evidence**
Mark van der Laan, University of California at Berkeley
Recipient of the 2026 **Lai Memorial Lecture Award**

Short Course Description

SC1: **Improving Treatment Evaluation by Integrating Randomized Clinical Trials and Real-World Data: Causal Inference and Machine Learning Approaches (Full-Day)**

Instructor(s): Shu Yang, Professor of Statistics, North Carolina State University, USA

Abstract: Motivation and Course Description:

The 21st Century Cures Act, enacted in 2016, highlights the importance of precision medicine and the utility of real-world data (RWD) to accelerate the development and evaluation of new treatments. It encourages the FDA and other regulatory bodies to consider real-world evidence (RWE) alongside traditional randomized controlled trials (RCTs).

RCTs are widely considered the gold standard for causal inference due to their internal validity. However, they often suffer from practical limitations such as restrictive eligibility criteria and limited sample sizes. In contrast, RWD offers broader population coverage and reflects clinical practice more realistically but is prone to confounding and other biases. Integrating RCTs with RWD offers the potential to combine the strengths of both data sources, achieving internal validity from RCTs and external validity from RWD, thereby enabling more generalizable, efficient, and timely treatment evaluations.

This short course will introduce statistical frameworks and methodologies that facilitate the integration of RCTs and RWD to: Improve the generalizability of RCT findings to broader patient populations, Enhance the estimation of treatment effect heterogeneity for precision medicine, and Address challenges such as covariate shift, unmeasured confounding, and model misspecification through robust statistical and machine learning techniques.

Simulated case studies and hands-on demonstrations using publicly available R packages will support the conceptual and methodological content. A foundational understanding of clinical trials and causal inference is recommended.

Course Roadmap

I. Introduction to Integrated Evidence Generation

- Overview of RCTs vs. RWD and their complementary strengths
- Motivation for data integration in regulatory and real-world settings
- FDA guidance on the use of RWD for treatment evaluation
- Research questions and causal inference roadmaps

II. Generating Real-World Evidence through RCT-RWD Integration

- Motivating example: non-small cell lung cancer
- genRCT framework for integrating RCT and RWD
- Empirical evaluation: simulation and case studies
- R package: genRCT

III. Assessing Treatment Effect Heterogeneity Using RWD

- Key sources of bias in RWD
- Classical and machine learning methods for treatment effect heterogeneity estimation

- Integrative methods to improve robustness:
- Test-then-pool strategies
- Bias function modeling
- Selective borrowing
- Conformal inference
- Randomization-based inference
- Empirical evaluation: simulation and case studies
- R packages: elasticIntegrative, inFRT

Target Audience and Learning Objectives:

Students, researchers, and professionals in academia, industry, or regulatory agencies who are interested in causal inference, health data science, and RCT/RWD integration. By the end of the course, participants will:

- Understand the respective roles of RCTs and RWD in treatment evaluation and the benefits of integration
- Be equipped to perform generalizable analyses of RCTs by leveraging RWD
- Apply advanced statistical methods for estimating treatment effect heterogeneity using RWD
- Gain hands-on experience with R packages developed for causal inference and data integration

Comments: About Instructors:

Dr. Shu Yang is Professor of Statistics, Goodnight Early Career Innovator, and University Faculty Scholar at North Carolina State University. She earned her Ph.D. in Applied Mathematics and Statistics from Iowa State University and completed her postdoctoral training at the Harvard T.H. Chan School of Public Health. Her research focuses on causal inference and data integration with applications to comparative effectiveness research. She has made significant contributions to methods for missing data, spatial statistics, and real-world evidence, and has served as PI on multiple NIH, NSF, and FDA-funded projects. She is a recipient of the 2024 COPSS Emerging Leader Award. Website: https://shuyang.wordpress.ncsu.edu

Dr. Yang has served as lead or co-lead instructor for numerous short courses at prominent professional conferences and academic events, including the Joint Statistical Meetings (JSM), New England Statistical Society (NESS), Duke Industry Statistics Symposium (DISS), Eastern North American Region (ENAR), International Chinese Statistical Association (ICSA), ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop (RISW), Society for Clinical Trials (SCT), and the Lifetime Data Science Conference (LiDS), as well as summer schools hosted by various universities.

A detailed list of these engagements is provided below.

- Statistical methods for Randomized Clinical Trials with Information Borrowing from External Controls, Full-Day Short Course, Joint Statistical Meeting (JSM), Nashville, TN, USA. August 2–5, 2025

- Statistical methods for time-to-event data from multiple sources: a causal inference framework, Full-Day Short Course, New England Statistical Society (NESS), New Haven, CT, USA. May 31, 2025
- Statistical methods for time-to-event data from multiple sources: a causal inference perspective, Half-Day Short Course, Duke Industry Statistical Symposium (DISS), Duke University, Durham, NC, USA. April 9–11, 2025
- Statistical methods for time to event data from multiple sources: a causal inference perspective, Half-Day Short Course, Eastern North American Region (ENAR) Spring Meetings, New Orleans, LA, USA. March 23–26, 2025
- Statistical methods for time-to-event data from multiple sources: a causal inference framework, Full-Day Short Course, International Chinese Statistical Association (ICSA) Symposium, Nashville, TN, USA. June 16, 2023
- Design and analysis of randomized clinical trials with real-world data using causal inference framework and Bayesian methods, Half-Day Short course, ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop (RISW), Rockville, MD, USA. September 27–29, 2023
- Unveiling the Power of Real-World Data: A Causal Inference Framework for Designing and Analyzing Randomized Clinical Trials, 2nd CANSSI-NISS sponsored Health Data Science Workshop, Half-Day Short Course, Waterloo, Ontario, Canada. August 2, 2023
- Statistical methods for time-to-event data from multiple sources: a causal inference framework, Full-Day Short Course, Lifttime Data Science (LiDS) Conference, Raleigh, NC, USA. May 31, 2023
- Design and analysis of randomized clinical trials with real-world data, Half-Day Preconference Workshop on the Society of Clinical Trial (SCT), Baltimore, MD, USA. May 21–24, 2023
- Improved causal inference combining randomized clinical trials and observational studies, Half-day lectures, Canadian Statistical Sciences Institute (Collaborative Research Teams) Summer School, University of Ottawa, Ontario, Canada. (Invited). July 5–8, 2022.

SC2: **Applied Longitudinal Data Analysis with AI for Pharmaceutical Science (Full-Day)**

Instructor(s): Depeng Jiang, Professor of Biostatistics, University of Manitoba, Canada

Abstract: This one day, AI-informed workshop bridges longitudinal data analysis with modern data-science approaches to advance biopharmaceutical statistics. The course offers an applied, hands-on introduction to modeling repeated-measures data in pharmaceutical contexts, with practical guidance on interpretability, reproducibility, and responsible AI in regulated environments. Through real-world public-health and pharmaceutical datasets, participants will translate theory into actionable analysis in R, from data preparation and model fitting to evaluation, reporting, and communication with non-technical stakeholders. Emphasis is placed on transparent workflows, robust handling of missing data and time-varying covariates, and the integration of AI-inspired techniques while maintaining interpretability and regulatory relevance.

Learning objectives

- Select appropriate longitudinal analytic approaches (e.g., linear and generalized linear mixed-effects models) for intervention evaluation in pharmaceutical settings.Fit, compare, and interpret mixed-effects models in R (lme4/nlme, with emphasis on random intercepts/slopes and model diagnostics).
- Manage time-varying covariates and missing data using principled strategies (e.g., multiple imputation, maximum likelihood approaches) within longitudinal analyses.
- Evaluate model performance using suitable criteria (AIC/BIC, likelihood ratio tests, predictive accuracy, calibration) and perform model validation.
- Communicate results to non-technical stakeholders with clear narratives, tables, figures, and reproducible code-based reports.
- Develop reproducible workflows in R (RStudio, version control, and R Markdown) to ensure auditability and transparency.
- Explore AI-informed enhancements to longitudinal analyses while preserving interpretability and regulatory suitability.

Target audience

- Biostatisticians, data scientists, and researchers in biopharmaceutical science
- Clinical trial statisticians, pharmacometricians, epidemiologists, and health outcomes researchers
- Regulators, policy researchers, and graduate students in biostatistics or pharmaceutical sciences
- Professionals seeking practical, reproducible skills for analyzing longitudinal data in pharmaceutical contexts

Notes for participants

- The course explicitly ties longitudinal data analysis to AI-for-health science goals, with emphasis on reproducible, interpretable results suitable for biopharmaceutical practice and regulatory contexts. If you have questions about fit or prerequisites, please reach out prior to registration.
- Prerequisites and recommendations
- Basic proficiency in R and RStudio (data frames, tidyverse familiarity recommended)
- Familiarity with regression or introductory mixed-effects concepts is helpful
- Prior exposure to simple longitudinal analyses is advantageous but not required
- Proposed benefits for ISBS and attendees
- A competitively designed short course that aligns with ISBS goals of standardizing statistical practice in biopharmaceutical settings.
- A practical, hands-on experience focusing on reproducibility, interpretability, and responsible AI in pharmaceutical data analysis.
- Access to a high-quality, code-based teaching approach that participants can reuse in their own organizations and collaborations.

Comments: Dr. Depeng Jiang, PhD, is a full professor of Biostatistics in the College of Community and Global Health, Rady Faculty of Health Sciences, University of Manitoba. He also leads the Biostatistics Group within the George and Fay Yee Centre for Healthcare Innovation (CHI). Dr. Jiang has extensive experience providing statistical consulting to diverse clients, training researchers and students in statistical methodology, and applying advanced statistical learning methods to health services research and program evaluation. His work spans local, national, and international venues, with recognized contributions to methodological development and practical application in biostatistics. He is renowned for bridging theory and practice, fostering multidisciplinary collaborations with researchers and policymakers, and delivering statistically rigorous, accessible training. His expertise includes modern statistical learning methods, mixed-effects modeling, AI-informed approaches, and the translation of complex analyses into actionable health insights. He has delivered short courses across both academic conference and university campus both locally, national and internationally.

What participants will gain (outcomes and benefits)

- A solid, implementable understanding of when and how to apply mixed-effects models to longitudinal pharmaceutical data.
- Practical, hands-on experience building, comparing, and validating mixed-effects models in R using real-world data.
- Competence in handling time-varying features and missing data in longitudinal analyses, with appropriate methodological choices.
- Skills to translate statistical results into actionable public health and pharmaceutical insights for stakeholders.
- Ability to present clear, reproducible results using well-documented R code, tables, and figures.
- A reproducible analysis workflow in R and RStudio, including code-enabled reporting for auditability and collaboration.
- Insight into integrating AI-inspired approaches with longitudinal modeling in a way that maintains interpretability and regulatory alignment.

SC3: **Implementing Bayesian Methodology in Clinical Trials: Navigating New FDA and EMA Guidance (Half-Day)**

Instructor(s): Yong Zhang, Associate Professor of Biostatistics, Indiana University, USA

Ying Yuan, Bettyann Asche Murray Distinguished Professor, Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

Abstract: In January 2026, the U.S. Food and Drug Administration (FDA) issued landmark draft guidance on the use of Bayesian methodology in clinical trials for drugs and biologics. This paradigm-shifting guidance signals a new era in which the Agency is increasingly open to Bayesian methods for primary analysis and as pivotal evidence for drug approval. Similarly, the EMA has issued a position paper echoing these sentiments while highlighting the key technical challenges involved.

This half-day short course provides a comprehensive overview of Bayesian methodology in clinical trials, closely aligned with the latest FDA and EMA regulatory frameworks. In the first part of this short course, we will explore the strategic opportunities and practical challenges of integrating Bayesian approaches into drug development, emphasizing real-world applications. Through clinical case studies and hands-on demonstrations of open-source software, attendees will gain the technical skills necessary to implement these designs and the communication strategies required to engage effectively with regulatory authorities. In the second part of this short course, we will focus on Bayesian information borrowing methods for clinical trials, a key methodological area highlighted in FDA guidance. We will review several widely used approaches, including the power prior, commensurate prior, and robust meta-analytic prior. We will then introduce the dynamic information borrowing strategy, which is recommended in FDA guidance and adapts the extent of information borrowing based on the degree of prior–data conflict between current and historical data. Finally, we will present recent methodological extensions in this area, many of which have been developed through collaborations with industry partners.

Key Topics Include:

1. Fundamental concepts of Bayesian statistics within a regulatory context.
2. Defining and calibrating Bayesian success criteria to ensure robust operating characteristics.
3. Strategic considerations for prior distribution selection and expert elicitation.
4. Advanced methods for information borrowing from historical and real-world data (RWD).
5. Best practices and common pitfalls in Bayesian regulatory submissions.
6. Bayesian information borrowing methods.
7. Dynamic information borrowing and its extensions.

Learning Objectives:

By the conclusion of this course, participants will be able to:

- Navigate the evolving regulatory landscape, including the features and challenges of contemporary Bayesian guidance.
- Demystify the intricacies of Bayesian decision rules, calibration, and information borrowing, while critically evaluating their respective advantages and limitations.
- Master the implementation of innovative Bayesian designs through hands-on training with freely available software, facilitating data-driven decision-making in drug development.

Comments: Yong Zang, PhD, is a Showalter Scholar Associate Professor in the Department of Biostatistics and Health Data Science at the Indiana University School of Medicine, where he also serves as Co-Director of Clinical Research for the Biostatistics and Data Management Core at the IU Simon Comprehensive Cancer Center. He received his Ph.D. in Statistics from the University of Hong Kong and completed postdoctoral training at The University of Texas MD Anderson Cancer Center. His research focuses on clinical trial design and health data informatics, with over 80 peer-reviewed publications supported by the National Institutes of Health, the Showalter Trust, and Eli Lilly. An expert in Bayesian adaptive clinical trial design,

Dr. Zang has taught a dedicated course on the topic annually since 2016 and has delivered numerous short courses at major international conferences including AACR, ICSA, New England Statistics Symposium and DahShu Data Science Symposium.

Ying Yuan, PhD, is a Bettyann Asche Murray Distinguished Professor and Deputy Chair in the Department of Biostatistics at the University of Texas MD Anderson Cancer Center.

Dr. Yuan is an internationally renowned researcher in innovative Bayesian adaptive designs with over 300 publications. The designs and software developed by Dr. Yuan's lab (www.trialdesign.org) have been widely used in medical research institutes and pharmaceutical companies. The BOIN design, developed by Dr. Yuan's team, is a groundbreaking oncology dose-finding design that has been recognized by the FDA as a fit-for-purpose drug development tool. Dr. Yuan is a member of the FDA Advisory Committee for Genetic Metabolic Diseases, an elected Fellow of the American Statistical Association, and the lead author of two books: Bayesian Designs for Phase I-II Clinical Trials and Model-Assisted Bayesian Designs for Dose Finding and Optimization, both published by Chapman & Hall/CRC.

SC4: **Graphical multiple comparison procedures: Combining flexibility with optimality (Half-Day)**

Instructor(s): Frank Bretz, Distinguished Quantitative Research Scientist, Novartis

Yao Chen, Statistical Consultant, Novartis

Dong Xi, Gilead Sciences

Abstract: Addressing multiplicity is essential in confirmatory clinical trials to ensure valid statistical inference. Various multiple comparison procedures (MCPs) have been developed, including fixed-sequence, fallback, and gatekeeping procedures, which allow trialists to reflect the relative importance and inter-relationships of study objectives in a tailored multiple test procedure.

This course focuses on graphical approaches which enable the construction and exploration of tailored MCPs to meet specific study objectives, such as comparisons of multiple treatments against a common control and multiple endpoint analyses. In these approaches, MCPs are represented by directed, weighted graphs, where each node corresponds to an elementary hypothesis. A simple algorithm then facilitates the sequential testing of hypotheses.

Optimizing MCPs to maximize the probability of success is often a key concern for clinical trial teams. We will discuss clinically relevant objective functions for optimization and introduce an efficient algorithm, based on constrained nonlinear optimization, to identify optimal graphs.

Case studies will illustrate the flexibility and practicality of these approaches in clinical trial settings. We will also introduce the graphicalMCP R package, which implements weighted Bonferroni tests, weighted parametric tests (accounting for correlations between test statistics), and weighted Simes' tests. We also briefly consider power and sample size calculation. Example code for optimizing graphs will be demonstrated and shared, providing participants with practical tools to implement these methods effectively.

Comments: About the instructors:

Dr. Frank Bretz is a Distinguished Quantitative Research Scientist at Novartis. He has supported the methodological development in various areas of pharmaceutical statistics, including dose finding, estimands, multiple comparisons, and adaptive designs. Frank is an Adjunct Professor at the Hannover Medical School (Germany) and the Medical University Vienna (Austria). Frank is a Fellow of the American Statistical Association. Frank Bretz has taught short courses at JSM, ASA BIOP RISW, and other statistical conferences.

Dr. Yao Chen is a Statistical Consultant in the Advanced Methodology and Data Science group at Novartis. He has supported development and implementation of innovative statistical methodologies in multiple comparisons and treatment effect heterogeneity. Yao Chen has taught short course at ASA BIOP RISW.

Dr. Dong Xi is a Senior Director in the Biostatistics Innovation Group at Gilead Sciences. He has supported development and implementation of innovative statistical methodologies in multiple comparisons, dose finding, group sequential designs, estimands and causal inference. He is an associate editor of Statistics in Biopharmaceutical Research and a committee member of the International Conference of Multiple Comparison Procedures. Dong Xi has taught short courses at ICSA symposium, JSM, ASA BIOP RISW, and other statistical conferences.

SC5: **Data Visualization with Applications to the Life Sciences (Half-Day)**

Instructor(s): Richard Zink, Principal Research Fellow, JMP Statistical Discovery

Abstract: (Bio)statisticians and data scientists are effective at utilizing cutting-edge methodologies to address the complexities of data in order to produce meaningful results. Despite this technical prowess, quantitative scientists struggle to communicate the story contained within these data to their non-statistical colleagues. Given the volume of data to review, the variety of analyses to perform, and the need to turn findings into actionable outcomes, it should come as no surprise that clear insight is often out of reach to the research team. The traditional means of data summary – static tables and listings – are ineffective for understanding the story hiding in plain sight; data visualization is the key to effective communication for the modern quantitative scientist. The goal of this short course is to describe visualization methodologies to aid in the interpretation and communication of data from applications in clinical trials and other life science topics. Categorical, continuous, and time-to-event endpoints are discussed, and numerous practical illustrations are presented.

Learning Objectives

1. Describe the transition from traditional methods of data analysis and summary to visual approaches
2. Explore and interpret life science data using one or more data visualizations
3. Assess the strengths and limitations of various graphical techniques
4. Communicate the "data story" through numerous examples

Topics covered in lecture

1. Features of study design

2. Distributional summaries
3. Signal detection
4. Data integrity
5. Patient journeys
6. Time trends, time-to-event, recurrence
7. Clustering, correlation and co-occurrence
8. Subgroups
9. Meta-analysis
10. Benefit risk

Comments: About the instructor(s):

Richard C. Zink is Principal Research Fellow at JMP Statistical Discovery and has spent 20+ years in and around clinical trials and medical product development. Richard is author, editor, and contributor to 10 books on statistical topics in clinical trials and clinical research. He holds a Ph.D. in Biostatistics from the University of North Carolina at Chapel Hill, where he serves as Adjunct Professor of Biostatistics and Adjunct Assistant Professor of Public Health Leadership and Practice. Richard was awarded the distinction of Fellow of the American Statistical Association in 2020.

My research and day-to-day work since 2011 has included data visualization as a necessary component for effective communication of statistical concepts. I previously offered this course at the 2017 ASA Biopharmaceutical Regulatory-Industry Statistics Workshop, multiple DIA Annual Meetings (2018-2021), and as part of the 2025-2026 ASA Council of Sections Traveling Courses. This material will serve as the foundation for a course data visualization that I will teach in the Department of Biostatistics at the University of North Carolina at Chapel Hill starting in Fall 2026, and I am currently writing a book for CRC Press based on this material, tentatively due in August 2027.

SC6: **Causal AI with Targeted Learning (Half-Day)**

Instructor(s): Susan Gruber, Co-Founder, Targeted ML Solutions

Alan Hubbard, Professor of Biostatistics and co-director of the Center for Targeted Machine Learning, University of California at Berkeley, CA, USA

Abstract: Sound decision-making is built on transparent, interpretable evidence. However, practical challenges can make it difficult to obtain efficient unbiased estimates from randomized controlled trial (RCT), observational, or real-world data (RWD) studies. Enrollment difficulties and smaller than expected effect sizes reduce power for answering the question of interest. Lack of baseline randomization or adherence to treatment and informative loss to follow-up can bias estimates. Targeted Learning (TL) unifies causal inference, machine learning (ML) and statistical theory to provide a framework for evaluating causal effects from data. Targeted maximum likelihood estimation (TMLE) is an estimator that incorporates powerful machine learning to address these challenges.

This short course will first describe the Targeted Learning Causal Estimation Roadmap that guides study design, analysis, and interpretation. The focus then shifts to theory and practice underling Targeted maximum likelihood estimation (TMLE) and super learning (SL). TMLE is a maximally efficient double robust estimator that relies on fewer assumptions than propensity score-based methods that are often used to analyze health care data. Unlike most machine learning approaches used in isolation, TMLE+SL provides accurate control of the Type-I error rate, and good confidence interval coverage.

Case studies of actual and simulated data in rare disease and oncology will illustrate the application of the TL roadmap for (1) study design and planning, (2) efficient and flexible estimation using TMLE and SL, (3) diagnostics that increase transparency, and (4) sensitivity analysis and interpretation of findings. Practical guidance on how to pre-specify the TMLE + SL procedure tailored to characteristics of the data and the estimation problem will be emphasized. The course will conclude with an overview of recent advances.

Comments: About the instructors:

Dr. Gruber is co-founder with Mark van der Laan of Targeted ML Solutions, a company working to deliver TMLE+SL based software to expert and novice users. A former Harvard Medical School faculty and Reagan-Udall Foundation leader, Susan is a pioneer in targeted learning and real-world evidence, She released the first open-source package for data analysis using TMLE in 2010, which now has over 100k downloads world-wide.

Dr. Hubbard is Professor of Biostatistics and co-director of the Center for Targeted Machine Learning at UC Berkeley. His research focuses on the application of statistics to population studies with particular expertise in semi-parametric models and the use of machine learning in causal inference, as well as applications in high dimensional biology.

SC7: **Baseline Covariate Adjustment in Randomized Clinical Trials (Half-Day)**

Instructor(s): Kelly Van Lancker, Assistant Professor of biostatistics, Ghent University and Vrije Universiteit, Brussels

Bohdana Ratitch, Bayer

Abstract: Background and Motivation: Baseline covariate adjustment in randomized clinical trials (RCTs) has received renewed attention following recent regulatory guidance, notably the 2023 FDA and the 2015 EMA guidelines on adjusting for covariates in RCTs. These documents encourage sponsors to adjust for baseline covariates associated with the outcome to improve statistical efficiency, while carefully distinguishing between conditional and marginal treatment effects. Despite this regulatory momentum, practitioners face a complex landscape of methods, each with distinct assumptions, strengths, and limitations — particularly regarding model misspecification, non-collapsibility, variance estimation, and performance in small samples. This short course aims to bridge the gap between the evolving methodological literature and practical usage in clinical trial settings.

Course Overview: This course provides a comprehensive, practically oriented overview of baseline covariate adjustment methods for estimating treatment effects in RCTs. The course will cover the following topics:

- Introduction and foundational concepts: The estimand framework, with a focus on the distinction between conditional and marginal treatment effects, and a causal roadmap for covariate adjustment; the role of randomization in providing validity versus balance, and why covariate adjustment can yield meaningful increases in statistical power; stratification versus analysis with covariate adjustment.
- Estimation: We briefly introduce the estimation of conditional estimands using direct covariate-adjusted regression, highlighting its role as a foundational component for estimating marginal estimands. Building on this, we focus on standardization (G-computation) and then turn to Augmented Inverse Probability Weighting (AIPW) and Targeted Maximum Likelihood Estimation (TMLE), which are particularly useful in more complex settings where machine learning methods are employed. We also consider prognostic scores that incorporate historical data.
- The phenomenon of non-collapsibility whereby certain effect measures such as the odds ratio yield numerically different conditional and marginal estimates even in the absence of confounding.
- Special topics and practical challenges for marginal estimation such as small-sample performance and finite-sample corrections, multicenter trials, variable selection strategies, and considerations for group sequential designs.
- Practical examples in R: Illustrations of key methods using R, demonstrating implementation and interpretation.
- Review of available evidence on methods' performance.

The target audience includes statisticians and data scientists involved in the design and analysis of RCTs in pharmaceutical, biotech, and regulatory settings. Participants should have a working knowledge of generalized linear models, hypothesis testing, and the fundamentals of randomized clinical trial design.

By attending this short course, the attendees will gain:

(1) understanding of the motivation for use and benefits of covariate adjustment in randomized clinical trials

(2) understanding of conditional versus marginal estimands and non-collapsibility;

(3) practical knowledge of a range of covariate adjustment methods and their assumptions;

(4) understanding of variance estimators aligned with regulatory expectations; and

(5) awareness of methods' performance trade-offs, including in small samples.

Comments: About the instrucotrs:

Prof. Kelly Van Lancker is an assistant professor in biostatistics at Ghent University and Vrije Universiteit Brussels. She obtained both her Master's degree in Mathematics and her PhD in Statistical Data Analysis from Ghent University. Prior to her current position, she was a postdoctoral researcher at the Johns Hopkins Bloomberg School of Public Health.

Her research focuses on causal inference in health sciences, particularly randomized clinical trials, with a strong emphasis on covariate adjustment to improve efficiency and power in treatment effect estimation.

Kelly is widely recognized for her expertise in covariate adjustment. She has delivered over 15 workshops and 30 invited talks on the topic at international conferences, pharmaceutical companies, and academic institutions. Her work includes numerous peer-reviewed publications, and her PhD research specifically focused on causal inference for clinical trials, with covariate adjustment as a core component.

Dr. Bohdana Ratitch works as Statistical Innovation Expert and is a Distinguished Science Fellow at Bayer. She has an advanced training in quantitative methods (applied mathematics and statistics and PhD with machine learning focus) and extensive applied clinical trials experience (over 20 years in pharmaceutical/biostatistics roles). Bohdana has a strong interest and expertise in causal inference and covariate adjustment (company-wide and cross-industry initiatives on covariate adjustment methods for RCTs; work spanning causal inference, estimands, missing data, and subgroup/heterogeneity evaluation).

Bohdana regularly provides trainings on advanced statistical methods to colleagues at Bayer and presents at conferences and workshops. In December 2025, Bohdana was invited to provide a short course "Baseline Covariate Adjustment in RCTs" at the FDA Statistical Association 2025 Statistical Webinar series.

SC8: **Hands-on Introduction to Trial Design with East Horizon(TM) – Methodologies and Practical Applications (Half-Day)**

Instructor(s): Martin Kappler, Director of Customer Success, Cytel

Pantelis Vlachos, VP of Customer Success at Cytel

Abstract: This short course provides a hands-on introduction to Cytel's East Horizon platform for clinical trial design. Participants will work primarily with the Design and Explore modules to construct, simulate, compare, and evaluate candidate trial designs across a range of use cases. A shorter section will introduce enrolment and event prediction aas well as early development Go-NoGo decision making.

The course will cover adaptive methods and their implementation in the software through practical examples from several case studies. Designs may include interim futility and efficacy stopping, sample size re-estimation, population enrichment, and multiplicity in designs with multiple endpoints and/or treatment arms, for event-driven trials as well as trials with continuous and binary endpoints. Emphasis will be placed on understanding how these features are specified, explored, and interpreted in practice.

The course is intended for biostatisticians involved in clinical trial design who want practical experience with software-supported design evaluation. A certificate of participation will be provided.

Comments: About the instructors:

Martin is Director of Customer Success at Cytel and worked for several years in Cytel's Strategic Consulting unit as Expert Innovative Statistics Consultant. Martin is an expert in adaptive clinical trial design methodology and uses Cytel's software, including East Horizon, on a daily basis. Martin has over 25 years of experience and has been working as a trial, lead

statistician and expert statistical consultant in different CROs, pharmaceutical companies and as independent consultant.

Each participant will receive a 14-day East Horizon license to allow hands-on work during the course and independent reproduction of the examples afterward. Participants will gain practical experience in defining and parametrizing trial designs in East Horizon, exploring a broader design and scenario space through simulations, customizing design and analysis features through R-code integration, comparing candidate designs side by side, and interpreting of graphical and tabular outputs.

Pantelis is VP of Customer Success at Cytel. Before joining Cytel in 2013, he was a Principal Biostatistician at Merck Serono and a Professor of Statistics at Carnegie Mellon University for 12 years. His research interests lie in adaptive designs, mainly from a Bayesian perspective, as well as hierarchical model testing and checking, although his secret passion is Text Mining. He has served as Managing Editor of the journal "Bayesian Analysis" as well as on editorial boards of several other journals and online statistical data and software archives.

Abstracts for Keynote Speeches

WK: **A Sad Farewell to the Study-Wise Type I Error?**
Considerations about Innovation and Conservativism in Regulatory Statistics

Speaker:

Armin Koch, Hannover Medical School, Germany

Abstract: Confirmatory clinical trials are the backbone of decision making about efficacy and a positive benefit/risk-ratio for new medications or licensed drugs in new indications. European Scientific Advice is the platform, where stakeholders meet to agree about what should constitute a sound basis for decision making in this context and agree on the future development plan. This is, as a consequence, the place to also directly discuss new approaches for planning, conducting and analyzing these late-stage clinical trials. Guidance should follow once experience with the interpretation of trials incorporating new approaches is available. Many "new" concepts like approaches to multiple testing, adaptive designs, synthetic co-variates found their place in regular study planning this way.

This presentation makes a plea that scientific advice has to be somewhat conservative in many instances, provides recommendations, how a safe way forwards can be found, and tries to explain the difference between innovation and lowering the hurdle for regulatory approval. In this context we argue for maintaining a strict position regarding the control of the study-wise type-1-error in confirmatory clinical trials to be precise about the amount of information that will be available for assessment..

Comments: About the keynote speaker:

Studied mathematics and chemistry at Heidelberg University and has worked as research assistant at the German Centre for the Research on Cancer (DKFZ) between 1984 and 1991. Thereafter he has been an employee at the Institute of Medical Biometry at Heidelberg University. In 1999 he joined the Federal Institute for Drugs and Medical Devices (BfArM) in Germany. From 2001 to 2008 he was head of the unit „Biostatistics and Experimental Design". Since 2008 he is Director of the Institute for Biostatistics at Hannover Medical School. Prof. Koch is a member of the Scientific Advice Working Party (SAWP) and has been a member of the drafting group for ICH-E17 (multi regional clinical trials) and for ICH-E20 (adaptive designs, until step 2). His research interests are centered around experimental design for clinical trials and the aspects of regulatory statistics with applications in rare and frequent diseases.

More tabular:

Doctorate: 1997 Thesis (Dr. sc. hum. for Medical Biometry, University of Heidelberg)

Academic Education

09/1980 - 07/1988: Study of Mathematics and Chemistry, first for becoming a teacher, then for a Diploma in Mathematics at the University of Heidelberg

Academic & Professional Experience

09/1984 - 04/1988: Research Assistant in the Department of Biostatistics at the German Center for the Research on Cancer in Heidelberg (DKFZ)

10/1988 - 09/1991: Research Assistant in the Department of Biostatistics at the German Center for the Research on Cancer in Heidelberg

10/1991 - 04/1999: Assistant at the Institute of Medical Biometry and Informatics, University of Heidelberg

04/1999 - 09/2008: Member (and head since 2001) of the Biostatistics Group at the Federal Institute for Drugs and Medical Devices, Germany (BfArM)

09/2009 – 10/2018: Member of the Biostatistics Working Party (BSWP) at the European Medicines Agency (EMA)

10/2010 – 10/2018: Member and (since 12/ 2015) Chair of the Scientific Advisory Committee of the Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWIG)

04/2015 – 11/2017: European Expert in the „Drafting Group for ICH-E17" of the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH)

11/2019 – 05/2024: European Expert in the „Drafting Group for ICH-E20" of the ICH (up to Step 2)

since 01/2007: Member of the Scientific Advice Working Party (SAWP) at the EMA

since 09/2008: Director of the Institute for Biostatistics at Hannover Medical School (MHH), Hannover, Germany

TK: **From Evidence to Decisions: How RWE, Causal Inference, and AI Are Reshaping Biopharmaceutical Statistics**

Speaker:

Dr. Weili He, AbbVie

Recipient the 2026 Heyse Memorial Lecture Award

Abstract: As real-world data become more widely available and decision-makers increasingly seek timely, relevant evidence across the product lifecycle, statisticians are being asked to expand their toolkit beyond traditional trial-based paradigms. This lecture will highlight how rigorous statistical thinking can help ensure that real-world evidence is fit for purpose, scientifically

credible, and decision-relevant for both regulatory and health technology assessment contexts. I will discuss key methodological considerations in formulating the research question, assessing data fitness, selecting appropriate study designs, defining estimands, and conducting sensitivity analyses, emphasizing how these elements work together to transform heterogeneous data into reliable evidence.

The lecture will also examine the growing role of causal inference and AI-enabled methods in evidence generation. Machine learning, natural language processing, and other AI-based approaches offer powerful opportunities to improve efficiency, scale, and the use of unstructured data, but they also introduce important challenges related to transparency, reproducibility, and interpretability. I will argue that these methods should be viewed as complements to, not substitutes for, sound study design and causal reasoning. Looking ahead, the future of biopharmaceutical statistics lies in integrating RWE, causal inference, innovative designs, and AI within a coherent framework that supports timely, robust, and trustworthy decisions across the product lifecycle.

Comments: About the keynote speaker:

Dr. Weili He is a Distinguished Research Fellow and head of Medical Affairs and Health Technology Assessment (HTA) Statistics at AbbVie, where she leads a team of over 40 statisticians supporting major business activities in Medical Affairs and HTA. Her extensive career spans clinical development at MSD and post-market life-cycle management at AbbVie. Renowned as a leading biostatistician in pharmaceutical statistics, her areas of expertise include early-phase oncology study designs, late-stage clinical trial design and analysis, adaptive trial methodologies, benefit-risk assessment, real-world data and evidence generation, causal inference, and statistical approaches for HTA.

Dr. He has authored over 60 peer-reviewed statistical publications and served as lead editor for three influential books: "Practical Considerations for Adaptive Trials Design and Implementation" (Springer, 2014), "Benefit-Risk Assessment Methods in Medical Product Development" (CRC Press, 2016), and "Real-World Evidence in Medical Product Development" (Springer Nature, 2023). In January 2026, she assumes the role of Editor-in-Chief for Statistics in Biopharmaceutical Research, following her tenure as Associate Editor from 2014 to 2025.

Her accomplishments are recognized by the American Statistical Association (ASA), where she was elected as an ASA Fellow in 2018. At AbbVie, she holds the distinction of AbbVie Distinguished Research Fellow—the company's highest honor for scientific excellence. Professionally, Dr. He co-founded and now co-chairs the ASA Biopharmaceutical Section (BIOP) HTA Scientific Working Group in 2023, and previously, the BIOP RWE Scientific Working Group (2018–2022). She has held leadership roles on the ASA BIOP Executive Committee, including Chair-Elect, Chair, and Past-Chair (2020–2022), served as Industry Co-chair of the Regulatory-Industry Statistics Workshop (2017), and co-chaired the QSPI Benefit-Risk Working Group (2012–2018). In addition to her professional impact, Dr. He has mentored numerous junior statisticians, fostering the next generation of leaders in the field.

FK1: **Europe in Action: Regulatory Considerations on Clinical Evidence and Data Science**

Speaker:

Peter Arlett

Head of the Data Analytics and Methods Taskforce, European Medicines Agency (EMA)

Abstract: In a rapidly changing world, do we face evolution or revolution? The world we live in is changing fast and in this keynote, Peter Arlett, head of data analytics and methods at the European Medicines Agency will present on five major trends that affect all our lives as well as, our work. We will explore huge data, artificial intelligence, population change, environmental destruction, and geopolitical upheaval. We will then explore how these five major changes impact the world of medicine regulation, drug development and biostatistics. Finally, we will debate how we can, collectively, navigate through this change, to a better world for patients, for science and for society.

Comments: Peter Arlett is Head of the Data Analytics and Methods Taskforce at the European Medicines Agency. In this role he leads on operations and transformation on clinical evidence at the EMA including clinical trials, real world evidence, safety reporting and data science including AI. He is Chair of the EMA Data Board, Co-Chair of the HMA-EMA Network Data Steering Group, Co-chair of the EMA AI Coordination Group, Co-chair of the Vaccine Monitoring Platform Steering Group, and Member of the ACT EU Steering Group. Prior to taking up this role in 2020, he held leadership roles within the EMA in the areas of pharmacovigilance, epidemiology, and risk management.

Prior to starting at EMA in 2008, Peter worked on new legislation and international collaboration for the European Commission, was the UK delegate to the European Committee for Human Medicinal Products, and was an assessor and manager at the UK's MHRA. He has a medical degree from University College London, and began his career as a hospital physician in Oxford and London.

In addition to his role at EMA, Peter is Honorary Professor at the London School of Hygiene and Tropical Medicine. He is also a Fellow of the Royal College of Physicians of Edinburgh and of the Faculty of Pharmaceutical Medicines of London.

FK2: **Targeted Learning for Generating Real World Evidence**

Speaker:

Professor Mark van der Laan, University of California, Berkeley

Recipient of the 2026 Lai Memorial Lecture Award

Abstract: Targeted Learning follows a general roadmap for 1) accurately translating the real world into a formal statistical estimation problem in terms of causal estimand, a corresponding statistical

estimand, and statistical model; 2) a corresponding template for construction of a targeted maximum likelihood estimator (TMLE) of the statistical estimand; and finally 3) a sensitivity analysis addressing the possible causal gap. The TMLE represents an optimal plug-in machine learning based estimator of the estimand combined with formal statistical inference. The three pillars of TMLE are super-learning, Highly Adaptive Lasso (HAL), and the TMLE-update step, where the latter has various regularizations such cross-fitted TMLE, collaborative TMLE, and adaptive TMLE (Lars van der Laan et al., 2023). Through super-learning it can incorporate high dimensional and diverse data sources such as images, NLP features, and state of art algorithms tailored for such data sources. To optimize finite sample performance, the precise specification of TMLE can be tailored towards the precise experiment and statistical estimation problem in question, while being theoretically grounded, optimal, and benchmarked. TMLE applies to any estimation problem including causal inference on survival outcomes allowing for time-dependent confounding of treatment and drop-out. We provide a motivation, explanation, and overview of targeted learning; the key role of super-learning and HAL; discuss some of the key choices and considerations in specifying the TMLE-step; and discuss (a priori specified) SAP construction based on targeted learning, incorporating outcome-blind simulations to choose a best specification of the SAP. We also discuss various case studies including a Sentinel and FDA RWE demonstration project of targeted learning demonstrating SAP specification on real data.

Comments: About the keynote speaker:

Dr Mark van der Laan is the Jiann-Ping Hsu/Karl E. Peace Professor in Biostatistics and Statistics at the University of California, Berkeley. He is co-director of the Center of Targeted Machine Learning and Causal Inference at UC Berkeley. Mark research interests include censored data, causal inference, genomics, observational studies and adaptive designs. Mark has led the development of Targeted Learning, including Super Learning, Highly Adaptive Lasso, and Targeted maximum likelihood estimation (TMLE). Targeted Learning improves on typical current statistical practice by avoiding reliance on wrong model assumptions, and its capability to target any question of interest. In 2005 Mark was awarded the Committee of Presidents of Statistical Societies (COPSS) Presidential Award in recognition of outstanding contributions to the statistics profession. He also received the 2004 Spiegelman Award and 2005 van Dantzig Award. He is co-founder of the international Journal of Biostatistics and Journal of Causal Inference. Mark has authored various books on Targeted Learning, Censored Data and Multiple Testing, published over 400 publications, mentored 65 Ph.D students and 30 postdoctoral fellows.

Abstracts for presentations in parallel sessions

Alphabetical Order of Presenter's Last Name

Title: **Weighted Cumulative Exposure modelling of the cumulative effects of time-varying drug exposures**

Speaker: Michal Abrahamowicz, McGill University

Abstract: "Weighted Cumulative Exposure (WCE) methodology allows for flexible modelling of the cumulative effects of time-varying exposures (TVE) to medications [1]. The method uses splines to estimate the weight function that quantifies how the relative importance of past exposures varies depending on how long ago they occurred [2]. WCE was recognized as addressing one of the most important challenges related to assessing the TVE effects in pharmacoepidemiology [3], and a recent narrative review of studies assessing safety of different drugs reported that it systematically improved models' fit to data [4]. Originally developed for time-to-event (survival) analyses, WCE modelling has been extended to competing risks, marginal structural models and mixed effects linear modeling of longitudinal changes in a quantitative outcome. The talk will focus on the concepts and real-world applications. It will start with a brief overview of the WCE model and its estimation. Then, the insights offered by WCE analyses will be illustrated by presenting real-world studies associating: (i) glucocorticosteroids and serious infections in rheumatoid arthritis, and (ii) didanosine, an antiretroviral drug, and coronary heart disease risks in people living with HIV.

[1] Abrahamowicz. Pharmacoepi & Drug Safety (PDS). 2024, e5746.

[2] Sylvestre & Abrahamowicz. Statistics in Medicine. 2009, 3437-53.

[3] Pazzagli et al. PDS. 2018, 148-160.

[4] Kelly et al. PDS. 2024, e5701."

Comments:

Title: **AI for group sequential oncology design with multiple hypotheses**

Speaker: Keaven Anderson, MSD

Abstract: Confirmatory oncology trials often evaluate multiple hypotheses, such as progression-free and overall survival in biomarker-positive and all-randomized populations, while incorporating group sequential monitoring and graphical multiplicity control. This presentation describes how large language model (LLM) coding agents have accelerated development of open-source R software for this workflow, including gsDesign, gsDesign2, simtrial, gsDesignNB, and gsDesignBroom, together with graphicalMCP for group sequential graphical testing. Since December 2025, enhancement requests captured as GitHub issues have been translated into code, documentation, and unit tests using AI-assisted development workflows. We highlight recent capabilities relevant to oncology design, including harm bounds, selective efficacy and futility testing, flexible calendar- and event-driven timing, and sequential p-values for non-

proportional hazards designs. Using a four-hypothesis oncology design, we show how power calculations can support planning across all alpha levels reachable through the multiplicity graph. We conclude that robust package foundations are essential for reliable, reproducible AI-assisted statistical software development.

Comments: Co-authors: Nan Xiao, MSD; Yujie Zhao, MSD

Title: **Holistic Decision Making**

Speaker: Zoran Antonijevic, Bioforum, the Data Masters

Abstract: Decision-making at master protocol and portfolio levels is a very complex process. It involves simultaneous consideration of multiple products and/or multiple indications. It is further complicated by involvement of multiple stakeholders. Decision-making at this broadest level results in largest improvements in efficiency, because spending more resources on one drug impedes the development of another.

The decision-making process is further complicated by involvement of multiple stakeholders. This presentation will conceptually describe Holistic decision-making, which encompasses not only simultaneous consideration of all sponsor's programs, but also interdisciplinary and transdisciplinary input from multiple stakeholders; including regulators, insurers, and patients and their advocates. Specifically, the following will be addressed.

1. Optimal decision making for master protocols and portfolios
1.1. Overpowering
1.2. Type III Error
1.3. Knapsack problem
2. Optimal interdisciplinary decision making in presence of multiple stakeholders
2.1. Decision analysis from the perspective of single and multiple stakeholders
2.2. Quantitative evaluation of options when multiple stakeholders are involved

Comments: Co-authors:

Title: **Clinical trial evaluation of digital medical devices for clinical decision support : a simulation study**

Speaker: Baptiste Archambaud, Inserm, Université Paris Cité, Inria, UMR 1346, HeKA, Paris, France

Abstract: Digital medical devices (DMDs) can assist physicians in clinical decision-making by tailoring therapeutic recommendations to individual patient characteristics. While DMDs require validation through randomized clinical trials (RCTs) and real-world studies, their evaluation raises specific statistical challenges, such as non-concordance when physicians disregard DMD recommendations. The objective of this study is to build a simulation framework to calculate power of RCTs evaluating such DMDs.

We consider the context of a two-arm cluster RCT in a pathology with several available treatments. In control arm, physicians follow standard practices. In experimental arm, physicians have access to DMD recommendations to help treatment decisions. Our hierarchical simulation framework first defines true patient response model, based on treatment and

covariates effects, then DMD model as a biased estimation of this model, and eventually physician decision model based on DMD output and physician prior knowledge. We compar empirical power against expected power under full concordance, using binary response rate differences in intention-to-treat and per-protocol populations.

Results show that partial concordance introduces a bias that should be accounted for when calculating sample size, demonstrating the potential value of our simulation framework. This would prevent from overestimating power in superiority trials and underestimating type I error in non-inferiority trials.

Comments: Co-authors:
Sandrine Boulet, Inserm, Université Paris Cité, Inria, UMR 1346, HeKA, Paris, France; Moreno Ursino, Inserm, Université Paris Cité, Inria, UMR 1346, HeKA, Paris, France; Rodolphe Thiébaut Université de Bordeaux, Inserm, Bordeaux Population Health Research Center, U1219, Service d'Information médicale, CHU de Bordeaux, Inria, SISTM, Bordeaux, France; Sarah Zohar, Inserm, Université Paris Cité, Inria, UMR 1346, HeKA, Paris, France

Title: **Adaptive population selection designs for clinical trials with multiple endpoints**

Speaker: Koko Asakura, National Cerebral and Cardiovascular Center

Abstract: In some clinical trial settings, there may be uncertainty about which patient populations are most likely to benefit from a new treatment. Benefit may be expected in a prespecified subgroup, defined by demographic, biomarker, or pathophysiological features, while the effect in the complementary subgroup remains unclear.

We consider confirmatory trials that allow adaptive selection of patient populations based on prespecified subgroups while assessing multiple endpoints. Following an interim analysis, enrollment may continue overall or be restricted to the targeted subgroup, with inference in either or both populations. Valid inference requires combining pre- and post-interim data while accounting for multiplicity from endpoints, populations, and adaptive decisions.

Although the broader framework covers multiple endpoint structures, including co-primary endpoints, this presentation focuses on hierarchically ordered endpoints, where secondary claims depend on the success of corresponding primary hypotheses. This setting is challenging because population selection and endpoint hierarchy must be incorporated simultaneously into confirmatory testing.

We propose a framework based on the closure principle and a marginal p-value combination approach. The procedure reflects interim population selection in sequentially rejective tests for hierarchically ordered endpoints. We examine Type I error control and power through simulations and illustrate the approach using an example.

Comments: Co-authors:
Toshimitsu Hamasaki, The George Washington University Biostatistics Centerp; Frank Bretz, Novartis Pharma AG

Title: **Evaluating causal inference methods for continuous outcomes: a power and performance analysis in small samples**

Speaker: Adrián Aurensanz Crespo, Universidad de Zaragoza

Abstract: Propensity score matching (PSM) is widely used to avoid confounding in observational studies. PSM performance may be sensitive to limited overlap between treatment and control groups and, especially, to small sample sizes. Alternative causal inference methods, as inverse probability of treatment weighting (IPTW), G-computation (G-comp), and doubly robust estimators like targeted maximum likelihood estimation (TMLE), may provide improved efficiency under these challenging settings. Previous studies about binary outcomes found larger differences in the performance of those methods if limited sample sizes appear.

In this work, we compare the performance of PSM, IPTW, G-comp, and TMLE when the outcome is continuous, focusing on small-samples. Different simulation settings are generated by varying sample size, treatment prevalence, and outcome variability, allowing the assessment of the estimators under a broad range of realistic conditions. For each scenario, methods are used to estimate both the average treatment effect (ATE) and the average treatment effect on the treated (ATT), then bias, mean squared error and confidence interval coverage are evaluated. A specific power analysis is conducted to assess the ability of each causal inference approach to detect treatment effects. Results provide practical recommendations regarding the appropriate choice of causal inference strategies with continuous outcomes and limited sample sizes.

Comments: Co-authors:
Cristóbal Manuel， Rodríguez-Leal Universidad Complutense de Madrid; Rosario Susi，Universidad Complutense de Madrid; Teresa Pérez, Universidad Complutense de Madrid
Jesús Asín, Universidad de Zaragoza; Jorge Castillo Mateo, Universidad de Zaragoza

Title: **The aggregate safety assessment planning (ASAP) process for multidisciplinary safety assessment**

Speaker: Greg Ball，ASAPprocess

Abstract: The Safety Management Team – regularly reviewing aggregate safety data throughout the development program – is vital not only for early signal detection but also for generating a better understanding of the accumulating data and context needed for decreasing false alarms.

The Aggregate Safety Assessment Planning (ASAP) process empowers the SMT to plan and coordinate the evolving product level safety dataset and set of analyses needed to comprehensively characterize the safety profile of the product. The strategic planning of the ASAP process provides the Safety Topics of Interest (STOIs) and pooling structure. This safety strategy guides the technical implementation, which provides the overall safety assessment (for learning) and the assessment of STOIs (for decision making).

The technical implementation supports early and ongoing characterization of the product safety profile and helps to prepare for integrated assessments presented in safety-related reporting documents and responses to inquiries from health authorities – consistently communicating the safety story.

Comments: Co-authors:

Title: **Generalized Randomized Basket Trials**

Speaker: Robert Beckman，Georgetown University Medical Center

Abstract: Basket trials pool indications sharing molecular pathophysiology, improving development efficiency. Generalized application of basket trials in the confirmatory phase, where most development resources are expended, could have a marked impact on efficiency. To date, basket trials have been confirmatory only for exceptional therapies and in end-stage patients. Our previous randomized basket design may be generally suitable in the confirmatory phase, maintains high power even with modest effect sizes, and provides nearly k-fold increased efficiency for k indications, but controls false positives for the pooled result only. Since family-wise error rate by indications (FWER) may sometimes be required, we now simulate a variant of this basket design controlling FWER at 0.025k, the total FWER of k separate randomized trials. Moreover, a variation of this design has been developed for non-inferiority studies. We also report the simulation of a specific application in autoimmune diseases, based on both the literature and on real world data from off-label use in the Georgetown-Medstar database. Even under FWER control, randomized confirmatory basket trials substantially improve development efficiency, and RWD is helpful to inform design simulations. Regulatory considerations affecting these designs will also be discussed.

Comments: Co-authors:
Cong Chen, MSD, Inc.; Linchen He, Novartis Pharmaceuticals; Valeriy Korostyshevskiy, Georgetown University Medical Center; Daphne Guinn, Georgetown University Medical Center; Sathvik Narayana, Albany Medical College

Title: **Applications to Proof of Concept Trials and Biomarker-Drug Co-Development at the Portfolio Level**

Speaker: Robert Beckman, Georgetown University Medical Center

Abstract: Efficiency is broadly defined as the maximum net utility per resource unit invested, where resources can be study participants (especially in rare diseases) or financial resources, potentially adjusted for the value of time. Maximizing efficiency of a portfolio in the face of a fixed budget can lead to counterintuitive optima. In particular, the optimal program design may be expensive and take resources from other programs, decreasing portfolio efficiency. "Type III error" is the opportunity cost when a potentially successful program is unfunded due to overinvestment in others. This concept will be illustrated with surprising results concerning optimal proof of concept study design and Go criteria at the portfolio level.

Putative biomarkers, like drugs, are uncertain assets, and their development involves investment at risk. Drugs and biomarkers also affect each other's' value through multiple mechanisms, such as altering market size and level of benefit. Yet, at the portfolio level, the optimal choice might be to forego biomarker co-development for some assets to allow it for others while staying within a fixed budget. Most portfolio optimization approaches do not explicitly consider biomarkers as separate assets nor their interactions with drugs. The talk will present quantitative solutions for evaluating joint portfolios of biomarker and drugs.

Comments: Co-authors:

Zoran Antonijevic, The Bioforum Group; Patrick Roney, Georgetown University Medical Center; Cong Chen, MSD

Title: **Optimizing Allocation-Probability Tests for Clinical Trials with Normal Outcomes and Unknown Variances**

Speaker: Philipp Besendorfer, University of Regensburg

Abstract: Response-adaptive randomization (RAR) improves patient benefit by allocating more participants to the better arm, but may invalidate standard inferential procedures in finite samples. We study a simulation-calibrated allocation-probability (AP) test for two-arm Bayesian RAR trials with normal outcomes and unknown variances under conjugate normal inverse-χ^2 priors, comparing it with the Welch t-test and the randomization test, the only other established finite-sample procedure in this setting.

In a phase~2b diabetes trial, the AP test outperforms the randomization test in approximately 94\% of simulated configurations; well-calibrated AP designs substantially increase the expected proportion of patients on the best treatment arm (p^*) while maintaining competitive power. The randomization test is competitive mainly in degenerate designs or under severe prior misspecification; the Welch t-test achieves competitive power only at the cost of inflated type~I error.

While BRAR hyperparameters are typically chosen ad hoc, we frame their calibration as a constrained optimization problem: maximize p^* subject to 80\% power and a minimax type~I error constraint guaranteeing α is not exceeded within any plausible scenario. Using Hyperopt, we identify configurations allocating 24 percentage points more patients to the superior arm than equal randomization, with stronger type~I error guarantees than a naively calibrated AP design.

Comments: Co-authors:
Thomas Jaki, University of Cambridge, University of Regensburg; Sofía Villar, University of Cambridge

Title: **Principal Component Analysis to identify multivariate patterns of patient-reported symptom change during adjuvant chemotherapy in early breast cancer: data from GEICAM/2003 02 study**

Speaker: Andrea Blasco, GEICAM Spanish Breast Cancer Group; PhD Program in Medicine and Surgery, Doctoral School, Universidad Autónoma de Madrid

Abstract: Purpose: Understanding how multiple symptoms jointly impact functioning is essential for interpreting patient-reported outcomes in oncology. We aimed to identify the main latent dimensions underlying change-from-baseline (CFB) in symptom scores during adjuvant chemotherapy for early breast cancer, and to examine their associations with functional outcomes.

Patients and Methods: We analyzed EORTC QLQ-C30 CFB scores from 820 patients enrolled in the GEICAM/2003-02 trial, assigned to FAC (n = 430) and FAC plus paclitaxel (n = 390) regimens. Principal component analysis was used to summarize correlated symptom changes

into a smaller number of components. Associations between components and functional scales were assessed using Spearman's correlations.

Results: Three components were identified in the pooled sample, explaining 53.6% of the variance: a dominant general symptom burden, a gastrointestinal, and a diarrhea-specific component. The first component was defined by fatigue (factor loading (FL) = 0.77), appetite loss (FL = 0.62), pain (FL = 0.60), insomnia (FL = 0.57), nausea/vomiting (FL=0.54), and dyspnea (FL = 0.52). Similar patterns were observed across treatment arms. The first component in each arm showed negative correlations with all functional scales.

Conclusion: Symptom changes during chemotherapy followed a multivariate structured pattern with a general symptom-burden component mainly driven by fatigue, showing the strongest relationship with functional decline.

Comments: Co-authors:
Roberto Pastor, National Center for Epidemiology, Instituto de Salud Carlos III; Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP); Pilar Zamora-Auñón, GEICAM Spanish Breast Cancer Group; Hospital Universitario La Paz; PhD Program in Medicine and Surgery, Doctoral School, Universidad Autónoma de Madrid; Miguel Martín, GEICAM Spanish Breast Cancer Group; Instituto de Investigación Sanitaria Gregorio Marañón, Centro de Investigación Biomédica en Red de Oncología, CIBERONC-ISCIII; Hospital Clínico San Carlos; Virginia Lope, GEICAM Spanish Breast Cancer Group; National Center for Epidemiology, Instituto de Salud Carlos III; Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP)

Title: **Estimands and estimation in platform trials with time trends**

Speaker: Marta Bofill Roig, Universitat Politècnica de Catalunya

Abstract: Platform trials are multi-arm trials with arms entering or leaving over time, comparing experimental treatments against a common control. Since these trials often run for extended periods, time trends can bias treatment effect estimates and must be adjusted for. Covariate adjustment in clinical trials requires correctly specifying the estimand: conditional treatment effects (specific to covariate strata) versus marginal treatment effects (averaged over the population). For marginal estimands, beyond unadjusted analysis, estimators such as standardization (G-computation) and augmented inverse probability weighting (AIPW) are available. In platform trials, model-based approaches mitigate bias from time trends. However, these methods typically condition on time, yielding treatment effect estimates specific to particular calendar periods. Investigators, however, seek an unconditional treatment effect -one that does not depend on calendar time and reflects the effect in the overall trial population. This raises fundamental questions: what is the appropriate target estimand when combining data from multiple periods? Which estimator should be used? We examine conditional and marginal estimands in platform trials with time trends, clarifying target populations of interest. We evaluate model-based approaches, G-computation and AIPW, comparing bias and variance. We discuss how the choice of estimand, target population and trial data used for estimation affects the estimator performance.

Comments: Co-authors:
Ekkehard Glimm, Novartis Pharma AG, Basel; Kelly Van Lancker, Ghent University / Vrije Universiteit Brussel; Martin Posch, Medical University of Vienna

Title: **Coherent and Efficient Subgroup Effect Inference Using the Estimands Framework and Causal Machine Learning**

Speaker: Philippe Boileau, McGill University

Abstract: In addition to inferring the marginal effects of the primary, secondary, and exploratory endpoints, statistical analyses of clinical trials frequently estimate these endpoints' effects in relevant subgroups. Common examples of subgroups include those defined by trial participants' biological sex, age, disease severity, and disease subtype. While there is growing consensus that causal machine learning estimators, such as targeted maximum likelihood estimators, should be used to infer marginal effects from clinical trial data owing to their robust statistical properties and improved precision over traditional, unadjusted estimators, there has been relatively little focus on causal machine learning estimators tailored to subgroup effect inference. This may be due, in part, to the lack of guidance on subgroup effect estimands in the International Council for Harmonization's estimands framework. To address these gaps, we (1) propose a rigorous approach for incorporating subgroup effect inference into the estimands framework, (2) provide a general formula for deriving subgroup effect estimands and accompanying causal machine learning estimators from a large class of marginal estimands and their own estimators, and (3) establish that these subgroup effect estimators can pool information across subgroups, thereby producing efficiency gains. We validate these results empirically through several simulation experiments and the re-analysis of a recently completed tuberculosis trial.

Comments: Co-authors:

Title: **Assessment of global evidence against homogeneity for exhaustive subgroup treatment effect plots**

Speaker: Frank Bretz, Novartis

Abstract: Subgroup analyses present important challenges in clinical trials. In exploratory settings, estimation of treatment effects within subgroups is often unreliable because of limited sample sizes and multiplicity. To address this problem, Muysers et al. (2020) introduced a graphical display that presents numerous subgroups within a single figure and may help visualize homogeneity or heterogeneity of treatment effects. We refer to this display as the exhaustive subgroup treatment effect plot.

Because the original plot does not incorporate inferential statistics, there remains a need for a global assessment of whether the observed heterogeneity is compatible with global homogeneity or exceeds what would be expected under that assumption. In this presentation, we introduce a computationally efficient method for constructing the exhaustive subgroup treatment effect plot and deriving confidence regions that control the family-wise error rate at a pre-specified significance level. These regions can also be used to perform a global interaction test. The proposed methodology is based on the doubly robust learner approach of Kennedy (2023) and uses high-dimensional integration to compute quantiles of the distribution of the maximum statistic, thereby obtaining the rejection region directly. We also present a

comprehensive simulation study to evaluate the validity of the approach and to benchmark it against alternative methods.

Comments: Co-authors:
Bjoern Bornkamp, Novartis; Jiarui Liu, Vertex

Title: **Regulatory Communication for Complex Trial Designs: Aligning Expectations Across Stakeholders – Industry Perspective (Panel Position Abstract)**

Speaker: Frank Bretz, Novartis

Abstract: Adaptive and other complex trial designs create communication challenges that extend beyond the validity of the testing procedure. Using estimation after adaptive testing as a motivating example, this contribution discusses how bias, variance, confidence intervals, and multiplicity affect the reliability and interpretation of treatment effect estimates, particularly when several endpoints, adaptation rules, or sequential testing procedures are involved. Drawing on ICH E20 as a common reference point, the discussion broadens from methodological considerations to the role of estimation in regulatory benefit–risk assessment and communication with clinicians, patients, health technology assessment bodies, and other stakeholders. Particular attention is given to situations in which desirable statistical properties cannot all be achieved simultaneously, raising questions about what constitutes a 'reliable' estimate and how methodological rigor, interpretability, transparency, and ethical considerations should be balanced in regulatory communication.

Comments: Panelists: Miya Saigusa, Pharmaceuticals and Medical Devices Agency; Florian Klinglmüller, Austrian Agency for Health and Food Safety; Frank Bretz, Novartis; Hiromi Sugano, Pharmaceuticals and Medical Devices Agency

Title: **A Decision-Theoretic Interpretation of Frequentist Error Rates in Rare Disease Trials**

Speaker: Nico Bruder, Institute of Medical Statistics, Medical University of Vienna

Abstract: Background: In rare disease trials, recruiting sufficient patients to demonstrate treatment effects is often impossible. Adjusting significance levels to improve power is frequently proposed, yet these adjustments are often not justified based on clinical trade-offs.

Methods: We propose a Bayesian decision-theoretic framework to find the implicit losses for Type I and Type II errors implied by specific frequentist designs. By assuming a frequentist boundary is the optimal solution to a Bayesian decision problem, we derive the underlying ratio of Type I to Type II error losses. We evaluate both single-stage and adaptive designs.

Results: We demonstrate in several examples that standard significance thresholds, such as one-sided $\alpha=0.025$, imply that Type I errors are penalized more heavily than Type II errors. Such a weighting may be clinically inappropriate for severe rare diseases, where only few patients are treated, and there is a high unmet medical need.

Conclusion: This framework makes explicit the decision-theoretic implications of frequentist testing procedures planned with conventional error rates. The error rates are translated into clinical trade-offs, which can support the transparent planning of rare disease trials.

Comments: Co-authors: Franz König, Institute of Medical Statistics, Medical University of Vienna; Martin Posch, Institute of Medical Statistics, Medical University of Vienna

Title: **Bayesian meta-analytic methods for leveraging external data in surrogate endpoint evaluation**

Speaker: Sylwia Bujkiewicz, University of Leicester

Abstract: Surrogate endpoints play an important role in drug development, particularly when they can be measured early compared to the final clinical outcome. When drugs receive regulatory approval based on effectiveness on a surrogate endpoint, the evidence base for health technology assessment (HTA) is limited in terms of effectiveness on patient relevant clinical outcomes.

HTA agencies require surrogate endpoints to be validated, as good predictors of clinical benefit. At the same time, the scope of the HTA submission tends to be limited to a narrow population; for example, of patients with a specific biomarker status. This often results in insufficient data for surrogate endpoint evaluation.

Bayesian statistics is a flexible tool for modelling and leveraging data from external sources. We extend Bayesian bivariate meta-analytic methods for surrogate endpoint evaluation to leverage data through hierarchical models and informative prior distributions to improve precision of surrogate relationship parameters. We will present a selection of methods allowing for synthesis of data from opposite biomarker subgroups or multiple indications. The analyses will be illustrated in examples in advanced colorectal cancer, with KRAS (Kirsten Rat Sarcoma viral oncogene homolog) or dMMR (deficient mismatch repair) biomarker groups and non-small cell lung cancer with EGFR (Epidermal Growth Factor Receptor) biomarker.

Comments: Co-authors: Ireoluwa Joshua Adeleke, University of Leicester; Lorna Wheaton, University of Leicester

Title: **When is Bayesian borrowing (not) needed?**

Speaker: Carl-Fredrik Burman, AstraZeneca

Abstract: Rare and ultra-rare diseases often make adequately powered confirmatory randomised clinical trials (RCTs) difficult or impossible. This raises the question of when Bayesian borrowing should be considered to complement a new Phase 3 trial.

This presentation discusses a practical framework for that decision. We first consider when a conventional RCT is likely to have inadequate power, even after improving efficiency through better endpoint selection, covariate adjustment, longitudinal analyses, and careful trial design. This is based on an empirical assessment of largest feasible sample size, as well as on a discussion on the choice of alternative hypothesis. We then place Bayesian borrowing in the broader set of options available in rare diseases, alongside alternative endpoints, cross-over designs, and possible adaptations of evidentiary standards.

The presentation also gives an accessible overview of Bayesian borrowing methods and discusses how suitable external data sources may be identified. The aim is to clarify when borrowing is justified, when it is unlikely to help enough, and which broader design considerations should guide evidence generation in rare diseases.

Comments: Co-authors:

Title: **Quantifying the impact of interim decisions in group sequential trials**

Speaker: Gianmarco Caruso, MRC Biostatistics Unit, University of Cambridge

Abstract: Reliable estimation of treatment effects is central to clinical research, as it shapes how trial results are interpreted and how confidently clinicians, regulators and HTA bodies assess new treatments. Group sequential designs are increasingly used, allowing interim analyses and early stopping when the accumulating data suggest that continuing is unnecessary or even unsafe. While this flexibility can shorten trial duration and reduce patient exposure to ineffective treatments, it may introduce bias in the final treatment effect estimates. Using a group sequential trial example, we show that ignoring the influence of interim decisions can result in overly optimistic conclusions.

We propose a new metric to quantify the cumulative impact of repeated interim decisions in group sequential trials. This metric summarises their effect across the whole parameter space, rather than at individual treatment effect values. By comparing alternative decision boundaries and prior choices, we show how the proposed metric supports the interpretation of trial results and informs future study design.

We also introduce a pre-trial version of the metric to support trial design choices, including interim schedules and stopping boundaries. This complements traditional strategies based on type-I error control by providing insight into how interim decisions affect treatment effect estimates at each interim analysis. We illustrate its use in a group sequential trial for nervous system disorders.

Comments: Co-authors: William Rosenberger, George Mason University; Pavel Mozgunov, MRC Biostatistics Unit, University of Cambridge; Nancy Flournoy, University of Missouri

Title: **Patient Reported Outcomes Mythbusters and other considerations**

Speaker: Yovanna Castro, Patient-Centered Outcomes in Novartis

Abstract: The presentation examined and discussed three myths taken from written regulatory feedback.

1. "PRO results are derived from an open-label study, and are subject to bias"

The validity of PRO results from open label cancer randomized clinical trials have been repeatedly demonstrated. Efficace et al. 2022 did not observe statistically significant association between treatment concealment (blinded vs open-label) with the proportion of trials favoring the experimental treatment.

2. "PROs are subjective"

FDA defines a PRO as a report coming directly from the patient, which provides a "holistic interpretation" of treatment benefits, making them objective measures of patient experience. Who is in a better position to report about "your" pain?

3. "The study is not specifically powered to assess the non-inferiority of the PROs"

When do we power the study for all the claims included in the label? Why are PROs judged differently compared to safety endpoints?

If the PRO is valid and reliable, provides clinically meaningful information regarding patient relevant concepts, and if the PRO compliance is high, there is no reason to believe there is bias.

Including at least some PRO endpoints in the testing hierarchy, may facilitate the inclusion of PRO results in the label.

Comments: Co-authors:

Title: **Open Source Drug Discovery in the AI Era**

Speaker: Jake Chen, University of Alabama at Birmingham

Abstract: Drug discovery remains slow, costly, and failure-prone, even as biomedical data and AI methods rapidly expand. In this talk, I will discuss how "open drug discovery" can be reimagined for the AI era by combining open evidence, AI-ready biomedical data, knowledge graphs, virtual-cell and digital-twin models, and human-supervised AI agents. Rather than presenting AI as a black-box shortcut from prompt to drug, I will focus on a more rigorous open discovery loop: reading the literature, integrating disease mechanisms, prioritizing therapeutic targets, generating candidate molecules or biologics, and identifying the next experiments needed for validation. I will illustrate this approach with case studies from the Systems Pharmacology AI Research Center, including AI-assisted pancreatic cancer target prioritization and a PHGDH allosteric modulator project for Alzheimer's disease that moved rapidly from a newly published biological insight toward candidate discovery and intellectual-property action. The broader vision is Open Source Drug Discovery 2.0: a model in which open evidence and open tooling accelerate early discovery, while pragmatic IP-gating and translational partnerships help promising candidates move toward real-world therapeutic development.

Comments: Co-authors:

Title: **Statistical Leadership in a Transforming Biopharma Landscape**

Speaker: Industry leadership panel discussion

Abstract: The biopharmaceutical industry is undergoing rapid and profound transformation, driven by an evolving regulatory environment, the growing use of real-world evidence (RWE), and the disruptive rise of artificial intelligence (AI) and machine learning across the drug development continuum. Against this backdrop, the demands placed on statistical leaders have never been greater — nor the opportunities more compelling.

This panel brings together senior statistical leaders from across the biopharmaceutical industry for a candid and forward-looking discussion on the most pressing challenges and exciting opportunities facing the profession today. Key topics to be explored include:

- The evolving regulatory landscape and its implications for statistical practice and submissions;

- The expanding role of real-world evidence (RWE) in evidence generation and regulatory decision-making;
- The strategic integration of AI and automation into biostatistics workflows; and
- The changing expectations of statistical leaders as scientific partners, strategic advisors, and organizational influencers.

Beyond technical considerations, the panel will also address the human dimensions of statistical leadership: how to attract, develop, and retain top quantitative talent in a highly competitive environment; how to foster a culture of scientific rigor, innovation, and psychological safety; and what it takes to build and sustain high-performing statistical teams capable of delivering impact at scale. Panelists will further reflect on the evolving identity and visibility of the statistics function within organizations and the key priorities for the next generation of statistical leaders.

Comments: Industry senior leadership panelists:
Keaven Anderson, MSD
Ivan Chan, Bristol Meyers Squibb
Margaret Gamalo, Pfizer
Elaine Hoffman, CSL

Title: **Beyond Go/No-Go: An Adaptive Recruitment Strategy for Basket Trials**

Speaker: Libby Daniells, MRC Biostatistics Unit, University of Cambridge

Abstract: Basket trials enable the evaluation of a single treatment across multiple patient populations that share a common characteristic, e.g. a genetic mutation or symptomatic response. Interim analyses in basket trials are challenging due to the small sample sizes typically encountered in early-phase studies. Motivated by the ongoing CRUK Phase 2b HTL0039732 basket trial, we developed an adaptive interim analysis framework that prioritises recruitment in the most promising baskets.

In addition to typical futility and efficacy analyses, the proposed design introduces a 'pause' decision for baskets demonstrating moderate efficacy. Decisions regarding continuation of paused baskets are informed by responses observed in other baskets, allowing resources to be focused on indications showing the greatest promise. We evaluate this design using simulation studies and compare independent analyses with a dynamic Bayesian borrowing approach.

Simulation results demonstrate that the proposed design efficiently prioritises recruitment while maintaining acceptable operating characteristics. Furthermore, the Bayesian borrowing model improves power relative to independent analyses by up to 7% when treatment effects are similar across baskets, with controlled family-wise error rate at 15% and modest increase in the type I error rate. These findings support the use of adaptive basket trial designs combined with dynamic Bayesian borrowing to improve the efficiency of early-phase basket trials.

Comments: Co-authors: Pavel Mozgunov, MRC Biostatistics Unit, University of Cambridge; Thomas Jaki, University of Regensburg

Title: **Impact of Missing Data Handling on Response Adaptive Randomisation in Rare Disease Trials**

Speaker: Rajenki Das, MRC Biostatistics Unit, University of Cambridge

Abstract: Response-adaptive randomisation (RAR) is increasingly used in rare disease trials as it allocates more patients to the better-performing treatment based on accumulating responses. Popular RAR algorithms include Thompson sampling, which maximises the probability of allocating the best treatment, and Trippa's algorithm[1,2], which incorporates additional constraints such as protecting the control arm. However, in rare diseases, the handling of missing data can substantially influence allocation probabilities thereby the resulting allocation ratios.

In this work, we 1) demonstrate the effects of missing-data handling in rare-disease RAR trials, 2) propose a modification of Thompson sampling that incorporates missingness into the Bayesian updating process, and 3) modify Trippa's algorithm in a similar manner. The methods are evaluated through simulation studies across a range of missing-data scenarios and sample sizes.

The results demonstrate the sensitivity of allocation probabilities to missing-data handling. The proposed methods address this issue in small samples, producing more stable allocations while converging to standard approaches as sample size increases. We show that appropriate handling of missing outcomes can mitigate extreme allocations and improve the robustness of RAR designs in rare disease settings.

[1] Trippa L, et al. J Clin Oncol. 2012;30:3258.

[2] Wason JMS, Trippa L. Stat Med. 2014;33:2206–2221.

Comments: Co-authors:

Title: **Not all Hierarchical Multi-Component Strategies are Created Equal: GPC, DOOR, and MOST**

Speaker: Mickael De Backer, UCB

Abstract: In clinical research, composite endpoints are increasingly used to capture multidimensional treatment effects, yet conventional analyses often rely on simplistic aggregations that obscure clinically meaningful priorities. Hierarchical multi-component strategies offer a principled alternative by incorporating clinical importance directly into the analysis and interpretation of outcomes.

This talk explores the landscape of three prominent approaches: Generalized Pairwise Comparisons (GPC), Desirability of Outcome Ranking (DOOR), and the emerging MOST framework. Through practical examples, we offer a comparative perspective on these methods, examining both their underlying philosophies, e.g. how they operationalize prioritization and shape interpretation, and their statistical properties, including the handling of data complexities and opportunities for covariate adjustment.

By navigating the similarities and differences among GPC, DOOR, and MOST, this presentation aims to provide guidance for statisticians and to stimulate further research in the rapidly evolving field of hierarchical multi-component methodologies, ultimately supporting

more statistically-sensitive and clinically-meaningful evidence generation in modern drug development.

Comments: Co-authors:

Title: **Assessing Surrogacy for a Binary True Endpoint and a Continuous Surrogate in Longitudinal Data Using Joint Modeling**

Speaker: Gökçe Deliorman, Complutense University of Madrid

Abstract: Surrogate endpoints are frequently used in clinical trials as substitutes for final clinical outcomes, since they can provide earlier evidence on treatment effects while reducing follow-up time, measurement burden, and costs. Their evaluation becomes considerably more challenging in longitudinal settings. Within the information-theoretic causal-inference (ITCI) framework, surrogacy is quantified by the Individual Causal Association (ICA), defined as the association between the individual causal treatment effects on the surrogate and on the true endpoint. In this work, we extend this framework to longitudinal data with a continuous surrogate and a binary true endpoint. We estimate the model in a full Bayesian framework using Markov chain Monte Carlo methods implemented in R via MCMCglmm. The surrogate endpoint is modeled using a Gaussian linear mixed effects model, whereas the binary endpoint is modeled using a probit link function based on an underlying latent continuous response. In the joint model, the association between the two endpoints is captured through shared subject-specific random effects. The proposed methodology is applied to randomized longitudinal data from patients with schizophrenia. Our motivating research question is whether the Positive and Negative Syndrome Scale (PANSS) can serve as a potential surrogate endpoint for the Clinical Global Impression (CGI) outcome.

Comments: Co-authors: Maria del Carmen Pardo, Complutense University of Madrid; Ariel Alonso, KU Leuven

Title: **General targeted machine learning for modern causal mediation analysis**

Speaker: Iván Díaz, NYU Grossman School of Medicine

Abstract: Causal mediation analyses investigate the mechanisms through which causes exert their effects, and are central to scientific progress. The literature on non-parametric definition and identification of mediational effects in rigorous causal models has grown significantly, with important progress addressing challenges in interpretation and identification of such effects. Despite great progress on the causal inference front, statistical methodology for non-parametric estimation has lagged behind, with few or no methods available for tackling estimation in the presence of multiple, continuous, or high-dimensional mediators. We show that identification formulas for six popular non-parametric approaches to mediation analysis proposed in recent years can be recovered from just two statistical estimands. We leverage this to propose an all-purpose one-step estimation algorithm that can be coupled with machine learning in any mediation study using any of these six definitions. The estimators have desirable properties, such as root-n-convergence and asymptotic normality. Estimating the first-order correction requires estimation of complex density ratios on potentially high-dimensional mediators, a

challenge solved using recent advancements in Riesz learning. We illustrate our methods in a simulation study and apply them to real data to estimate the extent to which pain management practices mediate the total effect of chronic pain disorder on opioid use disorder.

Comments: Co-authors:

Title: **External Control Data Augmentation for Randomized Controlled Recent-Onset Type 1 Diabetes Immunotherapy Trials**

Speaker: Aidong Ding, Northeastern University

Abstract: TrialNet established a uniquely consistent and high-quality clinical trial infrastructure for type 1 diabetes (T1D) clinical trials over years. Recent-onset T1D trials in TrialNet generally follow standard randomized controlled trial (RCT) designs, and plays a critical role in understanding progression and treatment for T1D. We propose to leveraging external control data from previous T1D trials, through propensity score matching, to enhance new studies. We found that leveraging external information enhances covariates similarity between treatment and control arms and enables reduction in required trial sample size while maintaining statistical power, thereby improving trial efficiency and lower study cost. The results support usage of hybrid clinical trial for future studies of T1D immunotherapy.

Comments: Co-authors: Hanzhi Gao, University of South Florida; Jeffrey Krischer, University of South Florida; Samuel Wu, University of South Florida

Title: **Oral microbial signatures identified in head and neck cancers can guide targeted therapies to mitigate oral mucositis**

Speaker: Kim-Anh Do, University of Texas M.D. Anderson Cancer Center

Abstract: Oral mucositis is a painful complication common in head andneck cancer care. Shifts in the oral microbiome are implicated in its pathogenesis, positioning microbial signatures as therapeutic targets. We aimed to (i) characterise longitudinal microbial patterns aligned with mucositis severity, (ii) identify clinically meaningful patient clusters by severity trajectory, and (iii) discover cluster-specific microbial signatures.

Methods: Weekly symptom ratings and buccal swabs (16S rRNA) were collected longitudinally from 140 head and neck cancer patients. A oral mucositis symptom composite score was derived by applying non-negative sparse principal component analysis to all symptoms. Functional data analysis captured individual severity curves, and hierarchical clustering grouped patients by trajectory. Across clusters, Prevotella spp. correlated positively with the OM symptom composite score; Alloprevotella was significantly enriched only in the rapid-progression cluster. Genera linked to oral health, including Haemophilus, Rothia, and Actinomyces, were negatively correlated with severity.

Conclusions: Distinct mucositis trajectories are underpinned by specific microbial and demographic profiles. The persistent association of Prevotella with higher scores suggests it reflects overall disease burden rather than cluster identity. These findings suppt ormicrobiome-targeted strategies.

Comments: Co-authors: Cielito Reyes-Gibby, Univeristy of Texas MD Anderson Cancer Center; Kim-Anh LeCao, University of Melbourne; Saritha Kodikara, University of Melbourne; Jiadong Mao, University of Melbourne; Erin Marie San Valentin, Case Western University

Title: **Comparing Type I Error Measures in Hybrid Clinical Trials**

Speaker: Teresa Engelbrecht, Medical University of Vienna

Abstract: Hybrid clinical trial designs incorporating external control data alongside randomized concurrent controls are increasingly used, especially in rare disease settings with limited patient availability. These settings are often analyzed using Bayesian approaches, which have received growing attention, as reflected by the recent draft guidance of the U.S. Food and Drug Administration on Bayesian methodology in clinical trials. The guidance distinguishes between designs that can be calibrated to control the type I error rate and those where this may not be feasible, emphasizing the need to evaluate alternative operating characteristics.

When borrowing external control information, challenges arise due to potential bias and non-exchangeability. Strict control of the conditional type I error rate may effectively preclude meaningful borrowing, as discussed by Kopp-Schneider et al. (2020). To address this, alternative definitions based on averaging conditional type I error rates have been proposed.

However, these approaches differ in the underlying averaging measure (e.g., over external data, latent parameters, or design priors), leading to distinct notions of error control. In this work, we compare these alternatives within a hybrid design framework. Through simulations and theoretical considerations, we investigate the relationships between these error definitions and discuss their implications for the design, calibration, and interpretation of hybrid clinical trials.

Comments: Co-authors:

Title: **Simulation Studies in Regulatory Science**

Speaker: Tobias Fellinger, AGES, Austrian Agency for Health and Food Safety

Abstract: Simulation studies play different roles in regulatory work. On the one hand, simulation studies have to be assessed if applicants use them to demonstrate properties of the methods or study designs they submit in regulatory proceedings. On the other hand, regulators use simulation studies to get insights about the properties of methods in developing regulatory standpoints, that may also be reflected in regulatory guidance.

Our consortium is currently involved in an EMA funded regulatory science research project which investigates the application of causal inference methods to handle intercurrent events in randomized clinical trials. Central to the project is a large simulation study. This presentation will cover lessons learned from the implementation of this simulation study: from development of the simulation study protocol to programming, running the simulations and reporting the results. Regarding disseminating the findings from simulation studies in the regulatory network, industry and the scientific community we draw additional experience from an already concluded simulation study on methods for time-to-event data under non-proportional hazards.

Finally, I will shortly discuss the development of methodology of and requirements for simulation studies to inform regulatory decisions or broader methodological guidance.

Comments: Co-authors: Florian Klinglmuller, AGES

Title: **Novel methods to address measurement error in safety and effectiveness studies of prescription-based time-varying exposures**

Speaker: Steve Ferreira Guerra, Université de Montréal

Abstract: Prescription databases are increasingly used to study medication safety and effectiveness in real-world settings. However, prescription-based exposure measures are prone to substantial misclassification because dispensed medications do not necessarily reflect actual drug consumption. Existing methods to correct for such measurement error generally require external information on the misclassification process, which is rarely available in pharmacoepidemiologic studies.

We developed pragmatic extensions of the SIMEX methodological framework to correct for measurement error in prescription-based time-varying exposures. The proposed approaches leverage information already available in the observed data, thereby avoiding the need for external validation data. Methods were evaluated using extensive plasmode simulations based on real-world prescription data under varying levels of exposure measurement error, treatment trajectories, and outcome associations.

Our proposed SIMEX corrected estimators substantially reduced bias and improved estimation accuracy compared to naïve estimates while remaining computationally feasible in complex longitudinal settings.

The proposed approach provides a practical strategy for addressing exposure misclassification in studies relying on routinely collected prescription data while extending measurement error methodology to complex real-world settings where conventional approaches fail.

Comments: Co-authors: Robert Platt, McGill University; Michal Abrahamowicz, McGill University

Title: **Bayesian meta-analysis of continuous outcomes for postoperative pain: Should treatment effects be modelled as additive or multiplicative?**

Speaker: Suzanne Freeman, University of Leicester

Abstract: Meta-analysis of continuous outcomes is commonly conducted on the additive scale and reported as mean differences (MD). However, an alternative option is to use a multiplicative scale, with treatment effects reported as ratio of means (RoM). We conducted an empirical study of previous meta-analyses of postoperative pain interventions to compare additive and multiplicative models.

We searched the Cochrane Database of Systematic Reviews for intervention reviews in the topic area of 'pain & anaesthesia', with further filtering to 'acute pain' and 'regional anaesthesia'. Eligible reviews included ≥3 trials and reported ≥1 MD model. We fitted fixed- and random-effects Bayesian pairwise meta-analysis models, with and without adjusting for pain scores on

the control arm. We compared model fit using the deviance information criteria (DIC, lower DIC preferred).

47 reviews were eligible for inclusion. Switching from MD to RoM reduced DIC in 30 (64%) and 26 (55%) reviews for fixed- and random-effects models, respectively. Adjusting for pain scores on the control arm in random-effects RoM and MD models reduced DIC in 17 (36%) and 13 (28%) reviews, respectively.

Multiplicative models imply constant relative effects but absolute treatment effects that depend on initial severity of pain. Nevertheless, multiplicative models improved model fit, reduced heterogeneity and should be routinely considered when synthesising continuous postoperative pain outcome measures.

Comments: Co-authors: Nicola Cooper, University of Leicester; Brett Doleman, University of Nottingham; Nicky Welton, University of Bristol; Alex Sutton, University of Leicester

Title: **Bridging the Gap: Utilising Clinical Trial Simulations as a Tool for Regulatory Dialogue and Funding of Platform Trials**

Speaker: Michaela Maria Freitag, Institute of Biometry and Clinical Epidemiology, Charité—Universitätsmedizin Berlin, Berlin, Germany

Abstract: Platform trials and other master protocol designs offer substantial gains in efficiency and flexibility but also introduce methodological complexities that can challenge regulatory assessment and funding decisions. Features such as non-concurrent controls, adaptive allocation ratios, and shared control arms may be difficult to evaluate and communicate to non-statistical stakeholders.

Clinical trial simulations can support transparent communication operating characteristics of complex platform trial designs. Using examples from the EU-PEARLDIVER platform trial in Major Depressive Disorder (MDD), we demonstrate how simulations informed early dialogue with the EMA Innovation Task Force by assessing design feasibility, temporal drift, and global Type I error control under adaptive settings and non-binding futility rules.

Beyond regulatory interactions, simulations also enabled a quantitative assessment of operational efficiency, including improved power-per-patient ratios and reduced development timelines. These analyses contributed to establishing an evidence base supporting the platform approach that supported discussions with academic clinicians and researchers, industry, funders including the Wellcome Trust, and patient representatives. Finally, publicly accessible simulation code strengthened transparency, reproducibility, and scientific credibility.

Comments: Co-authors: Stefan Gold, Charité - Universitätsmedizin Berlin, Berlin, Germany; Franz König, Medical University of Vienna, Vienna, Austria

Title: **Efficiency and interpretation of marginal versus conditional covariate adjustment for the Cox model**

Speaker: Laetitia Fritz, Medical University of Vienna, Center for Medical Data Science, Institute of Medical Statistics

Abstract: Generalized covariate-adjustment in randomized trials is an active research field of increasing regulatory interest, aiming to increase the precision of a marginal treatment effect estimator by explaining covariate-dependent variability. Recent publications show that marginal covariate adjustment for the Cox model hazard ratio may increase power compared to unadjusted analyses. However, the choice between marginal adjustment or a usual conditionally adjusted Cox model is not obvious due to non-collapsibility and the proportional hazards (PH) assumption. We considered survival endpoints under stratified randomization with event-driven design, possible interim analyses and hazard functions depending on stratification factors and covariates. Scenarios under covariate-conditional PH, non-proportional hazards (NPH) and marginally proportional hazards (MPH) were included. While the shape of hazard functions matching MPH in presence of covariate effects is not obvious, an according scenario was constructed by iteratively adjusting the hazard functions. Operating characteristics of unadjusted, conditionally and marginally adjusted Cox models and logrank tests were investigated by simulation. We further propose an estimator for the covariance of group sequential statistics based on stacked estimating equations. Our results favor conditional adjustment when (M)PH is approximately met. Under NPH, marginal adjustment increased power, however, other estimands may deserve future attention.

Comments: Co-authors: Robin Ristl, Medical University of Vienna, Center for Medical Data Science, Institute of Medical Statistics

Title: **Practical implementation of graphical multiple testing procedures in group-sequential designs**

Speaker: Michael Grayling, Johnson & Johnson

Abstract: Group-sequential designs (GSDs) combined with graphical multiple comparison procedures (MCPs) are increasingly used in late-phase oncology to support interim decision-making while controlling the familywise error-rate across multiple endpoints/populations. While the underlying methodology is well established, its practical implementation in complex trials remains non-trivial. Here, we focus on implementation considerations arising in real-world applications, including selection of alpha flows, choosing hypothesis-specific spending functions, leveraging delayed alpha recycling, and making event-driven adaptations. Our emphasis is on translating these design features into robust, reproducible, and auditable statistical specifications. To this end, we discuss appendMCP, an R package that provides a modular, configuration-driven framework for specifying, validating, and documenting graphical MCPs within GSD studies. Trial designs are decomposed into structured components—analyses, hypotheses, enrollment, outcome distributions, and the graphical testing plan—enabling standardized computations, simulation-based evaluation of operating characteristics, and SAP-ready outputs. Finally, we discuss AI-enabled layers built on top of this standardized infrastructure, including conversational and graphical interfaces that facilitate design exploration and documentation while preserving deterministic back-end validation.

Comments: Co-authors: Yevgen Tymofyeyev, Johnson & Johnson

Title: **The crucial role of Targeted Machine Learning to increase power in RCTs with small sample size**

Speaker: Susan Gruber, Targeted ML Solutions

Abstract: Little knowledge of natural history of disease, heterogeneity in disease presentation and response to treatment, difficulty in enrolling and retaining patients in randomized controlled trials pose challenges to evaluating treatment effects in patients with a rare disease. Targeted learning provides a roadmap for addressing these challenges, and sophisticated methodologies for estimating causal effects of treatment even when there is informative loss to follow-up. Simulation studies based on actual patient data will compare performance of the MMRM method often used with longitudinal data and targeted machine learning using targeted maximum likelihood estimation and super learning (TMLE+SL).

Comments:

Title: **TBEP-v2: Integrating Knowledge Graphs and Explainable AI for Traceable Biomedical Hypothesis Generation**

Speaker: Shivansh Gupta, Indian Institute of Information Technology Allahabad, UP, INDIA

Abstract: AI has emerged as a powerful tool for biomedical discovery; however, limited explainability often prevents researchers from validating and trusting AI-generated hypotheses. Evidence traceability remains a critical challenge for AI-assisted drug discovery. To address this limitation, TBEP-v2 extends the previously published Target and Biomarker Exploration Portal (TBEP) from protein interaction network visualization to a knowledge graph-driven framework for explainable biomedical AI. The platform integrates heterogeneous biological entities and relationships within a unified knowledge graph and combines interactive visualization with conversational AI workflows.

LLMs are integrated into TBEP-v2 to support natural-language query making it easier to explore biological knowledge. By linking LLM-generated insights to supporting evidence in the knowledge graph, TBEP-v2 helps users understand, review and validate the information behind their findings. User queries are translated into biologically relevant subgraphs, enabling AI models to operate on focused contexts. A central objective of TBEP-v2 is to connect AI-generated observations with supporting graph relationships and biological context, allowing researchers to visually inspect and validate findings. By integrating graph analytics, interactive visualization, and evidence-grounded reasoning, TBEP-v2 aims to support explainable, reproducible, and auditable hypothesis-generation workflows for target and biomarker discovery.

Comments: Co-authors (up to 5): Urvija Roy Chowdhury, EECS Department, University of Missouri, Columbia, MO, USA; Misti Garg, Indian Institute of Information Technology, Allahabad, UP, INDIA; Kyan Mahajan, Indian Institute of Information Technology, Allahabad, UP, INDIA; Abdul Azeem, Indian Institute of Information Technology, Allahabad, UP, INDIA; Sankalp Joshi, Indian Institute of Information Technology, Allahabad, UP, INDIA

Title: **Great Wall: A Generalized Dose Optimization Design for Drug Combination Trials Maximizing Survival Benefit**

Speaker: Yan Han, Indiana University School of Medicine

Abstract: Most phase I–II drug-combination trial designs assume that selecting the optimal dose combination based on early outcomes will also lead to maximum long-term survival benefits. However, this assumption is often violated in many clinical studies, generally due to high rates of relapse following the initial response. To address this problem, we propose the Great Wall design, a general dose optimization design for drug-combination trials. The Great Wall design employs a "divide-and-conquer" algorithm to address the issue of partial order of toxicity and uses early outcomes to eliminate dose combinations that are excessively toxic or less efficacious. It utilizes a dose randomization approach to construct a candidate set of the promising dose combinations balancing the toxicity and early efficacy outcomes. The patients assigned to the candidate set are followed to collect the survival outcomes and the final optimal dose combination is then selected to maximize the survival benefit.

Comments: Co-authors:

Title: **Treatment discontinuation and while-on-treatment estimation**

Speaker: Niels Hendrickx, Uppsala University

Abstract: As highlighted in ICH E9 R1 Addendum, preserving a causal interpretation of treatment effects requires a strategy for handling intercurrent events (ICEs) and the statistical analysis in randomized clinical trials. In this work, we consider trials in chronic diseases where Major Cardiac Adverse Event (MACE) endpoint needs to be considered, with discontinuation. An estimand with a while-on-treatment strategy is considered.

We simulated randomized chronic-disease trials with a treatment-related safety endpoint, assuming that treatment effects may persist after discontinuation during a predefined buffer window where adverse events are still counted. We varied the true and assumed buffer duration, event and discontinuation hazards, withdrawal probabilities and the dependence between discontinuation, withdrawal, and adverse-event risk. Treatment effects were estimated using standard Cox models, with and without inverse probability weighting to account for reduced observation after study withdrawal. Operating characteristics were compared across correctly specified and misspecified scenarios.

The Cox model was generally more powerful and produced the narrowest confidence intervals but was biased when the treatment effect differed during the post-discontinuation buffer period. The IPW model reduced this bias when the weighting model was correctly specified, particularly under informative withdrawal. It was, however, less powerful due to increased uncertainty (difference of 1-20%).

Comments: Co-authors:

Title: **Regulatory Update on Platform Trials**

Speaker: Benjamin Hofner, Paul-Ehrlich-Institut

Abstract: In this talk, we will provide an update of the current status of the EMA Reflection Paper on Confirmatory Platform Trials. We will given an overview of the current content and general directions of the Reflection Paper and specifically consider the impact of platform trial elements on estimands and estimated treatment effects.

Comments:

Title: **One-Class Support Vector Machines Integrated Bayesian Approaches for External Control Borrowing in Clinical Trials**

Speaker: Yi Huang, University of Maryland Baltimore County

Abstract: The integration of real-world evidence into regulatory approval processes for treatments has expanded globally, with an increasing emphasis on synthetic control methods. These methods enhance the assessment of treatment efficacy and safety in clinical trials, particularly randomized trials with limited concurrent controls. Currently, there is no superior synthetic control method capable of optimizing resource use with multiple and heterogeneous external datasets that mirror real-world conditions. Common approaches include propensity score (PS) methods for adjusting covariate balance, Bayesian approaches for controlling information borrowing based on outcomes, and various two-stage methods. We introduce a novel machine learning application: one-class support vector machines (OCSVM) to exclude external data with non-matching covariates, offering a potential alternative to PS methods. The comprehensive simulation study demonstrates that OCSVM integrated with Bayesian methods generally surpasses the performance of traditional PS methods integrated with Bayesian approaches. However, OCSVM may encounter challenges in other scenarios. This is a joint work with my student Ji Li, and my colleague Dr. Lilly Yue and her research group.

Comments:

Title: **Evaluating the Robustness of Firth-Type Logistic Regression Methods for Rare Events Under Model Misspecification and Complex Predictor Structures**

Speaker: MD Fakrul Islam, Department of Statistics, George Mason University

Abstract: Firth-type logistic regression and its variants, FLIC and FLAC, have been proposed to improve estimation and prediction in rare-event settings. However, existing evaluations have largely been conducted under ideal conditions with correctly specified models and simple predictor structures. This study extends the original simulation framework to assess the performance of maximum likelihood (ML), Firth penalization (FL), FLIC, FLAC, and ridge regression (RR) under more realistic conditions. Two challenging scenarios are considered: (i) model misspecification, where the true model contains nonlinear and interaction effects that are ignored in the fitted models, and (ii) complex predictor structures with correlated and noisy

covariates. The binary outcomes are simulated from logistic regression models with 20 predictors, event rates 0.02 and 0.05, and sample sizes 500 and 1500. Each of the 12 simulation scenarios is evaluated with 1000 replications. The performance is measured in terms of the prediction RMSE, bias and calibration measures. Results show that FL, FLIC and FLAC are more robust to model misspecification, with smaller increases in RMSE compared to ridge regression. Conversely, ridge regression works best in predicting outcomes when the predictors are complex. Calibration analyses reveal a trade-off, with ML and ridge regression yielding more stable calibration than Firth-type methods. None of the methods is uniformly better in general.

Comments: Co-authors: David Kepplinger, Department of Statistics, George Mason University

Title: **Causal inference methods for estimating treatment effects in the presence of rescue medication and treatment**

Speaker: Louise Jespersen, Medical University of Vienna

Abstract: The ICH E9(R1) Addendum emphasises precise specification of the clinical question of interest in randomised clinical trials. This work evaluates causal inference methods for handling intercurrent events(ICEs) and compares them with standard methods.

A diabetes trial setting with treatment discontinuation and rescue medication as ICEs was considered. Treatment policy and hypothetical strategies were investigated. Inverse probability weighting(IPW), de-mediation(DM), and the g-formula were compared with mixed models for repeated measures(MMRM) and multiple imputation(MI) in a simulation study.

Under the treatment policy strategy all methods produced approximately unbiased and similarly precise estimates under moderate missingness. Differences were more pronounced for the hypothetical estimand, particularly under high missingness or increased rescue medication use. MMRM generally showed lower bias and greater robustness. Type I error control was adequate under the treatment policy estimand. For the hypothetical estimand, DM and the g-formula showed slight inflation, only IPW maintained adequate control under increased rescue probability. Power was quite similar, although MMRM achieved the highest power in most scenarios.

Under favourable conditions, methods performed similarly. For the hypothetical estimand, MMRM provided the best balance of bias, precision, and power, whereas IPW offered better type I error control under increased rescue medication use.

Comments: Co-authors: Paula Starke, University Medical Center Göttingen; Susanne Urach, Austrian Agency for Health and Food Safety

Title: **Flexible and Robust Bayesian Information Borrowing for Clinical Trials with Longitudinal Outcomes**

Speaker: David Jesse, F. Hoffmann-La Roche AG

Abstract: The recent FDA draft guidance on the Use of Bayesian Methodology in Clinical Trials of Drug and Biological Products highlights the potential of Bayesian methods to improve the efficiency of randomized clinical trials (RCTs) through borrowing external information, particularly

relevant for trials in rare diseases and/or pediatric indications. Beyond the use of external data, there are several other options to increase the efficiency of trials. For instance, whenever possible, it can be advantageous to use a continuous endpoint (as opposed to a derived dichotomous one) and to leverage longitudinal data. It therefore seems natural to merge these ideas, yet only few of such approaches were considered so far. In this presentation, I will present ongoing work on developing a flexible, robust, and interpretable method for borrowing longitudinal historical control data in RCTs. The method is based on a Bayesian hierarchical model using natural cubic splines to model the longitudinal trajectories and robust meta-analytic predictive (MAP) priors to incorporate historical data on the control arm.

Comments: Co-authors: Diane Uschner, F. Hoffmann-La Roche AG; Markus Elze, F. Hoffmann-La Roche AG; Tim Friede, Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany; Francois Mercier, F. Hoffmann-La Roche AG

Title: **Beyond Means and Dispersions: Mixed-Effects Location-Scale Modeling for Cluster-Randomized Trials**

Speaker: Depeng Jiang, University of Manitoba

Abstract: Background: In pharmaceutical statistics, cluster-randomized trials and real-world data are increasingly pivotal for evidence generation. Traditional multilevel models (MLMs) emphasize mean changes and often treat dispersion as noise. Yet variability in both mean and dispersion can reveal differential treatment efficacy, potential safety signals, and inequities in care - information that is central to regulatory science and evidence-informed decision making.

Objectives: (i) through simulations, compare mixed-effects location-scale models (MELSM) with standard MLMs, focusing on the consequences of omitting a scale (variance) component; and (ii) apply MELSM to a cluster-based randomized controlled trial to quantify between-cluster heterogeneity, and test moderators (e.g., sex, socioeconomic status) of mean and variance effects.

Methods: We will simulate data under varying heteroscedasticity to evaluate bias, coverage, and power of MELSM versus conventional MLMs. The empirical analysis will fit MELSM and MLMs to a cluster-randomized trial (a) estimate mean and variance effects, (b) assess between-cluster variance, and (c) test moderation for both components.

Significance: By extending evaluation beyond mean outcomes to dispersion, this work informs regulatory risk assessment, labeling considerations, and equity implications in biopharmaceutical decision-making. The integration of MELSMs for trial and real-world data exemplifies statistical innovation aligned with evidence.

Comments: Co-authors:

Title: **Bayesian Sample Size Determination for Analysis of Covariance in Randomised Controlled Trials Incorporating Historical Data**

Speaker: David Jones, University of Bath

Abstract: Sample size determination is a fundamental step in designing clinical studies. In randomised controlled trials, regulatory guidelines endorse adjusting for baseline covariates through analysis of covariance. These calculations are traditionally developed from a frequentist perspective to maintain type I error and statistical power. However, these rely on an initial guess, often based on historical information. In this work, we develop a fully Bayesian approach that formally incorporates historical data into the study design and derive sample size formulae resembling frequentist formulations.

Consider settings with a continuous prognostic baseline covariate and normally distributed outcome, with correlation p. We represent historical data within a prior for the bivariate normal mean for each treatment group, obtaining a conditional posterior distribution for the treatment effect given the covariate for decision-making. For two-arm trials, these formulae ensure a probability of correctly deciding whether a new treatment is superior, or not, to the control by some clinically relevant difference. For a known covariance matrix, the formula shows a reduction by a factor of (1 - ϱ2), compared to standard approaches. For unknown covariance matrices, an Inverse-Wishart prior captures historical information on standard deviations and p. Simulations show that incorporating baseline covariates can achieve significant sample size savings while maintaining efficacy/futility thresholds.

Comments: Co-authors: Haiyan Zheng, University of Bath

Title: **Development and validation of a hierarchical composite endpoint for chronic kidney disease**

Speaker: Niels Jongs, University medical centre Groningen

Abstract: Background and Aims: A CKD progression hierarchical composite endpoint (HCE) was recently introduced but requires broader validation. Using the CKD-EPI consortium trial database, we assessed trial-level associations between treatment effects on the HCE and on true and established kidney endpoints.

Methods: We analysed individual patient data from 160,826 participants in 63 randomized trials. The HCE ranks six components by clinical severity: all-cause mortality, end-stage kidney disease, eGFR <15, 57% and 40% eGFR decline, and annual eGFR slope. The true endpoint was dialysis initiation or eGFR <15; the established endpoint additionally included 57% eGFR decline. Win odds (WO) for the HCE and hazard ratios (HR) were estimated per trial, with a bootstrap procedure (1000 replications) for within-trial variances and covariances. Bayesian meta-regression assessed trial-level associations.

Results: Intercept estimates (posterior median, 95% CrI) were −0.10 (−0.17, −0.04) and −0.02 (−0.08, 0.05) for the true and established endpoints, confirming a null HR corresponds to a null WO. Between-trial variation in WO was largely explained by both endpoints (R^2 88% [48%, 100%] and 87% [64%, 99%]), with RMSEs of 0.06 (0.01, 0.13) and 0.10 (0.02, 0.18).

Conclusions: Treatment effects on the CKD progression HCE are strongly associated with effects on established kidney endpoints, supporting its validity as a surrogate for CKD disease progression.

Comments: Co-authors:

Title: **Conditional Unbiased Estimation of the Selected Treatment Effects in a Three-Stage Drop-the-Losers Trials with Early Futility**

Speaker: Yogesh Katariya, Indian Institute of Technology Kanpur

Abstract: In multi-arm clinical trials with several experimental treatments and a common control, a primary objective is to identify the most effective treatment and estimate its effect relative to the control. We consider a setting with multiple experimental treatments under a three-stage drop-the-losers design with early stopping for futility, assuming normally distributed responses with unknown means and a common known variance. The design comprises Stage I screening with selection of a subset of promising treatments and a futility assessment. If the trial continues and only one treatment is retained, it proceeds directly to the final stage (Stage III); otherwise, the retained treatments enter an intermediate stage (Stage II), where the best treatment is selected based on cumulative data from Stages I and II. The selected treatment then advances to Stage III alongside the control for confirmatory sampling. We focus on estimating the selected treatment mean, the control mean, and their difference, conditional on continuation beyond the interim futility assessment. To address bias induced by adaptive selection and continuation, we derive uniformly minimum variance conditionally unbiased estimators (UMVCUEs) under a refined conditioning framework that accounts for the Stage I selected subset and the final selected treatment index. Simulation studies demonstrate reduced bias and improved or comparable mean squared error relative to standard estimators.

Comments: Co-authors: Neeraj Misra, Indian Institute of Technology Kanpur

Title: **Application of spatio temporal modeling in long term phenotypic screening of patient derived organoids**

Speaker: Maryam Kazemi Naeini, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK

Abstract: Patient derived organoids (PDOs) are powerful for preclinical discovery, but long term screens are confounded by measurement noise, within well spatial heterogeneity, and space–time correlation. Microcavity plates enable long term, single organoid tracking at scale, yet nonrandom positions within microcavities create microenvironmental hotspots that well level averaging cannot correct, inflating uncertainty and skewing hit calls.

We applied Bayesian generalized additive models (GAMs) to an internal oncology study combining a selective epigenetic modulator with a targeted pathway inhibitor, tracking treatment induced phenotypic change in colorectal cancer PDOs over 4–5 weeks. Using within well coordinates, we modeled image derived viability scores and organoid area readouts at the organoid level, fitting Bayesian GAMs to smooth and remove spatial and temporal artifacts. Integrating all time points in a single hierarchical analysis yielded artifact corrected plate maps, spatially adjusted trends, and calibrated uncertainty, enabling hit calls that reflect biology rather than plate position or time drift.

Bayesian GAMs separate biology from spatial and temporal artifacts, improving hit call accuracy and treatment interpretability. Using all time points in one spatio temporal analysis boosts power and data sustainability.

Comments: Co-authors: Joff Jones, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK; Yaxin Xue, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK; Ioritz Sorzabal Bellido, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK; Simon Vyse, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK; Yinhai Wang, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK

Title: **Bayesian dose–response mixed modeling: application to perturbation screening data**

Speaker: Maryam Kazemi Naeini, Discovery Sciences, R&D, AstraZeneca

Abstract: Dose–response (DR) analyses struggle to jointly model fixed effects (conditions/perturbations) and random effects (plate, batch, run, donor), especially when DR data are combined with CRISPR screening. We use an interpretable four-parameter logistic (4PL) model familiar to biologists and extend it to capture biological signal and technical noise, incorporating gene knockout (gKO) effects directly within the full DR curve.

We analyzed raw CRISPR DR data with a Bayesian mixed effects 4PL model. gKO effects are estimated on the full curve (bottom, top, EC50, slope), and plate within biological replicate is modeled as nested random intercepts, correcting technical offsets in-curve and propagating uncertainty. Each gKO has one curve with partial pooling across replicates, and contrasts for EC50, bottom, and top versus neutral controls are reported with 95% credible intervals and posterior hit probabilities. The joint model recovered hits missed under early viability loss and reduced false positives by shrinking isolated high-dose wells toward curve-supported values with appropriately wider intervals.

A Bayesian mixed-effects DR model recovers actives from incomplete or early signal-loss curves and yields more robust, less outlier-driven calls. It adjusts technical noise within the curve and enables hit decisions with transparent, uncertainty-aware summaries.

Comments: Co-authors: Joff Jones, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK; Maximilian Wechsung, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK; Salome Adam, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK; Elissavet Kentepozidou, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK; Bogumil Kaczkowski, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK

Title: **Estimating principal-stratum vaccine efficacy under incomplete regimen completion**

Speaker: Florian Klinglmueller, Austrian Agency for Health and Food Safety

Abstract: Incomplete adherence to multi-dose vaccination regimens may complicate estimation of vaccine efficacy in randomised trials. Principal stratum estimands provide a causal framework to evaluate efficacy among participants who would complete the vaccination schedule irrespective of treatment assignment, aligning with the notion of a biologic effect.

We developed a simulation framework for a randomised, placebo-controlled preventive vaccine trial with a two-dose prime–boost schedule and a binary infection endpoint. The data-generating mechanism includes treatment-dependent compliance, prognostic and predictive baseline markers (acting both on compliance and infection risk), vaccine efficacy, and infection incidence. Investigated methods include instrumental variable estimation, principal score weighting, and a conventional per-protocol analysis as comparator.

We assess bias, mean squared error, confidence interval coverage, type I error, and power across a wide range of parameter settings and different assumptions about observation status of baseline markers.

The per-protocol estimator was robust across settings, with moderate bias and near-nominal coverage, whereas the causal inference methods were less consistent, offering modest gains in some scenarios but performing substantially worse in others.

Comments: Co-authors:

Title: **European Regulatory Perspective (Panel Position Abstract)**

Speaker: Florian Klinglmueller, Austrian Agency for Health and Food Safety

Abstract: This contribution forms part of a panel discussion and reflects the European regulatory perspective on the assessment and communication of complex clinical trial designs. It will highlight practical regulatory experience and recurring challenges observed in interactions with sponsors.

From the European viewpoint, the discussion will focus on areas where interpretation and communication frequently require careful alignment. These include the assessment of results from adaptive and platform trials, particularly where design features complicate the interpretation of evidence; the handling of multiplicity and the specification and use of estimands; and the translation of complex findings into clear and meaningful regulatory conclusions. These aspects are not only technical but also closely linked to how evidence is communicated between stakeholders and ultimately conveyed in product information.

Recent European and international guidance reflects increasing regulatory attention to such designs and provides a framework for their evaluation. However, experience shows that applying these principles in practice still involves important judgment, especially when balancing statistical rigor, interpretability, and feasibility. In this context, the communication of results—both during regulatory interactions and in the final product information—remains a key challenge.

Comments: Co-authors:

Title: **Traditional Validation Pathways of a Biomarker and Examples of AI Created Efficiencies**

Speaker: Hope Knuckles, Abbott

Abstract: Traditional validation of an in vitro diagnostic (IVD) assay or biomarker requires a high degree of analytical and clinical rigor to ensure that the assay is reliable, reproducible, and fit for its intended clinical use. This process typically includes comprehensive analytical validation—such as assessments of accuracy, precision, sensitivity, specificity, linearity, limits of detection and quantitation, and robustness—followed by clinical validation to demonstrate clinical performance and utility in the defined patient population. Regulatory oversight of these activities is governed by established frameworks, including the In Vitro Diagnostic Regulation (IVDR) in the European Union and regulatory pathways in the United States such as 510(k),

Premarket Approval (PMA), and, for certain claims or biologic-linked diagnostics, Biologics License Applications (BLA). Typical study designs used to support these submissions—including method comparison studies, reproducibility studies, and clinical performance studies—will be discussed to provide the audience with a clear understanding of the foundational principles underlying traditional biomarker development and approval. In addition, emerging applications of artificial intelligence (AI) offer meaningful opportunities to improve the efficiency and effectiveness of biomarker development and validation. Examples of use of AI will be provided.

Comments: Co-authors:

Title: **Simultaneous testing of hypotheses and alternatives: theory and applications to problems of network analysis.**

Speaker: Petr Koldanov, NRU Higher School of Economics

Abstract: The hypotheses testing theory allows obtaining statistically significant conclusions only in the case of rejecting the null hypotheses. If it is desirable to obtain statistically significant conclusions about the truth of the null hypotheses, it is natural to proceed to testing alternative hypotheses. We propose to test null hypotheses and it's alternatives simultaneously constructing procedures that control the probability of at least one false rejection of true null hypothesis or true alternative. In this case the set of obtained conclusions can be divided into three parts: the set of statistically significant conclusions about the truth of the null hypotheses, the set of statistically significant conclusions about the falsehood of the null hypotheses, and the uncertainty zone (the set of insignificant conclusions), which corresponds to the set of conclusions where both null hypotheses and it's alternatives are simultaneously accepted. The quality of such procedures could be measured by the size of the uncertainty zone.

Such approach was proposed in [1] and developed in [2], [3], [4], [5], [6].

This paper briefly discusses a procedure for extracting statistically significant inferences when identifying threshold similarity graphs simultaneously at multiple thresholds. The specifics of such procedures are investigated.

Comments: Co-authors: Alexander Koldanov, NRU Higher School of Economics

Title: **Type 1 error rate and other metrics for randomized controlled trials borrowing external data**

Speaker: Annette Kopp-Schneider, German Cancer Research Center (DKFZ)

Abstract: In the context of clinical trials for rare diseases, the sample sizes of patients to be recruited often present a challenge. However, experience may exist with standard-of-care treatments to which novel therapies are compared. Consequently, external information borrowing can be considered to improve a trial's efficiency. The Bayesian paradigm is ideally suited for this aim, since external information can be incorporated into an informative prior for the analysis. A potential issue with this procedure is that external and current information may conflict. To protect against the harmful impact of potentially conflicting external information, the amount

of transferred information can be dynamically discounted based on the discrepancy between external and current data.

Type I error rate and power of trial designs are crucial for understanding how trial data support correct conclusions. It will be shown that, in general, no power gains are possible if strict control of the frequentist type I error rate is required, and this is also true for any robust dynamic method. However, the frequentist approach may be unrealistically stringent, and the Bayesian framework may be more appropriate, particularly in the rare disease setting, since benefits can be obtained if external information is reliable and consistent. In this talk, frequentist and Bayesian operating characteristics of randomized clinical trial designs with borrowing from external control data will be discussed.

Comments: Co-authors: Silvia Calderazzo, German Cancer Research Center (DKFZ); Vivienn Weru, German Cancer Research Center (DKFZ); Manuel Wiesenfarth, Cogitars GmbH

Title: **Modeling approaches to incorporate non-concurrent controls in platform trials with time-to-event endpoints**

Speaker: Pavla Krotka, Universitat Politècnica de Catalunya - BarcelonaTech

Abstract: Platform trials speed up drug development by testing multiple treatments in parallel and allowing new arms to enter over time, often using a shared control arm. For later-entering arms, control data include concurrent controls, enrolled while the arm is active, and non-concurrent controls (NCC), enrolled earlier. Using NCC can improve efficiency, but may bias treatment effect estimators when time trends exist.

For continuous endpoints, regression models can include NCC by adjusting for periods as factors, where periods are defined by treatment entry and exit times. Under equal and additive time trends, these models give unbiased estimates and asymptotically control type I error. Comparable methods for survival endpoints are less developed.

We extend the models for using NCC to time-to-event outcomes. Analogous to models for normal endpoints, Cox proportional hazards (PH) models with period effects could be considered. However, these models require the PH assumption to hold for hazards conditional on period and, in addition, equal conditional hazard ratios across periods for each treatment. These assumptions may be difficult to satisfy under time trends. As an alternative, we examine accelerated failure time (AFT) models with period adjustment. AFT models target a different estimand and yield collapsible effect measures. We compare both approaches, outline conditions for valid inference, and assess type I error, power, and robustness to assumption violations in simulations.

Comments: Co-authors: Martin Posch, Medical University of Vienna; Marta Bofill Roig, Universitat Politècnica de Catalunya - BarcelonaTech

Title: **A coherent assessment of safety through a contemporary application of the Bradford-Hill criteria**

Speaker: Jürgen Kübler, QSciCon - Quantitative Scientific Consulting

Abstract: In drug development, safety assessments must integrate data from heterogeneous sources including clinical trials, non-clinical toxicology, mechanistic evidence, and real-world evidence. At any given timepoint—such as protocol development, dose escalation, interim reviews, or regulatory safety reports—data may be incomplete, or contextually ambiguous. Teams often struggle to summarize the totality of evidence for a given safety topic in a way that is both scientifically rigorous and operationally consistent.

A structured, reusable framework based on the Bradford Hill Criteria (BHC) enables critical appraisal and consistent communication of causality reasoning.

This presentation presents the basic principles of the development of this framework and explores its utility over the product development lifecyle.

Comments: Co-authors:

Title: **Non-Linear Power for Cox Models: Design-by-Shape Sample Size Planning for Spline Effects**

Speaker: Stefania Lando, University of Padua

Abstract: Restricted cubic splines (RCSs) enable flexible modelling of non-linear covariate effects in time-to-event outcomes, but sample size planning often still relies on one-degree-of-freedom linear-effect formulas. This creates a design-analysis mismatch when the final Cox model uses a multi-parameter spline representation.

We propose Non-Linear Power (NlP), a sample size framework for Cox models with RCSs. NlP translates clinically interpretable hazard ratio trajectories into spline coefficients and estimates the events required for either the global spline association or the test of departure from linearity. The method accommodates uncertainty in the true effect shape, letting investigators incorporate plausible clinical assumptions directly into the design.

To support robust evidence generation, we incorporate the Riley et al. sample size framework based on the total number of predictor parameters, quantifying the additional stability requirement incurred when moving from a linear effect to an RCS representation. Simulations show that linear design calculations achieve nominal power when the final analysis is linear, but lose power when the final analysis uses RCS, particularly for non-linearity testing. In a Rotterdam breast cancer illustration, stability rather than power was the binding constraint.

NlP provides an approach for planning spline-based Cox analyses in biomedical studies where flexible modelling, interpretability, and regulatory credibility must be balanced.

Comments: Co-authors: Ester Rosa, University of Padua; Gloria Brigiari, University of Padua; Dario Gregori, University of Padua

Title: **When should the next clinical trial be conducted? A Bayesian borrowing approach for AI-based digital medical devices**

Speaker: Caroline Lawless, Inserm, Université Paris Cité, Inria, UMR 1356, HeKA, Paris, France

Abstract: AI-based digital medical devices (DMDs) evolve rapidly through software updates, challenging their clinical evaluation. While initial trials may establish safety and efficacy, these results may not generalize to later versions. Repeating randomized trials for each update is often infeasible.

We propose a Bayesian sequential framework to evaluate evolving DMDs by borrowing information across versions. Building on the temporal effects adjusted (TEA) prior, we introduce a fully data-driven extension, the adaptive TEA (A-TEA) model, in which the degree of borrowing is learned from the data. Posterior inference is performed using a sequential Monte Carlo algorithm, enabling efficient updating as new data accrue. The framework provides a novel Bayesian decision criterion to detect clinically meaningful changes in efficacy and formally determine when a new clinical trial is warranted.

Simulation studies show control of false positive decisions while improving efficiency. The method maintains power to detect clinically relevant changes while avoiding unnecessary re-evaluation when versions remain similar. Compared with non-borrowing approaches, it detects changes earlier and supports more timely decision-making.

Comments: Co-authors: Sandrine Boulet, Inserm, Université Paris Cité, Inria, UMR 1356, HeKA, Paris, France; Rodolphe Thiebaut, Université de Bordeaux, Inserm, Bordeaux Population Health Research Center, U1219, Bordeaux, France; Moreno Ursino, Inserm, Université Paris Cité, Inria, UMR 1356, HeKA, Paris, France; Sarah Zohar, Inserm, Université Paris Cité, Inria, UMR 1356, HeKA, Paris, France

Title: **TMLE for Regulatory Purposes: A Landscape Overview**

Speaker: Hana Lee, The US Food and Drug Administration

Abstract: Targeted Maximum Likelihood Estimation, also referred to as Targeted Minimum Loss Estimation (TMLE), is a statistical method that enables causal inference while incorporating flexible, advanced machine learning techniques. Within the drug development space, both drug sponsors and regulatory authorities have leveraged TMLE for a range of regulatory purposes. The goal of this presentation is to provide a holistic overview of TMLE use cases as they have been applied in submissions to the Food and Drug Administration (FDA) Center for Drug Evaluation and Research (CDER) and in CDER's research activities. This presentation will provide (1) an overview of regulatory submissions in which TMLE has been considered to support regulatory decision-making, and (2) a summary of FDA CDER's regulatory science initiatives that advance analytic capabilities and support the use of TMLE.

Comments:

Title: **Sample size reestimation in two-arm trials using doubly robust estimators for marginal treatment effects**

Speaker: Maria Lee Alcober, Universitat Politècnica de Catalunya - BarcelonaTech

Abstract: Clinical trials collect extensive baseline covariate data. Covariate adjustment methods can improve efficiency in estimating treatment effects, increasing power or reducing required sample size. Doubly robust estimators are useful for covariate-adjusted analyses, combining

outcome and treatment modeling while retaining consistency if at least one model is correctly specified.

Covariate-adjusted sample size calculations often involve nuisance parameters, such as covariate or censoring distributions, unknown at the design stage. To overcome this, conservative sample size estimates are used, resulting in a reduction in efficiency gains.

In this talk, assuming a two-arm trial targeting a marginal estimand in terms of RMST difference, we propose an adaptive design for survival trials addressing this challenge. The trial is planned using an unadjusted marginal sample size calculation, avoiding unknown nuisance information. At interim, nuisance parameters are estimated from blinded data and used to update an adjusted marginal sample size calculation based on doubly robust estimation. This allows the design to benefit from efficiency gains of using doubly robust estimators.

Through simulation studies, we evaluate the proposal across scenarios varying in covariate strength and prevalence, censoring proportion, and cutoff times. The study examines when adaptation improves power while preserving type I error control, and evaluates whether an unblinded approach may be a viable option.

Comments: Co-authors: Marta Bofill Roig, Universitat Politècnica de Catalunya – BarcelonaTech; Werner Brannath, University of Bremen

Title: **Statistical approaches and tools for identification of interpretable predictive subgroups with applications to consistency evaluation in multi-regional clinical trials (MRCTs)**

Speaker: Larry Leon, Statistical Methodology Research, MSD

Abstract: Multi-regional clinical trials (MRCTs) frequently face challenges demonstrating regional consistency of treatment effects, particularly when heterogeneous responses raise concerns about generalizability to specific subpopulations such as a specific region. We propose a principled framework for subgroup identification that directly addresses these regional consistency challenges through a dual Weibull AFT/Cox modeling structure — simulating data on the accelerated failure time scale while enabling causal interpretation metrics. The method evaluates three complementary estimands — the average hazard ratio (AHR), controlled direct effect (CDE), and marginal hazard ratio — to robustly characterize treatment effect heterogeneity. A train/test paradigm using non-regional data for discovery and regional data for confirmation guards against overfitting, while data-driven simulation calibrated to real trial data evaluates the full operating characteristics of the procedure, including false discovery rate, estimation bias, and confidence interval coverage. Simulation results demonstrate that our proposals can recover regional consistency under heterogeneous treatment effects while controlling false discovery under the null, offering a robust analytic strategy for navigating regional consistency assessments in modern MRCTs.

Comments:

Title: **Leveraging External Controls in Clinical Trials: Estimands, Estimation, Assumptions**

Speaker: Fan Li, Duke University

Abstract: It is increasingly common to augment randomized controlled trial with external controls from observational data, to evaluate the treatment effect of an intervention. Traditional approaches to treatment effect estimation involve ambiguous estimands and unrealistic or strong assumptions, such as mean exchangeability. We introduce a double-indexed notation for potential outcomes to define causal estimands transparently and clarify distinct sources of implicit bias. We show that the concurrent control arm is critical in assessing the plausibility of assumptions and providing unbiased causal estimation. We derive a consistent and locally efficient estimator for a class of weighted average treatment effect estimands that combines concurrent and external data without assuming mean exchangeability. This estimator incorporates an estimate of the systematic difference in outcomes between the concurrent and external units, of which we propose a Frish-Waugh-Lovell style partial regression method to obtain. We compare the proposed methods with existing methods using extensive simulation and applied to cardiovascular clinical trials.

Comments: Co-author(s): Laine Thomas, Duke University

Title: **An extended Bayesian semi-mechanistic dose-finding design for phase I oncology trials using pharmacokinetic and pharmacodynamic information**

Speaker: Yisheng Li, MD Anderson Cancer Center

Abstract: We propose a model-based, semi-mechanistic dose-finding (SDF) design for phase I oncology trials that incorporates pharmacokinetic/pharmacodynamic (PK/PD) information when modeling the dose-toxicity relationship. This design is motivated by a phase Ib/II clinical trial of anti-CD20/CD3 T cell therapy in non-Hodgkin lymphoma patients; it extends a recently proposed SDF model framework by incorporating measurements of a PD biomarker relevant to the primary dose-limiting toxicity (DLT). We propose joint Bayesian modeling of the PK, PD, and DLT outcomes. Our extensive simulation studies show that on average the proposed design outperforms some common phase I trial designs, including modified toxicity probability interval (mTPI) and Bayesian optimal interval (BOIN) designs, the continual reassessment method (CRM), as well as an SDF design assuming a latent PD biomarker (SDF-woPD), in terms of the percentage of correct selection of maximum tolerated dose (MTD) and average number of patients allocated to MTD, under a variety of dose-toxicity scenarios. When the working PK model and the class of link function between the cumulative PD effect and DLT probability is correctly specified, the proposed design also yields better estimated dose-toxicity curves than CRM and SDF-woPD. Our sensitivity analyses suggest that the design performance is reasonably robust to prior specification for the parameter in the link function, as well as misspecification of the PK model and class of the link function.

Comments: Co-author: Chao YangYishen, Eli Lilly

Title: **Semiparametric Regression Analysis of Interval-Censored Data**

Speaker: Danyu Lin, University of North Carolina at Chapel Hill

Abstract: Interval censoring arises frequently in medical, financial, and sociological studies, where the event or failure of interest is not observed at an exact time point but is rather known to occur within a time interval induced by periodic monitoring. We formulate the effects of potentially time-dependent covariates on the interval-censored failure time through semiparametric regression models, such as the Cox proportional hazards model. We study nonparametric maximum likelihood estimation with an arbitrary number of monitoring times for each study subject. We develop an EM algorithm that involves very simple calculations and converges stably for any dataset, even in the presence of time-dependent covariates. We show that the estimators for the regression parameters are consistent, asymptotically normal, and asymptotically efficient with an easily estimated covariance matrix. In addition, we extend the numerical algorithm and asymptotic theory to competing risks and multivariate failure time data. Finally, we demonstrate the desirable performance of the proposed numerical and inferential procedures through simulation studies and applications to real medical studies.

Comments:

Title: **Inference for ML-Driven Heterogeneous Treatment Effects in Randomized Experiments**

Speaker: Michael Lingzhi Li, Harvard Business School

Abstract: Randomized experiments provide credible evidence on average treatment effects, but many scientific and operational questions require understanding how treatment effects vary across individuals. This talk presents a line of work on inference for machine learning driven heterogeneous treatment effects in randomized experiments. The central challenge is that flexible algorithms can uncover rich treatment effect heterogeneity, but the resulting estimates, subgroups, and individualized treatment rules are themselves learned from data, so conventional inferential procedures can be misleading when this adaptivity is ignored. The talk develops design based approaches that combine causal machine learning with Neyman style randomization inference to validate discovered heterogeneity, quantify uncertainty for algorithm selected subgroups, and evaluate individualized treatment rules. These methods retain the credibility of randomized experiments while allowing researchers to use modern prediction tools to move beyond average effects. Applications in healthcare and operations illustrate how valid inference for heterogeneous effects can support personalized, policy relevant, and operationally actionable decision making.

Comments: Co-author(s): Kosuke Imai, Harvard University

Title: **Overview of modern approaches for evaluating treatment effect heterogeneity from clinical and observational data**

Speaker: Ilya Lipkovich, Retiree

Abstract: Modern approaches for the analysis of treatment effect heterogeneity (TEH) have emerged through the cross-pollination of causal inference, multiple hypothesis testing, and machine learning, resulting in powerful tools while also presenting methodological and implementation

challenges. These methods enable researchers to identify predictive biomarkers, subgroups with enhanced treatment effects, and optimal treatment regimens, as well as to conduct statistical inference for associated data-driven hypotheses.

This presentation, based on a recently published tutorial in Statistics in Medicine, will provide an overview of the typical objectives of TEH analyses and recommendations for a principled process for planning, conducting, and interpreting data-driven analyses. It will then introduce the main classes of methods for TEH evaluation, including traditional approaches for indirect estimation of conditional average treatment effects (CATE), direct estimation of CATE using meta-learners, individualized treatment regimes/rules (ITRs), and direct subgroup search and identification. We will review their key features, advantages, and limitations, as well as approaches to statistical inference for TEH. The use of selected methods will be illustrated through a case study analyzed using publicly available R packages.

Comments: Co-author: Alex Dmitrienko, Mediana.

Title: **A few cautionary notes for leveraging external data in clinical trials**

Speaker: Bo Lu, The Ohio State University

Abstract: The use of external control (EC) data to supplement clinical trials has become increasingly relevant in drug discovery. But the use of non-randomized data complicates causal inference. Propensity score (PS)-based adjustments are commonly used to remove confounding bias under ignorability assumptions, yet the variance behavior of the resulting inferential statistics with EC borrowing is less studied. Naive application can lead to erratic type-I error rates, which often causes confusion in practice. This presentation discusses several important issues in the practice of PS-based EC borrowing. First, we examine type-I error distortion in commonly used procedures and explain its underlying causes. Second, we propose weighting strategies with adequate variance estimation in both hybrid randomized controlled trial (RCT) and single-arm designs. Simulation studies show that the proposed weighting strategies outperform existing approaches and, when combined with bootstrap-based variance, effectively restore type-I error rates to nominal levels. Based on these results, we offer practical recommendations to ensure valid causal effect estimation when borrowing ECs.

Comments: Co-authors:

Title: **AI for Statistical Analysis Plan Authoring: A Retrieval-Augmented, Agent- Orchestrated Automation Framework**

Speaker: Zhaohua Lu, Daiichi Sankyo Inc.

Abstract: Preparing a statistical analysis plan (SAP) is a crucial and time-intensive process. The burden is especially high for junior statisticians and newly onboarded team members, particularly in studies where a high-quality SAP must be finalized before enrollment or randomization of the first patient. Even when standard templates are available, protocol-

specific content must be identified, interpreted, and translated into SAP-ready language, often resulting in avoidable variation, review rework, and inconsistencies that complicate downstream implementation.

We are developing an AI-assisted SAP authoring framework that integrates retrieval-augmented generation (RAG), deterministic document processing, and agent-orchestrated workflow control. The framework parses protocol and template documents into structured representations, extracts relevant protocol details and tables, detects template update signals, drafts targeted SAP content, and integrates approved outputs back into the Word-based SAP structure. A state-aware agent layer coordinates specialized workflow modules, including document parsing, evidence retrieval, template-signal interpretation, content drafting, validation, revision tracking, and human-review routing. These agents operate as controlled workflow coordinators rather than autonomous decision makers, enabling task decomposition, dependency management, provenance capture, exception handling, and reproducible execution across SAP drafting steps.

Comments: Co-authors: Cheng Zheng, Daiichi Sankyo Inc.

Title: **Evidence-Based Clinical Trial Design: When the Power of AI Meets Statistical Rigor**

Speaker: Will Ma, HopeAI

Abstract: Clinical trial design decisions — from endpoint selection to sample size estimation to comparator arm construction — are only as good as the evidence informing them. Yet in practice, development teams often face fragmented literature, incomplete data, and limited access to patient-level information from prior studies, leading to designs built more on convention than on evidence.

This talk presents a framework where AI and statistical rigor converge to enable truly evidence-based clinical trial design. The approach, developed at HopeAI, integrates three core capabilities: (1) systematic, human-curated synthesis of clinical outcome data across indications through the PURE Evidence platform; (2) generation of synthetic individual patient data (SynthIPD) from published aggregate results, enabling trial-level simulations grounded in real-world evidence; and (3) agentic AI tools — including an AI Clinician and an AI Statistician — that work alongside biostatisticians to translate synthesized evidence into actionable design parameters.

We illustrate this framework through oncology case studies, demonstrating how AI-curated evidence and synthetic patient-level data can inform adaptive designs, optimize randomization strategies, and strengthen regulatory submissions by grounding design assumptions in a more complete evidence base. Critically, each AI-generated output is validated through established statistical methods.

Comments: Co-authors:

Title: **A Master-Protocol Platform Trial to Enable Iterative Medical Device Innovation: The Example of Soft Contact Lenses**

Speaker: Cristiana Mayer, Johnson & Johnson Vision

Abstract: Platform study designs have accelerated early-phase pharmaceutical development by enabling evaluation of multiple interventions efficiently as multiple assets become available over time. This presentation focuses on the potential application of such statistical innovation in study design to MedTech by evaluating a platform trial designed under a master protocol to assess the feasibility of multiple prototypes for myopia-correcting soft contact lenses. The innovative approach supports consistent endpoint assessment across candidate materials, improves operational efficiency, and shortens development timelines by accelerating evidence generation for decision making in the device development plan. To our knowledge, platform trials remain uncommon in the Medical Device space, and this work is intended to stimulate the audience's interest in a broader application of innovative statistical study designs for therapeutic medical device development.

Comments: Co-authors: Wright Shamp, Johnson & Johnson Vision; Oksana Sova, Johnson & Johnson Vision; Augustine Nti, Johnson & Johnson Vision

Title: **Expert-centred evaluation of CURIE-AI in biosciences: from Gene Ontology narratives to biomedical benchmarking across psoriasis, Crohn's disease, and adverse food reactions**

Speaker: Yovani Torres, BIOST3, Section of Statistics, Faculty of Biology, University of Barcelona

Abstract: Artificial intelligence(AI) outputs require reproducible evaluation frameworks to be reliable in biomedical research. This work summarizes three in silico studies proposing a statistical framework for assessing AI-generated biomedical narratives, applied to CURIE-AI across psoriasis, Crohn's disease, and adverse food reactions. First, five dermatologists evaluated CURIE-AI-generated Gene Ontology narratives in psoriasis. Fleiss' kappa indicated moderate inter-rater agreement ($\varkappa$=0.466); Likert reliability was 0.76 for biological plausibility, 0.81 for clinical usefulness, and 0.735 for reference quality. Second, in Crohn's disease, CURIE-AI was benchmarked against ChatGPT, Gemini, and Claude under blinded and counterbalanced conditions by 13 evaluators, producing 2,600 Likert scores. Mean scores ranked Claude first(4.19±0.91), followed by CURIE-AI(3.60±1.26), ChatGPT (3.50±1.32), and Gemini (2.88±1.41), with highest reliability for specificity(ICC=0.92) and usefulness (ICC=0.90). Third, in adverse food reactions, eight specialists produced 400 Likert scores across ten conditions; 56.2% of ratings were fully aligned and 43.8% partially aligned, nominal agreement was very low Fleiss' $\varkappa$=0.02; Krippendorff's α=0.02), and topic specificity showed high reliability (ICC=0.809). Together, these studies propose a robust, replicable framework combining expert audit, inter-rater reliability metrics, and comparative benchmarking, applicable to AI evaluation across the biopharma lifecycle.

Comments: Co-authors: Giuseppe Tardiolo, Department of Biomedical, Dental, Morphological, and Functional Imaging Sciences, University of Messina, Messina, Italy; Jiajie Chen, BIOST3. Section of Statistics. Fac of Biology. University of Barcelona, Barcelona, Spain; Paula Pau,

BIOST3. Section of Statistics. Fac of Biology. University of Barcelona, Barcelona, Spain; Victor Madarnás, BIOST3. Section of Statistics. Fac of Biology. University of Barcelona, Barcelona, Spain; Qi Zhang, BIOST3. Section of Statistics. Fac of Biology. University of Barcelona, Barcelona, Spain; Antonio Monleón-Getino, BIOST3. Section of Statistics. Fac of Biology. University of Barcelona, Barcelona, Spain

Title: **Propensity Score Methods in Non-Randomized Clinical Studies: A Simulation Framework for Preterm Infant Growth**

Speaker: Atiye Nazari, Nestlé Research

Abstract: Introduction: In clinical research, randomization is not always feasible due to ethical, practical, or operational constraints. Studies using external or historical control groups are therefore prone to confounding and bias, requiring appropriate statistical methods to support valid comparisons.

Methods: Propensity score (PS) methods are widely used in non-randomized studies to reduce confounding. Common approaches include matching, stratification, inverse probability of treatment weighting (IPTW), covariate adjustment, and doubly robust methods. This work presents a simulation framework motivated by a study in preterm infants receiving a human milk oligosaccharide supplement compared with a historical control group receiving standard care.

Results: The simulation framework evaluates PS methods under scenarios with varying confounding strength and overlap. Matching achieves strong covariate balance with adequate overlap but may reduce sample size. Stratification improves balance across strata but can leave residual imbalance. IPTW uses the full sample but is sensitive to extreme weights. Doubly robust methods maintain performance across a broader range of scenarios.

Conclusion: The performance of PS methods depends on overlap, confounding strength, and model assumptions. This framework supports evaluation of PS approaches in non-randomized studies with external control groups in pediatric growth research.

Comments: Co-authors:

Title: **A Regulatory Perspective on Innovative Endpoints for ALS Trials**

Speaker: Stavros Nikolakopoulos, University of Ioannina

Abstract: Amyotrophic lateral sclerosis (ALS) presents unique challenges for clinical trial design, including high between-patient variability, informative dropout driven by mortality and small expected treatment effects. Despite over one hundred trials, no curative treatment exists, underscoring the urgent need for endpoints that sensitively and interpretably capture overall treatment benefit.

This talk examines the regulatory landscape surrounding innovative endpoints in ALS trials, focusing on generalized pairwise comparisons, win statistics, and joint function-survival approaches. Regulatory acceptability of such methods hinges on pre-specification, interpretability of the estimand, appropriate handling of intercurrent events such as death, and alignment with patient-relevant outcomes. We discuss how the estimand framework can guide

endpoint selection and analysis planning in a manner that is both scientifically rigorous and regulatorily defensible.

Drawing on regulatory experience, lessons from ALS trial methodology and bridging with other indications with similar challenges, we highlight common pitfalls when novel endpoints are introduced without adequate validation or clear estimand definitions. The talk aims to bridge methodological innovation and regulatory practice, offering practical guidance for statisticians working in rare disease drug development.

Comments: Co-authors:

Title: **A Unified Information–Time Paradigm for Single-Arm Time-to-Event Trials: The Window-Cohort with Calibrated Follow-Up Requirement**

Speaker: Haitao Pan, St Jude Children's Research Hospital

Abstract: Single-arm phase II trials with time-to-event endpoints are commonly designed using event-driven or enrollment-driven interim rules, which control statistical information or calendar time separately but not jointly. This limitation is particularly problematic in rare-disease and small-population settings: event-driven designs may lead to prolonged and unpredictable delays when events are sparse, whereas enrollment-driven designs may trigger interim decisions based on immature follow-up. In addition, the effective information at interim depends on the estimand of interest, such as milestone survival at a fixed time point or treatment effects under proportional hazards.

We propose a unified framework that jointly controls information and time by aligning interim timing with the estimand-specific information structure. The proposed Window-Cohort with Calibrated Follow-Up Requirement (WCR) design triggers interim analysis at the enrollment of the Nw-th patient, followed by a calibrated post-lock follow-up window.

Comments: Co-authors:

Title: **A Bayesian Enrichment Design for Dose Optimization in Dose-Ranging Basket Trials**

Speaker: Yeonhee Park, Sungkyunkwan University

Abstract: Basket trials provide an efficient framework for evaluating treatments across multiple indications and dose levels, yet patient heterogeneity and limited sample sizes within indications pose challenges for dose optimization. We propose a Bayesian enrichment design that integrates adaptive patient selection with a hierarchical Bayesian model to optimize doses across multiple dose levels in basket trials. The hierarchical structure enables adaptive information borrowing across indications while preserving indication-specific decision-making, and the enrichment strategy selectively recruits treatment-sensitive patients to improve efficiency. Simulation studies demonstrate that the proposed approach outperforms fully pooled and independent designs by achieving more accurate dose selection and higher decision reliability in heterogeneous settings. These results highlight the advantages of combining enrichment strategies with hierarchical Bayesian modeling for robust and efficient dose optimization in basket trials.

Comments: Co-authors: Kentaro Takeda, Astellas Pharma; Zhoujingpeng Wei, University of Wisconsin-Madison

Title: **Evaluating causal inference methods for dynamic deprescribing strategies: a simulation study**

Speaker: Teresa Pérez, Universidad Complutense de Madrid

Abstract: Deprescribing often involves dynamic treatment strategies in which medication doses are gradually reduced or stopped according to a patient's evolving clinical history. Evaluating these strategies using observational data is challenging because treatment decisions and clinical events occur over time, visits may be irregular, and subsequent treatment may depend on previous patient outcomes. Target trial emulation provides a framework for explicitly defining the treatment strategies, causal estimand and follow-up of interest before analysing observational data. We conducted a Monte Carlo simulation study to evaluate causal inference methods for estimating the effects of clinically relevant tapering strategies compared with treatment continuation. Data were generated using a multi-state model representing baseline characteristics, irregular visit processes, dose changes, withdrawal events and clinical outcomes, with treatment changes occurring only at visit times. Scenarios were designed to reproduce methodological challenges commonly encountered in real-world data, including limited treatment overlap and model misspecification. We compared the parametric g-formula and inverse probability weighting using marginal structural models, with targeted maximum likelihood estimation included as an additional approach. Performance was evaluated in terms of bias, empirical standard error, root mean squared error and coverage.

Comments: Ryan Muddiman, RCSI University of Medicine and Health Sciences; Elisa Medrano Buendia, Universidad Complutense de Madrid; Elliot McClenaghan, London School of Hygiene & Tropical Medicine; John Tazare, London School of Hygiene & Tropical Medicine; Fiona Boland, RCSI University of Medicine and Health Sciences; Emma Wallace, University College Cork; Tom Fahey, RCSI University of Medicine and Health Sciences; Li Wei, University College London; Carmel Hughes, Queen's University Belfast; Ann Doherty, RCSI University of Medicine and Health Sciences; Frank Moriarty, RCSI University of Medicine and Health Sciences.

Title: **AI-Powered Digital Ecosystem and Value Revitalization in Pharma Industry**

Speaker: Jingjun (Jeannie) Qiu, FosunPharma

Abstract: With the rapid evolution of cutting-edge technologies such as AI and ML, the pharmaceutical R&D landscape is embracing a pivotal opportunity for digital and intelligent transformation. FosunPharma has systematically built a comprehensive digital intelligence platform, anchored by the PharmAID intelligent platform, to integrate AI and Data across the entire product lifecycle—including information insight, decision support, drug discovery, clinical development, and post-marketing studies. This initiative drives a deep convergence from data asset management to intelligent decision-making. By sharing practical experiences in platform

architecture, scenario implementation, and value assessment, we aim to empower the industry to jointly explore innovative pathways and a collaborative ecosystem for AI-enabled pharmaceutical R&D.

There are 4 main parts in ""PharmAID Agent Platform"" = PharmaSight Decision Management Agent × Astrolabe Early Research AIDD Toolboxes × MedAlkaid Clinical Research Intelligence System × AquaVista (Data Integration Platform / Disease-Specific Databases / Knowledge and Skills Hub/ Algorithms and Models, etc.)

Comments: Co-authors:

Title: **Computational Prediction of Druggable Targets Using Temporally Validated Machine Learning**

Speaker: Mohammed Quazi, West Virginia University

Abstract: The quest for novel drug targets within the human proteome is a formidable challenge in pharmaceutical research, given the small fraction of proteins currently associated with approved drugs' modes of action (MoA). This study introduces temporally validated machine learning (ML) models that apply knowledge graph embeddings from various data sources to predict the druggability of human proteins. Based on single-class SVM and XGBoost, our models do not include recent scientific results. They are evaluated against a temporally validated external test set: This combined 8-model system demonstrates good predictive power in identifying potential drug targets (60% over seven years of new data) while utilizing features from the June 2018 Pharos database. Results indicate a broad spectrum of proteins predicted as druggable, extending the concept of the "druggable genome" beyond traditional confines. We critically evaluate model performance against external sets, highlighting the nuanced approach of using single-class SVM for anomaly detection in distinguishing Tclin proteins from the more extensive set of non-Tclin targets. Comparing our models with PINNED, DrugnomeAI, and phase 1-3 trial targets shows ML can expand the druggable genome. Our work underscores the importance of temporal validation in ML model development. It offers a promising addition to the computational biology efforts of ranking target-disease associations for drug discovery and development.

Comments: Co-authors: Suman Sirimulla, Expert Systems Inc.; Cristian Bologa, University of New Mexico School of Medicine; Alexei Pushechnikov, Expert Systems Inc.; Nikolay Savchuk, Expert Systems Inc.; Tudor Oprea, Dompe Farmaceutici Inc.

Title: **Assessing Treatment Effect Heterogeneity using Machine Learning: Questions, Methods, Challenges**

Speaker: Bohdana Ratitch, Bayer Inc., Canada

Abstract: Assessing treatment effect heterogeneity (TEH) is central to personalized medicine and to extracting the full value of clinical trial data. TEH poses a challenging problem at the intersection of causal inference methodology, machine learning (ML), and multiple hypothesis testing. In most applications, the target estimand is a conditional average treatment effect, and analyses must contend with potential model misspecification, high-dimensional covariates, and

the impact of data-adaptive subgroup or biomarker selection on valid inference. Over the past decade, methodological advances have increasingly leveraged modern ML to flexibly estimate nuisance functions and investigate effect modification while mitigating overfitting, multiplicity, and post-selection uncertainty. This talk introduces core concepts and key questions in ML-based TEH workflows: (1) testing for evidence against treatment effect homogeneity; (2) identifying predictive biomarkers while controlling false discoveries; (3) characterizing how biomarker(s) modify the treatment effect; and (4) estimating treatment effect magnitude within the identified subgroup(s). We review the evolution and current state of the methodological landscape, highlight recent approaches for addressing these questions, and discuss remaining challenges. While no universally superior and fully robust solution exists, ML-based methods increasingly offer a rich and practical toolbox for TEH analysis.

Comments: Co-authors:

Title: **Raising Standards for Patient-centered Evaluation of Benefit and Tolerability in Oncology Trials**

Speaker: Antoine Regnault, formerly Modus Outcomes, A THREAD Company

Abstract: Clinical trials in oncology primarily focus on survival, clinically defined endpoints, such as disease progression or clinical response, and clinician-reported treatment toxicities. However, these endpoints do not fully capture the patient's experience. Patient-reported outcomes (PROs) are essential in understanding how patients feel and function and how well they tolerate treatments. Despite the growing recognition of the relevance and importance of the patient perspective in the evaluation of new treatments, PRO evidence is still regularly challenged. This presentation will address two areas aiming to achieve high-quality PRO evidence from clinical trials: development of standards for PRO endpoint analysis and emerging rigorous methods for tolerability evaluation.

The SISAQOL-IMI consortium—comprising academics, regulators, industry representatives, and patients—recently published comprehensive, consensus-based recommendations for analyzing and interpreting PRO endpoints in oncology trials. These recommendations are grounded in the estimands framework and good scientific practice. In parallel, statistical and visualization methods have been developed to assess patient-reported tolerability in a robust manner.

Standardization of evidence for PRO benefits and tolerability will allow more consistent consideration of PRO results from cancer clinical trials across stakeholders, thereby facilitating true patient-centered evaluation of novel therapeutic options in oncology.

Comments: Co-authors:

Title: **Treatment switching in randomised trials with time-to-event outcomes – hypothetical estimand strategy via causal inference**

Speaker: Robin Ristl, Medical University of Vienna

Abstract: In many randomized oncology trials with a time-to-event endpoint, patients randomized to the control group may switch to experimental treatment in response to disease progression. In this setting, a hypothetical estimand strategy, assuming that switching was not possible, is often of interest. We performed a large scale simulation study on the application of causal inference methods for the estimation and inference under the hypothetical estimand strategy, utilizing information from baseline and time-dependent covariates. The studied methods included inverse probability weighting, rank-preserving structural failure time models, two-stage estimation and g-computation, and as comparator a covariate adjusted Cox model with event times censored at the time of switching.

When assumptions were met, and the sample size was not too small, the studied methods provided largely unbiased estimates and hypothesis tests controlling the type I error rate. Inverse probability weighting and two-stage estimation were found to provide the largest power. However, the simple covariate adjusted Cox model often performed similarly well. All studied methods were found to be sensitive with respect to their inherent model assumptions and violations of assumptions typically led to notable bias and potential type I error inflation. Challenges in an actual application may be the exact prespecification of typically complex procedures and the assessment of the plausibility of untestable assumptions.

Comments: Co-authors:

Title: **Causal inference in non-randomized research. The role of G-computation.**

Speaker: Cristobal Manuel Rodriguez-Leal, Universidad Complutense de Madrid

Abstract: Causal inference in non-randomised studies is based on four fundamental assumptions: no interference, consistency, conditional exchangeability and positivity. These studies are very valuable because they provide information about the real-world evidence of interventions and are sometimes the only research alternative. The main problem that a data analyst faces when trying to obtain estimates of interest is adjusting for confounding bias. This adjustment is only possible if the four fundamental assumptions are met. However, the first three assumptions are assumed to be true when the research design is examined, and only positivity can be empirically assessed. Unlike other adjustment strategies, such as inverse probability weighting and propensity score matching, G-computation does not require positivity to be examined prior to implementation. However, violation of this assumption can seriously bias the results of G-computation. We performed an extensive simulation study using Monte Carlo simulations to characterise the behaviour of G-computation with different degrees of positivity assumption violation, varying sample sizes, effect strengths, proportions of treated patients, and outcome frequencies, to compute the average treatment effect (ATE) and the average treatment effect on the treated (ATT). The results of the experiment, along with practical recommendations and limitations, are presented.

Comments: Co-authors: Adrián Aurensanz-Crespo, Universidad de Zaragoza; Jorge Castillo-Mateo, Universidad de Zaragoza; Jesus Asin, Universidad de Zaragoza; Rosario Susi, Universidad Complutense de Madrid; Teresa Perez, Universidad Complutense de Madrid

Title: **Increasing Statistical Efficiency in Clinical Trials using Ordinal Longitudinal Data**

Speaker: Maximilian Rohde, Bristol Myers Squibb

Abstract: Ordinal longitudinal outcomes, such as clinical status repeatedly assessed over time on an ordered categorical scale, are often used to characterize patient trajectories in clinical trials, particularly in critical care and infectious disease settings such as COVID-19. Despite the richness of these data, conventional analyses typically reduce the longitudinal structure to a univariate summary: a single-day proportional odds (PO) model evaluated at a pre-specified analysis day, a free-days endpoint analyzed via a PO model or the Wilcoxon test, or a time-to-recovery endpoint analyzed via Cox regression. The Markov Ordinal State Transition (MOST) model provides a principled alternative by jointly modeling the full sequence of state transitions, retaining the complete information available in ordinal longitudinal data.

We present a simulation study calibrated to the ACTT-1 trial of remdesivir in hospitalized patients with COVID-19, comparing the statistical power of MOST against single-day PO models, free-days analyses, and time-to-recovery Cox models across a range of treatment-effect scenarios. Across the scenarios studied, MOST yields meaningful power gains, while simultaneously removing the need to pre-specify the analysis day or the operational definitions of recovery and free-days. We additionally illustrate model diagnostics for MOST and characterize the settings in which the model may not be appropriate, providing practical guidance for its use in clinical trial design.

Comments: Co-authors: Frank Harrell, Vanderbilt University Department of Biostatistics; Benjamin French, Vanderbilt University Department of Biostatistics

Title: **Building Confidence in Generative AI: A Multi-Method Framework for Evaluating AI-Driven Document Automation in Drug Development**

Speaker: Camilo Rojas Cifuentes, Johnson and Johnson Innovative Medicine

Abstract: Background: Generative AI (e.g., RAG, multi-agent systems) is increasingly used in pharmaceutical statistics to automate regulatory documents (SAPs, summaries). However, key risks of hallucination, non-determinism, and context errors are not captured by traditional validation. In regulated settings (21 CFR Part 11, ICH E9(R1), FDA/EMA AI guidance), the lack of robust evaluation frameworks limits adoption.

Methods: We developed a multi-method evaluation framework within J&J Statistics and Decision Sciences to assess quality and regulatory trust across five AI initiatives. It combines: (1) automated metrics aligned with ISO/IEC 25023, including RAG-specific measures (faithfulness, context recall/precision, factual accuracy); (2) LLM-as-judge evaluation for consistency and hallucination detection; and (3) structured human review with inter-rater reliability (Cohen's κ) and compliant sign-off. Quality dimensions (accuracy, completeness, traceability, consistency, compliance) are mapped to ISO standards and aggregated into a weighted score. The framework is piloted on SAP automation using a RAG multi-agent pipeline with annotated gold-standard references.

Results/Conclusions: Preliminary results show generation-only metrics fail to detect retrieval errors, highlighting the need for multi-layer evaluation. This framework enables scalable, GxP-

compliant quality monitoring of AI in statistical workflows, supporting transparent and inspection-ready adoption.

Comments: Co-authors:

Title: **Exploring the Effects of Context on Biomarker-Drug Co- Development at the Portfolio Level using an R-shiny App**

Speaker: Patrick Roney, Georgetown University Department of Biostatistics, Bioinformatics, and Biomathematics

Abstract: Modern Phase III development requires simultaneous decision-making regarding both drug efficacy and biomarker validation. Extending the program level biomarker-drug co-development framework of Beckman, Clark, and Chen (2011), we propose a portfolio level approach that generates clear decision rules to maximize revenue given limited resources. By framing the process as a "knapsack" optimization problem, we account for logistical costs and varied economic outcomes. This talk will demonstrate an interactive RShiny application designed to simulate outcomes based on user-defined parameters. This tool empowers decision-makers to evaluate complex trade-offs and optimize resource allocation across a portfolio of biomarker-driven programs.

Comments: Co-authors: Robert Beckman, Departments of Oncology and of Biostatistics, Bioinformatics, and Biomathematics, Georgetown University Medical Center; Zoran Antonijevic, The Bioforum Group

Title: **A Framework for Sample Size Calculation in Kolmogorov–Arnold Network Prediction Models**

Speaker: Ester Rosa, University of Padua, Italy

Abstract: Background: Kolmogorov–Arnold Networks (KANs) are emerging machine-learning models with potential for clinical prediction because of their ability to capture complex nonlinear relationships. However, guidance on sample size planning for KAN development and validation studies is currently unavailable.

Methods: We propose a pragmatic framework combining analytical lower-bound calculations with simulation-based evaluation. Model complexity is summarized through an effective complexity parameter, representing the functional flexibility of the network. Minimum sample size is derived from shrinkage-based criteria adapted from clinical prediction modelling and from precision requirements for outcome prevalence. Estimates are further adjusted for anticipated missing data, clustering effects, and the proportion of observations allocated to model development. Sensitivity analyses explore a range of plausible assumptions regarding model complexity, outcome prevalence, and predictive signal strength.

Results:Preliminary analyses indicate that sample size requirements increase with model complexity, lower outcome prevalence, and weaker predictive signal. Under moderate assumptions, approximately 1500–2000 observations may be needed, with larger samples required in more complex settings.

Conclusion: the proposed framework provides a transparent starting point for sample size planning in KAN-based prediction studies, accounting for model complexity and key study assumptions.

Comments: Co-authors: Ester Rosa, University of Padua, Italy; Stefania Lando, University of Padua, Italy; Gloria Brigiari, University of Padua, Italy; Dario Gregori, University of Padua, Italy

Title: **Revisiting Adaptive Multi-Arm Two-Stage Designs**

Speaker: Daniel Sabanes Bove, RPACT

Abstract: We compare two established approaches to adaptive multi-arm multi-stage testing: Stagewise MAMS based on the inverse normal combination test, and cumulative MAMS based on the conditional error rate approach. Recent publications have highlighted the difference in power between the two approaches: The less treatment arms that are selected, the smaller the power difference between both approaches; if only one treatment arm is selected, the power difference is negligible. We show the results of a simulation study performed in R with the open source package rpact to investigate the reason for the power difference.

Comments: Co-authors: Silke Jörgens, University of Cologne; Gernot Wassmer, RPACT

Title: **The RobinCar Family: R Tools for Robust Covariate Adjustment in Randomized Clinical Trials**

Speaker: Daniel Sabanes Bove, RPACT

Abstract: Introduction: Recent guidance from the FDA provided recommendations and best practices for using covariate adjustment. However, there has existed a gap between the extensive statistical literature on covariate adjustment and software that is easy to use and abides by these best practices.

Methods: We have developed the RobinCar Family, which comprises RobinCar and RobinCar2 (Bannick et al, 2026). RobinCar2 provides clinical trial practitioners with a lite version of the more research oriented RobinCar package with the most essential and well validated statistical methods, following good practices for engineering high quality statistical software packages. These two R packages enable covariate-adjusted analyses for continuous, discrete, and time-to-event outcomes that follow best practices.

Results: We provide an accessible overview of the covariate-adjusted statistical methods in RobinCar and RobinCar2, and show a 7% (average) increase in effective sample size compared to a baseline method in 22 post-hoc analyzed oncology trials.

Discussion: The RobinCar Family provides a comprehensive suite of robust statistical methods for covariate adjustment in randomized clinical trials, in line with recent FDA guidance (U.S. Food and Drug Administration, 2024). The packages RobinCar and RobinCar2 fill an important gap, facilitating the practical implementation of powerful new statistical methods.

Comments: Co-authors: Marlena Bannick, Seattle Children's Therapeutics; Liming Li, AstraZeneca BioPharmaceuticals; Mark Baillie, Novartis; Dominic Magirr, Novartis

Title: **PMDA Perspective (Panel Position Abstract)**

Speaker: Miya Saigusa, Pharmaceuticals and Medical Devices Agency

Abstract: This presentation is part of a panel discussion and is intended to share the PMDA perspective in advance to facilitate discussion.

Interest in complex clinical trial designs—such as adaptive designs, platform trials, the use of external controls, and Bayesian approaches—has increased in Japan in recent years, particularly in clinical trial consultations at PMDA. Along with this trend, challenges have been observed in how such designs are explained, interpreted, and linked to regulatory decision-making.

Based on a review of such recent consultation and review cases, this presentation will provide an overview of key issues encountered in practice. These include how design objectives, assumptions, and uncertainty are communicated, as well as areas where differences in expectations between sponsors and regulators may arise.

We will also briefly refer to related PMDA initiatives, including recent PMDA Early Consideration documents.

By sharing these observations, we aim to contribute to a better understanding of how complex designs are assessed in regulatory practice and how communication can be improved.

Comments: Co-authors:

Title: **Non-Concurrent Controls in Platform Trials: Estimands, Identification, and Efficiency Across Endpoint Types**

Speaker: Michele Santacatterina, New York University

Abstract: Platform trials allow treatment arms to enter and exit over time while maintaining a shared control arm, creating both concurrent and non-concurrent controls. Although non-concurrent controls are often used to improve precision, their use raises central questions for evidence generation: what causal estimand is being targeted, under what assumptions is it identified, and when does borrowing actually help? We present an estimand-first framework for causal effect estimation in platform trials with non-concurrent controls. Across point outcomes, survival outcomes, and repeated-measures outcomes with missing data, we show that the most robust target is often the causal effect in the concurrent population. Borrowing non-concurrent controls can improve efficiency only under explicit but difficult-to-justify assumptions. Overall, the most robust route to improved precision is to target concurrent estimands and use covariate-adjusted doubly robust estimators based only on concurrent controls. Applications using data from the ACTT and ACTIV-2 platform trials corroborate these results.

Comments: Co-authors: Ivan Diaz, New York University; Federico Macchiavelli, New York University; Xinyi Zhang, New York University; Antonio D'Alessandro, New York University; Anushree Iyengar, New York University

Title: **Reframing Evidence: Biostatistical Innovations to Close the Women's Health Gap in Trials and Real-World Data**

Speaker: Grammati Sarri, Cytel, Inc

Abstract: Background: Persistent gaps in women's health research arise from inadequate integration of sex and gender in study design, analysis, and evidence generation. We outline key biostatistical practices to improve representation, validity, and equity in clinical trials and real-world evidence (RWE).

Methods: We synthesize recommendations aligned with SAGER guidelines, covering trial design, pharmacometric and biologic modeling, gender-variable quantification, advanced RWE methods, and data governance. Focus areas include adequate powering, sex-informed estimands, and outcomes that preserve clinically meaningful differences. Methods include interaction modeling, individual patient data meta-analysis, and robust handling of missing data. Mechanistic models (e.g., PBPK) incorporate female-specific physiology and life-course variability, while composite indices capture socio-cultural determinants. Emerging approaches—target trial emulation, digital twins, and causal inference—enhance generalizability.

Results: Sex- and gender-informed methods improve detection of treatment heterogeneity, reduce bias, and strengthen external validity. Standardized sex-disaggregated data collection, alongside AI bias mitigation and simulation, helps avoid perpetuating inequities in women's health.

Conclusion: Embedding sex and gender as core biostatistical variables across clinical and RWE studies is essential to produce equitable, high-quality evidence and advance women's health outcomes.

Comments: Co-authors: Michelle Hoiseth, Cytel, Inc.

Title: **Global Sensitivity Analysis of Operating Characteristics in Type I Error-Calibrated Quantitative Decision-Making Clinical Trial Simulations**

Speaker: Harry Saxton, Astellas Pharma

Abstract: Clinical trial simulations evaluate operating characteristics of quantitative decision-making designs using Bayesian Go/No-Go criteria. The FDA Draft Guidance on Bayesian Methodology mandates comprehensive exploration of assumptions across statistical and operational parameters. Current practice relies on one-at-a-time (OAT) sensitivity analysis, which misses parameter interactions, is computationally expensive, and lacks a quantitative basis for justifying parameter specifications.

We propose global sensitivity analysis (GSA) using variance-based Sobol indices as an efficient screening framework. This method decomposes the variance in operating characteristics (e.g., false decision probabilities) into individual parameter contributions and their interactions. This allows researchers to rigorously justify assumptions: parameters with negligible total-effect indices need less precision, whereas influential parameters and interactions require careful specification. GSA comprehensively explores the parameter space with substantially fewer runs than exhaustive OAT, providing superior computational efficiency.

We demonstrate this via a case study of a three-arm Phase II trial with Self-Adapting Mixture priors. The GSA framework successfully identifies the key parameters affecting operating characteristics, fulfilling FDA mandates while providing transparent, quantitative justification for trial design decisions.

Comments: Co-authors: Yusuke Yamaguchi, Astellas Pharma; Kentaro Takeda, Astellas Pharma

Title: **Causal inference via propensity scores for case-control and case-cohort studies**

Speaker: Mireille Schnitzer, Faculté de pharmacie, l'Université de Montréal

Abstract: The development and uptake of propensity score methods in outcome-dependent sampling schemes, such as case-control studies, remains limited. Under this design and its variants, propensity scores are estimable from control data under a rare disease assumption or, more specifically, the assumption that the controls are representative of the source population in terms of their exposure distribution conditional on covariates ("control exchangeability"). Notably, this "assumption" is automatically satisfied under case-cohort sampling. In this talk, we demonstrate that relative effects, such as causal risk ratios, are estimable under three variants of case-control designs using control-fitted propensity scores. We present three estimators: inverse probability of treatment weighting, a double machine learning estimator and a novel two-step propensity score caliper-matching procedure. We contrast these estimators with simulated and real data and introduce novel diagnostic criteria to verify propensity score overlap and covariate balance.

Comments:

Title: **Patient-Reported Outcomes in EU-HTA: From Considerations to Implementation**

Speaker: Anton Schoenstein, Boehringer Ingelheim

Abstract: Patient-reported outcomes (PROs) play a key role in EU Joint Clinical Assessments (JCA), where demonstrating patient-relevant benefit is essential. The initial implementation of JCA procedures focuses in particular on oncology indications.

This presentation addresses statistical considerations for integrating PROs at the planning stage of pivotal oncology trials to address HTA decision-making needs. Key topics include endpoint positioning, instrument selection, and assessment timing. Emphasis is placed on estimands aligned with HTA, handling of missing data, and the definition and justification of meaningful change thresholds, as well as potential requirements for general psychometric evidence, where current guidance remains non-specific.

In this context, EU JCAs also consider evidence from multiple sources, including indirect comparisons and external controls, introducing challenges related to scale comparability and measurement heterogeneity. From a practical perspective, the session offers statisticians an overview of key PRO-related methodological challenges, contrasting current guidance with early JCA experience.

Comments:

Title: **From Individuals to Effects: Control-group borrowing via individual exchangeability and disjunctive subgroup treatment-effect sharing**

Speaker: Darren Scott, AstraZeneca

Abstract: Control-group borrowing via individual exchangeability: When making decisions in a medical setting, we treat the individual rather than an entire group. Surely, borrowing control information from external trials is not immune to the importance of conditioning? Bayesian methods for borrowing external control information, such as the power prior, commensurate prior, and meta-analytic predictive prior, are widely used in clinical trials. These methods borrow information by treating external information as one large group and apply a global discount. I discuss a Bayesian semi-parametric LEAP prior, a borrowing framework that respects our clinical lesson of individual, patient-level decision-making.

Lessons from disjunctive subgroup treatment-effect sharing: In precision medicine, treatment effects may be expected to be larger in one subgroup than in its complement, while remaining positive in both. When the complementary subgroup is small, trials are often underpowered to establish efficacy within that subgroup alone, creating a natural setting for borrowing. We propose a Bayesian model that imposes non-zero ordered effects and places a prior on the ratio of subgroup treatment effects, yielding a simple and clinically interpretable form of incorporating external information. Borrowing within our model leads to gains in efficiency but encounters the recurring challenges of Bayesian borrowing, including prior sensitivity and differences from frequentist operating characteristics.

Comments: Co-authors:

Title: **Error-Controlled Effect Screening in Saturated Factorial Designs Using APCanalysis**

Speaker: Dr. Abu Zar Shafiullah, Thuenen Institute of Agricultural Technology, Germany

Abstract: Unreplicated two-level screening designs, including Plackett–Burman and fractional factorial designs, are widely used across scientific and industrial research because they enable rapid identification of influential factors while substantially reducing experimental cost, time, and resource requirements. These advantages make them particularly valuable in pharmaceutical and biopharmaceutical applications, where efficient early-stage screening is essential. However, saturated models in such designs lack sufficient degrees of freedom for conventional significance testing. To address this limitation, we present APCanalysis, a user-friendly R package implementing the All Possible Comparisons (APC) methodology for objective effect screening with controlled error rates. The package applies a tailored APC-criterion with support for individual error rate (IER), experiment-wise error rate (EER), and false discovery rate (FDR) control. These options provide flexibility across research contexts: IER control is useful in early-stage drug discovery where detection of true effects is prioritized, EER control is well suited for high-stakes drug safety assessment where strict error control is required, and FDR control is appropriate for high-dimensional exploratory research such as genomics and high-throughput screening. APCanalysis supports main-effect screening in Plackett–Burman designs, active effects and interactions in full and fractional factorial designs, and identification of active

Comments: Co-authors: Dr. Arden Miller, University of Auckland, New Zealand

Title: **MetaFemina: A Large Language-Model–Driven Meta-Analysis Platform for Studying Associations between Dietary Supplements and Breast, Ovarian and Uterine Cancer Risk**

Speaker: Yushu Shi, Weill Cornell Medicine

Abstract: We developed and evaluated an automated large language model (LLM)-based web platform, MetaFemina, for conducting meta-analyses of dietary supplements and breast, ovarian, and uterine cancer risk. MetaFemina scanned through the PubMed database for over two hundred dietary intakes, retrieve relevant articles through key word matching in titles and abstracts, and used LLMs to extract evidence. Then with extracted evidence, it automatically runs a random-effect meta-analysis and provides publication bias, heterogeneity and sensitivity assessment. It uses LLM with Joanna Briggs Institute Critical Appraisal Checklists to automatically evaluate the quality of the included articles. Additionally, MetaFemina provides automatic sample size calculation based on synthesized effect size from meta analysis. Compared with a recent peer-reviewed meta-analysis of folate intake and breast cancer incidence, the automated pipeline achieved a sensitivity of 91.67% for identifying eligible studies and identified 44 extra articles.

Comments: Co-authors: Margaux Delporte, Weill Cornell Medicine; Yongxian Zhang, Weill Cornell Medicine; Rulla Tamimi, Weill Cornell Medicine

Title: **Bayesian Expert Elicitation for Estimand Construction in Rare Disease: A Cystic Fibrosis Case Study**

Speaker: Woosub Shin, Université Paris Cité, Inserm, Inria, HeKA / KU Leuven

Abstract: Since its introduction in the ICH E9 (R1) guideline, estimands have been considered an important element in designing clinical trials. While the use of classical and well-known estimands is straightforward, the construction of new ones lacks detailed methodological guidance, and this may be attributable to the limited usage of the estimand framework in more complex clinical trials.

We designed a two-stage elicitation process to obtain expert opinions on organoids and on estimand attributes under four prespecified scenarios. We then applied a series of hierarchical joint Bayesian models to identify appropriate estimand attributes that show the highest relative rankings.

Ten individuals from the ORGESTRA consortium provided opinions, primarily regarding organoids, during the first-stage elicitation. Next, 39 clinicians, recruited mainly from the European cystic fibrosis networks, contributed input on estimand attributes. Across the scenarios, FEV1, LCI (lung clearance index), and pulmonary exacerbations were selected, either individually or in combinations of two, based on their rankings. Discontinuation of treatment due to lack of improvement was identified as the most likely, and death as the least likely, intercurrent event in all scenarios. Full estimand attributes were constructed based on the selections.

With our framework, we intend to support future clinical studies involving novel rare disease treatments and innovative biological platforms, such as organoids.

Comments: Co-authors: Moreno Ursino, Université Paris Cité, Inserm, Inria, HeKA; Danya Muilwijk, University Medical Center Utrecht; Cornelis K. van der Ent, University Medical Center Utrecht; Isabelle Huys, KU Leuven; Sarah Zohar, Université Paris Cité, Inserm, Inria, HeKA

Title: **What is a Virtual Patient? Toward Unifying Definitions in Model-Informed Drug Development**

Speaker: Laurens Sluijterman, Radboud University Medical Center

Abstract: Model-informed drug development increasingly relies on concepts such as virtual patients, virtual populations, and digital twins to inform decisions. These terms appear across pharmacometrics, statistics, and related fields, but are not used consistently. Within the European INVENTS consortium, we found that different researchers used the same terms to refer to different underlying concepts. This inconsistency hindered communication and raised a more fundamental question: what, precisely, do these terms represent?

To answer that question, we established a cross-disciplinary working group within the consortium. We reviewed existing definitions and used an iterative process of discussion, revision, and broader consultation to develop consensus definitions for the key terms.

This process revealed that the inconsistency observed within the consortium reflected a broader lack of agreement in the literature. We therefore propose a unified mathematical framework in which virtual patients, virtual populations, and digital twins are interpreted within a single structure that distinguishes models from their resulting outcomes. This framework clarifies how these concepts relate to one another and provides a more precise basis for communication and methodological development in model-informed drug development.

Comments:

Title: **Causal inference methods for specific estimands in the context of randomised trials – scoping review**

Speaker: Paula Starke, University Medical Center Göttingen

Abstract: While the ICH E9 (R1) estimand framework is increasingly adopted, selecting causal inference methods for specific estimands remains challenging. We present findings from a scoping review (n=325 included studies) assessing application of causal inference methods in randomised controlled trials (RCTs). The review reveals a growing trend in method usage. Inverse probability weighting (IPW) was most frequent, primarily for estimands using a hypothetical strategy in time-to-event endpoints with treatment switching, often via inverse probability of censoring weighting (IPCW). Rank-preserving structural failure time models (RPSFTM) were almost exclusively used in drug development (e.g., oncology), where switching and survival outcomes are central. Conversely, instrumental variable (IV) methods were predominantly applied in non-drug development contexts to estimate estimands using a principal stratum strategy, particularly the complier-average causal effect (CACE). G-estimation and G-computation were rarely applied despite methodological advances, due to complexity and implementation challenges. Notably, methods are often used in secondary or re-analyses, rather than primary analyses. Software and code availability remain poor, raising reproducibility

concerns. These findings underscore the need for greater transparency, guidance, and regulatory alignment to support consistent use of causal inference methods in trial analysis.

Comments: Co-authors: Lucia Schneider, University Medical Center Göttingen

Title: **Understanding Drivers of Treatment-Effect Heterogeneity via Interpretable ML: Lessons from Recent Papers**

Speaker: David Svensson, AstraZeneca

Abstract: Identifying predictive biomarkers remains a central challenge in precision medicine, largely because individual-level causal effects cannot be directly observed. Recent advances in methods for estimating Conditional Average Treatment Effects (CATE) have renewed interest in how to reliably quantify and rank treatment–covariate interactions in clinical data. However, several recent papers have examined the behavior of variable-importance measures within CATE frameworks, revealing inconsistent performance across methods and sensitivities to modeling choices.

This presentation reviews key insights from recent methodological studies, including work evaluating SHAP-based importance measures and permutation-based importance measures in extensive simulation settings. We examine their respective strengths and limitations across diverse scenarios. We compare how different ranking strategies perform under varying conditions and identify scenarios where certain approaches may give misleading signals. The accumulated empirical evidence demonstrates that no single variable-importance method performs optimally across all settings. Multi-stage workflows introduce additional complexities that can substantially affect variable rankings. We illustrate concrete examples where different ranking strategies diverge and characterize the conditions under which these divergences occur. Based on these findings, we provide practical recommendations for selecting and applying variable-importance methods.

Comments: Co-authors: Konstantinos Sechidis, Novartis, Basel, Switzerland

Title: **COMIC: A Bayesian Dose Optimization Design for Drug Combination in Multiple Indications with Application to CAR-T Therapies**

Speaker: Kentaro Takeda, Astellas Pharma

Abstract: Project Optimus, initiated by the US Food and Drug Administration (FDA), seeks to shift the focus of dose finding and selection from the maximum tolerated dose to the optimal dose that offers the most favorable risk-benefit balance. However, applying this paradigm shift to drug combination trials presents challenges, particularly due to limited sample sizes and a large two-dimensional dose exploration space. These challenges are amplified when trials involve multiple indications. To address this, we developed a two-stage Bayesian dose optimization design, called COMIC (Combination Optimization in Multiple IndiCations), to efficiently identify Optimal Biological Dose Combinations (OBDC) for multiple indications. The COMIC design follows a two-stage strategy: first, optimizing the dose for one indication based on a utility function that measures the risk-benefit tradeoff, and then using that data to inform and accelerate dose

optimization for additional indications. This approach significantly reduces the required sample size. Additionally, we incorporate a pharmacodynamic endpoint (e.g., receptor occupancy) to prioritize which component of the combination should be escalated, further enhancing the efficiency of dose optimization. Simulation studies demonstrate the strong performance and robustness of the COMIC design across various scenarios. We illustrate the method using a CAR-T therapy trial.

Comments:

Title: **Binomial Sum Variance Inequality correction of 95% CIs of percentages in multicentre studies ensures approximately 95% coverage with minimal width**

Speaker: Paul Talsma, Phastar, UK

Abstract: A method is presented for constructing a 95% confidence interval (CI) of an overall percentage events obtained from multicentre data which ensures approximately 95% coverage, has minimal width and is not overly complex: the Binomial Sum Variance Inequality (BSVI) correction.

In simulation study 1, coverage and width of CIs using the correction were compared with no correction. Data were generated using several by-centre population percentages. CIs were constructed using three methods, for varying numbers of centres, total N, and unequal number of participants between centres. Intervals constructed with no correction had coverage which is too high, which the BSVI correction corrected to the 95% level whilst reducing interval width.

Study 2 focused on suitability of applying the correction with increasing average number of participants per centre. Data were generated from 2 to 40 centres, event percentages 4 to 64, n per centre 2 to 20 and differences between centre percentages as in study 1. Results show that the correction leads to 95% confidence intervals with adequate coverage (≈95%) and reduced width for average n per centre ≥ 7.

Study 3 systematically investigated the relationship between Informative Cluster (Centre) Size (ICS) and coverage of the 95% CI, with factors as in study 1 and ICS being fully absent, present in 50% of centres, or fully present. Results indicate no association between ICS and coverage.

Comments: Co-authors: Francesco Innocenti, Maastricht University, Maastricht, The Netherlands

Title: **Beyond treatment effect drift in Bayesian dynamic borrowing: variance misspecification and type I error**

Speaker: Julien Tanniou, Saryga

Abstract: Bayesian dynamic borrowing (BDB) with robust mixture priors (RMP) is increasingly used to incorporate external information in clinical trials. Operating characteristics are usually evaluated by varying drift in treatment effect between source and target studies while keeping the sample standard deviation (SD) fixed. However, in these designs, the posterior borrowing weight also depends on the observed SD and may therefore affect type I error.

Using a real bridging case-study, a simulation study was conducted. Design parameters (sample size, initial mixture weight, and posterior decision threshold) were calibrated to achieve prespecified power and type I error at a reference SD. Operating characteristics were then

reassessed over a range of true SDs, and the impact of SD misspecification on the updated informative weight and the probability of declaring success under the null and alternative was quantified.

Under the RMP, type I error showed a non-monotonic relationship with the true SD. Inflation of the type I error occurred when the true SD was moderately larger than the reference value because the posterior weight on the informative component increased, whereas even larger SDs reduced type I error through greater posterior uncertainty.

Unlike standard frequentist tests, BDB designs are sensitive to SD assumptions. For designs intended to support regulatory decision-making, simulations should assess type I error jointly over mean and variance drift.

Comments: Co-authors: Marco Ratta, Saryga; Stefano Vezzoli, Chiesi Farmaceutici, Statistical Methodology

Title: **Powering Up: Enhancing Success Rates in Biomarker Discovery, Sample Size Calculation for Feature Selection**

Speaker: Fetene Tekle, Johnson & Johnson, Innovative Medicine R&D, Beerse, Belgium

Abstract: Identifying predictive features in high-dimensional omics data remains a major challenge, particularly when the number of features far exceeds the available sample size. We introduce the Sample Size for Feature Selection (SSFS) method, a statistical framework that estimates the sensitivity of feature selection procedures and determines the sample size required to achieve a target sensitivity under realistic data conditions.

SSFS uses the Area Under the Curve (AUC) as an effect size measure for individual features, with Tweedie shrinkage to improve estimation in high-dimensional settings. Analytical expressions for sensitivity are derived using nonparametric AUC estimators, and False Discovery Rate control is incorporated via the Benjamini–Hochberg procedure. Extensive simulations were conducted across a wide range of scenarios, varying feature dimensionality, sample size, effect size, number of predictive features, distributional assumptions, and correlation structures. Across all settings, SSFS is robust & closely approximated empirical sensitivity.

Using this framework, we provide practical guidance for biomarker plan in clinical trials by characterizing combinations of sample size and feature dimensionality required to achieve a specified sensitivity. The SSFS framework offers a principled and flexible tool for designing high-dimensional feature selection studies as it is implemented in R with configurable experimental scenarios to support reproducible biomarker research.

Comments: Co-authors: Alemu Takele Assefa, Johnson & Johnson, Innovative Medicine R&D, Beerse, Belgium; Bie Verbist, Johnson & Johnson, Innovative Medicine R&D, Beerse, Belgium; Olivier Thas, Data Science Institute, I-BioStat, Hasselt University, 3590 Diepenbeek, Belgium

Title: **Different estimands for function-related treatment effects and their relation to survival**

Speaker: Afroditi Tsatsi, University of Ioannina, Ioannina, Greece

Abstract: The evaluation of treatment effects in clinical trials for amyotrophic lateral sclerosis is particularly challenging, as functional decline, commonly measured using the ALSFRS-R, coexists with high mortality and informative loss to follow-up. The estimand framework provides a structured approach for clearly defining the research question, the treatment effect of interest, and the corresponding analysis strategy. A central component of this framework is the strategy for handling intercurrent events, such as death, which affect both the observation and interpretation of functional outcomes.

Using data from the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT), this study aims to examine how different estimand strategies for function-related treatment effects relate to survival, and how their choice affects patient ranking and the interpretation of treatment effects. Four approaches for the analysis of ALSFRS-R progression rate are considered, along with a survival-based approach. Agreement and association between function-related rankings and survival-based rankings will be assessed using visual diagnostics and statistical measures of concordance. A simulation study will be conducted to evaluate the performance of different analysis methods under alternative treatment-effect scenarios, according to the estimand and ranking strategy used.

Comments: Co-authors: Ruben van Eijk, University Medical Center Utrecht, The Netherlands; Stavros Nikolakopoulos, University of Ioannina, Ioannina, Greece

Title: **Beyond RMST: shape-weighted estimands based on survival curves**

Speaker: Hajime Uno, Dana-Farber Cancer Institute

Abstract: The hazard ratio is the default treatment-effect summary in time-to-event trials, but its interpretation can fail under non-proportional hazards. Model-free alternatives such as the RMST difference avoid this issue, but RMST uses a uniform weight over follow-up and cannot target early- or late-event effects. This talk introduces a beta-type shape-weight family for weighted survival-difference summaries. The RMST difference is recovered as the constant-weight special case. By mapping follow-up onto an event-progress scale, the two shape parameters gain direct interpretations as early-, middle- and late-event emphasis. This provides a pre-specifiable, clinically motivated alternative to post hoc weight selection. We establish coherence with stochastic ordering at the population, sample, and asymptotic levels, and present nonparametric inference, simulations under non-proportional hazards, and applications to recent oncology randomized clinical trials.

Comments:

Title: **Review of causal inference methods for randomised trials in European Public Assessment Reports and Scientific Advice**

Speaker: Susanne Urach, Austrian Agency for Health and Food Safety (AGES), Austria

Abstract: Regulatory assessment has increasingly embraced the ICH E9(R1) estimand framework, bringing causal inference considerations to the forefront of clinical trial design and analysis. This scoping review aims to systematically identify and map regulatory documents in which causal inference methods were discussed in relation to specific estimands, with particular

attention to handling of intercurrent events. We included European Public Assessment Reports (EPARs) for concluded marketing authorization procedures and EMA Scientific Advice Final Letters issued before August 2025 which discusse causal inference methods in relation to specific estimands, with particular attention to handling of intercurrent events. The scope of the review is limited to randomised controlled clinical trials. The initial search identified 450 procedures of which 87 (37 EPARs and 50 EMA-SAs) were found eligible for inclusion. Our review revealed a range of causal inference methods especially for handling intercurrent events and hybrid control designs in the context of randomised controlled trials (RCTs). Regulatory discussion of causal inference methods within the reviewed documents was limited, and the level of methodological detail provided is often sparse. Nevertheless, certain trial settings—particularly adjustment for cross-over in estimating effects in oncology trials—emerge as recurring topics of discussion.

Comments: Co-authors: Tobias Fellinger, Austrian Agency for Health and Food Safety (AGES), Austria; Florian Klinglmüller, Austrian Agency for Health and Food Safety (AGES), Austria; Angelika Geroldinger, Austrian Agency for Health and Food Safety (AGES), Austria

Title: **A look at safety and benefit-risk assessment in master protocols**

Speaker: Alessandro Vagheggini, MSD

Abstract: Safety evaluation and benefit-risk assessment (BRA) are central to medicinal product development and lifecycle management. In recent years, these assessments have shifted from descriptive, case-based review toward structured, aggregate-data approaches with greater emphasis on quantitative methods. This evolution has been influenced by CIOMS and ICH guidance and by collaborative input from regulators, industry, academia, and patients. Within this evolving landscape, master protocols, including basket, umbrella, and platform trials, introduce specific challenges for safety evaluation and BRA, including heterogeneity across cohorts, multiplicity, intercurrent events, and treatment switching. This chapter reviews general and technical considerations for safety assessment in these complex trial designs, with a focus on joint safety profiling, exposure-adjusted analyses, and time-to-event summaries. Advanced statistical approaches, including Bayesian hierarchical models, are discussed as methods to improve estimation and to borrow strength across related subpopulations while maintaining interpretability of subgroup-specific safety signals. The chapter also highlights the complementary role of qualitative and quantitative evidence in BRA and underscores the importance of incorporating patient experience data into an integrated decision-making framework.

Comments: Co-author(s): Sven Erik Ojave, MSD

Title: **Integrating Patient-Reported Preferences in the Joint Assessment of Function and Survival**

Speaker: Ruben van Eijk, University Medical Center Utrecht

Abstract: Amyotrophic lateral sclerosis (ALS) is characterized by substantial clinical heterogeneity, with patients presenting with diverse symptoms. Conventional ALS outcomes summarize these

symptoms into a single score; however, patients with the same score may nonetheless have markedly different prognoses or impacts on quality of life. This heterogeneity complicates the interpretation and clinical meaningfulness of treatment effects observed in ALS clinical trials.

The Patient-Ranked Order of Function (PROOF) was developed to prioritize symptoms based on what is most important to patients, using generalized pairwise comparisons. This approach yields a reweighted rank score that combines functional loss with patient priorities, providing a potentially more accurate reflection of the patient experience and the clinical benefit of a treatment. Because ALS clinical trials are often accompanied by non-trivial mortality, generalized pairwise comparisons also provide a natural framework for integrating PROOF with survival endpoints. In this talk, we outline the methodology underlying PROOF, illustrate its integration with survival, and discuss potential extensions for handling missing data.

Comments:

Title: **Automated, efficient and model-free inference for randomized clinical trials via data-driven covariate adjustment**

Speaker: Kelly Van Lancker, Ghent University and Vrije Universiteit Brussel

Abstract: In 2023, the U.S. Food and Drug Administration issued guidance for adjustment of covariates in randomized clinical trials, emphasizing its role in enhancing precision and power through prognostic baseline variables. Despite its potential, many trials underutilize this method partly due to challenges in pre-specifying optimal baseline covariates and their functional forms.

We explore the potential of automated, data-adaptive methods—including stepwise regression, Lasso and flexible machine learning algorithms—for covariate adjustment, addressing the challenge of pre-specification. Our approach ensures valid and interpretable treatment effect estimates and standard errors, even when outcome models are misspecified or biased outcome predictions are used.

This differs from most competing methods, which assume correctly specified models for consistent standard errors. Our estimators require cross-fitting for reliable standard error estimation, though it can be omitted when variable selection is used, provided the outcome model satisfies an ultra-sparsity assumption. As such, we arrive at simple estimators and standard errors for marginal treatment effects in randomized clinical trials, exploiting data-adaptive predictions from prognostic baseline covariates, with little (or no) bias in finite samples even when predictions are biased.

Empirical and methodological results demonstrate promise of automated covariate adjustment for improving statistical power of trial analyses.

Comments: Co-author(s): Iván Díaz, New York University School of Medicine; Stijn Vansteelandt, Ghent University

Title: **Beyond the Average: Leveraging Treatment Heterogeneity for More Powerful Testing in RCTs**

Speaker: Stijn Vansteelandt, Ghent University

Abstract: Data-adaptive methods, such as those based on targeted learning, are increasingly used in randomized trials to adjust for baseline covariates while avoiding model misspecification bias. However, most existing approaches focus exclusively on the Average Treatment Effect (ATE). This focus often leads to a significant loss in power when treatment effect heterogeneity dilutes the mean signal, potentially causing valuable therapeutic effects to go undetected.

In this talk, I present an optimal test for the null hypothesis that the conditional average treatment effect (CATE) is zero across all levels of observed baseline covariates. Our proposal builds upon the Projected Covariance Measure (PCM) test of Lundborg et al. (2024). We demonstrate that the original PCM test is suboptimal and provide an extension that effectively handles heteroscedasticity - a critical requirement for binary and count endpoints. Furthermore, by tailoring the method specifically to the randomized trial setting, we substantially relax regularity conditions and achieve greater efficiency.

We illustrate the practical advantages of this approach through simulations and re-analyses of clinical trials. The results demonstrate that our proposed test maintains robust finite-sample performance and offers substantially higher power than both the standard PCM and conventional ATE tests based on Maximum Likelihood Estimation (MLE) or Augmented Inverse Probability Weighting (AIPW).

Comments: Co-author(s): Feng Liang, Beijing Normal University; Kelly Van Lancker, Ghent University

Title: **Simulation-Driven Collaborative Trial Design with Regulators**

Speaker: Matteo Vestrucci, Berry Consultants

Abstract: The FDA's Complex Innovative Trial Design (CID) program provides a dedicated pathway for sponsors to develop clinical trials that use non-traditional methods. This talk will briefly describe the program and explain how, within this framework, simulations serve as the primary tool for communication between the sponsor and the regulator, allowing both parties to evaluate a trial design effectively. To illustrate this, I will present a case study showing how the interaction with the FDA led to specific design adjustments and helped refine the final study plan.

Comments:

Title: **Untangling the Comparator Web: A Blending Method for Individualised Treatment Comparisons in EU Joint Clinical Assessments**

Speaker: Alberto Vidal Diez, Cytel Inc.

Abstract: Implementing individualized treatment comparators within indirect treatment comparisons (ITCs) to support EU Joint Clinical Assessments (JCA) presents key methodological and interpretational challenges. A central issue is defining an appropriate "comparator bundle" when clinical practice reflects heterogeneous treatment patterns (e.g., physician's choice) rather than a single standard of care. Conventional approaches construct weighted averages; however, these introduce bias due to non-collapsibility of effect measures (e.g., hazard or odds ratios), where the effect of a mixture is not equal to the mixture of effects. Limited transparency on

trial-level treatment composition and reliance on imperfect external data (e.g., market shares) further increase uncertainty. These limitations conflict with JCA expectations for a single, clearly defined PICO and a single robust estimate.

We developed a Bayesian mixture-augmented model in which the comparator is specified as a finite mixture of component treatments. Mixture weights follow Dirichlet priors centred on observed proportions, while component effects use robust MAP priors informed by external evidence. Jurisdiction-specific estimands are obtained by re-weighting posterior outcomes.

This "blending" method disaggregates mixed comparators within a unified framework, avoiding invalid average

Comments: Co-author(s): Grammati Sarry, Cyten Inc.; Ignacio Parra, Cyten Inc.

Title: **Quantifying Effective Sample Size in Bayesian Clinical Trials**

Speaker: Chenguang Wang, Regeneron

Abstract: In clinical trials that leverage real-world data (RWD) and real-world evidence (RWE), it is critical to evaluate the impact of RWD/RWE on findings, particularly in the context of regulatory decision-making. Statisticians often use effective sample size (ESS) to summarize the impact of RWD/RWE. In this presentation, we propose an optimization-based ESS, defined by matching an informative prior to a family of power posteriors via reverse Kullback–Leibler divergence. This approach yields a single, interpretable scalar that quantifies how much data the RWD/RWE is equivalent to. Through examples, we demonstrate that the proposed ESS is both intuitive and easy to implement.

Comments:

Title: **Robust Bayesian Borrowing with Local Control of Type I Error and Power**

Speaker: Lingjiao Wang, University of Warwick

Abstract: When the sample size of a new clinical trial is limited, it may be attractive to incorporate external data from previous or similar trials in the analysis. A challenge in doing this is that we must ensure the use of the external data is appropriate. One approach to address this issue is adaptive borrowing, which includes both Bayesian and frequentist methods. Rather than always fully borrowing all external information, this approach uses a parameter or a set of parameters, δ, to determine the extent of borrowing based on the commensurability between the external and trial populations.

It is known that borrowing methods cannot simultaneously control the type I error rate and lead to a gain in power. In this case, we propose a framework that enables the incorporation of external data from one or both arms of two-arm trials with normally distributed outcomes, that focuses on choice of a value of δ to control the type I error rate over a restricted range for the true control treatment effect, enabling a simultaneous gain in power.

The magnitude of δ depends on the characteristics of the external data. Conditional on these characteristics, an increase in power compared to a trial without borrowing may be achieved. In addition, we propose methods to reduce the required sample size.

The results are discussed and compared with those obtained from proposed Bayesian borrowing methods.

Comments: Co-author(s): Nigel Stallard, University of Warwick; Martin Posch, Medical University of Vienna

Title: **Incorporating backfills in model-free dose-escalation designs**

Speaker: Nan Wang, AstraZeneca

Abstract: Using backfills to collect additional medical information is common in oncology phase I dose escalation studies. Occasionally, additional dose limiting toxicities (DLTs) may occur at previously tested doses. In traditional model-assisted designs, such as BOIN and mTPI2, dose-escalation decisions use data from the current dose only and do not account for these additional DLTs.

To address this limitation, BF-BOIN (Zhao, et al. 2024) was developed to incorporate additional DLTs arising from backfills into the decision-making process. In situations where the backfill result conflicts with the dose-escalation decisions without backfills, BF-BOIN pools the incidence of DLTs across all relevant doses from the conflicted dose up to the current dose to inform escalation decisions.

Following the BF-BOIN framework, we explore the modification to manage conflicts by isotonic regression. This procedure employs the pool-adjacent-violator algorithm (PAVA) and uses the DLT incidences adjusted by PAVA to make dose-escalation decisions. Simulations are conducted under the backfill strategy, which is to backfill whenever an escalation decision is made and backfilled dose is one level below the escalated dose. The PAVA procedure provides satisfying results compared to the BF-BOIN. Further evaluation of the PAVA-based procedure incorporating real enrolment patters will be performed to support this procedure as a practical and efficient alternative for integrating backfill data.

Comments: Co-author(s): Cindy Lu, AstraZeneca

Title: **Optimizing Pragmatic Clinical Trial Design for Medical Devices: an AI-Enabled Target Trial Emulation Framework Using Large-Scale Real-world Data**

Speaker: Xiaofeng Wang, Cleveland Clinic

Abstract: Clinical trials for medical devices must balance rigorous internal validity with real-world generalizability—a challenge uniquely addressed by pragmatic clinical trials (PCTs). However, defining eligibility criteria that maximize both safety and inclusivity often relies on conventional assumptions rather than data-driven evidence. This presentation introduces a methodological framework that leverages Target Trial Emulation (TTE) within high-dimensional real-world electronic health records (EHR) to prospectively optimize PCT design. By transforming complex, free-text eligibility constraints into computable phenotypes, we emulate the proposed

pragmatic trial across diverse, large-scale patient cohorts. Applying causal inference techniques, such as inverse probability of treatment weighting (IPTW), alongside machine learning interpretability tools like Shapley value analysis, we quantify the counterfactual impact of specific inclusion and exclusion criteria. This TTE-driven approach allows investigators to formally evaluate how modifying eligibility thresholds affects expected device performance, safety profiles, and cohort representativeness prior to patient enrollment. Ultimately, integrating TTE methodologies with real-world data reduces operational inefficiencies in PCTs and strengthens the evidentiary basis for both scientific and regulatory decision-making.

Comments:

Title: **Accelerating Clinical Trial Design with Agentic AI: From Fragmented Data to Evidence-Driven Development Strategies**

Speaker: Xiaoyan Wang, Tulane University

Abstract: Clinical trial design is becoming increasingly complex as biopharmaceutical development expands into precision medicine, rare diseases, immunotherapies, and rapidly evolving therapeutic landscapes. Designing statistically rigorous and operationally feasible studies now requires integration of fragmented evidence across scientific literature, prior clinical trials, regulatory precedents, real-world evidence (RWE), biomarker data, and clinical practice guidelines.

This session will explore how generative AI and agentic intelligence systems can accelerate and augment clinical trial design while preserving scientific rigor and regulatory defensibility. Rather than replacing statistical expertise, AI technologies can function as evidence synthesis and decision-support engines that rapidly integrate heterogeneous biomedical data to support endpoint selection, eligibility criteria optimization, comparator selection, and protocol feasibility assessment.

Drawing from real-world implementations in biopharmaceutical settings, the speakers will demonstrate practical frameworks for AI-enabled evidence synthesis, statistically informed trial planning, and transparent validation of AI-generated recommendations. The session will also address key methodological and governance considerations for responsible AI deployment in regulated environments, including reproducibility, auditability, human oversight, and alignment with evolving FDA, EMA, and ICH expectations.

Comments:

Title: **Learning from Literature: Integrating LLMs and Bayesian Hierarchical Modeling for Oncology Trial Design.**

Speaker: Wei Wei, Yale University

Abstract: Purpose: Designing modern oncology trials requires synthesizing evidence from prior studies for hypothesis generation and sample size calculation. Trial designs built on incomplete information or imprecise summaries can lead to mis-specified hypotheses and underpowered studies, resulting in missed opportunities and false positive or negative trial results. To address

this challenge, we developed LEAD-ONC (Literature to Evidence for Analytics & Design in Oncology), an AI-assisted platform that ingests results of published trials, extracts unstructured data, and performs Bayesian evidence synthesis to support well-informed design decisions for future trials.

Methods: Given expert-curated publications that meet prespecified eligibility criteria (population, treatment strategy, trial conduct), the proposed framework uses large language models (LLMs) to extract summaries of baseline characteristics from "Table 1" and reconstruct individual-patient data (IPD) from Kaplan–Meier (KM) curves. Trials are then clustered by similarity in baseline characteristics to define subpopulations more comparable for evidence synthesis. Within each cluster, we fit a Bayesian hierarchical model to the reconstructed IPD and generate the predictive distribution of survival curves for a new trial enrolling a similar population.

Results: We demonstrate the utility of LEAD-ONC using results from five recently completed phase III trials of non-small-cell lung cancer (NSCLC) testing PD-L1/PD-1 inhibitors

Comments:

Title: **Leveraging external information in adaptive platform trial designs: two case studies**

Speaker: Samuel Wu, University of South Florida

Abstract: To enhance efficiency and generalizability, clinical trials can leverage data sourced from outside the current study, such as previous clinical trials, real-world data from patient registries, or published literature. By integrating external data, researchers may need to enroll fewer patients in their trial, saving time and resources. Also, incorporating data from broader populations can help ensure the trial results are relevant to a wider patient group. In this talk, we will present two ongoing adaptive platform trials for the prevention of type 1 diabetes (T1D). Both studies used external information in the design and analysis to help reduce sample size where large internal control group is not feasible. We will also discuss challenges to appropriately account for potential differences between datasets.

Comments:

Title: **How early design and analysis choices of pivotal trials shape HTA usability**

Speaker: Stefanie Wuestner, AMS Advanced Medical Services GmbH

Abstract: The EU HTA Regulation has established a joint clinical assessment framework requiring comparative evidence on relative effectiveness and safety for Member State-requested PICO questions. In practice, pivotal trials are often planned primarily for regulatory purposes, while HTA-relevant evidence needs emerge later in development. This can limit the usability of trial evidence for JCA and downstream national assessments.

This presentation provides an introduction to the JCA evidence generation process and shows how early planning decisions on population, comparator, outcomes, and analysis strategy shape the later HTA usability of pivotal trial data. It focuses on how pivotal trial evidence is translated

into a JCA-relevant evidence base through identification of relevant PICO questions and mapping of direct and indirect evidence.

The presentation examines how early gap analyses can identify limitations of the pivotal evidence package once the evidence is translated into the broader JCA evidence space. This includes assessing whether relevant PICO questions can be addressed by available direct evidence, where indirect treatment comparisons (ITC) may become necessary, and how ITC feasibility depends on comparator availability, network structure, and study similarity. Examples include situations in which subgroup analyses, ITC, or non-randomized evidence are considered to address remaining evidence needs, together with the methodological assumptions and associated uncertainties.

Comments: Co-author(s): Kathrin Bogner, AMS Advanced Medical Services GmbH

Title: **Bayesian Optimization and Protein Structure Modeling for AI-Guided Peptide Therapeutic Development**

Speaker: Dong Xu, University of South Florida

Abstract: Peptide therapeutics are attractive alternatives to small molecules because they are relatively easy to synthesize, can be engineered for high target specificity, and are biodegradable. However, existing peptide design methods remain limited, in part because they often lack uncertainty-aware optimization and inaccurate structure-based evaluation. To address these issues, we present a unified framework for multi-objective, structure-guided peptide binder design that combines protein language model embeddings, Bayesian optimization, controlled sequence editing, and structure-based scoring. Peptides are encoded using ESM2 embeddings and optimized with Gaussian-process surrogate models and a Tree-structured Parzen Estimator sampler. A sequence control module enables biologically constrained substitutions, insertions, deletions, and masking, supporting both global latent-space exploration and local sequence refinement. The framework directly incorporates structural metrics, including interface confidence, solvent-accessible surface burial, and interaction energy. Candidate peptides are further filtered through conformational assessment and in silico toxicity screening. Overall, this platform supports refinement of known binders, de novo peptide generation, and data-driven optimization with limited experimental data, providing a flexible and practical approach for accelerating therapeutic peptide discovery.

Comments: Co-author(s): Negin Manshour, University of Missouri

Title: **Digital Twins: Opportunities and Challenges in Virtual Clinical Trials**

Speaker: Fangrong Yan, China Pharmaceutical University

Abstract: Digital twin-based virtual clinical trials bring both new opportunities and challenges to clinical research. Their core lies in using real-world data to construct "digital twins" of patient populations, thereby creating virtual control arms that can replace or supplement traditional control groups. To conduct digital twin-based virtual clinical trials, the first issue to address is how to integrate multimodal data (genomics, electronic health records, medical imaging) with

physiological mechanism models to generate virtual subjects with dynamic predictive capabilities. Subsequently, innovative trial designs, such as synthetic control arm designs, can be explored to improve statistical power and decision accuracy. The current generation of digital twins needs to focus on: model verifiability – i.e., ensuring the reliability of virtual subjects' behavior under out of distribution extrapolation compared to real patient populations; technical ethics, data privacy, and regulatory acceptance. In addition, exploring innovative design methods and applicable scenarios for digital twin based virtual clinical trials is also an important challenge that we need to address.

Comments:

Title: **Bayesian Simultaneous Credible Bands for Polynomial Regression**

Speaker: Fei Yang, The University of Manchester

Abstract: Quantifying efficacy uncertainty across the entire dose range is crucial in dose–response studies. Although the frequentist simultaneous confidence band (FSCB) is widely used for this purpose, it does not readily incorporate prior knowledge. The Bayesian simultaneous credible band (BSCB) offers a natural alternative, yet practical methods for constructing BSCBs remain scarce in the literature. In this paper, we propose a unified framework for constructing a BSCB for the regression curve in a univariate polynomial model over a finite covariate interval. An efficient simulation-based procedure is developed to determine the critical constant of a BSCB. The framework accommodates inference under different levels of prior information and can be implemented either analytically or via posterior sampling methods. Notably, we prove that under mild regularity conditions, the BSCB is asymptotically equivalent to the FSCB, thereby attaining the nominal frequentist coverage for a broad class of priors. Simulation studies confirm that the BSCB attains the exact posterior simultaneous coverage probability across various scenarios. An application to a dose–response study illustrates its importance in identifying the minimum effective dose in Phase II clinical trials. Software implementation of the proposed methods is available in an accompanying R package.

Comments: Co-author(s): Yang Han, The University of Manchester; Wei Liu, University of Southampton

Title: **The Patient Voice in Focus: Advancing Early Phase Oncology Trials Through PRO-Driven Innovation**

Speaker: Christina Yap, Institute of Cancer Research in London

Abstract: Patient-reported outcomes (PROs) are increasingly recognised for assessing tolerability in early-phase dose-finding oncology trials (DFOTs), but PRO objectives, analysis and reporting remain variable. The OPTIMISE programme addresses this gap through two resources. OPTIMISE-ROR provides international consensus guidance on defining patient-centred PRO objectives and key PRO concepts to capture, including patient-reported adverse events, side-effect impact and health-related quality of life. OPTIMISE-AR translates these objectives into practical analysis and visualisation recommendations, covering PRO endpoints over time, between timepoints, time-to-event endpoints and PRO summaries for dose decisions.

Emerging PRO-integrated trial designs show how patient experience can inform dose escalation, expansion, optimisation and final selection. These include approaches that incorporate clinician- and patient-reported toxicity into dose decisions, combine safety, PRO-derived tolerability and preliminary efficacy, or use pre-specified endpoint hierarchies to compare candidate doses and refine phase 2 dose recommendations.

Together, these developments offer a coherent pathway for embedding PROs in DFOTs: defining patient-centred objectives, linking analyses to those objectives, and using PRO-integrated designs to ensure patient-experience data inform dose-finding and dose-optimisation while aligning statistical models with real-world tolerability.

Comments:

Title: **On the Calibration of Bayesian Success Criteria and Operating Characteristics for Clinical Trials**

Speaker: Ying Yuan, UT MD Anderson Cancer Center

Abstract: Recently, the U.S. Food and Drug Administration (FDA) released draft guidance signaling a paradigm shift that facilitates the use of Bayesian methodology as the primary analysis and decision framework for drug approval. The cornerstone and fundamental challenge of this framework is the specification and calibration of Bayesian success criteria to control decision errors, ensuring reliable clinical and regulatory outcomes. In this work, we systematically investigate various Bayesian decision-error metrics, their theoretical interrelationships, and their alignment with conventional Frequentist counterparts. This investigation provides critical theoretical insights and practical guidance on calibrating Bayesian success criteria and operating characteristics to ensure robust decision-making and the integrity of public health decisions. We illustrate this framework using a clinical trial evaluating revascularization strategies for cardiogenic shock. A Shiny application will be available at www.trialdesign.org to assist sponsors and regulators in evaluating calibration strategies consistent with recent regulatory perspectives.

Comments: Co-author(s): Peng Yang, UT MD Anderson Cancer Center; Li Wang, AbbVie

Title: **Precision generalized phase I-II designs**

Speaker: Yong Zang, Indiana University School of Medicine

Abstract: A new family of precision Bayesian dose optimization designs, PGen I-II, based on early efficacy, early toxicity, and long-term time to treatment failure is proposed. A PGen I-II design refines a Gen I-II design by accounting for patient heterogeneity characterized by subgroups that may be defined by prognostic levels, disease subtypes, or biomarker categories. The design makes subgroup-specific decisions, which may be to drop an unacceptably toxic or inefficacious dose, randomize patients among acceptable doses, or identify a best dose in terms of treatment success defined in terms of time to failure over long-term follow-up. A piecewise exponential distribution for failure time is assumed, including subgroup-specific effects of dose, response, and toxicity. Latent variables are used to adaptively cluster subgroups found to have similar

dose-outcome distributions, with the model simplified to borrow strength between subgroups in the same cluster.

Comments:

Title: **Operationalizing Allocation Probability Tests for RAR designs**

Speaker: Stina Zetterstrom, MRC Biostatistics Unit, University of Cambridge, United Kingdom

Abstract: Response-adaptive clinical trials update allocation probabilities to treatment arms based on accumulating data, with the goal of improving patient benefit compared to equal randomization. However, the resulting allocation imbalance can substantially reduce statistical power for treatment comparisons. Recently proposed allocation probability (AP)-based tests, which use allocation probabilities rather than outcome-based sufficient statistics, have demonstrated improved efficiency for binary and normal endpoints. Despite this promise, practical guidance for implementing AP tests across endpoints remains limited.

We present a practical framework for operationalizing the AP test. Rather than introducing entirely new statistics, we develop alternative formulations that more fully exploit the allocation probability information. We extend the AP framework to survival endpoints under exponential models. We further show that a Bayesian decision rule is a special case of the AP test under a Bayesian response-adaptive randomization (BRAR) design.

Simulation studies using two-arm BRAR designs demonstrate that optimizing the AP test can substantially improve power, in some settings equal the theoretical maximum, while preserving patient benefit objectives and strict type I error control. The framework generalizes beyond BRAR to other response-adaptive designs, providing trialists with practical guidance for mitigating power loss without compromising statistical validity.

Comments: Co-author(s): David S. Robertson, MRC Biostatistics Unit, University of Cambridge, United Kingdom; Thomas Jaki, MRC Biostatistics Unit, University of Cambridge, United Kingdom, Faculty of Informatics and Data Science, University of Regensburg, Germany; Sofía S. Villar, MRC Biostatistics Unit, University of Cambridge, United Kingdom

Title: **An innovative statistical method for co-localization in super-resolution microscope images**

Speaker: Hui Zhang, Northwestern University

Abstract: Spatial data is common in many scientific disciplines. For example, single-molecule localization microscopy, such as stochastic optical reconstruction microscopy, provides super-resolution images to help scientists investigate the co-localization of proteins and hence their interactions inside cells, which are key events in living cells. However, there are few accurate methods for analyzing co-localization in super-resolution images. The current methods and software are prone to produce false-positive errors and are restricted to only 2-dimensional images. We will propose a novel statistical method to effectively address the problems of unbiased and robust

quantification and comparison of protein co-localization for multiple 2- and 3-dimensional image datasets. This method significantly improves the analysis of protein co-localization using super-resolution image data, as shown by its excellent performance in simulation studies and an analysis of light chain 3-lysosomal-associated membrane protein 1 protein co-localization in cell autophagy. Moreover, this method is directly applicable to co-localization analyses in other disciplines, such as diagnostic imaging, epidemiology, environmental science, and ecology.

Comments:

Title: **Leveraging external-trial information to optimise patient allocation in group sequential designs**

Speaker: Haiyan Zheng, University of Bath

Abstract: Leveraging external-trial information into the design and analysis of clinical trials has gained considerable interest in recent years. Such information, often available from historical studies conducted under the same or similar conditions, may be incorporated into a prior distribution within a Bayesian framework. A substantial body of statistical literature has explored the use of external-trial information in early phase drug development, providing a well-established foundation for its application in this setting. In contrast, considerably less attention has been paid to the topic for confirmatory clinical trials.

In this talk, I shall focus on phase III group sequential designs that allows a series of decision gates for early stopping, as well as for adaptive allocation of patients, so as to reduce the number receiving the inferior treatment. Our method builds upon Bayesian estimates of the treatment effect at the interim analyses, for which external-trial information is represented in a prior distribution. The sample size is quantified through Fisher information for the treatment effect, and the trial is monitored using standard group sequential boundaries, constructed to maintain type I error and power, with the cumulative Fisher information preserved. We propose a Bayesian decision rule that minimises the posterior expected loss to adapt the allocations, allocating fewer patients to the inferior treatment. This decision rule accounts for uncertainty in the treatment effect by evaluating the loss function over all possible values, weighting each according to the posterior probability, rather than substituting a single estimate. We further optimise the sequential design to weight the expected loss by conditional probabilities that the trial stops at the subsequent decision gate. The proposed sequential designs are then evaluated against a unified loss-based evaluation metric. The simulation results demonstrate improvements in the patient allocation, relative to designs that use a point estimate of the treatment effect (e.g., MLE or the posterior mean), while the error rates can be maintained to desirable levels.

Comments: Co-authors: Corey Voller, Christopher Jennison

Title: **Recurrence: An Analysis of Adverse Events Whose Time Has Come in the Evaluation of Patient Safety in Clinical Trials**

Speaker: Richard Zink, JMP Statistical Discovery LLC

Abstract: Most analyses of adverse events (AEs) are based on the presence or absence of events at the patient level. This analysis has important limitations. First, patients who experience one event versus many events are weighted identically. Second, the timing at which these events occur is completely ignored. Study teams have no insight into whether events occur early or late during exposure, or if the occurrence of events is constant, increasing, or decreasing over time. Non-parametric methods define the mean cumulative function to summarize the number of AEs experienced as a function of time under minimal assumptions. We discuss the importance of exploring AE recrudescence and applying recurrence analyses for greater insight into the safety of novel treatments. Further, we explore important considerations for identifying notable signals among numerous individual and grouped AEs, describe appropriate follow-up periods for the AEs under consideration, and discuss the importance of data visualization for the efficient summary, interpretation, and communication of safety findings. We illustrate methodologies with an analysis of patients with probable mild-to-moderate Alzheimer's disease.

Comments:

Title: **Using Simulations to Guide the Process of Trial Design**

Speaker: Dario Zocholl, NCT Berlin

Abstract: Simulations are an essential tool for biostatisticians in planning complex clinical trials when classical sample size calculations based on closed-form power formulas are not feasible. However, simulations can provide valuable insights far beyond numerical power calculations. For instance, they may be used to assess the operating characteristics of any element of a study design under specific conditions. This can even influence decisions about aspects of a study that are typically considered clinical decisions, such as the definition of endpoint assessment windows or inclusion criteria.

Other characteristics that may be evaluated using simulations include key logistical aspects resulting from study design specifications. For example, using early clinical endpoint readouts for decision-making may negatively affect the statistical properties of a study design, but in some situations the gains in study duration more than compensate for these drawbacks.

Weighing the pros and cons of a design requires not only the technical implementation of simulations but also effective communication to engage medical decision-makers. I will present use cases from phase I/II clinical trials and advocate for the widespread use of simulations, along with effective communication of their results, as an integral component of the trial design process.

Comments: