17 June 12:35

Title: A ballad of a Basket trial and historical information borrowing: application in neurodegenerative diseases

Libby Daniells[1], Alexander Murley[2], Pavel Mozgunov[1]

[1]MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom. [2]University of Cambridge, Cambridge, United Kingdom

Abstract

Recent advances in genetic and diagnostic testing have driven a shift towards personalised medicine, motivating the development of innovative clinical trial designs such as basket trials. In a basket trial, a single therapeutic intervention is evaluated across multiple disease subgroups that share a common biological or clinical feature. While basket trials have become fairly established in the oncology therapeutic area, their use in neurodegenerative diseases remains limited.

This presentation outlines the development of the SYNAPSE study, a symptom-based basket trial investigating the use of the selective serotonin reuptake inhibitor, citalopram, for managing challenging behavioural symptoms in frontotemporal lobar degeneration (FTLD)-related syndromes. Unlike traditional oncology basket trials, where the disease subgroups typically share a common genetic aberration, the syndromes in the SYNAPSE study are linked through shared symptoms targeted by the treatment.

We will discuss the full design process of the SYNAPSE study from conceptual development to the finalised protocol, outlining how the design itself evolved based on discussions with clinicians, conventions in the field, emerging obstacles from real-world constraints – such as wildly different prevalence rates and unexpected financial burdens – and the availability of external data. Through this discussion we provide practical guidance for design/implementing basket trials for neurodegenerative diseases, including the advantages of applying innovative Bayesian borrowing methods to leverage information between disease indications *and* from historic data sources.

15 June 13:30

Title: #FakeNews: Statisticians and the Challenge of Misinformation

Claire Brittain

Novartis, London, United Kingdom

Abstract

In the age of social media and instant "expertise," misinformation can spread faster than any dataset can catch it. From vaccine safety to medical trial results, distorted claims erode public trust in science and evidence. Statisticians, with our training in uncertainty, data interpretation, and critical thinking, are ideally placed to challenge these narratives. Yet too often, we stay silent.

In this talk, we'll examine how statisticians can push back against the rising tide of misinformation, as well as the barriers that stop us from speaking up more often. Using a small study created specifically for this conference we will interactively compare the general public with PSI attendees. We'll look at how people interpret (and misinterpret) data, where our communication falls short, and how we can harness our analytical skills to make truth as contagious as misinformation itself.

15 June 15:45

Title: Can NI margin calibration safeguard against invalid results in an evolving population?

Nuala Peter

Boehringer Ingelheim, Biberach, Germany

Abstract

Head-to-head trials are widely used to compare alternative interventions when standard treatments exist. Population shifts must be accounted for when estimating non-inferiority margins due to changes in population characteristics, to avoid an invalid assumption of transportability. Existing methods have practical and statistical limitations. We introduce a pragmatic conceptual approach and fully pre-specifiable procedure for calibrating margins that account for population shifts observed in the new (head-to-head) trial.

Our approach splits new trial and historical data into subgroups based on relevant effect-modifying covariates. Historical treatment effects are then reweighted to reflect the characteristics of the new trial, validating the assumption of transportability. This calibrated effect informs the NI margin, ensuring alignment between historical evidence and current clinical context. The procedure is transparent, reproducible, and supported by an implementation tool.

Applying this approach to a real-world case study demonstrates that subgrouping and reweighting produces a conservative treatment-effect estimate, albeit at the expense of the variance. It translates into a clinically acceptable NI margin, confirming the results of the consensus based approach.

This approach offers a simple, pragmatic alternative to existing methods for NI margin estimation in evolving populations. Its novelty lies in combining conceptual clarity with operational feasibility: it is fully pre-specifiable, statistically sound, easy to implement and comes with a downloadable calculator. By bridging methodological rigor and practical usability, this framework can enhance the credibility of head-to-head trials across diverse therapeutic areas, including Biosimilars.

17 June 12:35

Title: From Behind the Screen to Beside Your Team: Why the Office Still Matters

Justyna Mlynarczyk

Phastar, London, United Kingdom

Abstract

In a world reshaped by the pandemic, remote and flexible working has become more common. But as our office spaces grow quieter, it's important to remember the benefits of face-to face working. This talk explores why the office remains a powerful space for building relationships, especially for those early in their careers.

Beyond productivity, the office fosters collaboration, creativity, and the informal exchange of knowledge. It's where you can ask for help without booking a meeting, learn by observing others, and see that your work is part of a bigger picture. For young statisticians, the office gives an arena to build a support network outside of your direct line and explore other areas of the business, as well as find mentors within the industry and build friendships they can rely on to bolster them through difficult times. For individuals, these connections support mental and physical well-being, reduce isolation, and help manage stress. For organisations, they drive innovation, encourage knowledge transfer and a sense of community which improves retention.

Psychologists have long identified the desire to feel connected to others as a basic human need. Office-based working meets this need in ways remote setups often can't. This talk invites leaders and employees to reimagine the office not as outdated, but as a space for growth, support, and belonging.

15 June 13:30

Title: How about those estimands for my cross-over study?

Alexandra Jauhiainen

R&I Biometrics and Statistical Innovation, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

Abstract

In a cross-over study a subject receives multiple treatments in a sequence across several treatment periods, making the subject function as its own "control". Cross-over designs are common in bioequivalence (BE) settings but are also used in efficacy study scenarios.

Approaches for handling intercurrent events (ICEs), as described in the ICH E9(R1) estimand addendum, should in a cross-over study consider that an ICE during one treatment period may affect data availability for other treatments, which does not occur in a parallel trial."

For efficacy studies, the ICE approach used in a cross-over trial should ideally yield similar results as in a parallel arm trial. For a composite approach, a 'poor score' can be assigned to a measurement affected by an ICE. If the ICE occurs in the first treatment period and the subject withdraws, data that should not be imputed with a poor score will be missing for treatments in subsequent periods.

We show that the often-used random subject model to analyse cross-over data will lead to biased estimates for the composite and treatment policy approaches in scenarios with incomplete data, due to the differential weighing of within- and between subject information in such a model. We illustrate with simulation how the issue can be mitigated by additional imputation under a hypothetical approach, since measurements being imputed are unrelated to occurrence of the ICE. We compare ICE handling approaches for efficacy studies to those usually employed in BE trials and apply the approaches to a real data example.

16 June 10:30

Title: Multi-Study Causal Forest (MCF): Improving the estimation of heterogeneous treatment effects using auxiliary data

Ashwini Venkatasubramaniam[1], Julian Wolfson[2]

[1]GSK, Stevenage, United Kingdom. [2]University of Minnesota, Minneapolis, USA

Abstract

Recent advances in causal inference and machine learning have led to the development of flexible, nonparametric methods for estimating heterogeneous treatment effects from a single study. However, in many domains, auxiliary data may be available from additional similar studies, each with unique sampling designs and potential confounding mechanisms. Direct pooling of data from these auxiliary sources may lead to bias in estimation of conditional average treatment effects (CATEs) for a given primary source. In this work, we introduce *Multi-Study Causal Forest (MCF)*, a novel extension of the Causal Forest framework that is specifically designed to enhance CATE estimates in a given study population by borrowing information on effect heterogeneity from individual level data collected in other studies. Our method leverages the strengths of Causal Forests for capturing complex, nonlinear causal effects to assess the degree of concordance in the CATE functional form across studies, thereby determining the extent of borrowing. By allowing for both within and between study-level heterogeneity, MCF provides robust estimates of treatment heterogeneity and offer potentially improved precision over single-study approaches. We evaluate the MCF method through a simulation study and an empirical application to Breast Cancer data (primary RCT and auxiliary observational data), demonstrating its ability to enhance inference accuracy and account for nuanced patterns of effect variation across two distinct study settings.

16 June 10:30

Title: From Coders to Drug Developers: The Expanding Role of Statisticians in the Age of AI

Sofia Tapani

Astrazeneca, Gothenburg, Sweden

Abstract

This presentation highlights the rapidly evolving role of statisticians in medicine and technology, especially in an era where AI and automation are becoming increasingly important. Sofia Tapani, Executive Director Statistics at AstraZeneca, shares her journey from problem-solving coder to strategic drug developer, illustrating how statisticians today are key players in driving innovation and making informed decisions in drug development.

Main themes of the presentation include:

- **The evolution of the statistician**: From technical expert to strategic leader and innovator in drug development.
- **Interdisciplinary collaboration**: The importance of working in teams with biostatisticians, programmers, data scientists, and other experts to solve complex problems.
- **AI and automation**: How modern tools and AI are integrated into clinical trials, analysis, and reporting to streamline processes and improve outcomes.
- **Data storytelling and communication**: The need to convey complex data clearly and engagingly to diverse audiences.
- **Ethics and AI governance**: How robust frameworks ensure AI is used ethically and safely within the organization.
- **Future focus**: A call for continuous learning, innovation, and embracing new technologies to shape the future of healthcare.

In summary, the presentation shows how the role of statisticians is expanding and becoming increasingly central to driving development, innovation, and quality in the pharmaceutical industry.

15 June 15:45

Title: Optimising multiplicity adjustment in clinical trials using elicited functions of commercial value and clinical benefit

Alex Spiers, Graham Wheeler, Adrian Mander

GSK, London, United Kingdom

Abstract

A challenge in designing clinical trials with multiple hypotheses (e.g., endpoints, subgroups, or doses) is how to increase the likelihood of discoveries that reflect clinical and commercial priorities whilst controlling the familywise error rate. Fixed-sequence testing is easy to communicate but rarely optimal for what "success" truly means for a development programme. Graph-based closed testing procedures allow flexibility to align with clinical and commercial priorities, but choosing alpha allocation and recycling remains difficult.

Lisovskaja and Burman (2015) showed that multiplicity strategies can be optimised by maximising *expected utility*, a formal expression of the gain obtained by rejecting particular sets of hypotheses and enabling corresponding label claims. Translating this principle into practice requires both a structured way to elicit per-hypothesis utilities from cross-functional stakeholders and a method to optimise graphical tests accordingly.

We developed and piloted an elicitation-optimisation workflow at GSK to quantify the value of rejecting each hypothesis and identify graphical procedures that maximise expected utility conditional on a multivariate test statistic distribution. Building on ideas from decision theory and prior-elicitation frameworks (e.g. SHELF); clinical, regulatory, and commercial stakeholders allocate utilities in terms of the value of potential label claims. These are combined, creating a bespoke trial utility function, which serves as the objective for constrained optimisation of graphical procedures.

We demonstrate how this framework has been implemented across GSK Phase 3 trials using tailored utility functions capturing clinical benefit and commercial value. This approach has led to clearer, more transparent multiplicity strategies better-aligned to development priorities than fixed-sequence testing.

15 June 15:45

Title: Assessing covariate-adjusted risk differences in small-sample trials: A comparative evaluation of statistical methods

Martin Schnuerch[1], Christian Stock[2]

[1]Global Biostatistics & Data Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany. [2]Global Biostatistics & Data Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG, Ingelheim, Germany

Abstract

Binary endpoints are commonly used to measure clinical outcomes in randomized controlled trials. In this context, conditional odds ratios (ORs) based on logistic regression have been routinely used as population-level summary to quantify treatment effects. However, ORs have been criticized for a lack of interpretability, non-collapsibility, and sensitivity to model specification. In response, risk differences (RDs) have gained traction as a more interpretable and clinically relevant measure that better aligns with typical estimands of interest in clinical trials. However, assessing covariate-adjusted RDs, especially in small-sample settings ($N \leq 150$) typical of early-phase trials, remains methodologically challenging. Motivated by recent regulatory guidance and ongoing methodological discussions on covariate adjustment for unconditional estimators, we systematically evaluate a broad set of statistical methods for assessing RDs in a large-scale simulation study, including various g-computation approaches, Mantel-Haenszel methods, and unconditional tests. Our findings reveal that some g-computation variants with parametric variance estimators fail to maintain nominal Type I error rates in small samples. In contrast, bootstrap-based and Mantel-Haenszel methods may offer a more favorable balance between error control and statistical power. Based on our results, we provide practical recommendations to guide practitioners in selecting statistical methods that (1) target a desired estimand, (2) perform reliably under small-sample conditions, and (3) balance robustness, efficiency, and interpretability. Thereby, we hope to support more reliable and clinically meaningful inference from small-sample clinical trials.

16 June 16:15

Title: Comparing Conditional Mean (with Resampling) and Bayesian Imputation under MAR and Reference-Based Strategies in Rare Disease Trials

Imanol Zubizarreta

Denali therapeutics, South San Francisco/Zurich, Switzerland

Abstract

With the aim of handling missing data due to intercurrent events in RCTs under a treatment policy strategy, we evaluated and compared the performance of conditional mean imputation (CMI) with resampling-based inference and Bayesian multiple imputation in small-sample rare disease trials. While Wolbers et al. (2022) evaluated operating characteristics of CMI with resampling in larger trial settings, performance in rare disease context remains unexplored.

We simulated operating characteristics based on longitudinal continuous change from baseline in Vineland-3 ABRS-8 from COMPASS study (NCT05371613) in the rare disease MPSII (n = 42, 2:1 randomization) under a Missing At Random (MAR) mechanism with 15% dropout in the experimental arm. Imputation was performed under both a MAR and reference-based assumption using the copy increments in reference (CIR) strategy. For each, we compared standard Bayesian MI and CMI with bootstrap and jackknife, using the rbmi R package.

Under the null, Bayesian MI was overly conservative (type I error < 4 %), CMI with jackknife achieved nominal control, and CMI with bootstrap showed inflated type I error (> 6 %), consistent with prior findings.

Under the alternative, bootstrap yielded substantial power gains over Bayesian MI (approximately 6-9%) under CIR, while jackknife showed more modest gains than previously reported. Under MAR, all methods produced similar point estimates, but unexpectedly, jackknife showed slightly lower power while bootstrap showed still modest gains. These differences likely reflect finite-sample bias in standard error estimation.

Although CMI with jackknife performs well in large trials, finite-sample bias may alter operating characteristics in small-sample rare disease settings.

17 June 11:15

Title: Accelerating Alzheimer's Research: a modular framework for exploring Bayesian disease progression models

Oana Petrof[1], Dave Lunn[2], Aris Perperoglou[2]

[1]GSK, Wavre, Belgium. [2]GSK, London, United Kingdom

Abstract

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by considerable heterogeneity in clinical presentation and progression. Building upon the framework proposed by Lars Lau Raket et al. in *Statistical Disease Progression Modelling in Alzheimer Disease*, we use a statistical approach for modelling disease age using the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

The method provides a common time-scale representation of disease progression by aligning individual cognitive trajectories using a Bayesian hierarchical joint model that simultaneously operates on the response (cognitive scales) and the disease time (age). To capture the cognitive decline, we employ a nonlinear mixed-effects model.

Our primary contribution is a two-stage modularization approach that decouples the computationally intensive estimation of disease age from the subsequent covariate analysis. First, a single, covariate-free model is fit once to establish posterior distributions for each patient's disease age. The power of this modular design lies in the second stage: it enables the rapid and rigorous exploration of numerous, distinct covariate models—including joint models for multiple endpoints—without needing to re-run the expensive initial fit. This process formally propagates the parameter uncertainty captured in the disease age posteriors.

The second-stage model is specifically designed to address real-world data challenges, including covariate missingness, the non-normal distribution of biomarkers, and the harmonization of data from different assays. This framework provides a robust method for assessing biomarker influence in the context of cognitive decline. By improving personalized prognostication and facilitating more effective therapeutic stratification, our methodological refinements advance the statistical modelling of AD progression.

15 June 15:45

Title: PREDOSE : Pharmacometrically-Refined Early-phase Dose Optimization design for Oncology Study Enhancement

Damitri Kundu

Eli Lilly and Company, Bengaluru, India

Abstract

Traditional Phase I oncology trials have focused on identifying the maximum tolerated dose (MTD), operating under the assumption that higher doses increase both efficacy and toxicity. However, such monotonic dose-response relationship is often not observed with newer anti-cancer agents, leading to a paradigm shift towards determining the optimal biological dose (OBD) that offers a better balance between efficacy and safety. In response, the US Food and Drug Administration (FDA) now advise the inclusion of pharmacokinetic (PK) and pharmacodynamic (PD) data and recommends randomisation after dose escalation in trial designs for identifying the OBD. Despite advancements in utilising PK/PD data for OBD determination, practical adoption remains challenging due to complex analytical models and the lack of accessible tools.

This study introduces a novel dose-optimisation framework for early-phase oncology trials, utilising a two-stage process that integrates patient-level PK and PD data. In Phase 1a, joint modelling of drug concentration, biomarker, efficacy, and toxicity data is used to identify candidate doses via a utility-based method. Doses with favourable profiles are selected by maximising an expected utility function that considers safety, efficacy, PK, and PD outcomes. These candidate doses progress to Phase 1b, where randomised allocation and evaluation further refine indication specific OBD selection based on efficacy and safety. The framework's effectiveness will be assessed through simulation studies, and a user-friendly R-shiny application has been developed to facilitate clinical application.

Keywords : Bayesian Adaptive designs, Pharmacokinetics, Pharmacodynamics, Project Optimus, Optimal Biological Dose.

15 June 13:30

Title: All PICOs Great and Small; Dealing with small subpopulations in the EU HTA Landscape

Dave Gelb[1], Joel Baldwin[2], Ehsan Masoudi[3], Milana Ivkovic[4], Lovisa Berggren[5], Larry Leon[6]

[1]MSD, Zurich, Switzerland. [2]Cogent Skills Limited, Cheshire, United Kingdom. [3]Eli Lilly, Bad Homburg, Germany. [4]Novo Nordisk A/S, Søborg, Denmark. [5]UCB Biosciences GmbH, Monheim, Germany. [6]Merck, North Wales, USA

Abstract

Within the EU Joint Clinical Assessment (JCA), a subpopulation — patient population subset of the therapeutic indication — may be specified by Member States to reflect national clinical practices, patient demographics or local decision-making contexts. In the JCA process, subpopulations lead to distinct PICOs (population, intervention, comparator, outcome) and can be viewed as different from subgroups, which can be used to investigate effect modification within a PICO. Member States may request analyses in small subpopulations, which could lead to statistical challenges – notably increased variability in effect estimates and confidence intervals, leading to biased interpretation.

Objective: to demonstrate that simulations add rigour to current practice (rule-of-thumb) in evaluating appropriateness of analysing a subpopulation.

Simulations were run using German Breast Cancer Study Group dataset. Uncertainty introduced by small subpopulations was analysed by examining upper CI limits and point estimates across simulations on overall survival, to quantify risks of overinterpreting extreme values in small subpopulations, when treatment is beneficial.

Simulation results show that applying current rules-of-thumb (e.g. analysing subpopulations with N≥20 participants/arm) can lead to high upper CI limits. For example, results included a consistent treatment effect for a subpopulation (N=66), with mean observed upper CI limit = 3.84, compared to 0.87 (overall study population, N=686).

From our results, compared to rules-of-thumb, simulations better support the decision problem of whether to analyse small subpopulations in the HTA context. Examining behaviour of point estimates and CI limits provides a consistent, statistically rigorous approach to transparently assess appropriateness of analyses in a given (study-specific) subpopulation.

15 June 15:45

Title: A spectrum of causal estimands – differences in dosing adherence patterns

Anna Menacher[1], Yi Li[2], Nerissa Nance[1], Mark van der Laan[2], Maya Petersen[2]

[1]Novo Nordisk, Bagsværd, Denmark. [2]UC Berkeley, Berkeley, USA

Abstract

Randomized control trials often suffer from non-adherence to assigned interventions, including dose escalation protocols. We provide a spectrum of causal estimands to address these challenges, varying treatment adherence over time by following a hypothetical intervention versus leaving its natural value. Specifically, we define estimands by ensuring that the longitudinal cumulative probability of receiving the assigned treatment regime (i.e. the time-varying propensity score (PS)) stays below some user-specified threshold α, where α=1 represents intention-to-treat and α=0 corresponds to a static regime of perfect adherence. Increasing α between [0,1] increases the extent to which adherence takes its natural value, both shifting the true effect size (providing insights into the impact of non-adherence) and increasing data support. We use longitudinal targeted maximum likelihood estimation, and employ two thresholding strategies: 1) 'simple' thresholding by iteratively thresholding the time-varying PS from the static regime in each interval, and 2) 'sticky' thresholding by iteratively thresholding the time-varying PS, where the time-varying PS only updates if the product of PS exceed α. In simulation studies, both strategies are consistent for the intention-to-treat (α=1) and static perfect adherence (α=0) estimands; however, sticky thresholding results in effect sizes closer to the static regime. Bias is reduced and oracle coverage is recovered with both thresholding strategies, compared to the static regime. The known issue of under-estimation of influence curve-based variance estimation is circumvented with application of bootstrap-based variance estimation.

16 June 14:00

Title: Joint session by the Biomarkers ESIG and the Treatment Effect Heterogeneity SIG: Biomarker Discovery Across the Dimensionality Ladder

Marie-Karelle Riviere[1], Hugo Hadjur[1], Laura Schlieker[2], Mathias Cardner[3], David Svensson[3], Ashwini Venkatasubramaniam[4]

[1]Saryga, Tournus, France. [2]Staburo GmbH, Munich, Germany. [3]AstraZeneca AB, Gothenburg, Sweden. [4]GSK, London, United Kingdom

Abstract

This joint Special Interest Group session features four talks that trace biomarker discovery along the "Dimensionality Ladder"—from low-dimensional, hypothesis-driven markers to high-dimensional omics. We examine methodological rigour, practical considerations, and translational impact at each rung, highlighting how study design, statistical power, and validation requirements shift with increasing dimensionality. Predictive biomarkers are central to precision medicine, yet early-phase trials pose challenges due to limited samples and the need for robust evidence for regulatory decisions.

The four talks cover (1) a novel method for assessing single-biomarker predictiveness in small samples; (2) approaches for selecting optimal cut-offs for continuous biomarkers in small, unbalanced cohorts; (3) an AI/ML framework involving separate simulation and benchmarking teams to identify prognostic and predictive biomarkers in complex datasets; and (4) a Bayesian decision framework to integrate statistical evidence with biological plausibility and commercial considerations.

The session will illustrate typical problems encountered in the practice and ongoing co-work activities in the two SIGs.

Presentation 1: A novel approach to assess the predictiveness of a continuous biomarker in early phases of drug development (Marie-Karelle Riviere)

Identifying and quantifying predictive biomarkers is essential for personalised medicine and patient-centric drug development. However, in early-stage studies with limited data, assessing a biomarker's predictive value can be challenging and potentially misleading. Most existing methods assume moderate-to-large sample

sizes. We propose a novel, flexible approach inspired by the Kolmogorov–Smirnov distance to evaluate the predictiveness of continuous biomarkers in small-sample settings. Simulation studies show that our method achieves higher power than existing approaches across various scenarios while maintaining control of the type-I error rate at a pre-specified level.

Presentation 2: Evaluating methods for biomarker cut-off detection: a comparative study in small and unbalanced samples (Hugo Hadjur)

In personalised medicine and patient-centric development, biomarkers—measurable indicators of biological processes—are central to validating new treatments. Predictive biomarkers identify patients most likely to benefit and guide treatment assignment, but determining accurate cut-off thresholds is challenging, especially with limited, small, and unbalanced samples where some methods can be biased or imprecise. Most research on cut-off identification uses moderate or large samples. We evaluate methods for cut-offs in continuous biomarkers, analysing their theoretical properties and comparing performance via simulations across scenarios (including small, unbalanced samples and varied biomarker distributions). The results clarify trade-offs among methods and offer practical guidance for selecting robust approaches in challenging clinical settings.

Presentation 3: Machine Learning in Precision Medicine: A Collaborative Approach (Laura Schlieker)

Quantitative and qualitative assessment of prognostic and predictive biomarkers (treatment-effect modifiers) is essential for decision-making in precision-medicine drug development, demanding sophisticated analysis of complex, high-dimensional data with Artificial Intelligence (AI) and Machine Learning (ML). A dedicated ML sub stream within the PSI/EFPSI Biomarkers ESIG runs a hands-on project on treatment-effect heterogeneity using an agile two-team model—simulation and method benchmarking. The simulation team creates increasingly complex synthetic datasets, tuning p-to-n ratios, monotonicity/non-linearity, higher-order interactions, and gene network-like correlation structures. The blinded methods team analyses these datasets with diverse AI/ML tools to identify true prognostic and predictive biomarkers and infer the data-generating process, with cycle-end reviews guiding the next iteration. We will demonstrate how this interactive model engages contributors and fosters creativity, and we will showcase method performance—including ML-based meta-learners—in high-dimensional precision medicine, highlighting their potential to strengthen clinical development programs.

Presentation 4: A Bayesian precision-medicine decision framework for pursuing biologically plausible predictive biomarkers in early clinical development (Mathias Cardner)

In early clinical development, decisions to pursue predictive biomarkers must balance risk and benefit. In trials powered for overall efficacy, testing a predictive biomarker often has a much higher false-negative than false-positive risk, and sufficient sample sizes rarely arrive before later phases—delaying companion diagnostic development. Regulators acknowledge this and encourage methods that prioritise biological plausibility and replicability over strict significance. We propose a Bayesian decision framework based on the posterior conditional average treatment effect (CATE) in biomarker-defined subgroups, comparing posterior CATE to efficacy targets in the product profile and combining evidence across endpoints via Dempster–Shafer theory. Simulations show how priors on treatment-effect heterogeneity can suppress or reveal subgroup signals. The framework integrates evidence across endpoints and weights it by commercial opportunity, biomarker sampling difficulty, and diagnostic feasibility, enabling more informed early-phase decisions that can accelerate targeted therapies and companion diagnostics.

15 June 13:30

Title: Not Just Another Estimands Talk: Practical Strategies for Cross-Functional Engagement to Ensure Meaningful, Fit for Purpose Estimands

Emily Wood

Veramed, London, United Kingdom

Abstract

Six years after the release of the ICH E9(R1) Addendum, the pharmaceutical statistical community has made significant strides in developing training, publications, and guidance to support the application of estimands. However, beyond late-phase trials and large pharmaceutical organisations, implementation and reporting remain inconsistent. Cross-functional engagement is often absent, with decisions around estimands deferred to statisticians or estimands avoided altogether. Clinical teams may resist defining or addressing intercurrent events during protocol development. Regulatory feedback suggests that when estimands are included, detail and justification are often insufficient, with treatment policy strategies commonly adopted by default (EFSPI Regulatory Workshop, 2025).

Unlike most presentations on estimands, this session shifts the focus from technical implementation to practical communication with cross-functional teams. It is aimed at all statisticians, although some prior knowledge of estimands would be helpful. It will explore how statisticians can initiate and sustain meaningful dialogue with clinical colleagues. This is of particular use for those working with small biotech and pharmaceutical companies where statistical support may be limited and the estimands framework perceived as complex, inaccessible, or irrelevant. These obstacles and misconceptions will be addressed with practical advice on how to challenge assumptions, encourage cross-functional engagement and ensure that the downstream impact of decisions on dataset utility and interpretability is fully considered.

Participants will leave the session with practical tools and suggestions for actionable strategies to facilitate clearer communication around estimand components and to help foster collaborative decision making. These strategies will be illustrated using examples from early-phase oncology and CNS studies.

17 June 12:35

Title: Rigorous Type I error control for randomized BOP2-TE designs under minimal assumptions

Owen Leete[1], Alexander Ooms[2], Matt Psioda[1]

[1]GSK, Philadelphia, USA. [2]GSK, London, United Kingdom

Abstract

The Bayesian Optimal Phase 2 design for jointly monitoring efficacy and toxicity (BOP2-TE) is a pragmatic design for running single-arm Phase 2 studies. It allows arbitrarily many interim analyses for early stopping due to futility or toxicity via a look-up table whilst maintaining control of type I error across three null hypotheses: the drug is ineffective, unacceptably toxic, or both, through its Bayesian joint-modelling approach. However, no suitable extension to a randomized (particularly dose-optimization) setting exists that maintains strict type I error control and addresses across-arm termination considerations (e.g., if the highest dosing arm is stopped for futility, should all lower dosing arms also be stopped?).

Isotonic regression has been proposed to enforce monotonicity for estimated dose-response and dose-toxicity relationships to tackle across-arm termination considerations. When the BOP2-TE testing methodology is applied to estimates that have been smoothed via isotonic regression, type I error rates are no longer strictly controlled - even when monotonicity holds.

We propose a modification that requires the minimal assumptions that dose-response and dose-toxicity relationships are non-decreasing. This approach makes the problem of type I error rate control tractable, reducing it to the verification of a limited set of boundary scenarios that do not depend on the true dose-response or dose-toxicity relationships. The resulting design preserves the operational simplicity and straightforward decision rules of the original BOP2-TE while providing rigorous type I error rate control in a randomized setting, offering a practical design when evaluating efficacy and safety across doses in a Phase 2 study.

15 June 15:45

Title: Incorporating prognostic scores in time-to-event analysis

Harry Parr, Doug Thompson, Tasos Papanikos, Aris Perperoglou

GSK, London, United Kingdom

Abstract

Leveraging prognostic information in RCTs through covariate adjustment is well established. This can be further complemented by prognostic scores, which are growing in interest (Schuler, 2021; Siegfried 2023), where prognostic information is combined into an overall derived score (Hansen, 2008). The benefits and properties of these scores are well-founded, namely improving precision in estimating the treatment effect. However there have been fewer applications used in non-linear models, e.g. responder, rate, and time-to-event endpoints.

The estimate of interest, e.g. a time-to-event hazard ratio (HR) endpoint, can move further away from the null in the presence of adjusting for other covariates. The reason being the HR is a non-collapsable measure, hence the resulting marginal and conditional estimates are not equivalent. Often a marginal effect estimate is required, in which case alternate causal inference estimation techniques may be needed for such adjusted non-linear models (e.g. g-computation or TMLE).

We investigate the properties of prognostic scores under time-to-event outcomes via simulations, in simple cases of no censoring, and non-informative censoring mechanisms. We generate prognostic scores of varying strengths (from nil to strong) to assess impact on power and HR estimates. We also propose alternative estimands, such as the restricted mean survival time (which are collapsable) and whether additional efficiency gains can be made using causal inference methods. Full results will be presented.

Here we provide a framework to incorporate prognostic scores into time-to-event analyses, providing some recommendations for estimation and quantifying the expected precision improvements.

17 June 12:35

Title: Efficient modelling of complex responder endpoints to improve trial power

James Wason

Newcastle University, Newcastle upon Tyne, United Kingdom

Abstract

In some clinical areas, trials use composite responder endpoints that classify participants as responders based on several variables, some of which are continuous. Traditionally these endpoints are analysed as binary, resulting in a large amount of information being discarded through dichotomisation of continuous variables.

Augmented binary methods have been proposed to analyse these endpoints more efficiently by utilising the continuous variables to improve the statistical precision. Importantly, the method retains the clinically and regulatorily defined responder criterion while improving precision.

This talk will provide an overview of how the augmented binary method works, using a latent variable model to estimate the probability of response. Simulation studies and real trial analyses will be presented to demonstrate how the method increases power equivalently to increasing the sample size by at least 30%, compared to using a binary analysis. The efficiency gain can be much higher in some circumstances.

Finally, I will describe current case studies in which this approach is being implemented prospectively in academic-led clinical trials, particularly in rare diseases where efficient designs are crucial for feasibility. Use of the approach has led to substantial reductions in cost and duration of trials.

17 June 11:15

Title: On the interplay between prior weight and variance of the robustification component in Robust Mixture Prior Bayesian Dynamic Borrowing approach

Marco Ratta[1,2], Gaëlle Saint-Hilary[1], Mauro Gasparini[2], Pavel Mozgunov[1,3]

[1]Saryga, Tournus, France. [2]Politecnico di Torino, Torino, Italy. [3]University of Cambridge, Cambridge, United Kingdom

Abstract

Robust Mixture Prior (RMP) is a popular Bayesian dynamic borrowing method, which combines an informative historical distribution with a less informative component (referred as *robustification component*) in a mixture prior to enhance the efficiency of hybrid-control randomized trials. Current practice typically focuses solely on the selection of the prior weight that governs the relative influence of these two components, often fixing the variance of the robustification component to that of a single observation. In this study we demonstrate that the performance of RMPs critically depends on the *joint* selection of both weight and variance of the robustification component. In particular, we show that a wide range of weight-variance pairs can yield practically identical posterior inferences (in particular regions of the parameter space) and that large variance robust components may be employed without incurring in the so called *Lindley's paradox*. We further show that the use of large variance robustification components leads to improved asymptotic Type I error control and enhanced robustness of the RMP to the specification of the location parameter of the robustification component. Finally, we leverage these theoretical results to propose a novel and practical hyper-parameter elicitation routine.

16 June 14:00

Title: Including quantitative benefit-risk assessment in seamlessphase 2/3 designs with dose selection

Marco Ratta[1,2], Donia Skanji[3], Zhaoyang Teng[4], Gaëlle Saint-Hilary[1], Mauro Gasparini[5], Pavel Mozgunov[1,6]

[1]Saryga, Tournus, France. [2]Politecnico di Torino, Torino, France. [3]Servier, Gif-sur-Yvette, France. [4]Astellas Pharma, Boston, USA. [5]Politecnico di Torino, Torino, Italy. [6]University of Cambridge, Cambridge, United Kingdom

Abstract

Seamless phase II/III designs are increasingly employed to accelerate drug development, particularly in settings where multiple doses of the same compound need to be evaluated in parallel. In such designs, a restricted subset of doses is selected at the first stage of the trial and subsequently carried forward to the second stage, where efficacy is assessed against control. Dose selection is typically based on a single efficacy endpoint, ensuring consistency between interim decision-making and final hypothesis testing. However, safety is generally not incorporated formally into the selection procedure at the design stage, which may be problematic, specifically because potentially harmful doses could be advanced if deemed effective. In this work, we propose a novel multi-arm two-stage design in which the dose with the most favorable benefit-risk profile at the interim analysis is selected and continued into phase III. Benefit-risk assessment is conducted using probabilistic multi-criteria decision analysis (MCDA), where information from multiple efficacy and safety endpoints is combined into a single aggregated score. For the proposed design, we derive an analytical expression for the decision boundary, ensuring strong control of the type I error rate, and provide guidance on sample size determination. The operating characteristics are evaluated through extensive simulation studies and compared with existing designs from the literature, demonstrating that the proposed approach is particularly effective in terminating potentially harmful doses at an early stage.

15 June 11:00

Title: The impact of backfilling on early phase dose optimisation trials in oncology

James Willard[1], Thomas Jaki[1,2], Burak Kursad Gunhan[3], Christina Habermehl[3], Anja Victor[3], Pavel Mozgunov[1]

[1]MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom. [2]Faculty of Informatics and Data Science, University of Regensburg, Regensburg, Germany. [3]Merck Healthcare KGaA, Darmstadt, Germany

Abstract

Early phase dose escalation trials in oncology were originally designed for the development of cytotoxic chemotherapies, where larger doses often produced greater benefit. As such, these designs historically focused on identifying the largest safe dose, the so-called maximum tolerated dose (MTD). Recently, the FDA's Project Optimus highlighted how modern targeted therapies can provide benefit at doses lower than the MTD and so identifying these doses via dose optimisation has become a major objective of early phase trials. However, dose escalation designs target the MTD and may collect less information on lower doses, potentially making dose optimisation more challenging. To overcome this, backfilling throughout dose escalation has been proposed. During backfilling, patients are assigned to safe doses that are lower than the current highest dose used in testing. While growing in popularity, backfilling's impact on dose optimisation is still poorly understood. Several questions remain unanswered in the literature, including when to begin backfilling, how many patients to use for backfilling, and how to select backfill doses. In this work, we take a first step toward answering these questions by discussing findings from an extensive simulation study that investigates the impact of a variety of backfilling approaches on dose optimisation. We illustrate how backfilling affects parameter estimation, escalation decisions, and adaptive stopping rules and show that the impact of backfilling depends on the underlying data generating mechanism and objectives of the trial. We close by offering practical recommendations for practitioners who are considering using backfilling in their dose optimisation trials.

17 June 12:35

Title: Integrating Preclinical Insights for Adaptive Dose Escalation in Phase I Oncology Trials: A Methodological Framework for Enhanced Efficiency

Helen Barnett[1], Pavel Mozgunov[2,3], Mélanie Guhl[3], Fulvio Di Stefano[4], Donia Skanji[5], Sandrine Guilleminot[6], Gaëlle Saint-Hilary[3], Marie-Karelle Rivière[3]

[1]School of Mathematical Sciences, Lancaster University, Lancaster, United Kingdom. [2]MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom. [3]Department of Statistical Methodology, Saryga, Paris, France. [4]Dipartimento di Scienze Matematiche (DISMA) "Giuseppe Luigi Lagrange", Politecnico di Torino, Torino, Italy. [5]Department of Clinical Statistics, Institut de Recherches Internationales Servier, Paris-Saclay, France. [6]Department of Translational Statistics, Institut de Recherches Internationales Servier, Paris-Saclay, France

Abstract

Leveraging preclinical prior information has the potential to enhance the efficiency of Phase I oncology trials if used appropriately. Preclinical data are typically used to determine starting doses, but their potential to inform dose escalation remains underexplored.

A comparison of the meta-analytic predictive (MAP) prior approach and the power prior approach was undertaken in the context of the Bayesian Logistic Regression Model (BLRM). A novel methodology to determine the parameters of both approaches—based on external evidence of animal–human concordance—was introduced to support transparent specification of prior exchangeability. In addition, the commonly used escalation-with-overdose-control (EWOC) criterion was extended via an additional rule allowing less conservative escalation when supported by data. Simulation studies based on an oncology case study were conducted to evaluate operating characteristics.

Incorporating animal data through the flexible MAP prior provided greater robustness when concordance between animal and human data was uncertain, whereas the power prior performed best when strong agreement existed. The extended EWOC criterion enabled escalation efficiency without compromising patient safety.

The inclusion of preclinical information is recommended using the MAP prior with a justified specification of prior exchangeability. However, the benefit of including animal data must be weighed against potential losses. This framework provides practical guidance for efficient and safe design of early-phase oncology trials.

16 June 10:30

Title: Leveraging LLMs to navigate complex language in clinical trial informed consents forms

Mbangula Lameck Amugongo

Boehringer Ingelheim, Biberach, Germany

Abstract

Regulatory frameworks such as the Belmont Report, Common Rule, and the Declaration of Helsinki provide essential ethical guidance for research involving human subjects, mandating informed consent to ensure participants are properly briefed on study purpose, so they can make an informed decision about whether to participate. The Department of Health and Human Services further stipulates that informed consent forms (ICFs) should be concise, comprehensible, and accessible, but over time, these documents have become increasingly complex and difficult to read.

The advent of powerful large language Models (LLMs) present opportunities to navigate complex language in ICFs. LLMs have already shown effectiveness in biomedical tasks like information extraction and patient trial matching, yet they remain susceptible to hallucinations, generating inaccurate or misleading content. Retrieval augmented generation (RAG) has been identified as a method to reduce such errors, especially in high-stakes domains like healthcare. This study explores the integration of LLMs into clinical workflows for ICF review, evaluating their ability to understand and process complex linguistic structures. By employing RAG-like pipelines with prompt optimisation, we successfully extracted data re-use information, identified nuanced international regulatory requirements, and processed ICFs in multiple languages. Automated review reduced evaluation time from 15–20 minutes to about 5–7 minutes per ICF. This approach has the potential to significantly alleviate administrative burdens by automating labour-intensive processes, while also generating insights that could inform the standardisation of ICF creation. Ultimately, these advancements may contribute to reducing the complexity of ICFs, thereby improving their readability and comprehensibility for patients.

16 June 16:15

Title: From Nightingale to Now: Why Visualisations Are Still Essential in The Statisticians' Toolkit

Bethany George

UCB Pharma, Slough, United Kingdom

Abstract

In the 1850s, Florence Nightingale revolutionised medical statistics by using visualisations to advocate for reform in battlefield hospitals. Over 170 years later, statisticians have access to vast amounts of data and increasingly complex analytical methods to make sense of them. Yet in rare-disease research, where data are often sparse, inconsistent, and incomplete, these standard statistical approaches cannot be applied.

This presentation explores the challenges of characterising rare diseases with limited data. I will share insight from working with a dataset of 62 patients, derived from historical medical records, and describe how simple techniques, supported by visualisations, helped overcome key obstacles.

The dataset was both unstructured and sparse, lacking standard endpoints or consistent timepoints. I will outline the strategies used to clean and organise the data, including the use of assumptions to infer missing information, and how visualisation played a critical role in validating these assumptions. I will also describe the process of analysing the data focusing on navigating the constraints of sparse datasets, mitigating bias and favouring simple techniques over complex statistical models.

Finally, I will demonstrate how visualisations provided a clearer and more accessible way to communicate results than traditional tables and numerical summaries - making findings easier to interpret for clinical teams and offering a more transparent view of uncertainty and missingness.

This case study highlights how, even in the age of advanced analytics, visualisation remains a powerful and essential tool for statisticians - especially when working with imperfect data.

17 June 09:15

Title: Regulatory Hot Topics: Non-inferiority and equivalence comparisons in clinical trials; Use of external controls and real world evidence to support regulatory decision-making

David Wright[1], Helle Lynggaard[2], Florian Lasch[3]

[1]AstraZeneca, Cambridge, United Kingdom. [2]Novo Nordisk, Copenhagen, Denmark. [3]European Medicines Agency, Amsterdam, Netherlands

This session will cover two topics where recent developments are underway:

Non-inferiority and equivalence comparisons in clinical trials

In 2024 the EMA CHMP published a Concept Paper on the Development of a Guideline on Non-Inferiority and Equivalence Comparisons. The draft guideline has just been published: Draft guideline on non-inferiority and equivalence comparisons in clinical trials and the deadline for comments is the end of May 2026. This is perfect timing for the revised guideline and industry comments to be discussed.

Use of external controls and real world evidence to support regulatory decision-making

The need for guidance on use of external controls and real world evidence respectively have been acknowledged by regulatory authorities. With various initiatives in progress including development of an EMA Reflection Paper and workshop on the use of external controls for evidence generation in regulatory decision-making, and MHRA guidance on the use of real-world data in clinical studies to support regulatory decisions, it is timely for this broad topic to be discussed.

16 June 10:30

Title: What if We've Been Looking at the Wrong Data? Reimagining Clinical Trial Success Prediction using AI

Leo Fournier[1,2], Juan Martinez[3], Tarun Naithani[4], Krishna Sai Bellamkonda[5], Wenting Wang[6], Nils Ternes[1], Christelle Reynes[2]

[1]Sanofi R&D, Montpellier, France. [2]University of Montpellier, Montpellier, France. [3]Cytel, London, United Kingdom. [4]Sanofi R&D, Hyderabad, India. [5]Sanofi R&D, Frankfurt, Germany. [6]Sanofi R&D, Cambridge, USA

Abstract

Background: The prediction of clinical trial success has very recently mobilized increasingly AI-based sophisticated approaches. Models such as HINT or SPOT established benchmark performances (ROC-AUC 0.65), followed by complex architectures combining knowledge graphs and Transformers. Paradoxically, these methodological innovations have not produced substantial predictive gains, while their low explainability limits pharmaceutical industry adoption.

Objective: Faced with this performance plateau, we propose an alternative hypothesis: limitations stem more from data quality than model architecture. Our work aims to systematically evaluate this hypothesis and explore an AI-driven data enrichment-centered approach.

Methods: We conducted a critical audit of the main public benchmarks (TOP, CTO, TrialPanorama), assessingicompleteness, consistency, and label reliability. A feasibility study on 300+ clinical trials evaluated the possibility of automating outcome annotation using LLM-based techniques and web scraping. We developed an AI-based enrichment strategy integrating underexploited biological variables: pharmacokinetic data, preclinical biomarkers, and inter-phase links, culminating in a new machine learning predictive model that incorporates these data sources while prioritizing explainability.

Results: Our audit of 500 manually annotated trials reveals significant labeling error rates: 41% for CTO, 9.6% for TrialPanorama, and 8.3% for TOP, confirmingifundamental data quality issues. Automated annotation shows encouraging performance. Exploratory analysis suggests that missing biological variables constitute a significant limiting factor for current predictive models.

Conclusions: This work questions the current paradigm favoring algorithmic complexity in AI-based prediction of clinical trial outcome. We propose that rigorous data curation and AI-enhanced enrichment could unlock significant predictive gains with direct implications for development cost optimization and clinical trial ethics.

15 June 11:00

Title: Early Phase Dose-Finding Designs for CAR-T cell Therapies

Weishi Chen[1], Pavel Mozgunov[1], Xavier Paoletti[2]

[1]University of Cambridge, Cambridge, United Kingdom. [2]Institute Curie, Paris, France

Abstract

Chimeric Antigen Receptor (CAR)-T cell therapy is an immunotherapy that revolutionises the treatment of relapsed/refractory lymphoma and leukaemia. It demonstrates higher response rates, superior mid-to-long term overall survival, and lower toxicity compared to standard treatments. However, the lack of defined dose-limiting toxicity (DLT) and unclear dose-effect relationship mean that traditional phase I designs cannot accurately select the optimal dose (OD). Besides clinical outcomes, CAR-T cell expansion from serial blood samples is measured at various timepoints. Meta-analysis suggests that efficacy has a positive association with expansion but no association with toxicity. Hence, we propose a novel dose-finding design that utilises both toxicity and activity endpoints to locate the OD. CAR-T cell expansion is a more sensitive activity indicator than the short-term clinical responses traditionally used.

Two main profiles of cell-evolution have been reported: the injected cells either exhaust and are eliminated from the blood, or they proliferate and are maintained before elimination. A Bi-Exponential model is employed for the trajectories of CAR-T cell expansions, with parameters estimated under a Bayesian framework. The model is motivated by biological concerns and is flexible to accommodate different cell-expansion curve shapes. Since there are no clear efficacy criteria, three criteria are considered upon clinician discussions: 1) the number of cells at specific time points, 2) the duration before all cells are eliminated, and 3) the area under the cell-expansion curve. Simulation studies demonstrate high OD selection accuracy even under small sample sizes.

16 June 16:15

Title: Novel bayesian prediction of event times using mixture model for blinded randomized controlled trials

Jingyan (Janet) Fu[1], Dan Zhao[2], Donia SKANJI[3], Hua Liu[2], Rui (Sammi) Tang[4], Ying Yuan[5]

[1]Rice University, US, USA. [2]SERVIER, US, USA. [3]SERVIER, FRANCE, France. [4]ASTELLAS, US, USA. [5]Mdanderson, US, USA

Abstract

Accurate prediction of key milestone dates, such as the timing of interim and final analyses, is crucial in event-driven clinical trials with time-to-event endpoints. These predictions facilitate timely decision-making, enhance strategic planning and optimize resource allocation while minimizing patient exposure to potentially ineffective or harmful therapies.

Existing methods for predicting event timing typically assume identical time-to-event distributions for treatment and control arms in blinded randomized clinical trials (RCTs).

This assumption often fails to hold in practice, leading to biased predictions. To address this issue, we propose a novel Bayesian Prediction of Event Times (BayesPET) method that allows for distinct time-to-event distributions between arms in blinded RCTs. We employ a mixture Weibull model for the observed interim event times, while addressing the critical challenge of label-switching in mixture models through truncated priors.

Through extensive simulations and real world applications to phase 3 clinical trials, we demonstrate the BayesPET produces superior predictive performance in both blinded and unblinded settings, supporting effective trial execution and accelerating the development of new therapies.

A BayesPET package is currently in development and will be made publicly available

15 June 15:45

Title: Sensitivity Analysis of Missing Pharmacokinetic Samples in a Simulated Study of Rapid Acting Psychedelics

Nathan Patrick Burns[1,2], Pau Aceves Baldo[1], Rachael MacIsaac[1]

[1]GH Research, Dublin, Ireland. [2]University of Strathclyde, Glasgow, United Kingdom

Abstract

Pharmacokinetic (PK) sampling in psychedelic trials can be more demanding than in studies of non-psychoactive drugs. Most psychedelics exhibit rapid onset of psychoactive effects (generally within 30 seconds) and a similarly fast PK profile, requiring frequent blood sampling at short intervals. These potential operational challenges, particularly during intense psychoactive effects, may lead to sampling delays and missing data. At the same time, overly conservative protocol deviations can lead to the unnecessary removal of data. These factors may affect key PK parameter estimates, for example Cmax (peak concentration) and AUClast (total exposure). Understanding the impact of missing measurements and designing optimal protocol deviations is therefore essential for accurate PK analysis and interpretation of pharmacodynamic and safety endpoints.

This sensitivity analysis evaluated how missing samples could influence the estimation of PK parameters using simulated serum concentration data. Increasing numbers of samples were sequentially removed from complete profiles, and percentage errors in PK parameters were quantified by comparing the complete and modified profiles. The magnitude of the error varied depending on the timing and number of missing samples, and the parameter being estimated. Removal of samples in the early part of the PK profile (1-4 minutes post-dose) had a larger impact on Cmax and AUClast compared to the removal of late samples (7-240 minutes). These results informed protocol deviation planning and statistical analysis plans by providing a quantitative framework for assessing the impact of missing samples. Overall, this approach supports trial planning for drugs with rapid PK, particularly psychedelics.

15 June 15:45

Title: Multiple Endpoints in Early Phase Decision Making

Laura Barker[1], Belanger Em[2], Aleksandra Buchowicz[3], Chris Gibbs[1], Danny Lu[2], Paul Frewer[1]

[1]AstraZeneca, Cambridge, United Kingdom. [2]AstraZeneca, Mississauga, Canada. [3]AstraZeneca, Warsaw, Poland

Abstract

Clear and evidence-based decision-making frameworks are essential in early drug development. At AstraZeneca, an effective three-outcome framework has been developed, with criteria for a go, consider and stop decision being calculated based on a pre-specified target value and lower reference value for a target endpoint [1]. When using multiple endpoints, the approach was to assume independence and calculate the decision criteria separately, to generate joint operating characteristics. However, in the case where endpoints are correlated, this may lead to inaccurate operating characteristics. Therefore, we propose a new method for generating joint operating characteristics for multiple endpoints, which accounts for correlation by employing the Ruscio-Kaczetow method to rearrange independently generated samples so that they exhibit the desired Pearson correlation [2]. By reflecting the true joint behavior of endpoints, we evaluate the performance of the new approach and confirm its practical benefits for early development decision-making.

[1] Frewer, P., Mitchell, P., Watkins, C., & Matcham, J. (2016). Decision-making in early clinical drug development. Pharmaceutical statistics, 15(3), 255–263. https://doi.org/10.1002/pst.1746

[2] Ruscio, J., & Kaczetow, W. (2008). Simulating Multivariate Nonnormal Data Using an Iterative Algorithm. Multivariate Behavioral Research, 43(3), 355–381. https://doi.org/10.1080/00273170802285693

17 June 12:35

Title: From Classroom to Clinical Trials: How PSI Schools is inspiring the next generation of statisticians

Ciara Lucas-Garner

Amgen, Cambridge, United Kingdom. PSI Schools Committee, Cambridge, United Kingdom

Abstract

Inspiring the next generation of medical statisticians begins long before students make career decisions - it starts with curiosity in the classroom. The PSI Schools Committee is dedicated to nurturing that curiosity and showing students how STEM skills can shape the future of healthcare, particularly within the pharmaceutical industry.

As the committee evolves, we are expanding our outreach to provide inclusive, accessible opportunities for students and teachers. New initiatives include a STEM-themed escape room that brings problem-solving to life, along with updated resources to help members deliver engaging outreach experiences.

Our mission is to raise awareness of STEM careers and increase opportunities for students who may not otherwise have access to them. Many young people from under-resourced schools or lower socio-economic backgrounds still face barriers to engaging with inspiring, practical STEM learning. Our committee aims to develop interactive, real-world experiences that make STEM more visible, relevant, and attainable for all.

This presentation will share insights from our projects and showcase the creativity and commitment of our volunteers. Above all, it highlights the crucial role of PSI members in transforming ideas into classroom experiences. As an intern who only discovered the industry when applying for placements, I often think that earlier awareness could have helped me make different study choices - ones that built a clearer path into this field and shown how you can contribute to healthcare without becoming a doctor!

By increasing member engagement, every project - from concept to classroom - can inspire the next generation and strengthen our industry's future.

16 June 10:30

Title: Implementing ICH E20: Designing and Analysing Adaptive Clinical Trials

Christopher Jennison[1], David Robertson[2], Michael Grayling[3]

[1]University of Bath, Bath, United Kingdom. [2]MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom. [3]Johnson and Johnson, High Wycombe, United Kingdom

Abstract

Aim of the workshop

The *ICH E20: Guideline on Adaptive Designs for Clinical Trials* sets out expectations for the conduct and analysis of adaptive clinical trials.

Our aim is to:

- Explain the requirements stipulated in the ICH E20 guidance
- Describe the methods that can be applied to meet these requirements
- Facilitate discussion by participants of the implications for their trials

Workshop content

Context

We shall summarise the trial designs addressed by E20:

- Group sequential trials with early stopping for efficacy and futility
- Trials with sample size re-assessment for a nuisance parameter or estimated treatment effect
- Seamless Phase 2/3 trials and multi-arm multi-stage trials
- Enrichment trials
- Trials with response adaptive randomisation

Key principles

We shall describe the five principles of E20:

- Adequacy Within the Development Program: Justifying the selected dose, etc
- Adequacy of Trial Planning: Pre-planned, simple with some flexibility
- Limiting the Chances of Erroneous Conclusions: Type I error control
- Reliability of Estimation: Estimates for cost-benefit decisions
- Maintenance of Trial Integrity: Blinding, no information leakage, IDMCs

Brief overview of methods

We shall give a brief overview of methods that can be used to satisfy the E20 requirements:

- Error spending group sequential designs
- Combinations tests to combine data across stages of adaptive trials
- Multiple testing procedures using the closed testing principle
- Methods for estimation after an adaptive trial

We shall note the areas where current practices are liable to fall short of meeting the E20 requirements and where current methods may need further development.

We shall describe historical adaptive trials, indicating where they would have met or failed to meet the E20 requirements.

17 June 11:15

Title: A comparison of approaches to incorporate patient-selected and patient-ranked outcomes in clinical trials

David Robertson[1], Thomas Jaki[1,2]

[1]MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom. [2]University of Regensburg, Regensburg, Germany

Abstract

A key aspect of patient centered drug development is identifying and measuring outcomes that are important to patients in clinical trials. Many medical conditions affect multiple symptom domains, and a consensus approach to determine the relative importance of the associated multiple outcomes ignores the heterogeneity in individual patient preferences. Patient-selected outcomes offer one way to incorporate individual patient preferences, as proposed in recent regulatory guidance for the treatment for migraine, where each patient selects their most bothersome migraine-associated symptom in addition to pain. Patient-ranked outcomes have also recently been proposed, which go further and consider the full ranking of the relative importance of all the outcomes. This can be assessed using a composite DOOR (Desirability of Outcome Ranking) endpoint. In this talk, we compare the advantages and disadvantages of using patient-selected versus patient-ranked outcomes in the context of a two-arm randomised controlled trial for multiple sclerosis. We compare the power and type I error rate by simulation, and discuss several other important considerations when using the two approaches.

17 June 12:35

Title: A robustness assessment of the latent variable framework for composite endpoints: With application to late-stage trials

Paul Newcombe[1], Jasna Cotic[2], David Whitney[2], Lindsey Schader[3], Jane Bentley[2], Aris Perperoglou[1], James Wason[4], Dave Lunn[2]

[1]GSK, Stevenage, United Kingdom. [2]GSK, London, United Kingdom. [3]GSK, Denver, USA. [4]Newcastle University, Newcastle, United Kingdom

Abstract

Composite responder endpoints, which combine multiple clinical outcomes to determine a binary responder variable, are commonly used in clinical trials to capture various aspects of disease progression. Traditionally these endpoints are analysed as binary, which means a large amount of information is discarded as the continuous component variables are dichotomised and collapsed together. Various methods, including a latent variable framework proposed by McMenamin et al[1], enable more efficient analysis of composite endpoints through an expanded model that includes the underlying continuous endpoint information to improve precision, while inferring treatment effects on the same composite endpoint scale. Previous applications to academic trials, and post-hoc analysis of pharmaceutical trial data, have indicated up to 60% reductions in sample size can be possible.

Despite clear potential to enable smaller, shorter trials, thereby decreasing costs and delivering new medicines to patients faster, there are no examples to our knowledge of this methodology being used to design pivotal trials within the pharmaceutical industry. A key assumption underlying the model, and core to the precision gain, is that of a multivariate normal distribution for the latent outcome components. Since this assumption is not required in a traditional responder analysis, demonstrating robustness to its violation is expected to be critical for regulatory acceptance. We will present a simulation study evaluating the robustness of operating characteristics under mis-specification across a range of departures from multivariate normality, including misspecification of the marginal distribution, the copula family, or both. After discussing the conditions for robustness, we present the first application of this framework to large, late stage trials which indicates a substantial sample size saving could have been possible. We hope to raise awareness of this important efficiency-gaining methodology with statisticians working in drug development, and help de-risk its use in designing pivotal trials.

15 June 13:30

Title: Workshop - Navigating EU HTA: From pivotal trial to evidence networks based on first experiences

Lena Stein[1], Stefanie Wüstner[1], Anton Schönstein[2], Amelie Elsäßer[2]

[1]AMS Advanced Medical Services GmbH, Mannheim, Germany. [2]Boehringer Ingelheim, Ingelheim am Rhein, Germany

Abstract

The European Joint Clinical Assessment (JCA) has become a central milestone in the life cycle of medicinal products. For statisticians and evidence teams, it represents a new link between regulatory and health technology assessment (HTA) requirements. Clinical study data, statistical analysis plans (SAPs), and complementary evidence now need to serve both EMA and EU HTA purposes within aligned timelines and consistent methodological frameworks.

Preparations for the first JCA procedures have revealed key challenges: unanticipated PICO (Population, Intervention, Comparator, Outcome) definitions, evolving assessment scopes, differences in endpoint interpretation, and the integration of non-randomised evidence. This workshop builds on these early experiences to explore what statisticians and evidence scientists can do today to prepare for upcoming assessments and adapt their workflows for efficient dossier generation and reuse.

This interactive workshop will guide participants through the full JCA journey starting from pivotal trial planning and Joint Scientific Consultation (JSC), through evidence generation and synthesis, to dossier compilation and national reuse.

To make the experience tangible, two realistic case scenarios (a pivotal actively controlled randomised controlled trial [RCT] and a single-arm trial) will be used as thought experiments. Participants will be invited to explore analytical and strategic decisions within each scenario and reflect on how to anticipate and mitigate common challenges in the JCA process.

Key Discussion Themes

- **Early alignment:** translating JSC into feasible evidence plans.
- **PICO prediction vs reality:** dealing with unexpected comparators and outcomes.
- **Multiplicity & estimands:** transparent handling of analytical hierarchies, multiple data cuts, and intercurrent events such as treatment discontinuation or cross over.
- **Outcome alignment:** bridging regulatory endpoints and HTA-relevant outcomes.
- **Certainty assessment:** combining internal validity, external validity and statistical precision.
- **Strategic integration:** designing data workflows and visualisations that ensure traceability and reproducibility across EMA, JCA and national submissions.

Learning Outcomes

By the end of the session, participants will be able to:

- Describe the structure and timelines of the JCA process.
- Identify where and how statisticians contribute to evidence generation for HTA.
- Understand practical approaches for direct and indirect comparisons within the JCA framework.
- Manage PICO scope changes and communicate their analytical impact.
- Plan complementary analyses and reporting strategies consistent with regulatory and HTA submissions

15 June 11:00

Title: BOIN vs. BLRM: a systematic performance comparison in phase 1 dose escalation

Giulia Brunelli, Miguel Pereira, Erhard Quebe-Fehling, Oliver Schoenborn-Kellenberger

Cogitars Gmbh, Heidelberg, Germany

Abstract

The Bayesian Logistic Regression Model (BLRM) and Bayesian Optimal Interval (BOIN) design are two statistical approaches for estimating the maximum tolerated dose (MTD) in phase I trials. The BLRM explicitly models the dose–toxicity relationship, incorporates prior information, and allows inference across intermediate doses. The BOIN is a rule-based, model-assisted method that disregards actual dose magnitudes for escalation decisions. Additionally, the BOIN has increasingly seen regulatory pushback in target DLT rate selection. The aim of this study is to comprehensively compare the BLRM and BOIN designs and share our experience implementing the BLRM in several trials.

A comprehensive simulation study was conducted comparing the BLRM (with Escalation With Overdose Control) and BOIN across multiple dose–toxicity scenarios representative of single-compound development strategies. For each scenario, 500 simulated trials were generated per method, assuming cohort sizes of 3-6 patients. Performance was evaluated using seven metrics encompassing accuracy, safety, and reliability, with both methods benchmarked against the 3+3 design.

The BLRM achieved superior accuracy, with ~80% of doses in the target toxicity range correctly selected as the MTD, compared to ~75% for BOIN. The BLRM also demonstrated higher consistency and allocation efficiency, with reduced risk of overdosing, suboptimal dose allocation and lower variability in MTD selection. However, it exhibited a slightly greater tendency to select overly toxic doses. Both BLRM and BOIN substantially outperformed the 3+3 design.

Overall, the BLRM outperformed BOIN design which further shows the superiority of fully model-based methods.

17 June 12:35

Title: Beyond Dichotomization: Efficient Estimation of Response Rates using Continuous Outcomes

Michael Sweeting[1], Thomas Drury[1], Jessica Lim[2], Lindsey Schader[2], Michael Seath[1], Stephen Weng[1], David Whitney[1], Paul Newcombe[1]

[1]GSK, London, United Kingdom. [2]GSK, Upper Providence, USA

Abstract

Dichotomization of continuous outcomes remains widespread in clinical research, particularly when there is a clinically meaningful threshold considered important for the disease. While appealing for clinical interpretation, this practice discards substantial information and can lead to marked inefficiency and larger clinical trials than necessary.

We present a suite of modelling approaches that retain the continuous outcome but still target clinically meaningful estimands based on response rates. Using fitted predictive distributions, the response rate is obtained directly as an estimate of the tail probability, with standard errors estimated via either the delta method or nonparametric bootstrap. Normal linear regression, with optional quantile or Yeo-Johnson normalizing transformations, and skew-t regression are explored to flexibly accommodate departures from normality. Marginal estimands are naturally obtained through G-computation, facilitating covariate adjustment and population based interpretation.

Through simulation studies spanning a range of non-normal data-generating mechanisms, we assess robustness to model misspecification in terms of bias, efficiency, power and type-I error.

We illustrate the practical application of the methods in clinical trial case-studies, where modelling continuous change from baseline outcomes achieved large efficiency gains relative to a dichotomized analysis.

The proposed framework offers a statistically principled yet practical alternative to dichotomization. We discuss implementation considerations, including empirical versus model-based estimators, bias-variance trade-off, communication of results, and implications for regulatory acceptance.

16 June 14:00

Title: When futility is futile – an economic case for more pragmatism in late phase futility stopping

James Bell[1], Kevin Kunzmann[2]

[1]Elderbrook Solutions GmbH, High Wycombe, United Kingdom. [2]Boehringer Ingelheim Pharma GmbH & Co. KG, Ingelheim, Germany

Abstract

Aggressive futility stopping boundaries in late phase clinical trials can appear attractive due to the clear and easily communicated potential cost savings. Conditional power arguments seem to support this: Low conditional power values at interim might be perceived negatively by trialists and encourage stopping a trial early for futility. However, observing low conditional power at early interim analyses does occur in some trials that go on to be significant and should not necessarily lead to futility stopping. Stopping for futility in trials that would otherwise be successful can incur huge losses of future revenue.

We argue that futility boundaries should be set during the planning stage of a study relying primarily on marginal operating characteristics, like power. We make a simple economic argument deriving a pragmatic condition to evaluate the viability of proposed futility decision rules. We show, across typical phase III settings, that viable futility rules should not meaningfully affect marginal operating characteristics under the alternative hypothesis.

We thus recommend that, in typical phase III scenarios, teams first plan a trial without considering a futility analysis explicitly. A conservative futility rule might then be added afterwards. The primary motivation for the futility rule should be to provide a realistic chance of stopping the trial early under the null hypothesis without substantial negative impact on power under the alternative.

Our focus is on a pragmatic and easy to implement approach, without sacrificing analytical rigor, to help statisticians streamline the design process of late phase futility stopping rules.

15 June 15:45

Title: Disentangling Indirect Effects of Vaccine Assignment from Other Causal Pathways in Cluster-Randomized Trials with Noncompliance

Silvia Noirjean[1], Andrea Callegaro[2]

[1]GSK Vaccines, Siena, Italy. [2]GSK Vaccines, Wavre, Belgium

Abstract

Vaccines against infectious diseases not only provide protection to individuals who receive vaccination but may also offer indirect protection by increasing "herd" immunity. Cluster-randomized trials (CRTs) can capture indirect effects by randomizing the vaccine assignment to clusters of individuals (e.g., individuals going to the same hospital or living in the same geographical area). In CRTs, noncompliance typically arises because not all individuals in vaccine-assigned clusters receive the vaccination, and some individuals in clusters not assigned to receive vaccination may still access the vaccine. The standard approach for estimating indirect effects involves comparing disease risks for unvaccinated individuals in clusters assigned to the vaccine (noncompliers) with unvaccinated individuals in clusters not assigned to the vaccine (compliers). However, this comparison is susceptible to selection bias because, while the cluster assignment is randomized, individual receipt of the vaccine is not. To address this issue, we propose blending principal stratification (PS) and mediation analysis. Specifically, we leverage PS to address noncompliance, defining subpopulations of individuals based on their potential vaccination under the two possible cluster assignments. We also view the proportion of vaccinated individuals within the cluster as a mediator, formally defining “principal direct” and “principal indirect effects”. These effects are, respectively, natural direct and indirect effects within the principal strata. Finally, we demonstrate, through a toy example, how these measures can be estimated and interpreted to disentangle indirect effects from other causal pathways. We adopt a Bayesian approach for inference.

16 June 16:15

Title: Validating Shiny Apps in Regulated Environments with the Litmusverse

Pedro Silva

Jumping Rivers, Warsaw, Poland

Abstract

Shiny apps are an essential essential part of clinical research and healthcare, but once they're used in regulated settings, validation is no longer optional. This talk introduces a practical, risk-based approach to Shiny validation using the Litmusverse, a suite of R packages that assess code quality, score risk, and generate audit-ready evidence.

We'll look at common challenges such as traceability, documentation, and reproducibility, and show how litmus can automate assessments, highlight issues early, and make Shiny apps easier to validate from the start. You'll see how litmus fits directly into your development workflow, supporting both internal reviewers and external regulators.

By the end of the session, you'll understand:

- Why validation matters for Shiny apps in regulated contexts;
- How litmus supports structured, risk-based validation;
- What makes a Shiny app more (or less) "validatable.

This session is ideal for R users in pharma, clinical research, and healthcare who want to deliver trustworthy, regulator-ready Shiny applications without slowing down development.

16 June 16:15

Title: Which CRAN Packages Pharma Can Actually Rely On

Colin Gillespie

Jumping Rivers, Newcastle, United Kingdom

Abstract

CRAN is full of useful R packages, but only a fraction of them are suitable for clinical and regulated work. Statistical programmers and data scientists in pharma often rely on the same small group of well known packages, yet many others could be used safely if teams knew what to look for. This talk shares an analysis of CRAN packages through a pharma lens and highlights the features that make a package appropriate for clinical use.

The session walks through practical signals of suitability: clear maintenance patterns, consistent releases, active issue tracking, readable documentation, stable dependencies, and evidence of testing. These indicators may sound simple, but they quickly separate packages that can support clinical pipelines from those that introduce avoidable risk.

To support this review, we use tools such as Diffify and Litmusverse to surface version changes and package health indicators, but the focus remains on the packages themselves. Real examples from CRAN will be shared to show how these signals appear in practice, including cases where popular packages are reliable long term and cases where subtle gaps become more visible once examined closely.

The talk ends with a straightforward checklist that teams can apply when deciding whether a package is fit for pharma use. The goal is to give programmers and data scientists a clear, practical way to judge CRAN packages rather than relying on habit or guesswork.

15 June 13:30

Title: Estimands for the Percentage Change from Baseline: Guidance for Clinical Trials

Tanja Högg[1], Wenyue Zhu[1], Björn Holzhauer[2], Tobias Mütze[2], Simon Wandel[2], Tim Morris[1]

[1]Novartis Pharmaceuticals UK Ltd., London, United Kingdom. [2]Novartis Pharma AG, Basel, Switzerland

Abstract

The percentage change from baseline is a common endpoint in clinical trials with positive-valued biomarkers, with the estimands' summary measures often expressed as the difference in mean percent change from baseline between active and placebo arms.

In practice, however, its asymmetric and heteroskedastic distributional properties can conflict with the assumptions required for direct linear modelling of the participant-level percentage change from baseline, ultimately rendering this ad-hoc estimation approach inadequate. Modelling of the logarithmically transformed ratio to baseline is a regularly used alternative in these circumstances, whilst also providing estimates that are easily mistaken for the true inferential target of the mean percentage change from baseline.

This talk reviews these two principal analysis strategies and clarifies the difference in estimands arising in their context — the mean percentage change from baseline and the geometric mean ratio to baseline. Drawing on recent cardiovascular trials in the literature, we highlight examples where these distinctly different estimands are targeted yet reported using similar verbiage, risking inadvertent misinterpretations from non-statistical audiences as a result. Additionally, we provide technical details for post-processing steps by which to recover an estimate of the mean percentage change from a model of the log ratio to baseline, such that the intended estimand is targeted whilst simultaneously acknowledging the modelling assumptions needed for valid inference.

Together, this talk underscores the need for clear reporting language and aims to equip statisticians with practical tools and conceptual clarity for selecting, implementing, and explaining analysis strategies for percentage change endpoints in clinical research.

15 June 11:00

Title: Dose finding in late phase Bayesian trials

Connor Fitchett

MRC Biostatistics Unit, Cambridge, United Kingdom

Abstract

Project Optimus is an FDA initiative which encourages early phase dose optimisation: going beyond traditional methods which tend to focus on a single maximum tolerated dose and proceed to later phases of clinical development. However, dose finding can also be done in later phases, notably in non-inferiority trials if we want to change or reduce doses. For example, acute lymphoblastic leukaemia (ALL) is currently treated using a long period of chemotherapy. Recent studies investigating Blinatumomab, a novel drug being used to fight ALL, have suggested that it may be possible to 'deintensify' the chemotherapy treatment following the introduction of Blinatumomab. Using Bayesian design and adaptations, we investigate through simulations a late phase trial framework that is able to consider alternative doses efficiently whilst protecting patient benefit metrics, with data being based on the ALL deintensification study. We then compare this design to common early phase dose finding methods to evaluate its strengths and weaknesses.

15 June 13:30

Title: AI & ML SIG: Practical AI and Developments in Machine Learning

Sam Hadlington[1], Lesedi Ledwaba-Chapman[2], Paola Berchialla[3], Harry Parr[4], Jason Nicholas[4,5]

[1]Plus-Project Partnership, Hook, United Kingdom. [2]MMS, London, United Kingdom. [3]University of Torino, Turin, Italy. [4]GSK, London, United Kingdom. [5]University of Bath, Bath, United Kingdom

Abstract

AI & ML SIG Updates

A quick discussion on developments within the SIG and our goals for the coming year.

SAP-GPT: Using Modern AI for Document Generation

Large language models (LLMs) could reshape how technical documents are produced in clinical development. This presentation examines the use of LLMs to draft and iterate Statistical Analysis Plans (SAPs) while meeting regulatory and quality requirements, and improving efficiency. We discuss the potential of a practical, end-to-end workflow that grounds generation in the protocol and in a company standard template. Prompt templates and retrieval-augmented generation (RAG) are used to anchor claims to specific protocol sections, preserving traceability.

We will outline the steps we have come up with to generate a SAP and we will talk about the reasoning behind these steps and some issues we ran in to during prompt development. We will show how these steps can be used to build efficiencies into the SAP generation process and how they can be used to make LLM's work for you by having them perform to their strengths. We will also discuss the limitations of LLM's for this particular task that we discovered while building these prompts.

We will then go into a discussion of different LLM's and their relative strengths and weaknesses pertaining specifically to SAP creation.

Finally we will give an assessment of where and when to use this approach and in which situations it could potentially improve development efficiency. We will also discuss how this approach could be applied to other documents and the potential drawbacks of this.

Patient Preference Research with AI-Powered Evidence Generation

Background. Patient Preference Studies (PPS), including Discrete Choice Experiments and other choice-based elicitation methods, are increasingly used to quantify how patients value treatment benefits and risks. However, the synthesis of PPS evidence remains largely manual and fragmented. Heterogeneity in study design and reporting prevents the direct comparison of preferences and limits their integration into regulatory and health technology assessment frameworks.

Objective We developed an AI–based system for the automated extraction, structuring, harmonization, and retrieval of data from published PPS, enabling scalable and reproducible evidence synthesis.

Methods The system integrates OCR and Large Language Models (LLMs), including multimodal architectures, to extract categorical and numerical data from documents, like PDFs, tables, and figures. Categorical variables include study-level descriptors (study type, model specification, disease area), while numerical variables comprise preference weights, p-values, standard errors. Extracted content is validated, harmonized via ontology alignment, and converted into a machine-readable format suitable for meta-analytic aggregation. A real-time module receives natural language queries, interprets intent through vector search and LLM-based semantic parsing, and retrieves the corresponding harmonized data.

Results The system replaces manual data extraction and expert-driven normalization, achieving consistent, reproducible structuring of PPS outputs. Our system was validated using the SIGMA

framework, which classifies outputs as correct, incomplete, wrong, missing, or hallucinating. Overall, 89% of extracted data were correct, 6.5% incomplete, 4.2% wrong, no missing or hallucinating outputs.

Conclusions This framework represents the first end-to-end AI solution for automated extraction and on-demand retrieval of PPS evidence, enabling scalable, transparent, and regulatory-ready meta-analysis of patient preferences.

A Tutorial on Tidymodels

The development of prediction models, e.g. statistical and machine learners (ML), are fast becoming standard practice within life sciences and medicine. These learners aim to predict outcomes / prognosis which can guide clinical decision making, e.g. support further treatment interventions. Widely used examples in clinical practice include the Framingham risk score for estimating 10-year cardiovascular risk . Predictive models are also being integrated into prospective clinical trials, to improve precision in estimating the treatment effect by complementing standard covariate adjustment. For instance, ML-derived prognostic scores can better capture nonlinear relationships, supplementing trial efficiency.

At the same time, the life sciences sector is rapidly adopting more open-source software, leveraging ecosystems like the tidyverse as well as more bespoke packages under the pharmaverse, which promotes best practices within the industry. This momentum in open-source adoption and the utility of robustly developed prediction models brings recent work in tidymodels into sharp focus. We will present a concise primer, introducing a multi-step unified framework for developing and validating prediction models using the tidymodels R package.

Tidymodels offers a comprehensive ecosystem for building predictive models in R. The workflow spans data preprocessing via ‘recipes’, models specification (including ensembles), and resampling & hyperparameter tuning, where we combine these steps into a unified pipeline for evaluating the performance metrics (e.g. R^2 and other loss-functions). Visual and explainable methods will also be demonstrated. The session will be useful for attendees who wish to see practical implementations of complete modelling frameworks within R.

Panel Discussion

All of the speakers will come together at the end on a panel to field any questions from the audience.

15 June 15:45

Title: Evaluation of Z-tests to compare fixed time survival probabilities using stratified Kaplan-Meier estimates with different variance estimators and weights

Maria P. G. Blanco[1], Stephan M. Bischofberger[1], Bernhard Haller[2], Gabriele Bleckert[1], Eva Hoster[3], Hannes Buchner[1]

[1]Staburo GmbH, Munich, Germany. [2]Technical University of Munich, School of Medicine and Health, Institute of AI and Informatics in Medicine, Munich, Germany. [3]Institute for Medical Information Processing, Biometry, and Epidemiology (IBE), Ludwig-Maximilians-Universität München, Munich, Germany

Abstract

Time-to-event variables are among the most relevant primary efficacy endpoints in clinical trials, particularly in later phase oncology trials. When the proportional hazards assumption is expected to be severely violated, an alternative to the log-rank test is needed. Testing for differences in survival probabilities at a pre-defined time point offers one such option and has already been employed in some studies. However, in trials with stratified randomization, using the Kaplan-Meier estimate poses particular challenges: many common variance estimators are not evaluable or estimate a value of zero in strata with no events or in strata where all patients had the event, and the optimal stratum weighting strategy remains unclear.

Through a simulation study mimicking clinical trials with stratified randomization and various non-proportional hazards scenarios, we compared the stratified log-rank test and Kaplan-Meier-based Z-tests using different variance estimators and stratum weights, evaluating type-I error and power. While the log-rank test remained optimal under proportional hazards, some Z-tests provided robust performance across a broad range of scenarios and outperformed the log-rank test in situations with delayed treatment effects or crossing survival curves.

For stratified analyses under non-proportional hazards, we recommend a Kaplan-Meier-based Z-test with Borkowf's adjusted variance estimator and Mantel-Haenszel type weights, as inverse variance weights can lead to alpha-inflation. For non-stratified analyses, Greenwood-based or complementary log-log Z-tests are viable alternatives when zero variance can be excluded.

15 June 13:30

Title: Evaluating Estimand Implementation in Clinical Trials in the UK and Beyond

Morgaine Stiles[1], Fay Cafferty[1], Rodrigo Nunes[1], Jack Skinner[1], Maggie Qiao[1], Jessica Maudsley[1], Beatriz Goulao[2], Victoria Shepherd[3], Christina Yap[1]

[1]The Institute of Cancer Research, London, United Kingdom. [2]The University of Glasgow, Glasgow, United Kingdom. [3]Cardiff University, Cardiff, United Kingdom

Abstract

Background

Estimands clarify the precise research question to be answered within a clinical trial, ensuring that factors impacting the interpretation of the treatment effect are considered and documented at an early stage. To improve uptake of the estimand framework, it is important to understand its current use, and barriers to its wider implementation.

Methods

A systematic scoping review examined the extent of estimand use in clinical trials registered between 2021 and 2024.

To capture uptake in academic-sponsored trials, a survey of UKCRC (UK Clinical Research Collaboration) academic clinical trials units (CTUs) assessed current estimand framework implementation and barriers to its use.

Results

The scoping review found that estimands were described in 19% (387/2055) of trials. Factors associated with estimand use included the trial having an industry funder, being a non-cancer trial, running in multiple countries, being randomised and being a later phase.

The survey indicated that, while most CTUs (40/46, 87%) used estimands in the last three years, fewer (27/46, 59%) use them routinely and have them included in internal SOPs/guidance. Where they are used, it is predominantly in phase III trials. Statisticians were involved in estimand development in all CTUs using estimands, and clinicians were involved for most units. Other roles, including patient partners, were much less likely to be involved.

Conclusion

The estimand framework is an important tool to ensure that clinical trials are asking the correct questions, yet remains underutilised. Further work is needed to improve its adoption and deepen understanding of estimands across research settings.

16 June 16:15

Title: Assessing Multiple Endpoints Using a Novel Software Solution in a Late-Stage Oncology Study

Valeria Mazzanti[1], Kyle Wathen[2]

[1]Cytel Inc, Geneva, Switzerland. [2]Cytel Inc, Boston, USA

Abstract

Evidence generation in clinical research has seen a transformation over the past decade, with an ever-increasing demand for a variety of data from clinical trials to satisfy both regulatory requirements, and back commercial claims. This demand is particularly acute for indications with many competitor products. Researchers, therefore, are increasingly looking to detect statistical significance for more than one primary endpoint in a phase III study setting. We propose a novel implementation to simulate a variety of multiple endpoints efficiently and test this generated data rigorously for trial planning purposes.

Native workflows in software will be used to design and optimize an adaptive oncology clinical trial with three concurrent endpoints: Overall Response Rate (ORR), Progression-Free Survival (PFS), and Overall Survival (OS). We will show how software and programming tools could be used to initially generate data for each of these endpoints by leveraging multi-state models as well as correlation-based assumptions. These approaches both rely on prior knowledge of the treatment effects, which inevitably will have uncertainties tied to the information they bring. The impact of this uncertainty can be evaluated through simulating several scenarios with each approach.

Using a case study from an oncology trial, this presentation will show how to use simulation-based tools to optimize trial designs based on the information available at the planning stage. Future research based on the findings will also be discussed, including the introduction of RWD into the simulation inputs to create the original assumptions that are the foundations of the patient-level data generation.

16 June 10:30

Title: Developments in early phase dose-finding trials

Sam Hinsley[1], Pavel Mozgunov[2], Andrew Hall[3], Matt George[1], Anaïs Andrillon[4], Sandrine Micallef[5]

[1]Phastar, London, United Kingdom. [2]MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom. [3]Leeds Institute of Clinical Trials Research, Leeds, United Kingdom. [4]Department of Statistical Methodology, Saryga, Paris, France. [5]Debiopharm International SA, Lausanne, Switzerland

Abstract

Introduction to the new early phase ESIG

In this inaugural session, we will introduce the new Early Phase ESIG and discuss it's aims and objectives. The group has been established to advance knowledge sharing, foster collaboration, and improve education in the rapidly developing area of early phase clinical trials.

Objectives of the group include:

- Improving understanding and implementation of statistical methodologies within early phase trials via training,meetings or publications
- Sharing knowledge and real-life experiences between key stakeholders in both academia and industry, and across different geographies
- Allowing space for collaboration between and across different companies and institutions, fostering a community of support rather than competition

Our ultimate goal is to improve the use of innovative trial methodologies to improve standards within the early phase clinical trials community and reduce substandard treatments being taken into later phase trials. We welcome you to join us!

Recent advances in dose-finding: an introduction

In this talk, I would briefly introduce the recent advances in the setting of early phase dose-finding oncology trials. I will start with the typical Phase I oncology setting and the challenges that we often face. I will then introduce the class of model-based dose-escalation designs and how, in the light of the FDA's project Optimus, we can use them to include multiple toxicity and efficacy endpoints to define the optimal dose. I will then talk about various dose optimisation strategies including backfilling and how it should (and should not) be used. Finally, I will talk about the extension for the setting of the combination trials where two drugs can be escalated simultaneously. This talk will be a bridge to the methodological innovations presented by the rest of the speakers in the session.

Designing phase I trials with backfilling to mitigate screening and treatment delays

There is growing interest in backfilling in Phase I dose-escalation trials. Conventional staged designs open and close recruitment with each cohort, leading to long pauses and extended timelines. Backfilling offers an alternative, allowing continued recruitment by enrolling additional patients at lower doses outside the dose under evaluation. This approach can strengthen safety and endpoint data while maintaining trial momentum. We describe a novel application of backfilling to address challenges posed by extended screening periods and delayed treatment initiation.

Two examples involve an experimental treatment following autologous stem cell transplant in multiple myeloma. The primary objectives are to identify suitable dose levels for further study based on dose-limiting toxicity (DLT) endpoints over a short assessment period. In both scenarios, patients were recruited before transplant while dose allocation occurred five months later, with variability due to treatment delays. Conventional staged designs would have extended timelines beyond five years and were infeasible. The proposed designs specified continuous recruitment and incorporated backfilling

principles to allow dose allocation for patients who had undergone procedures beyond standard care access to novel therapy. Safety was ensured through constraints on initial recruitment and oversight by an independent safety review committee with a statistical model and limits on concurrent DLT assessments.

Simulation studies under varying recruitment scenarios demonstrated feasibility in an acceptable timeframe compared with a sequential design. Across scenarios, the design-maintained safety, preserved operating characteristics, and achieved modest improvements in correct selection for a fixed sample size.

A novel method for inserting dose levels mid-trial in early phase combination studies

The use of combination treatments in early phase oncology trials is growing. The objective of these trials is to search for the maximum tolerated dose combination from a pre-defined set. However, cases in which the initial set of combinations does not contain one close to the target toxicity level pose a significant challenge. There is uncertainty around how to handle these situations effectively in practice and the literature does not fully evaluate potential solutions.

To address this, we propose a novel method for inserting dose levels mid-trial. The idea is based on evaluating contours that partition the set of combinations into ones above and below the target toxicity. Dose insertions are made only if a single contour is highly probable, indicating an absence of combinations to explore around the target toxicity.

We examine our proposed approach applied to two established designs, although any model-based or model-assisted design is an appropriate candidate. Results from our comprehensive simulation study demonstrate that the insertion method can increase the probability of selecting combinations close to the target toxicity, whilst controlling for selecting overly toxic combinations. These methods can be extended to more complex settings, such as trials with joint toxicity and efficacy endpoints.

U-DESPE: a Bayesian Utility-based methodology for dosing regimen optimization based on Dose-Exposure, Safety, Pharmacodynamics, Efficacy in early-phase oncology trials

For targeted therapies and immunotherapies, the standard cytotoxic chemotherapy paradigm “more is better” may no longer apply, which has motivated the U.S. FDA’s Project Optimus to reform the dose optimization in oncology. In light of this, U-DESPE, a Bayesian dose-finding design that recommends an optimal dosing regimen by integrating pharmacokinetics (PK), pharmacodynamics (PD), safety and early efficacy data, has been proposed. Bayesian models are built to characterize the relationships between exposure and the probability of relevant endpoints on safety, efficacy, and PD. The dose-exposure derived from PK data and the exposure-endpoints models are then combined to obtain the dose-endpoints relationships. Finally, a utility function is proposed to quantify the trade-off between these endpoints and to determine the optimal dosing regimen. U-DESPE can be used after a dose-escalation phase, to randomize patients among safe candidate optimal dosing regimens.

The U-DESPE has been implemented in a phase I study evaluating the combination of a WEE1 inhibitor (Debio 0123) with carboplatin in adults with advanced solid tumors (NCT03968653). The design supported the selection of the optimal dose and schedule by jointly considering safety, PK/PD, and efficacy information, providing a coherent decision-making framework for dose optimization aligned with current regulatory expectations in oncology development.

16 June 16:15

Title: Fast, Fresh & Interactive: R Dashboards for Statisticians in 30 Minutes or Less!

Martin Brown

PPD, Cambridge, United Kingdom

Abstract

Interactive tools now don't have to take days, large teams—or deep coding expertise—to create. In this session, we'll explore how statisticians can build practical, engaging tools and interactive content in under 30 minutes using R.

We'll start with simple Quarto data tables and Plotly figures, showing how to make analyses dynamic and visually engaging directly from your R session. Next, we'll see how an AI-generated Shiny app can be created from a simulation prompt in minutes—demonstrating how artificial intelligence can act as a coding companion to accelerate ideas from concept to prototype. Finally, we'll use the teal framework to assemble a polished, interactive dashboard designed to support review, exploration, and reporting in a pharmaceutical setting.

Along the way, we'll highlight how these "recipes" not only create useful tools but also help statisticians build confidence and skills in R. Each example demonstrates how small, fast steps can lead to impactful tools that make your work easier to share, explore, and understand. Whether for internal review, simulation studies, or data communication, these "recipes" show that with just curiosity and a bit of creativity, you can serve up interactive insights in no time. Attendees will leave with inspiration, practical examples, and the confidence to bring their own analyses to life through interactivity.

17 June 11:15

Title: Advancing the Implementation of Safety Methodologies

Dooti Roy[1], Matthias Trampisch[2], Florence Le Maulf[3]

[1]Boehringer Ingelheim, Ridgefield, USA. [2]Boehringer Ingelheim, Ingelheim, Germany. [3]Cytel, Nantes, France

Abstract

This session presents the initial outputs from a dedicated working group, established last year within the EFSPI/PSI Benefit-Risk SIG, with a core focus on advancing the implementation of safety methodologies. Recognising the persistent challenges in translating innovative safety science into routine practice, this session brings together three distinct but complementary perspectives on enhancing drug safety across the pharmaceutical development lifecycle.

Firstly, we will explore the broader strategic and practical hurdles in accelerating the adoption of quantitative drug safety inventions into impactful innovations, examining regulatory, organisational, and technical factors. Following this, we'll delve into the statistical and operational aspects of aggregate safety reporting, particularly for anticipated serious adverse events in investigational trials, offering practical strategies for regulatory compliance and trial integrity. Finally, we will provide an update on the working group's efforts to develop a practical 'Toolkit' for Integrated Summary of Safety (ISS) development, aiming to standardize best practices for pooling, grouping, and adjusting safety data.

Quantitative Evolution of Drug Safety: Adapting Inventions to Innovations

Ensuring drug safety is a cornerstone of pharmaceutical development, yet the adoption of advanced analytical tools and methods often lags behind scientific invention by years. This work examines the strategic and practical challenges of transforming innovative safety approaches into widely adopted solutions. We review the current regulatory landscape (FDA, EMA) and highlight limitations in standard safety analyses, such as reliance on incidence rates and aggregate reporting.

Our strategic analysis explores the value of advanced statistical techniques, the importance of regulatory acceptance, and the organizational factors—such as infrastructure, training, and stakeholder engagement—required for successful implementation. We emphasize the need for open-source tools and collaborative development to accelerate uptake.

To ground these concepts, we present a series of real-world case studies, including Bayesian analytics, systematic safety surveillance, and interactive safety graphics. These examples illustrate both the opportunities and barriers encountered during implementation, offering insights into what enables inventions to become impactful innovations.

We conclude by identifying persistent gaps and proposing strategies to bridge them, aiming to foster a culture of innovation in drug safety. Our findings provide actionable guidance for researchers, industry leaders, and regulators seeking to translate scientific advances into improved patient outcomes.

Beyond Individual Cases: Operationalizing FDA IND Safety Reporting for Anticipated Serious Adverse Events

How should sponsors assess serious adverse events that are anticipated in the study population and therefore not informative as isolated cases? This presentation addresses a key challenge in FDA IND safety reporting: the aggregate assessment of anticipated serious adverse events (anticipated SAEs) in ongoing blinded clinical trials. Rather than relying on individual-case review, the session shows how sponsors can operationalize aggregate approaches when clinically meaningful treatment-arm imbalances must be assessed while preserving trial integrity.

The talk presents a practical framework for translating FDA expectations into operational practice. It highlights how sponsors can prospectively define roles, responsibilities, review cadence, triggers, escalation pathways, and documentation standards through an Aggregate Safety Assessment Planning (ASAP) process and/or a Safety Surveillance Plan (SSP). The framework supports both trigger-based unblinding after blinded rates exceed projected thresholds and periodic unblinded review at prespecified intervals.

A major focus is governance. Attendees will learn how routine blinded surveillance, unblinded output generation, and clinical interpretation can be separated through a model involving a Statistical Data Analysis Center (SDAC) and a firewalled Safety Assessment Entity. The presentation compares internal and external implementation options, including use of an internal independent Statistical Analysis Team (iSAT) or an external CRO for controlled unblinded output generation, and illustrates safeguards that support compliant, timely, and audit-ready decision-making.

Using practical examples, the session will show how different operating models can help sponsors maintain blinding, support regulatory decision-making, and strengthen the credibility of ongoing trials.

Pooling, Grouping, Adjusting: A Statistician's Toolkit for Integrated Summary of Safety development

The Operationalising ISS reporting sub-team was formed to recommend best practices for safety data analysis in Integrated Summary of Safety (ISS) projects. Noting variability in company approaches, the team seeks to develop standardized recommendations for methods and data presentation and potentially provide code to facilitate analysis.

The team has identified topics of interest for ISS development (e.g. exposure adjustment, study-size adjustment, grouping of terms, pooling strategy) and collated guidelines, literature, and examples for each topic. While considerable published information is available on these topics, clear cohesive recommendations are not readily available for statisticians and clinicians working on an ISS.

The team is therefore developing a toolkit for ISS development. For each topic, a summary will be available summarizing literature and guidance, providing best practice recommendations along with examples or recommendations for the ISS SAP text and shells. Code to produce the recommended analyses and displays may also be provided.

This talk will provide a status update of this work to date.

17 June 11:15

Title: Does your PRO sum it all up? Investigating the variability in item specific PRO effects using random item slopes regression

Tom Booth, Lena Hubig, Meg Fluharty, Sarah Acaster

Acaster Lloyd, London, United Kingdom

Abstract

Patient Reported Outcomes (PRO) are commonly analyzed to demonstrate patient benefit of treatments. When doing so, the focus will be on PRO sum scores, comprised of all or a subset of items. These scores may then be used in psychometric evaluations (e.g. criterion or known-groups validity) and to evaluate treatment arm differences. Rarely, is it investigated whether the constituent items of a scale score show consistent validity coefficients, or treatment differences. This is despite the increasing use of PRO instruments in new settings, where the psychometric evidence for their use is limited.

We discuss the application of random item slope regression (Donnellan, Usami, & Murayama, 2023), an application of linear mixed effect models, to evaluate whether items show consistent effects to the sum score. Results from a simulation study are presented to demonstrate how the method identifies differential item effects. First, we show how the average item effect (fixed effect) is equivalent to the criterion association estimated for the sum score if all items have consistent criterion associations. Next, we show that the effects at the level of the sum score become inaccurate under differing degrees of item specific associations (increased item slope random effect). Finally, we demonstrate how trial design, disease or sample specific variables can be included in the model to help explain variance in item criterion associations.

The application of random slopes item analysis has the potential to significantly improve our understanding of nuanced PRO effects, and as a result, patient benefit.

17 June 12:35

Title: Forecasting and cost-efficient designing restricted enrolment in clinical trials

Vlad Anisimov[1], Matt Austin[2]

[1]Amgen, London, United Kingdom. [2]Amgen, Thousand Oaks, CA, USA

Abstract

This talk presents a rigorous and practical statistical framework for forecasting and optimizing patient enrolment in multicentre clinical trials under complex operational constraints. As a baseline, the framework builds on a previously developed Poisson-gamma model of patient enrolment, which accounts for heterogeneity in recruitment rates across sites and stochastic variability over time. We extend this model to handle real-world challenges such as country-level enrolment caps and minimum quotas, staggered site activation schedules and operational cost constraints.

Key contributions include:

Country-level process approximation: development of a novel analytical technique to approximate country-level enrolment using aggregated Poisson-gamma processes, enabling precise predictions even with few sites per country.

Global enrolment forecasting under constraints: development of methods to forecast global enrolment under country-specific restrictions and quantify the Probability of Success (PoS) of meeting enrolment targets within specified timelines.

Cost-optimal site allocation: formulation of an optimization problem to identify the cost-minimizing allocation of sites across countries subject to constraints on timelines, PoS, and country-specific enrolment limits; solution approaches include sequential stepwise linear programming (simplex), exhaustive search for small dimension, and evolutionary or other metaheuristic algorithms for high-dimensional problems.

The methodology is statistically rigorous, computationally efficient, and operationally impactful. It empowers statisticians to design evidence-based enrolment strategies that reduce trial delays and optimize resource allocation. Implemented as internal R-package, this technique provides a powerful toolset for clinical and operational teams in the pharmaceutical industry, enabling data-driven, cost-effective, and timely execution of trials. This work received 2025 RSS/PSI Award for Statistical Excellence in the Pharmaceutical Industry.

15 June 15:45

Title: Integrated Decision-Theoretic Optimisation of Phase II/III Oncology Trials

Haotian Wang[1], Peter Kimani[1], Michael Grayling[2], Josephine Khan[2], Nigel Stallard[1]

[1]Warwick Clinical Trials Unit, Coventry, United Kingdom. [2]Johnson & Johnson Innovative Medicine, Buckinghamshire, United Kingdom

Abstract

Traditional oncological drug development typically follows a sequential process in which phase II and phase III are planned separately. The phase III design is not directly linked/optimised in response to phase II outcomes. This may lead to inefficiencies and misalignment of early decisions with utility-related long-term development objectives.

We propose a novel decision-theoretic framework for an integrated phase II and phase III development programme, explicitly linking early binary endpoints with long-term time-to-event outcomes. This framework utilises the phase II results not only for a go/no-go decision but also for determining the expected size of phase III, optimising the entire development pathway.

The approach jointly optimises the phase II sample size and decision threshold by maximising an expected utility, defined as the balance between potential financial gain and total development cost. Prior beliefs about response and survival parameters are incorporated to reflect uncertainty in treatment performance, which guides decisions before any phase II data are observed.

A simulation study, motivated by oncology development scenarios reported in the literature, compares the proposed design with more conventional (separated) designs in terms of expected utility and decision robustness. Sensitivity analyses across different parameter settings further highlight how optimal strategies can vary — in some cases recommending proceeding to phase III regardless of phase II outcomes, with a phase II conducted primarily to determine the size of phase III to reduce overall cost. This highlights the framework's ability to reveal cost-efficient decision strategies that are not captured by conventional designs.

16 June 16:15

Title: A Unified Inference Framework for Risk Difference and Risk Ratio: Enhanced Performance in Small-Sample, Low-Incidence Binary Endpoints

Jingxin Yan[1], Gregory Chen[2], Margarita Donica[2], Yujie Zhao[3], Larry Leon[3], Linbo Wang[1]

[1]University of Toronto, Toronto, Canada. [2]MSD, Zurich, Switzerland. [3]Merck & Co., Inc., Rahway, USA

Abstract

Binary endpoints in clinical trials are commonly summarized using the risk difference (RD) or risk ratio (RR), with RD more frequently used in regulatory settings and RR more common in HTA. Some concerns arise in practice: statistical significance may disagree between confidence intervals and hypothesis tests due to non-duality of the employed estimation and testing procedures, and conclusions may differ between RD and RR even when based on the same data and stratification factors. Small samples due to ethical, operational, or rare-disease constraints, and low or zero incidence in safety analyses, further complicate inference. Current practices vary from pragmatic continuity corrections to more complex approaches such as Firth correction or penalized likelihood. We investigate the extent of the incoherence issues above, examine the statistical properties of several commonly used procedures for estimation and/or testing across diverse data settings, and introduce an innovative modeling framework that unifies inference on RD and RR while employing Blaker's exact method for confidence intervals and tests. The framework also handles zero-incidence settings via a Bayesian correction that generalizes continuity corrections by assigning Beta priors to event probabilities. Simulations show improved estimation accuracy and near-nominal confidence interval coverage in small-sample and low-incidence settings. Extensive numerical experiments provide practical guidance for selecting appropriate inference strategies across data conditions. An R package facilitates implementation, with code and materials available at: https://github.com/hta-pharma/brm-plus.

16 June 10:30

Title: PSI SIG Vaccines: Innovative Statistical Approaches, Designs and Predictive Models in Vaccine Clinical Trials

Giulia Zigon[1], Joshua Havumaki[2], Xinxue Liu[3], Federico Francone[4], Seth Seegobin[5], Andrea Callegaro[2], Fabian Tibaldi[2]

[1]GSK, Siena, Italy. [2]GSK, Wavre, Belgium. [3]Oxford University, Oxford, United Kingdom. [4]Astrazeneca, Barcelona, Spain. [5]Astrazeneca, Stevenage, United Kingdom

Abstract

The session, organized by the PSI Vaccine SIG, focuses on advanced statistical methods and designs for vaccine clinical trials, featuring four presentations on topics such as Targeted Maximum Likelihood Estimation (TMLE) for Vaccine Efficacy trials, test-negative design validity, forecasting seasonal disease waves, and a Threshold of Protection Model for monoclonal antibody efficacy against SARS-CoV-2 variants. The session aims to provide insights into innovative methodologies and their practical applications in vaccine research.

1) Applying Targeted Maximum Likelihood Estimation (TMLE) to Vaccine Efficacy Studies

***Presenters**:* Giulia Zigon and Joshua Havumaki, GSK Vaccines

Abstract: Targeted Maximum Likelihood Estimation (TMLE) is a semiparametric, doubly robust approach that combines flexible nuisance-function estimation (including machine learning) with a targeted updating step to produce consistent, asymptotically efficient causal effect estimates. We applied machine-learning–based TMLE in the context of vaccine efficacy trials, comparing its operating characteristics with simpler approaches such as unadjusted Kaplan–Meier and Cox model–based G-computation. We considered as primary estimand the risk ratio at time t.

Using both simulations and data from a real vaccine efficacy study, we provide practical implementation guidance to facilitate adoption in vaccine research.

Our simulation study was motivated by a phase 2 therapeutic vaccine trial of roughly 250 participants and was supplemented by an application to a large phase 3 trial.

Results show that the potential benefit of machine-learning–based TMLE depends on the data-generating setting: it increases with stronger prognostic baseline covariates, larger numbers of relevant covariates, and the presence of nonlinearities or interaction effects that are difficult to capture with parametric models. In the real trial data the TMLE produced a small precision improvement relative to the simpler methods; the gain was limited because covariates had only modest prognostic value in that setting.

Overall, this work highlights when and how machine-learning–driven TMLE can improve precision in causal estimation for vaccine studies and practical recommendations for implementation.

2) The validity of test-negative design for assessment of vaccine protection

***Presenter**:* Prof. Xinxue Liu, Oxford Vaccine Group, Department of Paediatrics, University of Oxford, Oxford, UK

***Abstract**:* The test-negative design (TND) is a useful tool to evaluate vaccine protection following deployment. We compared different observational and experimental study designs for assessing VE by re-analysing data from the TyVAC Bangladesh trial, a participant- and observer-blinded cluster randomised controlled trial (CRCT). We compared estimates of VE from the CRCT analysis, which assessed the risk of blood-culture-confirmed typhoid fever among TCV recipients compared with JE recipients, to estimates from the cohort and TND analyses, which compared TCV recipients and non-vaccinees in the TCV clusters. We further conducted a negative-control exposure (NCE) and a

negative-control outcome (NCO) analyses as bias indicator. The VE estimates were 89% (95% CI: 81-93) in the CRCT analysis, 79% (95% CI: 70-86) by the cohort design, and 88% (95% CI: 79-93) and 90% (95% CI: 75-96) by the TND with two definitions of test-negative controls. Using NCE analysis, JE vaccination was associated with an increased risk of typhoid fever in the cohort design (IRR: 1.98, 95% CI: 1.56-2.52), but no significant association was found in the TND. Similarly, an increased risk of non-typhoid infections was observed in the cohort NCO analyses when comparing vaccinees with non-vaccinees in both JE and TCV clusters, but not in the TND NCO analyses. The TND provides reliable estimates of TCV VE, while the cohort design can bias the VE estimates likely due to unmeasured confounding effects, such as healthcare-seeking behaviours. NCE and NCO approaches are useful tools for identifying such biases.

3) Winter is coming; forecasting the next season's wave in advance.

***Presenter**:* Federico Francone, AstraZeneca (Clinical Data Science, Vaccine & Immune Therapies)

Abstract: The incidence of many infectious diseases varies seasonally, which complicates recruitment planning and study monitoring activities. An accurate, time-resolved forecast of incidence which reacts to changes in the present season would be highly valuable.

We will describe a methodology for monitoring publicly available global surveillance data to forecast incidence in target countries, and show how that forecast can be integrated with live study data to continually refine expectations of events in the study. Our approach employs the discovery of leading indicators, building time series regression models, simulation-based modelling, and the automation of this cycle using cloud infrastructure.

A real example will be shown in which we accurately forecasted the RSV season in the United States six months in advance, primarily using leading indicators identified in South America.

4) A Threshold of Protection Model for Monoclonal Antibody Efficacy Against SARS-CoV-2 Variants

***Presenter**:* Seth Seegobin, AstraZeneca (Vaccine and Immune Therapies)

Abstract: Clinical development of monoclonal antibodies (mAbs) against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is challenging due to rapid changes in the variant landscape. Using efficacy data from the phase 3 PROVENT pre-exposure prophylaxis trial of tixagevimab–cilgavimab (NCT04625725), individual nAb titres were predicted by dividing serum mAb concentration by prevalence-adjusted tixagevimab–cilgavimab potency (from in vitro IC50 values combined with viral surveillance data) and related to efficacy with a Cox model. The Threshold of Protection (ToP) Cox model was externally validated using data from the phase 3 SUPERNOVA trial (NCT05648110), which assessed sipavibart efficacy against symptomatic COVID-19 in immunocompromised participants. This novel approach integrates predicted nAb titres against multiple SARS-CoV-2 variants into a ToP Cox model that can be applied across different variants and could serve as a surrogate endpoint in immunobridging studies to expedite clinical evaluation and regulatory approval for mAbs targeting SARS-CoV-2.

16 June 10:30

Title: Trustworthy AI in Medicine: A Unified Bayesian Approach to Uncertainty, Performance and Fairness

Bruno Boulanger[1], Nils Boulanger[2]

[1]Sanaitio, Louvain-la-Neuve, Belgium. [2]Saniatio, Louvain-la-Neuve, Belgium

Abstract

There is increasing focus on how to assess the reliability and uncertainty of predictions produced by AI‐based medical devices used for clinical decision making and decision support. At the same time, the literature has introduced a range of related terms—uncertainty, accuracy, personalised uncertainty quantification (PUQ), AI fairness, conformal prediction and explainable AI—without clear definitions or a common conceptual framework. As a result, it is difficult to understand how these notions relate to one another or how they should inform evaluation of AI tools in practice.

This presentation will introduce a formal definition of the uncertainty associated with a prediction or decision produced by an AI based system, using the Bayesian framework. In this view, the predictive distribution is the central object from which uncertainty and performance metrics are consistently derived. Treating AI systems as measurement instruments, the emphasis will be on how uncertainty emerges from this predictive distribution, rather than from model complexity or data dimensionality. The constraint on reliability is the domain covered by the training data.

Building on this foundation, a unified scheme will be proposed to connect core concepts in medical AI—precision, accuracy, uncertainty, PUQ, predictive values, conformal prediction and AI fairness—into a single coherent framework. Links to explainable AI will also be outlined, with a particular focus on their relationship to causal reasoning and clinical interpretability. This integrated perspective aims to support transparent evaluation and communication of the performance of AI‐based medical solutions, systems that learn and adapt over time.

17 June 11:15

Title: Timepoint selection for long-term PRO data modelling in oncology trials

Anna Rigazio[1], Konstantina Skaltsa[1], Pavol Kral[2], Carolina Pestana[3], James Bell[4]

[1]IQVIA, Barcelona, Spain. [2]IQVIA, Bratislava, Slovakia. [3]IQVIA, Lisbon, Portugal. [4]Elderbrook Solutions GmbH, High Wycombe, United Kingdom

Abstract

In oncology studies, it is common for patient-reported outcome (PRO) objectives to be supportive and of particular interest to clinicians and health technology assessments (HTAs). Traditionally, the mixed model for repeated measures (MMRM) has been used to model these data assuming missing data after events like treatment discontinuation or death are missing at random (MAR). While this may be reasonable for fixed duration studies with low rates of intercurrent events and/or missing data, oncology presents unique challenges: 1) difficulty in defining a specific timepoint of interest, 2) focus on a long time-horizon, and 3) variable number of assessments across patients due to continued PRO collection until treatment discontinuation. This leads to substantial missing data that may make estimates unstable and biased under MAR. The number of timepoints to include in the model becomes then a statistical issue, with arbitrary thresholds for the number of patients providing valid data at various timepoints, often set at 10%, 30%, or 60% of patients in both treatment arms that may lack a solid empirical basis.

In this talk, we report a simulation study to inform these thresholds. By simulating different scenarios of numbers of timepoints and missingness, we aim to identify more robust and evidence-based thresholds that can improve the reliability and interpretability of PRO data in oncology trials. This approach addressed a critical gap in the literature and offers a practical solution to a common problem in the analysis of oncology studies.

16 June 16:15

Title: Developing a simulation-based decision framework for interpretation of interim survival data in oncology trials

Alessandra Bisquera[1], Andrew Mills[1], Nevine Zarifa[2], Mark Stewart[3]

[1]MMS, Belfast, United Kingdom. [2]NMD Group, Philadelphia, USA. [3]Friends of Cancer Research, Washington, USA

Abstract

Interim analysis of overall survival (OS) presents ongoing challenges in oncology trials, especially when progression-free survival (PFS) shows clear benefits but OS data suggest harm in the presence of an unknown influence like immature data or true harm, for instance. While OS is the gold standard for assessing long-term patient outcomes, delays in robust OS results create a strong amount of uncertainty around early findings. This complicates decision-making, trial design, regulatory discussions, and patient access to new therapies.

To address these issues, Friends of Cancer Research brought together a multi-stakeholder consortium to develop practical frameworks to improve interpretation of interim OS data. In August 2025, the FDA released Guidance for Industry on Approaches to Assessment of Overall Survival in Oncology Clinical Trials.

This session details the development and implementation of these simulation-based methodologies that mimic real-world trial scenarios and evaluate various analytical strategies. The work focuses on cases where PFS is positive, but a complete OS analysis is required to define and subsequently rule out harm, guiding sponsors and regulators by quantifying uncertainty under pre-specified truth scenarios (simulation parameters). The consortium's frameworks will be presented and include simulation of correlated endpoints, efficient analysis and estimation of operating characteristics, and approaches for effectively presenting complex, multi-dimensional results.

Attendees will learn from practical case studies demonstrating how these tools inform trial decision-making, improve consistency and transparency in interim analyses, and facilitate the timely and reliable approval of innovative oncology treatments for patients.

17 June 12:35

Title: How Statisticians Can Use the Growth Mindset Framework for Stronger FSP Success in Pharma

Amy Spencer

MMS Holdings, Belfast, United Kingdom

Abstract

In a recent report featured on Forbes.com from TalentLMS, 89 percent of senior leaders shared that future success depends on having a growth mindset. This includes key principles like embracing challenges, learning from feedback, valuing effort, resilience in the face of failure, and inspiration from others.

Drawing from real-world experience, we will explore how statistical and technical capabilities must evolve beyond core analysis skills to include strong communication, regulatory awareness, and leadership within cross-functional environments. Flexibility remains essential, as

statisticians may be called upon to shift between teams, take on new responsibilities, support urgent deliverables, and balance multiple project priorities while maintaining accuracy and compliance.

The presentation highlights the critical “soft” skills required for success in FSP settings, such as explaining statistical methods to non-statisticians, representing expertise in sponsor meetings, and being proactive in finding solutions when regulatory or project needs shift. Leadership development through a growth mindset is also key: statisticians embedded in pharma teams are frequently asked to lead small programming teams, contribute to strategy, or act as domain experts in high-level discussions.

Emphasis will also be placed on how transferable skills, like document interpretation and teamwork across departments, help bridge the gap between technical excellence and operational impact.

Attendees will leave with actionable strategies to support FSP success, whether they are statisticians seeking to sharpen their influence or organizations looking to structure effective cross-functional collaborations.

16 June 14:00

Title: Assessing the impact of interim decisions in group sequential trials

Gianmarco Caruso[1], William F. Rosenberger[2], Pavel Mozgunov[1], Nancy Flournoy[3]

[1]MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom. [2]George Mason University, Fairfax, USA. [3]University of Missouri, Columbia, USA

Abstract

Reliable estimation of treatment effects is central to clinical research because it shapes how trial results are interpreted and how confidently clinicians and regulators can assess a new treatment. Group sequential designs are increasingly used because they allow interim analyses and the option to stop a trial early 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 can also introduce a bias in the final treatment effect estimates. Using a group sequential trial example, we show that neglecting 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 evaluating individual treatment effect values. By comparing alternative decision boundaries and prior choices, we illustrate how the proposed measure can support the interpretation of trial results and inform the planning of future adaptive studies.

We also introduce a pre-trial version of the metric to support investigators in trial design choices, including the timing of interim analyses and stopping boundaries. This guidance complements traditional strategies based on type-I error control, such as Pocock or O'Brien-Fleming boundaries, by offering insights into the influence of interim decisions on the treatment effect estimates at each interim analysis. We illustrate its practical use in a group sequential trial evaluating a treatment for central nervous system disorders.

15 June 11:00

Title: The river keeps moving: How statisticians can survive (and flourish!) in a world that never stands still

Kimberley Hacquoil[1], Lucy Rowell[2], Frances Denny[3], Sam Ruddell[4]

[1]Veramed, Jersey, Jersey. [2]Impactful Authenticity, High Wycombe, United Kingdom. [3]MMS Holdings Europe Ltd, Belfast, United Kingdom. [4]Chiesi Ltd, Manchester, United Kingdom

Abstract

Heraclitus, the ancient Greek philosopher, is best known for his phrase "No man ever steps in the same river twice, for it's not the same river and he's not the same man". This is a reminder that change is the only constant in life and that's certainly true of the pharmaceutical industry.

Industry context: An overflowing river

The industry is amid rapid transformation driven by AI integration which is projected to grow from more than $4 billion in 2025 to $25.7 billion by 2030. Some pharma leaders believe this effort will require a complete reimagining of drug discovery and development workflows. There are also shifts in the industry with regards to recent large pharmaceutical layoffs, restructuring, increased M&A and licensing deals, and uncertainty around payer landscape and pricing pressures. For statisticians, these changes are not abstract trends; they shape our roles, responsibilities, and career trajectories.

The emotional side: Riding the rapids of change

The emotional side of managing change often determines whether we merely survive or truly flourish. Change evokes a spectrum of responses from anxiety, resistance, excitement, and growth. The Kubler-Ross change curve was originally developed to describe grief. It maps emotional stages related to shock, denial, frustration, depression, experimentation, decision, and integration. Recognising these phases in ourselves and others can help individuals and teams normalise reactions, be resilient and move toward acceptance during periods of change.

Personal insights: Reflections on the water

Statisticians often find themselves at the intersection of stability and disruption as our work demands rigour and precision, yet the environment around us is fluid. How do we thrive in this paradox?

During a facilitated panel discussion, a diverse group of presenters will provide theories and models, personal examples, lived experiences, and insightful reflections related to managing change both from an individual perspective and in supporting others. It will cover aspects relating to change as a statistician ranging from large change initiatives, restructuring, and everyday curve balls! It will address these from two angles:

- when change happens *to us*
- when we have the chance to *shape it*

By blending industry context with personal lessons learned, we hope to spark a conversation on how we can embrace change as a catalyst for growth and impact, both professionally and personally. Attendees will be invited to share their own reflections as well, creating a collaborative exploration of what it means to thrive in the pharmaceutical world that never stands still.

17 June 11:15

Title: A Comprehensive Self-Adaptive Mixture Prior Approach to Dynamic Borrowing from External Data

Alfredo Farjat[1], Ming-Dauh Wang[2]

[1]Bayer B.V., Hoofdoorp, Netherlands. [2]Bayer HealthCare Pharmaceuticals, Whippany, NJ, USA

Abstract

Mixture priors have been proposed for borrowing information from external data sources to complement the analysis of clinical trial data. A mixture prior is constructed by combining two distributions, a vague prior expressing no effect and an informative prior that reflects the observed effect in the external data sources. The mixing weight can be pre-specified, which often involves a tipping point analysis to assess the robustness of the findings relative to the chosen weight. Alternatively, the weight could be estimated from the data based on the similarity between the external and current data, as recently suggested through self-adaptive mixture priors. In this study, we propose a comprehensive dynamic mixture prior approach. The informative prior is derived from external data using the propensity score integrated power prior method, which accounts for baseline differences between trial participants and those from external sources. To address the potential subjectivity in pre-specifying the mixture weight, a data-driven approach is proposed based on the posterior predictive probability of observing the current data given the external data; this method employs a class of arctangent elastic functions to adaptively adjust the mixture weight. The properties of our approach are examined through numerical simulations. Furthermore, the proposed approach is illustrated in a phase II randomized controlled trial, where the control arm is augmented with real-world data. Our simulation study shows that the proposed method exhibits desirable operating characteristics allowing for adaptive information borrowing while effectively managing baseline and outcome differences.

16 June 14:00

Title: Beyond Chance: Randomisation Designs for Innovative Clinical Trials (Randomisation SIG)

Diane Uschner[1], Johannes Krisam[2], Peter Jacko[3], Ayon Mukherjee[4]

[1]F. Hoffmann-La Roche, Basel, Switzerland. [2]Boehringer Ingelheim Pharma GmbH & Co. KG, Ingelheim, Germany. [3]Lancaster University, Lancaster, United Kingdom. [4]Eli Lilly, Basingstoke, United Kingdom

Abstract

From Design to Estimand: Stratification, Precision, and Randomisation-Based Inference (Speaker: Diane Uschner)

Randomisation is the cornerstone of causal inference in clinical trials, but the choice of restricted randomisation scheme—particularly stratification—is often oversimplified in analytical practice. This presentation addresses the critical methodological link between controlling covariate imbalance at the design stage and ensuring valid quantification of uncertainty during analysis. We will first explore the practical implications of stratification, demonstrating how the proper inclusion of important prognostic covariates improves precision and bias control, while cautioning against the risks associated with omitting a key covariate or the pitfalls of over-stratification, which can complicate logistics and reduce efficiency. We then connect stratification directly to the modern concept of the target estimand, examining how the defined stratum-specific effects influence the overall treatment effect being estimated. Crucially, we will discuss how the stratification variables and the trial's target population relate to the generalisability and transportability of the treatment effect to the broader patient community. Finally, we provide principled guidance on performing Randomisation-Based Inference (RBI). We will detail practical approaches for incorporating stratum information into the RBI framework (e.g., in the construction of the reference set for randomisation tests) to ensure the analytical method respects the design restrictions. This yields valid p-values and confidence intervals that accurately reflect the restricted sampling space, offering a robust, model-free pathway to align trial design and statistical inference.

How the Chosen Randomisation Method Affects the Individual Clinical Trial Participant (Speaker: Johannes Krisam)

Randomisation is fundamental to clinical trials, yet its impact on participants is often overlooked. Stemming from the need of proper informed consent, informational material such as videos have been created by health authorities, which aim at bringing the concept of randomisation to the trial participant. These instructional materials simplify the randomisation process by equating it to "a fair coin flip" - suggesting equal probability of assignment to each treatment arm, and state that randomisation is applied in order to ensure that the allocation is not predictable. In this session we will dive into Permuted Block randomisation, the most common form of randomisation applied in clinical trials and see how closely that method comes to fulfilling what the participant-facing materials promise regarding a "coin flip" allocation and unpredictability of randomisation. We will review and evaluate the class of so-called Maximum Tolerated Imbalance (MTI) randomisation methods, that are currently under-used in clinical trials, by means of their conditional allocation probabilities. We will derive performance characteristics which measure the randomness/fairness of both block-based and MTI procedures, showing that MTI methods represent a more ethical and unbiased approach to randomisation in clinical trials. Additionally, we will review the regulatory landscape to assess the feasibility of implementing these alternative randomisation strategies and offer concrete suggestions on which method to choose based on the trial design.

Designing Master Protocol Trials for Single-Arm Studies (Speaker: Peter Jacko)

In this talk we propose and examine randomisation and allocation procedures in master protocol trials where each subtrial is performed as a single-arm study, which is motivated mainly by rare diseases. The subtrial analysis is done by comparing the single-arm observations data to prespecified historical rates without using data from any other study in the master protocol trial. The objective of this paper is to examine relevant options for participant allocation procedures in a variety of patterns in which the subtrials enter the master protocol trial, such as umbrella, platform and perpetual patterns. Our simulation study shows that our proposed allocation procedures using the predictive probability of current study success may bring substantial efficiency gains in the master protocol trial, making a clever allocation of the in-trial participants to achieve study success declarations sooner, benefiting the out-of-trial population.

This is joint work with Guenter Heimann and Tom Parke.

Covariate-Adjusted Response-Adaptive Designs for Semiparametric Survival Models (Speaker: Ayon Mukherjee)

Covariate-adjusted response adaptive (CARA) designs aim to increase the likelihood that patients in a clinical trial receive the superior treatment based on their individual covariate profiles. While recent research on CARA designs often relies on parametric assumptions about patient responses, this limits their practical application in real-world clinical settings. Although Sverdlov et al. (2013) demonstrated that CARA designs based on the exponential model can yield valid inference when the final analysis uses the appropriate accelerated failure time (AFT) model, most survival trials do not utilise the AFT model as their primary analysis. This presentation introduces the proposed innovative CARA design methodologies that do not depend on specific distributional assumptions for survival outcomes; instead, they rely solely on the proportional hazard assumption between treatment arms. To address multiple trial objectives, these designs employ an optimal allocation strategy. Patients are randomised using covariate-adjusted doubly adaptive biased coin and efficient-randomised adaptive designs to achieve expected allocation targets, which are derived as functions of Cox regression coefficients updated sequentially with each participant. The advantages of these approaches will be illustrated using extensive simulation studies evaluating their operating characteristics, and the session will conclude by showcasing the practical implementation of these designs through the re-design of an actual confirmatory clinical trial.

15 June 11:00

Title: Patient-Focused Drug Development SIG: Tolerability PROs across the drug development lifecycle

Emily Alger[1], Lorenz Uhlmann[2], Antoine Regnault[3], Konrad Maruszczyk[4]

[1]The Institute of Cancer Research, London, United Kingdom. [2]Boehringer Ingelheim, Biberach an der Riß, Germany. [3]Modus Outcomes, Lyon, France. [4]University of Birmingham, Birmingham, United Kingdom

Abstract

1. **Introduction by session chair (Alexandra Lauer)**
2. **Inclusion of PROs in Dosing Finding Trials (Emily Alger)**: There is growing scientific interest in incorporating Patient-reported Outcomes (PROs) in early phase dose-finding oncology trials (DFOTs) to assess tolerability, inform dose selection, and guide later stage trial design. However currently, PRO objectives, analysis, and reporting within DFOTs is often unclear and inconsistent. OPTIMISE-ROR (incOrporating PaTIent-reported outcoMes In doSE-finding trials–Research Objectives Recommendations) and OPTIMISE-AR (OPTIMISE-Analysis Recommendations) were established to provide foundational recommendations for trialists wishing to leverage the early insights provided by PROs within their DFOT. The OPTIMISE-ROR project identifies critical PRO research objectives relevant to DFOTs. Building upon these recommendations, OPTIMISE-AR provides a practical toolkit supporting the statistical analysis, visualisation, and reporting of PRO data within publications in line with such critical objectives. International, multidisciplinary, cross-sector statistical analysis and data visualisation working groups identified analytical and visualisation approaches addressing these key DFOT PRO research objectives. Informed by existing literature, case studies and recommendations are provided for analysing binary, ordinal, and continuous PRO data to assess tolerability across dose levels and timepoints and integrating PROs within interim and final dose-decision processes. This talk will detail how OPTIMISE-ROR and OPTIMISE-AR facilitate the systematic integration of PROs in DFOTs.
3. **A meta-analytic approach to the assessment of tolerability in early phase oncology trials (Lorenz Uhlmann):** Tolerability is defined as the degree of symptomatic side effects patients are willing to endure in order to adhere to their assigned treatment. Unlike adverse events, tolerability is solicited from patients by means of PROs such as the Patient-Reported Outcomes Common Terminology for Cancer Adverse Events (PRO-CTCAE) or generic measures of Overall Side Effect Bother/ Burden, such as FACT-GP5/ EORTC-IL46 (Q168). Tolerability data plays a crucial role in the dose characterization in early phase oncology trials. Beamion LUNG-1 is an ongoing Phase Ia/II study assessing efficacy and safety of zongertinib in HER2 mutant NSCLC patients. Several cohorts of rather small sample sizes were included which led to limitations in terms of the interpretability of cohort-specific results. In support of a holistic characterization of the dose's tolerability profile the question arises if patients' tolerability ratings across cohorts are sufficiently homogeneous, allowing researchers to pool the data. We approach this question from a meta-analytic perspective. Tolerability data is ordinal in nature and longitudinally collected, allowing us to apply existing methods based on proportional odds models. We introduce a meta-analytic framework for the statistical evaluation of the degree of heterogeneity between cohorts. Several approaches pertaining to modeling and inference are presented and discussed.
4. **Confirmatory comparative tolerability endpoints in cancer clinical trials: Strategic opportunities and realisation challenges (Antoine Regnault):** Comparing the tolerability of therapeutic options is increasingly considered a relevant confirmatory endpoint in cancer clinical trials. This presentation will illustrate the underpinning principles of such comparative tolerability endpoints as well as the challenges with their operationalization and interpretation. Trial endpoints for this purpose should summarize the multifaceted notion of tolerability with a single

metric allowing a simple, but fair comparison. The assessment of overall side-effect bother, evaluated using a single, patient-reported item, over a set period of exposure is currently the main option for comparative tolerability endpoints: proportion of participants with severe side-effect bother and proportion of time with severe side-effect bother. Comparative tolerability endpoints should address the question of baseline assessment of tolerability, which is often controversial. They should also clearly define the targeted exposure period, and account for the parts of this period when the participant burden of side-effect is unknown, either because it was not collected by design (e.g., in the case of it being assessed once by cycle) or because the assessment is missing. Finally, the interpretation of these endpoints should discuss the possible effect of the different length of exposure for the compared treatment options (e.g., in case of differential time to treatment discontinuation).

5. **Advancing Tolerability Assessment Through Real-World Patient-Reported Outcomes (RW-PROs) (Konrad Maruszczyk)**: Real-world evidence (RWE) plays an increasingly important role in global regulatory and reimbursement decision-making. Incorporating patient-reported outcomes (PROs) into RWE can enrich the understanding of treatment effectiveness, safety, and tolerability from the patient perspective. Real-world patient-reported outcomes (RW-PROs) are direct reports from patients, captured without interpretation by clinicians or others, and collected outside the artificial setting of the traditional clinical trial environment. RW-PROs offer opportunities to assess the tolerability of healthcare interventions as delivered in routine care. Their use enables the inclusion of broader and often underrepresented populations, such as individuals with multimorbidity, pregnant women, and underserved groups. Although regulators, health technology assessors, and policymakers acknowledge the value of PROs in real-world studies, formal guidance on their optimal use remains limited. This session will explore the value of RW-PROs and identify opportunities and facilitators for their wider implementation in real-world studies. Key barriers and methodological and operational challenges that must be addressed to maximise their impact will also be highlighted. Overall, the session provides a timely opportunity to reflect on the evolving landscape of PROs in RWE and to discuss their growing role in supporting patient-centred evidence generation.
6. **Panel discussion of opportunities and challenges of tolerability PROs throughout the drug development lifecycle.**

15 June 11:00

Title: Don't Get Distracted by Noise: Simulations Done Right

Isabelle Smith[1], Sam Miller[2], Andy Grieve[3], Tim Friede[4]

[1]Veramed, London, United Kingdom. [2]MMS, Belfast, United Kingdom. [3]Weatherden, London, United Kingdom. [4]University Medical Center, Göttingen, Germany

Abstract

Simulations are useful in a statistician's toolkit, but without proper training, can be underutilised, or poorly implemented which could lead to inappropriate or unclear recommendations for the conduct of a study. When done well, simulations can support critical tasks such as:

- Sample size determination
- Analysis method comparison
- Quantifying benefits and risks of design options
- Scenario planning (not only for trials but also for whole development programmes)

Unlike clinical trials, simulation studies are not governed by formal regulations. However, best practice guidance exists and in 2017 the EFSPI Modelling and Simulation SIG published recommendations for pre-specifying goals, assumptions, methods and outputs[1]. The Clinical Scenario Evaluation (CSE) framework is an early example providing some guidance on how to structure a simulation study and to share information between stakeholders[2]. Frameworks such as ADEMP[3] have also since emerged to guide planning and conduct of simulation studies to ensure they are fit for purpose.

Regulatory expectations are increasing. Where simulations are the main method used to quantify operating characteristics, the draft ICH E20 guidance for adaptive designs calls for a detailed report including the simulation plan, results, and code to enable verification. This makes programming language choice and computational efficiency critical for regulatory review.

This workshop is designed for statisticians who want to strengthen their simulation skills, whether you're new to simulations, experienced but seeking best practices, or responsible for overseeing simulation work.

Supported by the PSI Professional Development Committee, this workshop will be facilitated by academic and industry experts, and will take participants through each step of conducting a simulation study including:

- When and why a simulation study is required
- How to plan and structure simulations
- Practical considerations, such as timelines and scale of computing resources required
- How to QC and check assumptions
- Pitfalls of repurposing existing code
- Communicating results to a non-statistical audience.

Participants will work in breakout groups using real examples, exploring strategies to improve design and implementation using best-practice principles. By the end of the workshop, participants will be able to design, implement, and critically review simulation studies, prepare documentation for regulatory submissions, and communicate results effectively to diverse stakeholders.

1. *O'Kelly, M., Anisimov, V., Campbell, C. and Hamilton, S., 2017. Proposed best practice for projects that involve modelling and simulation. Pharmaceutical statistics, 16(2), pp.107-113.*
2. *Benda N, Branson M, Maurer W, Friede T (2010) Aspects of modernizing drug development using scenario planning and evaluation. Drug Information Journal 44: 299-315.*
3. *Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Statistics in Medicine. 2019; 38: 2074–2102.*

17 June 11:15

Title: Information borrowing in Bayesian clinical trials: choice of tuning parameters for the robust mixture prior

Vivienn Weru[1], Annette Kopp-Schneider[1], Manuel Wiesenfarth[2], Sebastian Weber[3], Silvia Calderazzo[1]

[1]German Cancer Research Center (DKFZ), Heidelberg, Germany. [2]Cogitars GmbH, Heidelberg, Germany. [3]Novartis Pharma AG, Basel, Switzerland

Abstract

External data borrowing in clinical trial designs has increased in recent years. This is accomplished in the Bayesian framework by specifying informative prior distributions. To mitigate the impact of potential inconsistency (bias) between external and current data, robust approaches have been proposed. One such approach is the robust mixture prior arising as a mixture of an informative prior and a more dispersed prior inducing dynamic borrowing. This prior requires the choice of four quantities: the mixture weight, mean, dispersion and parametric form of the robust component. To address the challenge associated with choosing these quantities, we perform a case-by-case study of their impact on specific operating characteristics in one-arm and hybrid-control trials with a normal endpoint. All four quantities were found to strongly impact the operating characteristics. As already known, variance of the robust component is linked to robustness. Less known, however, is that its location can have severe impact on test and estimation error. Further, the impact of the weight choice is strongly linked with the robust component's location and variance. We provide recommendations for the choice of the robust component parameters, prior weight and alternative functional form for this component.

16 June 16:15

Title: The estimand conundrum - is ICH E9 R1 crystal clear or are there still areas of confusion?

David Wright[1], Frank Petavy[2]

[1]AstraZeneca, Cambridge, United Kingdom. [2]European Medicines Agency, Amsterdam, Netherlands

Abstract

Introduction (talk 1): (David Wright (AstraZeneca)

ICH E9 (R1) came into effect nearly 6 years ago. However, there is a lack of alignment in the statistical, clinical and regulatory communities on a number of key issues. A sub-team of the EFSPI/EFPIA Estimand Implementation Working Group have identified key topics where clarification is needed and these will be shared.

One of the main motivations for this session is the observation that Estimands in both pharmaceutical and regulatory settings are too often considered as a statistical topic. This was never the intention and this presentation will explore ways in which the statistical community can change their status quo in leading these discussions and allow others, in particular clinicians, to drive conversations on estimands. Only then will we be able to fully realise the potential the estimand framework offers in designing, conducting and analysing trials that are better aligned to key clinical questions of interest and achieving alignment with key stakeholders on treatment effects when trials are designed.

Another motivation is the recent paper in Statistics in Medicine by Fleming et al: https://onlinelibrary.wiley.com/doi/epdf/10.1002/sim.70104. Areas of disagreement between the message of this paper and letters sent to the editor will be summarised too

Talk 2: Reflections from Industry representatives involved in writing ICH E9 (R1) (Novartis clinical and statistical representatives have been invited to present if this session is accepted. If they are unable to attend alternative presenters will be found from EIWG)

Given the challenges with implementation highlighted above this presentation will focus on key areas where change (either in terms of how estimands are implemented, communicated or indeed the choice of estimand) is needed.

Talk 3: Reflections from regulatory representatives (Frank Petavy (or another EMA representative) will present)

This presentation will focus on areas regulators feel sponsors could improve in how estimands are being defined, how chosen estimands are justified in submission documents (e.g. protocols, summary of clinical efficacy etc) and in how estimands are positioned for feedback in scientific advice procedures.

Panel discussion: There will be time for Q&A from the audience.

16 June 16:15

Title: Enhanced reconstruction of pseudo-individual patient data using quadratic programming

Andrew Titman

Lancaster University, Lancaster, United Kingdom

Abstract

Extracting information from digitised Kaplan-Meier plots in published papers is increasingly common particularly within health technology assessment, for instance to allow secondary analysis for the purposes of survival extrapolation. It can also enable the use of patient-level meta-analysis techniques or facilitate the use of external data into clinical trial designs. Guyot et al's algorithm for obtaining pseudo-individual patient data (IPD) is by far the most used approach, but its ad hoc construction makes it difficult to extend.

In this presentation, reconstruction of IPD is framed as a constrained optimisation problem for which well-established quadratic programming methods can be applied. Under this framework additional information such as marked censoring times can easily incorporated when reconstructing IPD from a Kaplan-Meier curve while also ensuring the IPD is consistent with the reported numbers at risk and total events. Moreover, the same approach can be extended to allow reconstruction of IPD from cumulative incidence functions from competing risks analysis, reconstruction of left-truncated survival data and to ensure IPD reconstructed from progression-free survival and overall survival curves are mutually admissible. An R package, CIFresolve, has been developed to implement the methods, and will be illustrated on a range of examples.

17 June 11:15

Title: Real-world data – do you know all the opportunities? The key questions they can answer and how (RWD SIG)

Eleanor Ralphs[1], Josie Wolfram[2], Rima Izem[3]

[1]IQVIA, London, United Kingdom. [2]Astellas, Leiderdorp, Netherlands. [3]Novartis, Basel, Switzerland

Abstract

Is there an unmet need? What is happening in clinical practice? Are clinical trials feasible? Are there safety concerns? Is the treatment efficacious? Is the treatment cost-effective? What is the impact on the patient, beyond key outcomes? These are all questions that studies on real-world data can help answer. The session, sponsored by the RWD SIG, will touch on the rationale for asking each of these questions and will dive into applied examples of real-world studies, with a focus on statistical and epidemiological considerations.

Eleanor Ralphs (IQVIA): Leveraging RWD to assess unmet need and insights into clinical practice

Clinical trial design and regulatory/HTA decisions often rely on limited evidence from controlled settings, which may not fully reflect real-world patient populations or long-term outcomes. Real-world data can address these gaps by providing insights into disease characteristics, treatment patterns, adherence in routine clinical practice, clinical outcomes, health care resource utilisation and patient-reported outcomes. This information informs an array of decisions that will be outlined in this session.

Additionally, applied examples of retrospective cohort studies will be presented, including collaborative projects across diverse geographies and tumor types. The first case study will examine treatment patterns and clinical outcomes in patients with non-small cell lung cancer across Spain, Germany, Canada, and England, demonstrating how adherence can be evaluated over extended follow-up periods beyond traditional trial timelines. The second case study will focus on high-risk endometrial cancer patients using England's national cancer registry to inform trial eligibility decisions and explore opportunities for potential label expansion.

These case studies will illustrate the value of collaborative approaches to leverage real-world data for decision-making in oncology.

Josie Wolfram (Astellas): Leveraging RWD for safety evaluation through the lifecycle

Safety monitoring is a key component of the conduct of clinical trials. The use of real world (or historical trial) data can provide important context to any emerging safety signals and enhance safety evaluation through the development phase. Furthermore, though assessing safety using pre-marketing clinical trials is necessary, it is often not sufficient to evaluate adverse events due to the limited patient numbers. This can lead to post-approval studies that go beyond routine safety surveillance to facilitate careful evaluation of potential or identified risks. Reflecting a number of years' experience with such studies, recently the ICH M14 Guideline on "General Principles on Plan, Design and Analysis of Pharmacoepidemiological Studies That Utilise Real-World Data for Safety Assessment of Medicines" entered the implementation phase and will come into effect in March 2026. This talk will firstly share some example practises that can support safety evaluation during development, and then highlight the key methodological and statistical considerations in the design and analysis of post-approval safety studies by use of a real completed example of a multi-country, multi-data source collaborative safety study program.

Rima Izem (Novartis): Leveraging external data to the clinical trial for efficacy evaluation, where we are and a look forward

This presentation will review case studies for the use of external data along with clinical trials in generating evidence in drug development and discuss remaining challenges and future opportunities for these hybrid approaches. The use of historical data to complement or contextualize single arm trials at the time of approval is illustrated in multiple case studies in oncology or rare diseases (e.g., as presented in the recent EMA workshop on external controls (November 3rd 2025)). These case studies illustrate when, whether, and how one can mitigate potential bias in a non-randomized comparison to inform the treatment effect estimation.

Looking forward beyond these case studies, there are multiple opportunities for using hybrid approaches combining information from RWD with clinical trials. Those include linkage to fill gaps in the journey of clinical trial participants with external sources, transporting or generalizing the treatment effect from the clinical trial to another population, and extrapolating the treatment effect beyond the time period observed in the trial. Some of main challenges with all these approaches include assessing feasibility, external data reliability, and access to sufficient individual patient data. Statisticians are uniquely positioned in generating evidence but also guiding decision making in the presence of potential bias to benefit patients.

17 June 11:15

Title: Impact of interval-censored data on comparative time-to-event endpoints: a simulation study applied to patient-reported outcomes in oncology

Joel Sims, Emma Martin, Mike Greenwood, Rachael Lawrance

Adelphi Values, Manchester, United Kingdom

Abstract

Patient-reported outcome (PRO) time to deterioration (TTD) endpoints are assessed at scheduled visits, so true event times are only known to lie between visits. This interval censoring is rarely accounted for when using Cox proportional hazards (PH) or Kaplan-Meier (KM) methods, and its impact on treatment effect estimation is not routinely quantified. We conducted a simulation study to evaluate bias of the Cox PH hazard ratio (HR) under interval censoring. Three main data generating mechanisms were evaluated: 1) varying on-treatment visit intervals (0–16 week intervals); 2) asymmetry of visits (e.g. 3-weekly vs 4-weekly); and 3) off-treatment assessments with reduced frequency, triggered by intercurrent events (ICE). True event times were assigned to the first visit at which they would be observed or imputed as the mid-point of the visit interval. Weibull-distributed TTD times were simulated to reflect realistic deterioration patterns. HR bias, 95% CI coverage, and KM estimates were summarised.

Under symmetric assessment schedules (≤16-week intervals), HR estimates showed negligible bias and acceptable coverage but KM median estimates increase with longer visit intervals. Asymmetric assessment frequencies on the other hand induce substantial bias favouring the arm with less frequent assessments. Differential transition between arms to reduced-frequency off-treatment assessments results in biased HRs (e.g. estimated HR~1.2 even with true HR=1). Mid-point imputation attenuated these biases. In conclusion, the selection of a suitable PRO assessment schedule impacts interpretation of the treatment effect on TTD endpoints due to interval censoring; regular PRO assessment frequency continuing post-ICE is the least biased approach.

17 June 12:35

Title: Mircroleadership: The habits of building leadership behaviours

Emma May

Independent, Cambridge, United Kingdom

Abstract

Traditionally we associate leadership with a job title; someone who manages departments, budgets, teams. 'Big Leadership' is a role defined by authority and hierarchy.

In contrast, the proposed new framework is defined by clarity, connection, and consistency. 'Microleadership' is defined by small, intentional actions that build trust, clarity, and momentum, without needing the authority of a leadership job title.

Small changes in everyday behaviours such as the way we listen and ask questions, or how we respond to a challenge, if practiced regularly, become transformative habits. These microsteps, over time, develop into influential and impactful behaviours that enable everyone, from new graduate statisticians to technical experts, and those with a job title, to lead from any seat and in multiple directions.

In this talk, you be be introduced to the six foundational components of Microleadership: Microenergy, Microculture, Micropresence, Microtrust, Microclarity, and Microcourage.

You will leave the talk with an actionable first step on your Microleadership journey to develop trust, bring clarity, and act courageously.

The Microleadership framework is a digestible and useable format based on discerning the content of multiple professional development books, podcasts, blogs and courses, drawing on behavioural science, communication research, and organizational psychology.

15 June 13:30

Title: Academic-Industry Collaboration and Connection

Ian Wadsworth[1], Tom Burnett[2], Sue Todd[3]

[1]Phastar, Macclesfield, United Kingdom. [2]University of Bath, Bath, United Kingdom. [3]University of Reading, Reading, United Kingdom

Abstract

The work of statisticians in the pharmaceutical industry has its foundations in statistical research. In some instances, this is research conducted in house by companies but often this research comes from academia. Industry/academic collaboration can drive research that is relevant to real world problems and accesses appropriate experts from academia benefiting both parties. These benefits span both the wider objectives of the institutions involved and the individuals working together.

In this session we shall draw together the benefits of these collaborations sharing examples from across a spectrum of activities, discussing how these connections are formed, how they contributed to the work and the benefits experienced by both parties.

The first half of this session will consist of a selection of bite size talks from people working in such collaborations. Topics covered include direct collaborations, academic/industry interactions for large scale grants and bids, and joint initiatives for PhD supervision.

In the second half, there will be a panel discussion with people from both industry and academia, across different career stages. Topics to be explored include the benefits of such collaborations to both industry and academia, and the challenges and opportunities associated with pursuing joint collaborations.

17 June 12:35

Title: From Priors to Partnership: A Bayesian View of CRO-Sponsor Collaboration

Kimberley Hacquoil

Veramed, Jersey, Jersey

Abstract

Effective collaboration between CRO and sponsor biostatisticians is essential for delivering high-quality clinical trials, yet these partnerships can become strained due to misaligned expectations, unclear responsibilities, and communication gaps. As drug development has evolved towards greater complexity and faster decision-making, traditional collaboration models have struggled to adapt.

Drawing on experience working across both pharma and CRO settings, this presentation reframes collaboration as a Bayesian updating process. Teams begin with prior beliefs shaped by past experiences, generate evidence through day-to-day interactions and deliverables, and update trust and behaviour over time. When expectations, roles, and objectives are unclear, this evidence becomes noisy, leading to friction, rework, and defensive behaviours. When well aligned, collaboration becomes proactive, transparent, and resilient.

The talk explores two key levers for improving collaboration: robust processes that reduce ambiguity in the evidence, and aligned mindsets that support shared ownership. Attendees will gain practical insights into how clearer ways of working and a shift towards collective success can transform siloed interactions into effective, sustainable partnerships that support better trial outcomes.

16 June 14:00

Title: Framework for timing interim analyses in longitudinal trials with missing data: the role of blinding and sample size

Neža Dvoršak[1], Jianmei Wang[2], Thomas Burnett[1], Christopher Jennison[1], Robin Mitra[3]

[1]University of Bath, Bath, United Kingdom. [2]Roche, Welwyn Garden City, United Kingdom. [3]UCL, London, United Kingdom

Abstract

Interim analyses are key decision points in adaptive clinical trials; yet, determining when sufficient information has accrued to justify an interim inspection remains challenging. This is particularly true in longitudinal settings, where missing data are inevitable due to ongoing recruitment and data accrual.

We propose a framework that leverages all available data to guide the timing of interim analyses. Central to our approach is the concept of *Equivalent Sample Size,* introduced as an alternative to Fisher information, providing a principled way to quantify and predict information growth over time.

Two practical challenges arise at the interim stage: being blinded to treatment allocations, and the sample for estimating correlations may be small. In our framework, the Equivalent Sample Size depends only on the number of observed missing-data patterns and a correlation estimate. Counting missing-data patterns is straightforward, but the quality of the correlation estimate may vary depending on whether treatment assignments are known and the amount of observed data. We therefore investigate how these factors influence the Equivalent Sample Size – and, in turn, assess the robustness of the framework.

Through a simulation study, we compare blinded and unblinded correlation estimates and evaluate their bias relative to the theoretical value. We incorporate these estimates into the Equivalent Sample Size calculation to assess the resulting bias and variability as a function of the available interim data. The results provide practical guidance on how blinding and interim sample size affect the robustness of information assessment and, ultimately, the timing of interim analyses.

16 June 14:00

Title: JCA Insights Unleashed: What Statisticians Can Learn from the first JCA Procedures

Katrin Kupas[1], Lara Wolfson[2]

[1]Merck Healthcare KGaA, Darmstadt, Germany. [2]MSD, Brussels, Belgium

Abstract

Step into the heart of the Joint Clinical Assessment (JCA) revolution for the 2025 HTA Town Hall! This plenary session invites statisticians from regulatory and HTA domains to dive deep into the inaugural JCA procedures—where methodological rigor meets real-world impact. Discover pivotal lessons from the scoping process and final PICOs, and hear firsthand insights and practical implications from statisticians who shaped the first dossier submissions. We'll unravel the evidence and assessment strategies that defined the initial JCA reports, spotlighting both challenges and breakthroughs. Expect dynamic, though-provoking presentations and a lively panel discussion that brings together diverse viewpoints on how to navigate this new process. Join us to exchange ideas, debate solutions, and help chart the future of JCA through the lens of statistical innovation.

15 June 15:45

Title: Data Monitoring Committees - Best Practices and Future Development

Martin Jenkins[1], Tim Friede[2], Sue Todd[3], Chrissie Fletcher[4]

[1]AstraZeneca, Cambridge, United Kingdom. [2]University Medical Center Göttingen, Gottingen, Germany. [3]University of Reading, Reading, United Kingdom. [4]GSK, Stevenage, United Kingdom

Abstract:

Data Monitoring Committees (DMC) plays a critical role in safeguarding participant safety and ensuring trial integrity throughout clinical research. To foster consistency and best practices across DMC activities the Biopharmaceutical Statistics Leadership Consortium (BSLC), a cross-industry think tank, formed a DMC workstream to actively engage in discussions on current DMC practices and regulatory guidance as well as to advocate for developing the next generation of DMC experts.

This session will feature insights from the BSLC DMC workstream initiative including a landscape review of DMC best practices, regulatory expectations, including recent updated FDA draft guidance, and evolving industry standards for DMCs.

Key challenges will be discussed in terms of the proposed best practice positions suggested by the cross-industry workstream with responses from the varying perspective of the sponsor, DMC member and regulator. Topics to address include clarifying guidance on provision of efficacy and safety clinical trial data to DMCs, optimizing communication flow between DMCs and sponsors to ensure trial integrity, the relation of DMC decisions to non-binding decision boundaries, and ensuring transparency in DMC recommendations. Additionally proactive collaboration and targeted training to meet the growing demand for qualified DMC experts are emphasized.

The session aims to foster meaningful dialogue amongst esteemed leaders drawing on their diverse experiences in academia, industry, and government on best practices, training, and future directions for DMCs in clinical trials.