IBIG FORUM

22–23 OCTOBER 2026

University of Rome, "La Sapienza"

Day 1

Time	*Topic*	*Chair*
8.45 – 9.30	**Registration**	
9.30 – 9.40	**Welcome & Introduction**	Marco Costantini Stefania Gubbiotti
Enhancing Early Phases of Clinical Development		Giulia Zigon Stefano Vezzoli
9.40 – 10.10	**A ballad of a Basket trial and historical information borrowing: application in neurodegenerative diseases**	Pavel Mozgunov
10.10 – 10.40	**Bayesian Optimal Phase II design**	Maria Vittoria Chiaruttini
10.40 – 11.10	**Coffee Break**	
11.10 – 11.40	**Optimus Project**	Luca Genetti
11.40 – 12.10	**Anchoring the Winner: A Refined Conditional Test for Enhancing Efficiency in Multi-Arm Selection Trials**	Marco Ratta
12.10 – 12.40	**Gender Medicine: A Statistical Matter in preclinical and Phase I trial**	Clelia Di Serio
12.40 – 13.00	**Speed poster presentation**	
13.00 – 14.00	**Lunch**	
Causal Estimation in CTs: External Comparisons and Synthetic Data		Luca Grassano Lucia Simoni
14.00 – 14.30	**Effectiveness of Cannabidiol in Patients with Rare Epilepsies Compared with External Placebo Control: A Post Hoc Analysis from the Expanded Access Program**	Teresa Greco
14.30 – 15.00	**Informing MenB Vaccine for Adolescent Booster Immunization: Strategies through External Comparisons and Persistence Modeling**	Meike Adani
15.00 – 15.30	**Generative AI–Derived Synthetic Cohorts for Robustness Assessment and Sensitivity Analysis in Target Trial Emulation: A Framework for Real-World Treatment Effect Estimation**	Paolo Messina
15.30 – 16.00	**Coffee Break**	
16.00 – 16.30	**Evaluation of Artificial Intelligence-Generated Synthetic Data for Clinical Research in Secondary Cardiovascular Prevention of Patients with Dyslipidemia**	Alice Bonomi
16.30 – 17.00	**Evaluating historical benchmark Targeted Single-Arm Trials with External Controls single-arm platform trial in glioblastoma**	Ajsi Kanapari
17.00 – 17.30	**Targeting Clinically Relevant Causal Estimands in Trials with Latent Exposure-Induced Intermediate Confounding through Structural Equation Modeling and Propensity Score Weighting**	Domenico Iervolino

Day 2

Time	*Topic*	*Chair*
Issues in RCTs and Observational Studies		Andrea Nizzardo Elisa Rizzo
9.00 – 9.30	**Robust (Exchangeability) Bayesian Methods for Vaccines Targeting Multiple Serotypes**	Margherita Annaratone
9.30 – 10.00	**Operating characteristics of Bayesian testing procedures for clinical trials**	Fulvio De Santis
10.00 – 10.30	**Design of Single-Arm Time-to-Event Studies Using Simulation: Median, 95% CI and Optimal Sample Size using Shiny application**	Bianca Bracco
10.30 – 11.00	**Event Prediction in clinical trials**	Bastiaan Jansen
11.00 – 11.30	**Coffee Break**	
11.30 – 12.00	**Behind the Scenes of Real-World External Comparator Studies: Operational Issues and Lessons from the Field**	Alessandra Mignani
12.00 – 12.30	**RCTs vs Real-World Studies: Operational Insights from Academic Experience**	Lorena Torroni
12.30 – 13.00	**The importance and pitfalls of surrogate endpoints**	Tom Cattaert
13.00 – 13.10	**Forum Closure**	

ABSTRACTS

Bayesian Optimal Phase II design

Maria Vittoria Chiaruttini, Italfarmaco

Bayesian adaptive designs are increasingly used in early-phase oncology trials for their flexibility in interim decision-making and ability to incorporate multiple endpoints. The Bayesian Optimal Phase II (BOP2) design provides a framework for jointly evaluating efficacy and toxicity, although its application in multi-arm, multi-stage randomized trials remains limited.

A generalized BOP2 design was applied in a Phase II randomized clinical trial in patients with basal cell carcinoma, evaluating multiple treatment strategies across three arms. The design jointly assessed confirmed Best Overall Response Rate and grade ≥3 treatment-related adverse events, with adaptive stopping rules based on posterior probabilities and dynamic thresholds.

Simulation results showed appropriate control of the family-wise error rate under the global null hypothesis, high probabilities of early stopping for ineffective or unsafe arms, and substantial patient savings. Incorporating both efficacy and toxicity endpoints enabled a more comprehensive assessment of treatment benefit–risk profiles.

This application illustrates the feasibility and practical value of implementing the BOP2 design in Phase II trials, supporting adaptive and evidence-based decision-making while maintaining robust operating characteristics and regulatory alignment.

Anchoring the Winner: A Refined Conditional Test for Enhancing Efficiency in Multi-Arm Selection Trials

Marco Ratta, Saryga & Politecnico Turin

Multi-arm multi-stage (MAMS) trials with treatment selection are increasingly used to accelerate drug development by comparing multiple experimental treatments against a shared control within an adaptive framework. In “pick-the-winner” designs, one treatment is selected at interim, creating multiplicity and selection bias that require strict control of the family-wise error rate (FWER).

The proposed Rank-Based (RB) method conditions inference on the observed runner-up arm, which serves as a stochastic floor. A conditional tail probability is calculated for the selected arm, reflecting both its interim ranking and final cumulative evidence, and is used to construct a formal hypothesis test.

Strong control of the FWER is theoretically proven under global and partial null scenarios, and Monte Carlo simulations comparing RB and the Stallard and Todd framework across several configurations show that RB maintains error control while improving efficiency in competitive settings.

By conditioning on the realized runner-up, the RB method uses interim competition as inferential information rather than treating it solely as a feature of the selection process, improving power without sacrificing Type I error control.

Effectiveness of Cannabidiol in Patients with Rare Epilepsies Compared with External Placebo Control: A Post Hoc Analysis from the Expanded Access Program

Teresa Greco, Jazz Pharmaceuticals

Real-world evidence suggests that cannabidiol (CBD) may have utility in rare treatment-resistant epilepsies, including developmental and epileptic encephalopathies (DEEs). This post hoc analysis evaluated CBD effectiveness outcomes in patients with rare epilepsies from an expanded access program compared with an external placebo control arm derived from randomized clinical trials.

Patients received plant-derived, highly purified CBD. Effectiveness was assessed as percent reduction from baseline in median 28-day convulsive and total seizure frequency and compared with the external control using inverse probability weighting based on propensity scores to balance baseline characteristics.

The support region included 79 CBD expanded-access participants and 152 external-control participants with balanced baseline characteristics. Comparative analysis showed a 43% greater seizure reduction with CBD versus external control, with no changes in estimates when hidden bias and unobserved confounders were assessed.

These findings suggest that CBD may be an effective medication for a range of DEEs and other rare epilepsies and warrants further investigation. The use of external-control evidence in single-arm clinical trials may help strengthen the evidence base for informed clinical decision-making.

Informing Men B Vaccine for Adolescent Booster Immunization: Strategies through External Comparisons and Persistence Modeling

Meike Adani, GSK

Neisseria meningitidis serogroup B (MenB) causes significant morbidity and mortality, especially among infants and adolescents. While the rMenB+OMV NZ vaccine (Bexsero) is used in national immunization programs (NIPs) for infants, waning immunity raised the need to evaluate an adolescent booster dose. Phase 3 Study 220030 compares the immune response to a single booster dose in individuals aged 10–20 years who were vaccinated in infancy (primed group) versus the first of two doses in those never vaccinated against MenB (naïve group).

In that study, no comparisons with the full two-dose schedule are possible because no blood samples are collected after the second dose in naïve participants. Also, primed participants only have a single blood draw at one month after booster vaccination and persistence of immune response cannot be directly measured.

This collaborative work between RWE & HO, Clinical Statistics and Clinical group is intended to inform national immunization technical advisory group (NITAG) recommendations addressing the limitations mentioned above and stakeholders' concerns.

The first objective is to compare the immunogenicity of a booster dose in primed adolescents with the currently used two-dose regimen in naïve adolescents, using two completed studies as external comparators, where blood samples were obtained at one month after the second vaccination. At this scope, propensity-score based methods will be leveraged to deal with non-randomized groups. Establishing comparability would support broader NIP adoption, addressing concerns about schedule simplification and program feasibility.

The second objective is to model and predict the persistence of antigen-specific immune response after the booster dose in primed adolescents, integrating previously estimated titer-evolutions models from naïve adolescents and observed one-month post-booster data. Combining this model with real-world MenB typing data will generate long-term protection curves against actual MenB disease. The resulting immune-persistence estimates would support recommendations on optimal booster timing and assess overall public health impact.

Evaluation of Artificial Intelligence-Generated Synthetic Data for Clinical Research in Secondary Cardiovascular Prevention of Patients with Dyslipidemia.

Bonomi Alice, Centro Cardiologico Monzino IRCCS, Milan

Access to real-world clinical data is essential for advancing medical research and developing statistical and Artificial Intelligence models, but it is often constrained by privacy requirements and operational barriers. This study evaluates the use of synthetic data in secondary cardiovascular prevention among patients with dyslipidemia, using two datasets from Centro Cardiologico Monzino. A generative framework based on Large Language Models was trained on the original clinical data to produce synthetic patient records. Dataset quality was assessed through measures of fidelity, clinical utility, and privacy, and was benchmarked against traditional anonymization techniques. The results showed a more favorable balance between utility and confidentiality, with good preservation of the main distributions and clinical associations observed in the real data, including predictors of statin-associated muscle symptoms and LDL cholesterol reduction. These findings were further supported by internal validation methods, including cross-validation and bootstrap analyses.

Evaluating historical benchmark Targeted Single-Arm Trials with External Controls single-arm platform trial in glioblastoma

Ajsi Kanapari, University of Padua

Single-arm oncology trials depend on external benchmark reflecting the survival enrolled patients would have experienced under standard of care. In our motivating recurrent glioblastoma platform trial, eligibility depends on confirmed drug brain penetrance, a selection factor unobserved in historical sources.

We compared full external datasets with those reconstructed by propensity and prognostic score matching across scenarios. Benchmark error was decomposed into registry-selection, matching, trial-

sampling, estimation components and propagated to the operating characteristics. Under prognostic enrichment the naïve registry was too pessimistic, inflating apparent effect and type I error. Matching reduced but did not eliminate benchmark error, the residual was dominated by matching error and depended on the correlation between unobserved factor and observed covariates. Prognostic score matching performed better with larger registries, propensity score with limited data. Matching is a diagnostic not a universal solution: it quantifies residual benchmark error and its effect on go/no-go decisions. Benchmark validity should be assessed at design stage.

Targeting Clinically Relevant Causal Estimands in Trials with Latent Exposure-Induced Intermediate Confounding through Structural Equation Modeling and Propensity Score Weighting

Domenico Iervolino, Biostatistics and Clinical Trial Methodology Unit, CRC DEMeTra, University of Naples Federico II

Evaluating therapeutic efficacy in heart failure with reduced ejection fraction is challenging because improvements in outcomes such as left ventricular ejection fraction may be influenced by complex mechanistic pathways. Congestion may act as an exposure-induced intermediate confounder of the relationship between rescue therapy and the primary outcome, requiring clear causal estimands aligned with ICH E9(R1).

This study evaluates the performance of Structural Equation Modeling (SEM) in estimating the Controlled Direct Effect in the presence of latent intermediate confounding and measurement error, using a Marginal Structural Model with inverse probability weighting as a benchmark.

A simulation study assessed estimator performance across sample sizes from 100 to 1000 and different target-effect scenarios. The comparative framework contrasted SEM, which incorporates latent variables and measurement error for congestion indicators, against marginal structural modeling, with performance evaluated using bias, RMSE, statistical power and Type I error.

SEM demonstrated lower bias than weighting-based approaches as intermediate confounding increased, supported stable estimation under practical positivity challenges, and improved statistical power as sample size increased. The results suggest that SEM can provide a robust framework for estimating clinically meaningful causal effects in complex trials with intercurrent events.

When combined with a well-specified causal graph and explicit identification assumptions, SEM is a powerful tool for pre-specified causal estimands rather than merely an exploratory path-analysis method. Treatment–mediator interaction will be addressed in future studies.

Robust (Exchangeability) Bayesian Methods for Vaccines Targeting Multiple Serotypes

Margherita Annaratone, GSK

Multivalent and combination vaccines have been extensively used in recent decades, including pneumococcal, human papillomavirus, rotavirus, influenza, meningococcal and poliomyelitis vaccines. They offer protection against multiple diseases or subtypes of bacteria or viruses and help simplify immunization schedules.

When these vaccines are extended by adding new components, regulatory agencies often require immunological studies demonstrating superiority for the new components and non-inferiority for the shared components. Sample sizes are usually large because they depend on the number of co-primary outcomes.

With application to a case study, this work considers robust exchangeability Bayesian methods and compares them with the frequentist benchmark to evaluate the potential increase in power for studies targeting multiple serotypes.

Design of Single-Arm Time-to-Event Studies Using Simulation: Median, 95% Confidence Intervals and Optimal Sample Size using Shiny application

Bianca Bracco, IQVIA

This Shiny application provides a flexible, simulation-based framework for designing single-arm time-to-event clinical trials. It estimates median survival precision, confidence intervals, survival probabilities at fixed time points and restricted mean survival time (RMST), explicitly accounting for accrual, follow-up and dropout dynamics. It also supports precision-driven and optimal sample size determination under both monthly and annual dropout assumptions.

In particular, these functionalities are not available for single-arm designs in commonly used software such as PASS and East/East Horizon

---

Event Prediction in Clinical Trials

Bastiaan Jansen, SGS

Accurate prediction of event accrual is central to the planning and execution of analyses in clinical trials. A broad range of statistical techniques is available for predicting the number and timing of events (naïve protocol-based estimates, simple extrapolation and more sophisticated modelling and simulation approaches).

In practice, the performance of event prediction techniques can be substantially improved through careful monitoring and incorporation of additional expert input. Mechanistic features of a trial—such as changes in patient mix, evolving standards of care, competing risks, or deviations event patterns—can have a profound impact on event rates and, consequently, on the accuracy of predictions.

We present two case studies with highly similar trial designs that nonetheless demonstrated markedly different prediction accuracy due to underlying mechanistic differences. These examples illustrate how reliance on a single method or static assumptions can be misleading, and how contrasting multiple approaches can yield valuable insights.

Behind the Scenes of Real-World External Comparator Studies: Operational Issues and Lessons from the Field

Author: Alessandra Mignani (IQVIA)

Studies using an external comparator involve methodological and operational challenges, as they require alignment between separate studies that may be designed and conducted by different teams.

Alignment should be considered from study planning through data collection and analysis to support the validity, comparability, and interpretability of results.

Early definition of key design features, shared standards, harmonized derivation rules, and clear documentation within the Study Protocol and Statistical Analysis Plan are essential to enable meaningful cross-study comparisons. In settings where the external comparator is derived from an ongoing retrospective observational study, regular assessment of data completeness and comparability may further support the identification of variables suitable for comparative analyses. Drawing on practical experience and lessons learned, this contribution highlights approaches to improve alignment, data comparability, and operational efficiency in EC studies.

RCTs vs Real-World Studies: Operational Insights from Academic Experience

Author: Lorena Torroni (*Saint Camillus International University of Health and Medical Sciences)*

Randomized controlled trials (RCTs) and real-world evidence (RWE) provide complementary insights but differ in operational demands, data characteristics, and susceptibility to bias. While RCTs offer greater control and standardization, RWE provides broader insights into everyday clinical practice but faces greater heterogeneity and potential bias.

Drawing on academic experience across multicenter cohorts, environmental epidemiology, surgical registries, and outcome-development studies, this contribution explores how methodological rigor can be maintained despite limited funding, heterogeneous infrastructures, and non-dedicated personnel. Examples include data harmonization, modeling strategies to address exposure misclassification, propensity score methods to mitigate confounding by indication, and Delphi-based validation to strengthen outcome quality. These cases illustrate how appropriate biostatistical methods can address operational challenges and strengthen the credibility of real-world data.

The presentation provides a practical comparison of RCTs and RWE, emphasizing methodological choices, operational feasibility, and the role of biostatistics in bridging controlled research and everyday clinical practice.

The Importance and Pitfalls of Surrogate Endpoints

Tom Cattaert, SGS

Surrogate endpoints are biomarkers or intermediate used in clinical trials to predict clinical benefit. They can be chosen for different reasons (e.g. speed up the clinical trial process) or out of pure necessity (e.g. ethical reasons). While essential for accelerating access to therapies for severe diseases, they possess significant pitfalls that can be overlooked. We present a few real-life examples illustrating where things may go wrong but also offer some design and statistical methodology to reduce risk and evaluate the effectiveness of the surrogate endpoint.
