MBSW 2026
BOOK OF ABSTRACTS

RENAISSANCE INDIANAPOLIS NORTH HOTEL, CARMEL, IN

WEDNESDAY MORNING 8:30AM- 12:00PM MAY 27 – SHORT COURSES

Practical Large Language Models (LLMs): Foundations and Applications
Xing He, Haining Wang, Jiang Bian

This introductory, practice-oriented course provides participants with a clear understanding of how modern large language models (LLMs) work and how to apply them effectively in real-world contexts. The course begins with an accessible overview of LLM fundamentals, including how models are trained, how they process information, and their current capabilities and limitations. Building on this foundation, participants will explore practical techniques such as retrieval-augmented generation (RAG), structured outputs using JSON schemas, and the use of tools and agentic workflows. Realistic examples will be drawn from pharmaceutical and biomedical domains, such as answering questions from scientific documents with citations. By the end of the course, participants will have the skills to build simple and interpretable LLM-based applications.

Learning Objectives

- Understand the core principles of modern LLMs and recent advances such as long-context reasoning and agent frameworks.
- Learn how to build reliable RAG pipelines and structured-output workflows.
- Explore practical applications relevant to pharmaceutical research and development.
- Gain hands-on experience through guided exercises.

Target Audience

Professionals in the pharmaceutical and biotechnology industries, including R&D, clinical, pharmacovigilance, and data/IT functions. Basic programming background is required.

Format

3–4 hours, combining lectures, live demonstrations, and guided mini-labs.

Brief Course Outline

1. Introduction and Motivation

- What are LLMs?
    - Milestones and taxonomy
- Why do LLMs matter for pharmaceutical and biotechnology research?

2. LLM Fundamentals

- Transformer basics
  - What is a Transformer?
  - Architectural variants of Transformers
    - encoder-only (BERT)
    - encoder-decoder (T5)
    - decoder-only (GPT)
  - The core component: Attention mechanism
    - A quick recap of "Attention is All You Need"
  - Tokenization (subword tokenization and positional encoding)
  - "Which architecture fits my research scenario better: BERT, T5, or GPT?"
- How GPT models are trained and used
  - Pre-training via next-token prediction (causal/autoregressive language modeling)
  - Post-training:
    - Supervised finetuning and instruction tuning,
    - Reinforcement finetuning (Reinforcement Learning from Human Feedback, RLHF)
  - Generation (sampling and temperature parameter)
  - "For my research scenario, should I":
    - "Train a GPT from scratch"
    - "Fine-tune one, or"
    - "Use as-is"
- Current capabilities and limitations of state-of-the-art LLMs
  - Capabilities (long-context model, complex instruction following, agent and tool use)
  - Limitations (e.g., true creativity, planning, and hallucination)

3. Building Reliable Workflows (Key components)

- Prompt engineering
  - System instructions and user prompts
  - Context management
    - Exemplars (e.g., "few-shot")
    - Role playing
    - Reasoning (e.g., chain-of-thought)
- RAG workflow
  - Index construction
  - Common retrieval methods
- Structured outputs
  - JSON Schema
  - Validation and hinted retry

4. Hands-on Practice

- In the practical session, participants will work through a **simplified end-to-end retrieval-augmented question answering workflow** using **Google Colab** and the open-source **PaperQA2** library. They will upload or access a small set of biomedical or pharmaceutical PDFs, build a lightweight document index, and issue natural language queries. Through this exercise, they will see how PaperQA2 performs evidence retrieval, summarization, and generates structured answers with citations. Participants will then inspect and discuss the outputs to understand how retrieval and structured generation work together to produce interpretable results. No local setup or paid tools are required.

5. Instructor

- **Jiang Bian**, PhD, Walther and Regenstrief Endowed Chair in Cancer Informatics, Chief Data Scientist, Indiana University Health, Chief Data Scientist, Regenstrief Institute, Associate Dean for Data Science, School of Medicine, Indiana University, Chief Research Information Officer, IU Melvin and Bren Simon Comprehensive Cancer Center, Regenstrief Institute Deputy Director, Indiana Clinical and

Translational Sciences Institute (CTSI), Professor and Vice Chair for Translational Informatics Biostatistics and Health Data Science, School of Medicine, Indiana University

- **Xing He**, Ph.D., Assistant Professor, Department of Biostatistics and Health Data Science at the Indiana University School of Medicine, Faculty Lead for Strategic Data Science Infrastructure at the Regenstrief Institute.
- **Haining Wang**, Ph.D. Haining Wang, Ph.D., is a Postdoctoral Fellow, Department of Biostatistics and Health Data Science at the Indiana University School of Medicine.

RStudio & Positron Hands-on Workshop: Advancements in AI for Statistical Programming
Phil Bowsher, POSIT

Posit/RStudio will be providing a hands-on session on using AI to enhance open-source statistical programming. This session will discuss opportunities and applications for AI to empower statistical programmers and data scientists. This talk will explore and discuss GenAI to support programmers in the process of writing code and provide a hands-on opportunity to test it out.

This is a gentle introduction to the ecosystem for AI, discussing different models and integrations with RStudio and Positron. This session will also highlight open-source packages as well as working with AI and Shiny. This session is great for users coming from Excel or SAS and looking to understand the various AI models for programming, specifically for data analysis and programming. Topics discussed include Positron, Claude Code, RStudio, Ellmer, MCP and more.

AI innovation for programming is changing quickly. This workshop will explore various areas of AI for statistical programming, including IDEs, packages, Shiny, and AI orchestration/architecture. This session will have something for everyone—whether you're a statistical programmer looking to learn R, a seasoned R programmer looking to learn AI, or an AI sage looking to create and drive agents. We will explore the new IDE, Positron, while also highlighting tricks for RStudio users. Finally, we will cover how to integrate AI with Shiny in both R and Python. Posit/RStudio will be presenting an overview of AI in Positron and RStudio for the open-source user community at PharmaSUG. This session will provide an opportunity to use AI in Positron and see how it can empower statistical programmers using R and the Pharmaverse. This workshop will explore and discuss areas such as ADaM data and TLG creation, while providing examples for getting started in your own environment after the session. You will gain hands-on insights using R for clinical workflows and discover how these emerging technologies can enhance your approach to pharmaceutical data science. This deep dive ensures you remain at the forefront of innovation within the rapidly shifting landscape of open-source pharmaceutical AI.

Clinical: Adaptive Bayesian Methods for Dose Finding and Evidence Borrowing

Longitudinal self-adaptive borrowing
Saijun Zhao, Indiana University

Borrowing information from historical controls can improve the efficiency of randomized controlled trials (RCTs), but its application to longitudinal outcomes with informative dropout remains limited. We propose two Bayesian mixture priors for longitudinal data information borrowing: the mixture prior for longitudinal data borrowing (MLB) and its self-adapting extension (SLB). Both approaches use a shared-parameter model to handle the informative dropout and apply a mixture prior framework to incorporate historical control data while accounting for possible prior-data conflict. Simulation studies show that the proposed priors yield desirable operating characteristics enabling efficient and rigorous information borrowing. In particular, the SLB prior demonstrates the best overall performance. The R code to implement the proposed priors is also provided to facilitate the practical applications.

Combination therapy dose finding
Jeremiah Jones, Lilly

Borrowing placebo data across studies can reduce reliance on large control arms, improve efficiency, and support better trial decisions, particularly in master protocols where standardized trial elements across sub-studies enhance comparability. However, existing dynamic borrowing methods rely mainly on outcome-based similarity at a single primary time point and often ignore baseline covariates, longitudinal information, and missingness, all of which can confound placebo response. We propose a causally inspired doubly robust self-adapting mixture (DR-SAM) prior that extends the SAM framework by integrating outcome regression, covariate adjustment, and inverse probability weighting to address measured confounding and intercurrent events while leveraging repeated longitudinal data. DR-SAM adaptively reduces borrowing when residual discrepancies suggest unmeasured confounding, thereby improving robustness in heterogeneous settings. Simulation studies comparing DR-SAM with Adaptive Targeted Maximum Likelihood Estimation (A-TMLE) show that DR-SAM maintains robustness under covariate imbalance, dropout, and non-ignorable missingness, while standard borrowing approaches fail.

Adaptive Bayesian methods and recent topic on evidence borrowing
Zhen Zhang, Otsuka Pharmaceutical

TBD

Integrated Framework for Rapid Oncology Clinical Trial Evidence Synthesis
Bruno Larvol, Judith Pérez-Granado, Mark Gramling,. The Larvol Group LLC, San Francisco, CA.

Introduction The increasing volume and complexity of oncology clinical trial data pose substantial challenges for statistical evaluation, particularly in the identification, digitization, harmonization, and preparation of outcomes data prior to analysis. These upstream steps remain time-consuming, error-prone, and difficult to reproduce, often limiting the feasibility of exploratory and comparative analyses. Methods LARVOL CLIN is an integrated analytical framework built around a large-scale, oncology outcomes database structured using a common data model. The database integrates highly curated clinical trial metadata with pre-processed time-to-event and non-time-to-event outcomes, enabling direct application of statistical methods without manual data reconstruction. The current dataset spans approximately +104k trials, +100k digitized Kaplan-Meier (KM) curves, 3.5M pseudo-individual patient data (IPD), and +17K hazard ratios (HRs). This structure enables survival pooled analysis, network meta-analysis (NMA) and other statistical methods while preserving traceability to the original trial sources. Results As a case study, first-line (1L) EGFR-mutated metastatic non-small cell lung cancer (mNSCLC) was examined using CLIN framework. Reconstructed pseudo-IPD from phase 3 trials, obtained from digitized KM curves and their corresponding at-risk tables, enabled cross-trial survival comparisons. Pooled KM overlays comparing first-generation EGFR tyrosine kinase inhibitors (TKIs) (gefitinib, erlotinib, icotinib) with third-generation TKIs (including osimertinib, lazertinib, furmonertinib, aumolertinib and befotertinib) demonstrated clear separation in progression-free survival (PFS). Pooled survival estimates derived from pseudo-IPD yielded an HR of 0.53 (95% CI 0.48–0.58; p<0.001), favouring third-generation EGFR TKIs. Pairwise meta-analysis using a random-effects DerSimonian–Laird model across trials evaluating third-generation versus first-generation TKIs confirmed consistent treatment effects, with a pooled HR of 0.47 (95% CI 0.42–0.52) for PFS. The reconstructed dataset also supported exploratory NMA across multiple EGFR inhibitor regimens and combination strategies. The resulting treatment network connected first-generation TKIs with multiple targeted combinations and later-generation inhibitors. Heterogeneity across comparisons was moderate (I^2 = 44.5%) and global inconsistency testing did not indicate significant disagreement between direct and indirect evidence (p = 0.166). Ranking analysis indicated higher relative PFS benefit for combination strategies such as osimertinib plus chemotherapy and lazertinib plus amivantamab, with third-generation TKIs generally outperforming first-generation inhibitors Conclusions This work highlights how curated structured oncology data can reduce analytical burden and support iterative, exploratory biostatistical workflows. By lowering barriers to survival reconstruction, meta-analysis, and NMA, such a framework may broaden the practical use of established statistical methods in early-stage evidence synthesis and hypothesis generation.

BIOMARKERS/PRE-CLINICAL/DISCOVERY: Analytical frameworks for assessing digital biomarkers in biomedical research

Neural network-based approaches to modeling subject-specific activity distributions from wearable devices
Caihong Qin*, Lan Xue, Roger Zoh, Carmen Tekwe, Indiana University

Dense repeated measurements from wearable devices and continuous monitoring technologies are increasingly used in biomedical and biopharmaceutical studies. These data are often summarized by scalar metrics, such as mean levels, total activity, or time spent above clinical thresholds. While such summaries are interpretable, they may miss important within-subject distributional patterns, including variability, skewness, multimodality, and tail behavior. This work develops a flexible framework for estimating subject-specific density functions from repeated measurements, with extensions to settings where the observed measurements are affected by measurement error. We illustrate the approach using NHANES wearable physical activity data. In the applications, subject-specific density representations are used as features for downstream prediction of health-related outcomes. The results suggest that distributional summaries provide information beyond standard covariates and scalar mean summaries. Simulation studies further support the utility of the framework for recovering latent subject-specific densities under both error-free and measurement-error-contaminated settings.

Improved Phenotyping Between Narcolepsy Type 1, Type 2 and Idiopathic Hypersomnia Through Classification and Clustering Methods Applied to Polysomnography
Marta Karas, Takeda Health

Authors: Marta Karas[1], Marco Vilela[1*], Brian Tracey[1], Francesco Onorati[1], Lucie Barateau[2,3,4], Alice Cai[1], Melissa Naylor[1], Derek L. Buhl[1*], Dmitri Volfson[1], Yves Dauvilliers[2,3,4]

Affiliations:

1. Takeda Development Center Americas, Inc., Cambridge, Massachusetts, USA.

2. Sleep-Wake Disorders Center, Department of Neurology, Gui-de-Chauliac Hospital, CHU Montpellier, France.

3. National Reference Network for Narcolepsy, Montpellier, France.

4. Institute for Neurosciences of Montpellier (INM), INSERM, University of Montpellier, France.

*At the time the study was conducted.

Presenting Author: Marta Karas
Presenting Author email: marta.karas@takeda.com

Introduction: Differentiation between central disorders of hypersomnolence can be challenging due to overlapping clinical features, especially between narcolepsy type 1 (NT1) and type 2 (NT2), and between NT2 and idiopathic hypersomnia (IH). Recent debates even questioned whether NT2 is a distinct disease entity. This work seeks to better characterize disease phenotypes by studying nocturnal polysomnography (PSG) characteristics.

Methods: First, we developed an automated classification method to improve differentiation between patient groups. A rich set of sleep features, including quantitative electroencephalography (qEEG), was derived from routine PSG of 358 drug-free adults, with 114 NT1, 90 NT2, 105 IH and 49 clinical controls (CC) diagnosed according to ICSD-3 criteria at a French Reference Center for Rare Hypersomnias. Features were computed by whole night and by quarter-night. Classifier performance was evaluated on resampled training and test sets. Next, we used cluster analysis to quantify how sleep features varied within and across diagnostic groups. Unsupervised k-means clustering was performed within each diagnosis to identify subject sub-groups (clusters); the number of clusters per diagnosis (between 2-10) was selected by minimizing the Calinski-Harabasz index. We then tested (Kruskal-Wallis) for between-cluster differences and compared sleep characteristics across clusters.

Results: Random forest classifiers performed best, achieving AUCs of 0.78, 0.85, and 0.82 in discriminating NT2 versus IH, NT2 versus CC, and IH versus CC, respectively. Adding qEEG and quarter-

night features improved NT2 versus IH classification (AUC was 0.67 for whole-night hypnogram-derived features, versus 0.78 for all features). Clustering revealed two NT1 subgroups, with more extreme (shorter REM sleep onset, higher wakefulness) and less extreme phenotypes. Two main IH clusters were found, with similar hypnogram-derived features but some qEEG differences. Two NT2 subgroups emerged: one more similar to the NT1 clusters, another to the IH clusters.

Conclusion: Classification results show the value of quarter-night features, with qEEG being especially helpful in NT2 versus IH classification. Clustering results provided insights into potential heterogeneity within each diagnosis and suggest the existence of one NT2 subgroup similar to NT1 and another similar to IH. These results may help improve patient diagnosis and identify novel biomarkers for clinical trials.

Support: Funded by Takeda

Actigraphy-Based Assessment of Daily Rest-Activity Rhythms
Kaysen Lu, Merck

The advancements in wearable technology have significantly enhanced personalized monitoring of various physiological and environmental signals. Rest-activity rhythms (RARs) are internally regulated to align with external light-dark cycles, and disruptions, such as excessive light exposure at night, have been linked to cardiometabolic and cognitive outcomes. We aim to quantify rest–activity rhythms (RARs) and light exposure and evaluate how their misalignment is associated with diabetes risk.

Our motivating dataset is the National Health And Nutrition Examination Survey (NHANES) 2011-2014, where participants wore the ActiGraph devices for seven days, 24 hours a day. The device sensors collected actigraphy data (i.e., activity movements) and ambient light exposure at 1-min epochs. We applied a harmonic hidden Markov model (HHMM) to characterize intensity levels (low, moderate, high) and rhythmic patterns in both activity and light exposure. Alignment in intensity and phase between activity and light exposure was quantified using Dynamic Time Warping (DTW).

Disrupted RARs were found to be associated with increased odds of diabetes. This disruption is characterized by reduced probability of low activity during the day and increased probability of high activity at night. Greater misalignment in intensity and phase between activity and light exposure was also associated with higher odds of diabetes. These findings suggest that disrupted RARs and misalignment between activity and light exposure may adversely impact diabetes risk.

A Functional Accelerated Failure-Time Model with Varying Effects Correcting for Measurement Error: An Application to REGARDS Data
Joseph Yang[1], Roger S. Zoh[1], Lan Xue[2], Carmen D. Tekwe[1]

[1] Department of Epidemiology and Biostatistics, Indiana University, Bloomington

[2] Department of Statistics, Oregon State University, Corvallis

Physical activity patterns are often recorded as high-dimensional functional data, motivating studies of its association with time-to-event health outcomes. Existing scalar-on-function survival models, such as Cox or accelerated failure-time (AFT) models, typically assume a common functional effect across individuals. However, differences in demographic factors can lead to heterogeneity in the functional effect of physical activity, which may be further obscured by measurement error in the functional covariates arising from inconsistencies in device calibration or wear compliance. We develop a Bayesian approach for estimating a functional varying-coefficient in a AFT model while accounting for measurement error. The varying functional effect is modeled with basis functions over time with coefficients that vary smoothly through a

single-index projection of scalar covariates. In the absence of replicate measures, we assume an instrumental variable to estimate the latent functional covariate. An application to the Reasons for Geographic and Racial Differences in Stroke study reveals subgroup differences in the relationship between step count activity and time to ischemic stroke.

REAL WORLD EVIDENCE: RWE Strategies for Clinical and Regulatory Success

Statistical Methods and AI Enabled Analytics in Real World Evidence Supporting FDA Drug Safety Labeling Changes
Miles Xi, AbbVie

Patients in clinical studies often exhibit heterogeneous treatment effect (HTE). Classical subgroup analyses provide inferential tools to test for effect modification, while modern machine learning methods estimate the Conditional Average Treatment Effect (CATE) to enable individual level prediction. Each paradigm has limitations: inference focused approaches may sacrifice predictive utility, and prediction focused approaches often lack statistical guarantees. We present a hybrid two-stage workflow that integrates these perspectives. Stage 1 applies statistical inference to test whether credible treatment effect heterogeneity exists with the protection against spurious findings. Stage 2 translates heterogeneity evidence into individualized treatment policies, evaluated by cross fitted doubly robust (DR) metrics with Neyman-Pearson (NP) constraints on harm. We illustrate the workflow with working examples based on simulated data and a real ACTG 175 HIV trial. This tutorial provides practical implementation checklists and discusses links to sponsor oriented HTE workflows, offering a transparent and auditable pathway from heterogeneity assessment to individualized treatment policies.

Advances on time-to-event prediction using machine learning methods
Lingsong Zhang, Purdue University

TBD

Digitally Interrupted: A Difference-in-Differences Analysis of a Hospital Cyberattack and Its Impact on Operational Performance
Xiaochun Li, Indiana University

A hospital cyberattack on August 4, 2021 raised concerns about disruption to clinical operations, prompting an evaluation of its actual impact. This study uses real-world evidence to quantify the impact of that event on key hospital performance metrics. Data were assembled for July through October in 2020 and 2021 for both the affected hospital (“index hospital”) and all other hospitals in Marion County. For analysis, July and August were defined as the pre-attack period and September and October as the post-attack period; August was included in the pre-period because operational disruption was not immediate and required time to manifest. Using a semi-experimental design, we used difference-in-differences analysis to evaluate changes in emergency department (ED) availability, ED boarding counts, and overall hospital capacity within the index hospital between

2021 and 2020 where the latter served as the control year, as well as between the index hospital and other hospitals unaffected by the attack in year 2021. The cyberattack was associated with significant deterioration across multiple operational metrics. These results demonstrate how real-world data can be leveraged to assess the downstream impact of cyberattack on clinical operations and highlight the importance of resilient clinical systems and proactive cybersecurity preparedness.

Pushing the Boundaries: Innovations in Indirect Treatment Comparisons and Health Technology Assessment
Haitao Chu, Pfizer

Health technology assessment (HTA) is increasingly pushing beyond traditional evidence paradigms, with indirect treatment comparisons (ITCs) playing a critical role when head-to-head trials are unavailable. The European Union's Joint Clinical Assessment (JCA) has further elevated the prominence of ITCs, intensifying the need for statisticians to address emerging methodological and evidentiary challenges at the frontier of decision-making. Among contemporary approaches, population-adjusted indirect comparisons (PAICs) have become central to HTA submissions, with matching-adjusted indirect comparison (MAIC) being the most widely applied. MAIC leverages individual participant data (IPD) to reweight trial populations so that their covariate distributions align with aggregate data (AgD) from comparator studies, enabling comparisons within a specified target population. However, this apparent flexibility reveals a key limitation—the "MAIC paradox." When effect modifiers differ in both distribution and strength across treatments, MAIC can yield conflicting conclusions depending on whether analyses are anchored to IPD or AgD populations. This paradox underscores the importance of clearly defining target estimands and advancing transparency, including broader access to de-identified IPD by HTA agencies. Looking beyond MAIC, this talk examines three foundational statistical principles that are essential for pushing the boundaries of valid ITCs: transportability, which governs the extrapolation of evidence across populations; collapsibility, which shapes the interpretability of effect measures; and transitivity, which ensures coherence between direct and indirect evidence. Together, these principles offer a unifying framework for developing innovative, robust, and policy-relevant evidence to support HTA in an increasingly complex and evolving landscape.

Robust Causal Inference for Skewed and Outlier-Prone Data: A Doubly Robust Rank-Based Approach
Ruohui (Matt) Chen, Northwestern University

TBD

STAT PROGRAMMING/AI: Practical Use of Open-Source Tools and Infusing Statistical Programming and Visual Analytics Using Large Language Models

FDA Experience with R Pilots
Hye Soo Cho, FDA

Open-source software, particularly the R language, has gained significant traction in the pharmaceutical industry for clinical trial data analysis and reporting. In collaboration with the R Consortium's R Submission Working Group, the FDA has been actively engaged in a series of R Submission Pilot Projects designed to evaluate the feasibility of R-based regulatory submissions. These pilots have progressively expanded in scope: Pilot 1 tested whether R could support a standard clinical trial submission with tables and figures; Pilot 2 explored the integration of an R Shiny interactive application into a submission package; and Pilot 3 evaluated R's ability to produce ADaM datasets from SDTM data and generate Tables, Listings, and Figures (TLFs) for regulatory submission. In the most recently completed Pilot 4, WebAssembly and Docker containers were tested as methods to bundle the Pilot 2 Shiny application into a fully self-contained submission package. Pilot 5, currently under FDA review, focuses on delivering an R-based submission using SDTM and ADaM datasets in Dataset-JSON format — a modern alternative to the traditional .xpt format. This session will share key findings and feedback from the FDA's review of these pilots, highlight practical challenges encountered, and discuss what the FDA prioritizes when evaluating R-based submissions — offering valuable insights for sponsors and statisticians considering R for future regulatory submissions.

Beyond the Keyboard: Voice-Driven Analysis for Clinical Data Science
Phil Bowsher, Posit

This presentation explores a new frontier: using AI and voice modality to accelerate clinical data science analysis with open-source tools. We demonstrate how a data scientist can use a conversational interface to ask questions and receiving real-time insights based on natural speech. The session explores the underlying architecture—including tools like ggbot2, Shiny and OpenAI's Realtime API—that make these voice-driven interactions possible for clinical analysis. We will showcase practical use cases, such as exploratory data analysis (EDA) with clinical data. Attendees will gain insight into how this approach empowers statistical programmers and biostatisticians to engage with open-source data through a new conversational partner.

{pmrm}: an R package for fitting progression models for repeated measures
Will Landau, Lars Lau Raket, Kasper Kristensen. Lilly

Progression models for repeated measures (PMRMs) are continuous-time nonlinear mixed-effects models for longitudinal data in progressive diseases. Unlike mixed models for repeated measures (MMRMs), which estimate treatment effects as linear combinations of additive effects on the outcome scale, PMRMs characterize treatment effects in terms of the underlying disease trajectory. This framing yields clinically interpretable quantities such as average time saved or percent reduction in decline due to treatment. {pmrm} is an open-source R package for fitting frequentist PMRMs. Under the hood, {pmrm} leverages {RTMB}, a high-performance user-friendly R framework for fitting complex statistical models, achieving orders-of-magnitude speedups over equivalent implementations with nlme::gnls(). {pmrm} provides first-class functionality for

simulation, post-processing, visualization, marginal mean estimation, and S3 methods for standard generics, making PMRMs accessible to clinical statisticians and analysts.

CVARS - Transforming Drug Safety Analytics and Deliverables for Enhanced and Comprehensive E2E Review, Analysis, Signal Detection, Reporting, and Regulatory Submissions with the Aid of AI and LLMs
Michael Schuhler, AbbVie

TBD

CMC: Process Development & Assay Stability

Monitoring Assays: Control Samples, Reference Stability, Assay Stability, Bridging References
Steven Walfish, Iovance Biotherapeutics

After assay validation, a control strategy should be implemented to ensure that the assay and control samples are behaving consistently with the validated state. Typically trending analysis with statistical limits are utilized at this stage. This talk will present other statistical methods in addition to traditional control charting for assay monitoring. Establishing a baseline with inherent assay variability can be utilized to set acceptance criteria for assay monitoring.

Combining Bayesian Optimization and DOE to Define Design Space
Andrea Biasi Coombs, JMP

Bayesian optimization is an iterative learning technique that starts with existing data and uses a model to generate new candidate runs for assessment. Bayesian optimization learns from the responses with each iteration and gives a clearer guidance than traditional approaches about when we can stop experimenting. The iterative nature has the potential to dramatically reduce the time and resources required for process and analytical development. However, traditional DOE models are still necessary for QBD approaches such as setting ranges on critical process parameters. Visual, point-and-click tools in JMP software allow for a hybrid approach to better understand Design Space using Bayesian Optimization and DOE.

A Bayesian based Stability Platform Framework for Oligonucleotides
Meng John Zhao, Lilly

Conventional ICHQ1 based stability studies are often time-consuming and resource-intensive, especially during early development stages when data are limited. We propose a Bayesian modeling framework for oligonucleotide stability that enables data-driven prediction of degradation. This bottom-up approach integrates stability data from early drug substance and drug product studies that are considered very similar to, but not identical to the compound in question. In our approach, oligonucleotides are categorized according to key molecular attributes, such as

sequence composition and linkage chemistry, allowing compounds with similar degradation behavior to be grouped into stability platforms supported by model-based kinetic estimates.

The Bayesian framework naturally incorporates data from both accelerated stability assessment programs (ASAP) Accelerated Predictive Stability (Arrhenius based) and long-term stability studies, facilitating hierarchical borrowing of information across related molecules and storage conditions. This prior knowledge-platform-based strategy supports the establishment of predictive stability profiles in early development, even when empirical data are sparse. By providing probabilistic retest period and shelf-life estimates for drug substance and drug product respectively with explicitly quantified uncertainty, the approach enables risk-based decision-making for clinical supply planning.

Overall, the proposed platform stability framework offers a systematic and scientifically rigorous method for predicting oligonucleotide stability, supporting extended clinical dating while maintaining assurance of product quality.

CLINICAL: Applied Machine Learning for Clinical and Regulatory Decision-Making

Clinical endpoint adjudication by AI
Zhili Qiao, Lilly

Clinical Event Committee (CEC) adjudication is the gold standard in cardiovascular outcome trials but is labor-intensive and time-consuming. We evaluated three open-source AI models—Clinical Longformer, GPT-4o, and LLaMA 3.1—to adjudicate the MACE-3 endpoint (cardiovascular death, myocardial infarction, stroke) using 2,881 patient dossiers from the REWIND trial. LLaMA 3.1 achieved the strongest performance, with each of CV death, stroke, MI. Replicated Cox\/Kaplan–Meier analyses produced treatment effect estimates closely mirroring CEC-based results, supporting AI-enabled adjudication as a scalable complement to traditional workflows.

Comparative Effectiveness of Treatment Regimens in Moderate-to-Severe Atopic Dermatitis: Integrated Analyses Across Heterogeneous Study Designs using Causal Techniques
Prosenjit Kundu, Pfizer

In multi-arm clinical trials, naïve pairwise comparisons of proportions or means fail to provide causally interpretable treatment effect estimates in the presence of confounding and covariate imbalance. Using pooled data from multiple atopic dermatitis trials with outcomes assessed at weeks 2, 4, 8, and 12, we implemented a hierarchy of analytical approaches of increasing causal rigor. Generalized Linear Mixed Models (GLMM) for binary endpoints and Mixed Models for Repeated Measures (MMRM) for continuous endpoints were fitted with baseline characteristics as fixed effects, with random intercepts capturing arm-specific heterogeneity and within-subject correlation, and target estimands derived by marginalizing model-based predictions over the covariate distribution. To strengthen causal identification under the potential outcomes framework, G-computation was applied to GLMM\/MMRM predicted values to estimate population-average treatment effects. Inverse Probability of Treatment Weighting (IPTW) was employed using generalized boosting models where covariate balance was not achieved (standardized mean differences ≥ 0.2) — with stabilized weights and target-population-specific reweighting. Augmented IPTW (AIPTW) was further implemented, combining outcome modelling with propensity score estimation to yield a doubly robust estimator that remains consistent if either the outcome model or the propensity score model is correctly specified, while also achieving semiparametric efficiency. Targeted Maximum Likelihood Estimation (TMLE) was additionally performed, providing a doubly robust, semiparametric efficient estimator to check for improvement in finite-sample performance by iteratively updating the initial outcome model to satisfy efficient score equations. A semi-parametric GEE combined with G-computation served as a sensitivity analysis, relaxing normality assumptions and assessing robustness of causal estimates. Convergence of estimates across these approaches strengthens causal conclusions, while divergence exposes sensitivity to modelling assumptions, offering a principled and transparent template for causal analysis in complex multi-arm dermatology trials.

A general, flexible, and harmonious deep-learning-assisted framework to construct interpretable functions in regression analysis
Tianyu Zhan, AbbVie

An interpretable model or method has several appealing features, such as reliability to adversarial examples, transparency of decision-making, and communication facilitator. However, interpretability is a subjective concept, and even its definition can be diverse. The same model may be deemed as interpretable

by a study team, but regarded as a black-box algorithm by another squad. Simplicity, accuracy and generalizability are some additional important aspects of evaluating interpretability. In this work, we present a general, flexible and harmonious framework to construct interpretable functions in regression analysis with a focus on continuous outcomes. We formulate a functional skeleton in light of users' expectations of interpretability. A new measure based on Mallows's Cp-statistic is proposed for model selection to balance approximation, generalizability, and interpretability. We apply this approach to derive a sample size formula in adaptive clinical trial designs to demonstrate the general workflow, and to explain operating characteristics in a Bayesian Go/No-Go paradigm to show the potential advantages of using meaningful intermediate variables. Generalization to categorical outcomes is illustrated in an example of hypothesis testing based on Fisher's exact test. A real data analysis of NHANES (National Health and Nutrition Examination Survey) is conducted to investigate relationships between some important laboratory measurements. We also discuss some extensions of this method.

A Pragmatic Statistician's Case Reviews of AI Use in Drug Development
Jack Knorr, Lilly

Artificial intelligence (AI) are reshaping drug development across the full product lifecycle. The FDA's CDER discussion paper (2025) and the joint EMA-FDA guiding principles (2026) establish a regulatory foundation emphasizing transparency, data governance, risk-based validation, and lifecycle management across target identification, clinical trial design and execution, pharmacovigilance, and manufacturing.

AI applications such as endpoint adjudication through natural language processing (NLP) enable systematic mining of medical records to define outcomes of interest. In decentralized settings, continuous data capture through wearable and remote collection devices enables functional and patient-centric endpoint assessment with greater sensitivity than traditional approaches. Beyond trial design, AI is increasingly automating critical execution tasks — including tables, figures, and listings (TFLs), Clinical Study Reports (CSRs), and regulatory submission deliverables — compressing timelines and reducing manual error.

Drawing on the Veridical Data Science framework (Yu & Barter, 2024) — anchored in Predictability, Computability, and Stability (PCS) — alongside the FDA regulatory landscape, this presentation evaluates AI use cases across key workflow stages, assessing efficiency gains, validation rigor, interpretability, and regulatory alignment.

References:

FDA/CDER. Using Artificial Intelligence & Machine Learning in the Development of Drug & Biological Products: Discussion Paper. May 2023, Revised February 2025.

EMA/FDA. Guiding Principles of Good AI Practice in Drug Development. January 2026.

Yu B, Barter R. Veridical Data Science: The Practice of Responsible Data Analysis and Decision Making. MIT Press, 2024. (vdsbook.com)

REAL WORLD EVIDENCE: Digital twins, external control, AI

XYZ – A Unified Framework Modernizing Drug Development with Real-World Data and AI
Yong Chen, U of Pennsylvania

We introduce XYZ, a unified framework that advances clinical evidence generation by jointly modeling the treatment (X), outcome (Y), and population (Z) dimensions. This framework is built to support AI-driven drug repurposing, outcome profiling, and multi-dimensional optimization of eligibility criteria using real-world data. By integrating lossless federated target trial emulation for drug discovery, negative control–based debiasing for robust outcome inference, and AI-guided simulation to evaluate alternative inclusion criteria across real-world populations. XYZ enables principled and scalable evidence generation. Applications in repurposing GLP-1 RA, drug identification for AD\/ADRD, and trial design for advanced non-small cell lung cancer (NSCLC) illustrate how XYZ enhances generalizability, supports regulatory alignment, and ensures reliable insights generated from distributed research networks.

Accounting for unobserved confounders via negative control
Hongwei Wang, AbbVie

There has been increasing usage of real-world evidence in regulatory submission, reimbursement decision-making, scientific communication, and more efficient drug development. Related to its non-randomized nature and data collection constrain in routine clinical practice, unmeasured confounding is a major threat to the validity of real-world studies. Among the active research on addressing this challenge, negative control (defined as variables associated with the unmeasured confounders but not causally related to either the treatment or outcome variables of primary interest) have been widely used to detect residual confounding and more importantly account for them. Considering its importance, FDA PDUFA VII Commitment for 2023-27 specifically includes negative control methodology development and its application in the use of RWE. In this talk, we report the latest development in employing negative control to detect, empirically calibrate, and account for unmeasured confounders. Best practices in formulating study design and analytic strategy, identification of negative control are also presented.

Adaptive Bayesian Borrowing with Prospective Specification: Ensuring Congruent Integration of Historical Controls
Saurabh Mukhopadhyay, AbbVie

Integrating external data sources into clinical trials is increasingly important for addressing major challenges in drug development, especially concerning rare diseases, pediatric cases, and novel therapies. Bayesian methodologies are powerful for leveraging historical control data, but practical and statistical challenges arise when prospectively specifying designs to apply these methods. This presentation introduces a novel adaptive Bayesian borrowing (ABB) method with a prospectively specified strategy. The ABB method dynamically borrows information from historical controls only when current control data are sufficiently aligned with historical data, based on predefined criteria. If congruence criteria are not met, no borrowing occurs. This approach reduces risks of bias or misspecification from inconsistencies between historical and current controls. Additionally, ABB enables adaptive sample size increases to accommodate uncertainty in endpoint variability, strengthening study robustness.

From Averages to Individuals: Multi-task Learning for Heterogeneous Real-World Data:
Satrajit Roychoudhury, Pfizer

Real world data (RWD) are increasingly used to inform clinical and regulatory decisions, yet standard analytical approaches often focus on population average effects and struggle to accommodate patient level heterogeneity, irregular sampling, and missingness. Such limitations can obscure meaningful individual level variation in treatment response and disease trajectories. We propose a multi-task learning (MTL) framework that shifts inference from averages to individuals by jointly modeling multiple related prediction tasks (e.g., predict all-cause mortality and predict risk of hospitalization at 6 month) across heterogeneous real-world data sources. By sharing information across tasks while allowing tasks specific deviations, the proposed approach balances population-level learning with individualized modeling, effectively borrowing strength where appropriate while preserving heterogeneity. The framework naturally accommodates high-dimensional covariates, variable follow-up, and incomplete data common in RWD settings. The approach will be evaluated using empirical simulation and real-data example.

CMC: Medical Device

Navigating Ambiguity, Complexity, and Cross-Functional Needs
Aili Cheng, Kang Liu, Ke Wang, Pfizer

The increasing complexity of drug–device and combination products has amplified the need for rigorous statistical strategies across the product lifecycle. Statistics play a critical role in demonstrating the evidence used to support drug and combination product development. Yet In practice, the unique nature of R&D devices and combination products presents - challenges in implementing statistical analyses and establishing appropriate criteria during product development. This presentation highlights several common statistical challenges observed in cross functional CMC environments, illustrated through practical examples. It also encourages earlier, more deliberate, and collaborative interactions between statisticians and engineers. By building stronger partnerships and defining statistical responsibilities more clearly, teams can improve product understanding, regulatory preparation, and overall lifecycle reliability.

Enhancing Patient-Centered Care: Bridging Pharma and Medical Device Technologies Through Collaborative Statistical Practices
Yenny Webb Vargas, Genetech

As drug-delivery technologies evolve, the pharmaceutical and medical device sectors are merging to offer patients greater flexibility in their care. This convergence requires nonclinical statisticians and quality engineers to collaborate more deeply on combination products, navigating their distinct professional methodologies through shared statistical practices. This presentation introduces an industry workstream dedicated to bridging these two fields, fostering the cross-functional expertise necessary to advance patient-centered innovation.

Sampling Rationale Challenges in Combination Products
Craig Bernier, J&J

Sampling Rationale Challenges in Combination Products Craig Bernier, Johnson & Johnson Innovative Medicines There are some unique challenges with production sampling plans and rationale for combination products. The device component is typically tested by traditional acceptance sampling approaches and sampling plans need to consider the trade-offs between consumer risk and producer's risk. This talk reviews practical and statistical issues in designing acceptance-sampling plans. We will define

and apply AQL and RQL concepts and illustrate how the RQL\/AQL ratio frames acceptable risk tradeoffs. The session will contrast PPQ versus post-PPQ routine sampling objectives, and discuss the role of ISO 2859 inspection levels and rationale to support various levels. We will question the roles of switching rules or if skip-lot testing might be applicable to reduce sampling costs when quality is consistently good. Finally, we discuss the details that should be included in a statistical rationale for a sampling plan and add some thoughts on where the rationale might sit within the quality system.

Closing the Quality Loop: Integrating Statistics and Data Science Within Quality Assurance for Medical Device Lifecycle Improvement
AJ McKechnie, Cook Medical

Medical device organizations generate quality data across the full product lifecycle, including product development, manufacturing, complaints, audits, inspections, regulatory activities, field actions, and external surveillance sources. These data are often reviewed in separate workflows, which can limit an organization's ability to identify emerging risks early and systematically translate insights into improved product and process decisions. This poster describes an integrated framework in which statistics and data science are embedded within Quality Assurance to support a closed-loop model for continuous improvement from R&D through post market. In this model, statisticians provide non-clinical analytical support across the device lifecycle, including design optimization (DOE), verification and validation, test method assessment, process characterization, process improvement, and quantitative decision support for engineering and quality teams. In parallel, data scientists integrate internal and external quality data sources, enable signal detection and signal management, and develop tools that improve monitoring, prioritization, and visibility of quality risks. Together, these capabilities support a practical quality decision pathway of identify, monitor, assess, and act. Data science strengthens the organization's ability to connect disparate data streams and surface meaningful patterns, while statistical methods provide the rigor needed to investigate issues, quantify performance, optimize solutions, and evaluate effectiveness. Insights generated through this framework inform current product and process decisions while also feeding forward into future product requirements and next-generation development. By integrating statistics and data science within Quality Assurance, organizations can create a more proactive, connected, and evidence-based quality system. Expected benefits include improved regulatory readiness, enhanced customer satisfaction, greater operational efficiency, and continuous improvement of product quality and risk management across the medical device lifecycle.

Biomarker

Interpretable Deep Learning-Based Multi-biomarkers Integration Model for Response Prediction
Hong Wang, Sanofi

Combining multiple biomarker information and clinical information can offer a more complete and comprehensive information to predict treatment response and reveal the underlying disease mechanisms than using single biomarker or clinical data alone. We developed novel interpretable deep learning (DL)\/machine learning (ML)-based approaches for integrating multi-dimensional biomarker data to differentiate and predict treatment response. By leveraging the power of DL\/ML, we seek to uncover complex patterns and relationships within heterogeneous data types, including high dimensional genomic data and clinical pharmacodynamic biomarkers. Furthermore, by incorporating the DeepSHAP\/SHAP algorithm, which explains the predictions of DL\/ML models by attributing importance to each input feature, we enhance the ability to understand specific biomarker characteristics and patterns present in treatment responders. Our innovative approach has been successfully implemented to understand the underlying biological mechanism of action in real clinical trials.

A Hybrid Two-Stage Workflow Linking Biomarker Evidence to Individualized Treatment Policies
Miles Xi, AbbVie

Patients in clinical studies often exhibit heterogeneous treatment effect (HTE). Classical subgroup analyses provide inferential tools to test for effect modification, while modern machine learning methods estimate the Conditional Average Treatment Effect (CATE) to enable individual level prediction. Each paradigm has limitations: inference focused approaches may sacrifice predictive utility, and prediction focused approaches often lack statistical guarantees. We present a hybrid two-stage workflow that integrates these perspectives. Stage 1 applies statistical inference to test whether credible treatment effect heterogeneity exists with the protection against spurious findings. Stage 2 translates heterogeneity evidence into individualized treatment policies, evaluated by cross fitted doubly robust (DR) metrics with Neyman-Pearson (NP) constraints on harm. We illustrate the workflow with working examples based on simulated data and a real ACTG 175 HIV trial. This tutorial provides practical implementation checklists and discusses links to sponsor oriented HTE workflows, offering a transparent and auditable pathway from heterogeneity assessment to individualized treatment policies.

Next-generation AI for Ki67 Image Analysis: Foundation Models, Synthetic Images, and Knowledge Transfer
Xuhong Zhang, Indiana University

This talk presents next-generation AI methods for Ki67 image analysis, focusing on two major challenges in computational pathology: rare-cell recognition within a dataset and knowledge transfer across datasets with limited annotation. Ki67-stained images contain rich cell-level information that is important for tumor grading, biological interpretation, and clinical assessment, but manual analysis is labor-intensive, subjective, and difficult to scale. To address the first challenge, we develop a SAM-assisted supervised learning framework that uses Segment Anything Model (SAM)-derived shape priors to improve cell detection and classification, particularly for rare and morphologically ambiguous cell types. Our ShadoNet framework formulates cell recognition as a structured regression problem by predicting class-specific proximity maps that jointly encode cell location, class identity, and morphology. To address the second challenge, we propose a structure-aware diffusion strategy for cross-domain knowledge transfer. By generating synthetic images that preserve cellular content from a well-annotated source dataset while adapting visual style to a target dataset, our method enables label transfer under domain shift and limited supervision. Together, these approaches improve accuracy, robustness, and data efficiency for Ki67 cell analysis, and suggest a broader path toward scalable and generalizable AI tools in digital pathology.

FRIDAY MORNING 8:30AM- 11:30AM MAY 29

CLINICAL: Innovations in Clinical Trial Methodology: From Digital Health to AI-Enhanced Analysis

Advancing the Analysis of Daily Digital Health Data in Clinical Trials Using Two-Level MMRM and Functional Data Analysis Methods
Liuqing (Jasmine) Yang, AbbVie

TBD

Decentralized clinical trial
Wenjia Wang and Tuo Wang, Lilly

A Faster Algorithm for the Finkelstein-Schoenfeld Test and Composite Outcome Measures
James Austrow, Cleveland Clinic

More than 95% of the runtime of your next power analysis simulation could be eliminated. We overturn the previously-unchallenged assumption that the win ratio requires explicit comparison of every pair of subjects by presenting a novel, sub-quadratic algorithm. We also demonstrate empirical speedup against incumbent software of 20-50x for mid-size trials, and over 100x for very large trials. Hierarchical composite endpoints, and the win ratio in particular, have been growing in popularity over the conventional composite endpoint due to their ability to prioritize outcomes by clinical importance. However, the computational expense of these methods has long been recognized as a weakness, and various workarounds have been proposed. Our technique is the first to directly confront the fundamental scaling problem. We present the algorithm, a proof of correctness, and an implementation in pure Python. We also present quantitative benchmarks showing that our implementation, despite having no compiled code, significantly outperforms existing R packages that use C++ internally.

Med-ICE: Enhancing Factual Accuracy in Medical AI through Autonomous Multi-Agent Consensus
Tom (Zhiyuan) Chen, Hill Research

The integration of Large Language Models into high-stakes clinical workflows is critically hampered by their lack of veri\u0002fiable reliability and tendency to generate hallucinations. This paper introduces Med-ICE, an autonomous framework de\u0002signed to enhance the reliability of LLMs for medical appli\u0002cations. Med-ICE adapts the Iterative Consensus Ensemble paradigm, enabling a group of peer LLM agents to collabora\u0002tively converge on a final answer through iterative rounds of generation and peer review, thereby eliminating the need for an external arbiter and its associated scalability bottleneck. Our work makes three key contributions: (1) a novel seman\u0002tic consensus mechanism that determines agreement based on semantic similarity, crucial for nuanced clinical language; (2) demonstration of state-of-the-art performance, where Med\u0002ICE significantly outperforms both direct single-LLM gener\u0002ation and the Self-Refinement technique on challenging med\u0002ical benchmarks; and (3) a highly efficient and scalable archi\u0002tecture, as our Semantic Consensus Monitor is computation\u0002ally lightweight. This research establishes a new standard for developing safer, more trustworthy LLM systems, paving the way for their responsible integration into medicine

STAT PROGRAMMING/AI: Practical Use of Open-Source Tools and Infusing Statistical Programming and Visual Analytics Using Large Language Models

Retrieval-Augmented Generation in Regulated Biopharmaceutical Environments: A Perspective on Knowledge-Grounded AI
DJ Penix, Pinnacle Solutions and John LaBore, SAS

Biopharmaceutical organizations operate in one of the most data intensive and tightly regulated domains, spanning discovery, nonclinical and clinical development, regulatory affairs, pharmacovigilance, medical affairs, quality, and manufacturing (GxP). Decision making depends on large volumes of heterogeneous information—including protocols, data management plans, statistical analysis plans, study reports, assay validations, SOPs, regulatory guidance, deviations, CAPAs, inspection findings, manufacturing records, and historical submissions—distributed across siloed systems and formats. Although large language models (LLMs) demonstrate strong performance in natural language understanding and generation, their

probabilistic behavior, opacity, and propensity for hallucination limit direct use as trustworthy artificial intelligence in settings that demand traceability, reproducibility, and auditability. Regulatory expectations from authorities such as FDA, EMA, and MHRA, alongside internal quality management requirements, necessitate transparent linkage between generated content and authoritative source records. Retrieval Augmented Generation (RAG) offers a pragmatic, statistically grounded framework for deploying LLM based artificial intelligence under these constraints. By coupling probabilistic generation with deterministic retrieval from curated, version controlled, and access controlled knowledge sources, RAG constrains outputs to be explicitly conditioned on relevant documents—such as clinical and nonclinical protocols, validated SOPs, labeling, regulatory and ICH guidance, inspection responses, and prior submissions—retrieved via vector, symbolic, or hybrid search. These artifacts are incorporated into the model's context, enabling evidence backed, consistent, auditable, and therefore more trustworthy responses. This perspective outlines RAG as an applied statistical and machine learning architecture for trustworthy artificial intelligence in regulated life sciences, emphasizing design elements critical for compliance and reliability: document governance and provenance, access control, retrieval and prompt evaluation, continuous performance monitoring, validation and change control, and human in the loop review. Use cases include regulatory intelligence, clinical trial design and operations, medical affairs and medical information, pharmacovigilance and safety, manufacturing and quality (GxP), and research and discovery. By embedding LLMs within a governed retrieval layer tied to controlled knowledge, RAG provides a pathway to realize the benefits of trustworthy generative AI in biopharmaceutical environments while maintaining the rigor, transparency, and governance required for regulated use.

Hepatotox Tool for Analysis and Displays Associated with Hepatotoxicity
Peigang Li, AbbVie

The PHUSE Safety Analytics Workstream on Analysis and Displays Associated with Hepatotoxicity focused on the development of two white papers, a Stage 1 paper and a Stage 2 paper. The Stage 1 White paper considered a set of aggregate analyses and outputs that can be used, with reasonable medical probability, to rule out potential drug induced liver injury (DILI). The Stage 2 White paper considered additional aggregate analyses and outputs in the event that potential DILI had not been ruled out in Stage 1, along with clinical and medical discernment. These are additional analyses and outputs that would be generated to help characterize potential DILI and in the assessment of causality. A value-added proposition was to also develop software tools to help with the generation of outputs for Stage 1 and 2 White Papers. This presentation will highlight Stage 1 and Stage 2 White Papers and the Hepatotox tool that was developed to help generate outputs specified in the two white papers. A demo of the Hepatotox tool will also be provided. Some ongoing potential extensions of the tool will also be highlighted.

Rebooting Independent Data Monitoring Committee Decision-Making Through Targeted Data Visualization Strategies
Lava Timsina, IDDI

The increasing complexity of contemporary clinical trials-characterized by breadth and depth of interim analyses, adaptive elements, and high volume, multiregional and faceted data streams-has intensified the analytical demands placed on Independent Data Monitoring Committees (IDMCs). Since the issuance of the 2006 FDA guidance on IDMC structure and practices, the role of these Committees has expanded to encompass adaption recommendations, safety surveillance per investigational new drug (IND) reporting requirements, and oversight of large, multi-trial clinical development programs. IDMCs review accumulating evidence under compressed timelines and within geographic limitations to determine whether emerging data suggests potential safety concerns within the context of the intervention's efficacy. However, interim data monitoring reports often rely on extensive tabular outputs that may obscure important patterns or clinically meaningful deviations requiring particular attention. To address these

challenges, and best protect the interest of trial participants, this abstract presents a structured set of graphical and interactive visualization methodologies designed to support and modernize IDMC review by transforming complex datasets into interpretable and navigable visual insights. While rebooting traditional static plots, the proposed framework, through case examples, highlights the value of interactive displays - such as dynamic longitudinal plots, linked AE selection features for specific laboratory parameters, patient profile, underlying conditions for outliers, etc. - that enable reviewers to interrogate data from multiple point of views. Each graphical approach is aligned with its methodological purpose and its contribution to IDMC decision making, including accelerating signal detection, improving interpretability, and facilitating communication of accumulating evidence. As clinical trial designs and data environments continue to evolve, integrating robust and interactive visualization strategies into IDMC processes is essential for maintaining scientific rigor, participant safety, and ethical oversight fundamental to contemporary clinical research.

An Overview of the SASSY System of R Packages
David Bosak, Archytas Clinical Solutions

TBD

BIOMARKERS/PRE-CLINICAL/DISCOVERY

Change Point Mixed-Effect Model Identifies Early Plasma Proteomic Changes in Alzheimer's Disease
Xiaoqing Huang, Indiana University

Longitudinal Plasma proteomic biomarkers for Alzheimer's disease (AD) require advanced methods that capture nonlinear changes at disease onset. We developed changepoint mixed-effects models (CPMMs) to identify plasma proteins with trajectory shifts surrounding clinical conversion in a discovery cohort from the Indiana Alzheimer's Disease Research Center and two independent replication cohorts from the Global Neurodegeneration Proteomics Consortium. Aligning longitudinal protein trajectories to individual onset age, we identified 55 proteins with concordant directional changes across cohorts. These proteins are enriched for immune signalling, vascular dysfunction, and extracellular matrix pathways, with network analysis revealing 10 hub proteins. Nine proteins showed prognostic associations with time to onset, including PECAM1, PCDHB2, PEX14, PSMA1, FABP9, CRIPT, PELI1, TNFRSF8, and SPDEF. A three-protein panel (ECRG4, PCDHB2, BTLA) achieved improved disease classification performance (AUC = 0.78). Onset-aligned CPMM reveals dynamic plasma proteomic transitions in early AD and identifies biologically coherent candidate biomarkers for clinical development.

An LLM-Powered Virtual Agent for Genomics Analysis to Accelerate Drug Discovery
Keela Dai, Lilly

Single-cell genomics assays produce high-dimensional, heterogeneous datasets central to drug discovery, yet their analysis demands intensive computational workflows that bottleneck biological insight. We present an LLM-powered virtual agent that autonomously executes single-cell RNA-seq pipelines — including cell type annotation, differential expression, and biological interpretation — through multi-step reasoning and dynamic code generation. To rigorously assess performance, we developed a structured evaluation framework that scores agent outputs against expert-curated references for accuracy, reproducibility, and analytical best practices. Our agent achieved high concordance with expert results while completing routine analyses in a fraction of the typical time. This evaluation framework also serves as

a reusable benchmark for assessing future AI tools as new models emerge, supporting rigorous adoption of agentic workflows in genomics-driven drug discovery.

Detecting Precise Adverse Drug Events from Real-world Data
Pengyue Zhang, Indiana University

Drug responses cannot be comprehensively revealed in premarketing clinical trials due to limitations in patient heterogeneity. As real-world data (RWD) include diverse subpopulations, RWD can be used to identify precise drug responses in subpopulations. First, a trajectory-informed model (TIM) will be presented. TIM can use drug label-supervised learning to improve the performance of adverse drug event (ADE) detection, and can identify adverse drug combinations and their corresponding risk factors. Particularly, TIM can identify drug combinations that remain at a low risk in the general population, but possess a higher risk in certain subpopulations. Second, a blood biomarker-based PhenoAge-specific pharmacoepidemiologic study will be presented. In this study, we identify metformin compared to sulfonylureas is associated with a reduced risk of Alzheimer's disease (AD) only in individuals with a lower PhenoAge, but was not associated with AD in individuals with a higher PhenoAge and in individuals with a lower age. In conclusion, extension of RWD-based pharmacoepidemiologic study with respect to patient characteristics and biomarkers could identify new and precise drug response.

AdmixSuSiE: Local Ancestry-Aware Fine-Mapping for Biomarker and Target Discovery in Admixed Populations
Boran Gao, Purdue University

Fine-mapping in genome-wide association studies (GWAS) is a critical step for translating association signals into putative causal variants and actionable biological hypotheses. In preclinical research and drug discovery, improved fine-mapping resolution can help prioritize functional variants, nominate target genes, and clarify whether genetic effects are shared across populations or enriched in specific ancestral backgrounds. However, most existing fine-mapping methods are designed for homogeneous populations and are not well suited for admixed populations with mosaic local ancestry. Here we present AdmixSuSiE, a probabilistic fine-mapping framework for admixed populations that leverages local cross-ancestry genetic architecture. AdmixSuSiE addresses three key limitations of existing methods. First, it accommodates ancestry-specific variant architectures beyond the common subset of SNPs shared across ancestries. Second, it explicitly models the effect induced by local ancestry proportion through a random-effect component. Third, it distinguishes between shared causal effects and ancestry-dependent causal effects, enabling more interpretable fine-mapping in admixed populations. Methodologically, AdmixSuSiE extends the sum of single effects framework with a mixture prior on ancestry-specific SNP effects and uses variational inference for scalable posterior approximation. Through extensive simulation and real data applications, we show that AdmixSuSiE substantially improved fine-mapping performance relative to existing approaches. It improved fine-mapping resolution by 63% compared with SuSiE with local ancestry adjustment while maintaining comparable credible set coverage. Under FDR control at 0.05, it increased power to detect shared causal signals from 9% to 18% and increased power to detect ancestry-dependent causal signals from 9% to 24% for African ancestry. We further applied AdmixSuSiE to six traits in the All of Us v8 dataset. Across 55 genome-wide significant regions, AdmixSuSiE identified more high-confidence signals than competing methods and achieved better calibrated credible sets, supporting improved balance between resolution and coverage in real data, which suggest a practical path toward more equitable variant prioritization for downstream functional studies and therapeutic discovery.

EPIDEMIOLOGY: Unresolved Issues and Methodological Challenges in the Analysis of Real-World Evidence (Special Invited in-person and Virtual Session)

In-person: Target Trial Emulation for Real-World Evidence: A Bridge Between Observational Studies and Randomized Trials

Tianming Gao, AbbVie

This presentation will discuss the conceptual foundations and practical implementation of target trial emulation for real-world evidence studies. The talk will examine how aligning observational study design with the protocol of a hypothetical randomized trial can improve transparency and reduce bias while also highlighting potential pitfalls and practical limitations.

Integrating Target Trial Emulation and the Estimand Framework
Gerd Rippen, IQVIA Germany

TBD

Regulatory Perspectives on Methodological Frameworks for Real-World Evidence
Hana Lee, FDA

TBD

Targeted Learning Approaches for Causal Inference in Real-World Data
Lauren Dang, Amgen

TBD

Spline-Based Joint Modeling of PFS and Cumulative Incidence with Interval and Right Censoring
Whitney Su, Duke University

Progression-free survival (PFS) is a common endpoint in oncology trials, yet its estimation becomes challenging when progression is interval-censored and death is right-censored. Standard approaches, such as the Aalen–Johansen estimator fail to respect the clinical ordering constraint that progression must occur prior to death. We propose a nonparametric maximum likelihood estimator (NPMLE) for the joint cumulative distribution function of progression and death, employing a sieve approach with I- and M-splines to ensure monotonicity and clinical interpretability. The proposed method is fit via convex optimization and accommodates both right- and interval-censoring mechanisms. Simulation studies under a Clayton copula framework demonstrate that the proposed estimator gives unbiased estimates of PFS and cause-specific cumulative incidence functions with substantially reduced bias as compared to standard approaches. This work provides a practical and theoretically rigorous tool for analyzing interval-censored progression endpoints in oncology and offers a foundation for extensions to evaluation in the context of competing risks.

Precision Mapping of Tissue Microenvironments: Spatially-Regularized Deconvolution via Empirical Bayes Total Variation
Hongyi (Tom) Liu, University of Michigan

Spatial transcriptomics is a critical tool in biopharmaceutical research for profiling gene expression within its native context. However, standard per-spot deconvolution methods often ignore spatial coherence, yielding noisy maps that obscure boundary definitions essential for assessing tumor margins or drug penetration. We present STT, a statistically rigorous framework that formulates spatial deconvolution as an empirical Bayes penalized-likelihood problem to recover sharp, biologically interpretable tissue architectures. STT introduces a weighted anisotropic total-variation (TV) penalty that acts as a prior on spatial gradients, smoothing noise while preserving edge-specific discontinuities. We solve this non-smooth, non-convex problem via an efficient alternating optimization scheme: FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) for cell-type mixtures and dictionary factors, coupled with a linear program to update adaptive edge weights. nInspired by many computational biology tools, STT is backed by three tiers of theoretical guarantees:**Inner-Loop Optimality:** We prove an convergence rate for the FISTA subproblems.**Global Convergence:** Using the Kurdyka-Łojasiewicz (KL) inequality, we establish global convergence of the full alternating sequence to a critical point, with linear convergence rates under standard non-degeneracy conditions. **Approximation Bounds:** We derive strict error bounds for our smooth soft- relaxation, ensuring the solution remains faithful to the exact TV objective. We demonstrated the utility of STT on mouse cerebellum Slide-seq data , where it successfully recovered complex laminar organizations with significantly sharper boundaries than convex baselines. By combining rigorous statistical theory with an open-source, high-performance R/Rcpp implementation, STT offers a robust solution for researchers seeking precise characterization of tissue heterogeneity in therapeutic development.

Critical success factors for study success with correlated primary endpoints
Carly Middleton, University of Louisville

Bayesian decision rules are often used to guide development decisions in clinical trials. However, when primary endpoints are correlated, failing to account for this correlation can reduce statistical power and mask true treatment effects. We introduce a stepwise procedure that first decorrelates endpoints using a whitening transformation and then evaluates study success using a Bayesian critical success factor (CSF). In simulation, whitening substantially improves the probability of correctly identifying study success across a range of clinically relevant scenarios. We further demonstrate the approach in a hypothetical weight loss study with correlated endpoints of fat and lean mass reduction, illustrating the role of endpoint decorrelation in settings where dual outcomes jointly inform development.

SRT: A Statistical Framework for the Integrative Analysis of Spatial TCR-seq and Spatial Transcriptomics Data
Arkobrato Gupta, The Ohio State University

T cells play a critical role in the adaptive immune system, e.g., protecting our body against cancer and viral infection. Spatial T cell receptor sequencing (TCR-seq) is an emerging technology that measures T cell receptors, which can be used as a barcode to track each T cell, at the close-to-cell level with spatial information on the tissue. Spatial TCR-seq data can provide invaluable insights about clonal relationships among cells, e.g., studying tumor microenvironment during the cancer progression or response to cancer immunotherapy, yet it lacks functional characteristics, which can be studied using high-throughput spatial transcriptomics (HST) data. Hence, it is critical to integrate spatial TCR-seq and HST data to fully understand the T cell activities on the tissue. Motivated by this exciting opportunity, we developed SRT, a statistical framework for the integrative analysis of spatial TCR-seq and HST data. SRT generates spatial maps of local TCR diversity using a moving window approach and implements an association analysis with HST data using Earth Mover's Distance. We applied SRT to the glioblastoma spatial TCR data and it revealed distinct immune niches and key genes associated with clonal heterogeneity. It improves inference from emerging biomedical technologies & supports cancer research & public health.

A Bayesian Method for Pattern Discovery in Longitudinal Vaginal Microbiome Data
Souvik Paul, University of Illinois Chicago

Analysis of microbiome data remains challenging due to high dimensionality, sparsity, compositional constraints, batch effects, and inter-subject variability, challenges that become even more pronounced in longitudinal studies. Our motivating application comes from vaginal microbiome data measured on women/girls in Kenya. We propose an approach based on Bayesian nonparametric modeling that performs stochastic two-level clustering in a longitudinal setting. Our method targets vaginal microbiome trajectories, modeling per-subject taxa counts with a multinomial likelihood to infer on the simplex, preserving compositional constraints without ad hoc transformations. Sparsity is inherent and addressed by incorporating zero inflation via an explicit structural-zero component in the model, distinguishing truly absent taxa from low-abundance taxa. This approach provides a flexible framework for modeling longitudinal dependence, allowing the clustering structure to evolve over time. We evaluate performance through extensive simulation and demonstrate how learned longitudinal microbial patterns can support pattern recognition to predict sexually transmitted disease risk. Keywords: Biclustering, Longitudinal, Compositionality, Pattern Recognition

Doubly Robust Estimation of Treatment Effects with Missing Outcomes in Longitudinal Studies
Asteria Herbert Chilambo, Carnegie Mellon University

Longitudinal studies are central to understanding dynamic treatment effects, but their analysis is complicated by within-unit temporal dependence and sequentially missing outcomes. Standard methods, such as outcome regression or inverse-probability weighting can address missingness, but may be biased under model misspecification or when nonparametric models are used. Although doubly robust estimators offer protection against such misspecification, existing theory largely focuses on cross-sectional data or single-time-point missingness. We develop a doubly robust framework for estimating mean potential outcomes under a sequential missing-at-random (SMAR) assumption. We derive the efficient influence function for the mean potential outcome under a fixed treatment regime and construct a doubly robust estimator that is root-n consistent and asymptotically normal, provided that cross-fitting is used and the nuisance functions are estimated at n^(-1/4) rates. The estimator allows flexible, data-adaptive estimation of nuisance components. Our methodology is motivated by and applied to a longitudinal randomized clinical trial of mindfulness-based interventions for irritable bowel syndrome, in which outcomes are collected via smartphone surveys and occasional nonresponse induces complex missingness structure. To address this missingness, our method relies on a strong SMAR assumption that only depends on the observed history (i.e., prior non-missing outcomes). In future work, we explore relaxing this assumption to allow missingness to depend on past outcomes, even if they are not observed.

Screening and Validating Drug Efficacy in MCI-to-AD Progression through EHR-Based Analysis
Nabasmita Talukdar, Michigan State University

Background: Alzheimer's disease (AD) affects over 6.9 million U.S. adults aged ≥65, with annual costs exceeding $360 billion. Individuals with mild cognitive impairment (MCI) progress to AD at rates far higher than the general population, making them ideal for studying early therapeutic effects. Despite over 570 proposed repurposed drugs (2012-2022), none identified through EHR-based studies have achieved clinical approval. We applied an analytic framework to assess associations between medication use and MCI-to-AD progression, finding that many drugs appeared significant. Method: We conducted a systematic review of EHR-based studies (2020-2024) to identify repurposed medications previously reported with reduced AD risk or improved survival. I assembled a retrospective cohort of adults aged ≥65 with incident MCI from Truveta (2018-2024). Inclusion criteria included ≥2 MCI diagnoses, ≥1 year of prior healthcare history with ≥2 encounters, and no prior AD or related dementia (Fig. 1). We developed an automated drug-screening pipeline to define drug-specific exposure groups and estimate odds of incident AD using multivariable logistic regression. Analyses were conducted for both new-user and prevalent-user designs and replicated in discovery and validation cohorts defined by index-date splits to mimic real-world conditions. Results: We analyzed 57,410 individuals with incident MCI (9,236 cases vs 48,174 controls; 57% female; 34% age 65-74, 46% 75-84). In Phase 1, several drugs were significantly associated with lower odds of AD and replicated in the temporal split. Signals that agree across designs (new-user and prevalent-user) include albuterol, amoxicillin, dexamethasone, fluticasone, gabapentin, glucagon-like peptide agents (GLP), loop diuretics, losartan, and proton pump inhibitors (PPI); statins, and hydrochlorothiazide showed non-significance in prevalent-user (Fig. 2). Statins, atenolol, metformin, and sildenafil were inconsistent. Conclusion: Our results are consistent with prior real-world studies and demonstrate that the proposed pipeline enables efficient large-scale screening of drug candidates using EHR data. This work prioritizes AD repurposing candidates using EHRs. However, these findings should be interpreted as preliminary screening signals. The widespread significance observed raises concerns about false discoveries, underscoring the need for more rigorous approaches. Next, we will apply target trial emulation with balance diagnostics to adjust for confounding, quantify false discovery rates, and further validate promising candidates, including sensitivity analyses with alternative exposure definitions.

Extended Fiducial Inference for Statistical Models with Unknown Noise Distributions
Wen-Hung Wang, Purdue University

Gaussian noise assumptions underpin much of classical statistical inference and machine learning, yet real-world data often deviate substantially from this ideal, rendering Gaussian-based inference unreliable. Existing approaches for unknown error distributions—such as robust statistics, flexible parametric families, and conformal prediction—typically fail to simultaneously provide valid parameter inference, automatic adaptation to complex noise structures, and principled uncertainty quantification. To address this gap, we propose Nonparametric Error-Extended Fiducial Inference (NE-EFI), a multi-stage framework that generalizes Extended Fiducial Inference (EFI) to settings with unknown error distributions. NE-EFI leverages EFI's error recovery property to impute latent errors, reconstructs the noise distribution via kernel density estimation, and iteratively refines inference. We establish theoretical convergence guarantees and demonstrate through simulations and real data that NE-EFI achieves accurate inference and valid uncertainty quantification, while consistently outperforming robust statistical methods and conformal approaches across a broad range of non-Gaussian error distributions.

Approximate Multi-Source Exchangeability Based Borrowing on Hybrid Control Trials
Tong Zhang, University of Illinois at Chicago

Conventional randomized clinical trials (RCTs) remain the gold standard for evaluating experimental interventions. However, the considerable financial burden and extended duration of conventional designs often undermine the efficiency and variability of RCTs. Hybrid control trials (HCTs) provide a contemporary methodological alternative by combining data from a concurrent control arm with external or historical control data to form a more efficient and robust comparison group. Multiple-source exchangeability models (MEMs) provide a flexible Bayesian framework that is designed to integrate multiple control arm data sources while accommodating varying degrees of exchangeability. These models have been further developed with propensity score weighting (PW) to achieve alignment in characteristics with the current trial data. However, exact exchangeability can be difficult to justify in practice. We introduce a flexible Bayesian framework that relaxes the strict conditions of the MEM framework and demonstrates this approach in various settings. The performance of the proposed approach is evaluated through extensive simulation studies.

Fusion Design to Assess Placebo-controlled Effects in Non-Inferiority Trials
Quynh Long Khuong, University of Pennsylvania

Non-inferiority (NI) trials have become increasingly popular and serve as a critical design to address the practical and ethical challenges of modern medicine. However, assessing placebo-controlled effects (e.g., assay sensitivity) in NI trials is challenging due to the absence of a placebo arm. In this paper, we propose a new framework for assessing placebo-controlled effects using fusion design methodology that integrates information from both the current NI trial and historical superiority trials without depending on the constancy assumption (i.e., the assumption that NI trial and historical studies are sufficiently similar in all important design and conduct features). This assumption is strong and difficult to satisfy in practice, as differences in patient populations, clinical practices, and outcome definitions over time can substantially alter the apparent efficacy of the active control. We derived a fusion estimator based on the inverse odds of sampling weights that enables direct comparison of the experimental treatment against a transported placebo arm. This already addresses a key limitation of traditional NI trial designs. To evaluate our proposed framework, we provide a numerical example and a simulation study to illustrate the estimation approach and evaluate estimator performance. Simulation results across a range of sample sizes demonstrate that the fusion estimator is unbiased (with absolute bias ranging from 0.02% to 0.5%) and achieves appropriate 95% confidence interval coverage when the assumptions are satisfied, in contrast to the crude estimator whose bias ranges from 9.15% to 10.61%. Finally, we discuss the conditions required for

nonparametric identification and practical considerations in applied settings. Overall, our method offers a straightforward and effective approach for combining trial data to estimate placebo-controlled effects in NI trials.

High dimensional Bayesian Variable Selection with Applications to Cancer Data
Shanta Ghosh, University of Illinois at Chicago

Variable selection has become an essential element of statistical modeling to yield parsimonious models while keeping high prediction accuracy. However, in high-dimensional data, it can be challenging in the presence of collinearity among covariates. Classical variable selection methods often fail when predictors exhibit complex dependence structures. To address these challenges, we propose a Bayesian Clustering and Structured Selection (BCSS) framework that jointly learns covariate clusters and performs variable selection. We perform a model-based clustering using Dirichlet-Process without assuming the correlation structure to be known. It clusters similar covariates that share similar latent correlation structure in X to be grouped together. Within this clustering framework, BCSS performs variable selection at both the cluster and predictor levels, improving stability by treating correlated predictors as units. Posterior inference is performed using Markov chain Monte Carlo (MCMC) sampling. Through simulation studies, we demonstrate improved selection accuracy under high collinearity. The proposed method is applied to cancer biomarker datasets to identify molecular predictors associated with survival and disease progression. The goal of this analysis is to identify genes associated with breast cancer prognosis while evaluating the performance of a Bayesian clustering-based variable selection method relative to Lasso and Elastic Net in high-dimensional genomic data.

Comparison of Disproportionality Analysis for Detection of Signals for Adverse Drug Reactions Using Zoloft Data
Karen Lu, Adelaide University (Australia)

Background: Zoloft is one of the most prescribed antidepressants in the world. But the adverse effects (AEs) on the label were based on results from 3606 clinical trial participants, which may not be enough to detect important but rare adverse reactions. By analyzing data from the Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS), which contains a wider array of information about adverse effects, it is possible to potentially discover new adverse drug reactions (ADRs) for Zoloft. Method: Analysis began by downloading all FAERS data from 2024, processing it into R using the package faersquarterlydata, filtering for indications Zoloft is prescribed for. A two-step duplicate removal process is conducted, first removing completely identical items and second demographic matches reported on the same date. Afterwards two new datasets are created, filtered to include or exclude Zoloft. AEs with both a proportional reporting ratio (PRR) above 2 and reporting odds ratio (ROR) with a lower 95% confidence interval bound above 1 are tentatively considered potential new ADRs specific to Zoloft. Results: Only data from 2024 was used to test the methodology. 798 patients prescribed Zoloft reported 2,507 AEs, while 6417 patients with the same indications not prescribed Zoloft reported 23,349 AEs. Comparing demographic data between these two groups showed that Zoloft users were significantly younger, and indication proportion significantly differed between populations (p=0.000499). Using this method, the most prominent potential ADRs were eyelid ptosis, Bell's palsy, and abnormal gastrointestinal sounds. Since gastrointestinal disorders are already listed on the label as an ADR, this supports the method's ability to discover ADRs. Conclusion: The FAERs database can be used to discover new ADRs. This can provide evidence for future clinical trials to further explore the prevalence of these ADRs. The method used in this study gives more specific and consistent results compared to mass analysis over multiple decades of FAERs cases and stronger evidence in favor of the new ADRs discovered, as many are described on the label or other publications.

Applying Event-Aligned Time Scale in Latent Class Trajectory Analysis of Cardiovascular Disease Cohorts
Jiafeng Zhu, Northwestern University

Latent class analysis is widely used to identify heterogeneous patient subgroups in cardiovascular disease (CVD) studies based on longitudinal risk factors, such as systolic blood pressure (SBP). Latent class methods uncover underlying subgroups in which individuals who experience CVD events and those who are censored may coexist. In most longitudinal CVD cohort analyses, follow-up time is aligned by enrollment date. However, patients approaching their first CVD event often exhibit shared clinical patterns, such as increasing SBP trajectories prior to the event. In this study, we propose an event-aligned time scale (EATS) that aligns longitudinal trajectories by each patient's first CVD event before conducting latent class analysis. To address the presence of censoring, we develop a method that incorporates censored observations in a stable manner within the EATS framework. Using both simulated data and real-world data, we demonstrate the advantages of the EATS approach, highlight the importance of incorporating censored cases, and evaluate the performance of the proposed method under various settings. This work introduces the EATS framework to CVD research and provides an initial methodological approach for integrating censored observations into latent class analysis under event-aligned time scales.

BMW: Bayesian Model-Assisted Adaptive Phase II Clinical Trial Design for Win Ratio Statistic
Di Zhu, Indiana University Indianapolis

The win ratio (WR) statistic is increasingly used to evaluate treatment effects based on prioritized composite endpoints. However, existing Bayesian adaptive designs are not directly applicable because the WR is a summary statistic derived from pairwise comparisons and does not correspond to a unique data-generating mechanism. In this paper, we propose a Bayesian model-assisted adaptive design for randomized phase II clinical trials based on the WR statistic, referred to as the BMW design. The proposed design uses the joint asymptotic distribution of WR test statistics across interim and final analyses to compute posterior probabilities without specifying the underlying outcome distribution. The BMW design allows flexible interim monitoring with early stopping for futility or superiority and is extended to jointly evaluate efficacy and toxicity using a graphical testing procedure that controls the family-wise error rate (FWER). Simulation studies demonstrate that the BMW design maintains valid FWER control, achieves power comparable to conventional methods, and substantially reduces sample size. An R Shiny application is provided to facilitate practical implementation.

Who are we studying in TriNetX? The first concordance assessment of atopic dermatitis cohorts against 11 US data sources
Jessica Wong, University of Pennsylvania

TriNetX is a federated electronic health record (EHR) network enabling queries across 100+ US healthcare organizations (>100 million patients), increasingly used to generate dermatologic real-world evidence (RWE). While federated EHRs provide substantial scale and clinical detail (diagnosis, medication, procedure codes), aggregated data without patient-level access introduces limitations: selection bias (affiliated systems only), heterogeneous coding practices, and lack of enrollment information (uncertain observation windows). Before generating RWE, we must determine whether TriNetX cohorts are comparable to established cohorts in prior research. We conducted the first concordance assessment of TriNetX atopic dermatitis (AD) cohorts, comparing demographics and atopic comorbidities against 18 studies (2015-2025) from 11 US sources: claims (MarketScan Commercial, MarketScan Medicaid, Optum), registries (PEDISTAD, PROSE, PEER), surveys (GfK, NHANES, NHIS), integrated systems (Kaiser), and trials (TRuE-AD). Using a validated ICD-10 algorithm (L20.8, L20.9, L20.84), we identified AD patients treated 10/2024-10/2025 receiving topical corticosteroids or dupilumab. Concordance was strong. Among pediatric dupilumab patients, median age (9 vs 7.6-10.9 years), sex (54% vs 51-53% male), asthma (37% vs 38%), and allergic rhinitis (42% vs 43%) matched PEDISTAD/PROSE. Among adult dupilumab patients, median age (42 vs 38-44 years), sex (43% vs 40-44% male), asthma (22% vs 25%), and allergic rhinitis (36% vs 35%) matched

PROSE/MarketScan/Optum. Black representation was 2.3x higher (pediatric: 28% vs 12% PEDISTAD; adult: 19% vs 8% Optum). Internal consistency supported severity-based prescribing: asthma 26%-37% and eosinophilic esophagitis 1%-8% from topical corticosteroids to dupilumab. These findings support TriNetX for dermatologic RWE and underscore the importance of concordance assessment prior to federated EHR analyses.

Revisiting Pediatric Echocardiographic Z-Scores: Addressing Sex Bias in the Pediatric Heart Network Model
Elliott Shi, Purdue University

Background: Echocardiographic Z-scores are important tools used for diagnosing and managing pediatric heart disease. Lopez et al. (2017) used the Pediatric Heart Network (PHN) dataset to develop BSA-based z-scores and concluded that age, sex, and race were statistically significant but clinically insignificant. They later wrote a follow-up in 2021 that compared their model with those from Boston, Italy, and Detroit, but did not mention sex differences. Emerging evidence has revealed small but systematic differences that may influence diagnostic thresholds. Methods: Data from the PHN (n = 3,566) were used to develop a new regression-based model that accounts for these covariates. The PHN model proved difficult to reproduce consistently with discrepancies in significant interactions. Z-scores were then compared between models, with statistical testing of distributional shifts. Results: The PHN z-scores, when plotted against the regression-based scores, showed a distinct bimodal distribution by sex, with females underestimated and males overestimated across most variables. Sex explained roughly 85% of the variance in the difference between PHN scores and regression-based scores for Aortic Root. In addition, regression-based z-scores for females were +0.19 higher than those from the PHN's calculator for Aortic Root. The regression-based model reduced systematic bias while maintaining accuracy. Conclusion: While the PHN study remains foundational and influential, there are distributional shifts that need to be addressed. Our regression-based framework builds upon this foundational resource by offering a more reproducible and equitable approach. Future guidelines should incorporate demographic factors into their models to improve reproducibility and fairness in pediatric echocardiography.

Bayesian Methods for Determining Average Bioequivalence in 2x2 Cross-over Studies
Clara Andersen, Baylor University

To assess bioequivalence (BE) between two drugs, the FDA recommends the use of the two one-sided t-test (TOST), however there is no Bayesian method currently included in the FDA guidance. Kruschke (2013), Peck and Campbell (2019), and Peck et al. (2022) consider an empirical Bayesian approach for assessing BE using the BEST R package (Bayesian Estimation Supersedes the t-Test); we call their BE application of BEST "Empirical Bayesian Bioequivalence" (EBBE). We propose a fully Bayesian approach (BBE) to determine average BE between a test and a reference drug, that uses relatively non-informative priors. In simulation studies without outliers, BBE, EBBE and TOST exhibited similar pass rates. However, when outliers were introduced, both Bayesian methods outperformed TOST, demonstrating improved robustness, especially in small sample size scenarios. We also propose a Bayesian sample size determination approach based on a two-prior framework and a predictive probability criterion. Across 1,000 simulated datasets, both BBE and EBBE required 40 subjects to achieve 80% predictive power, demonstrating that the fully Bayesian approach achieves performance comparable to EBBE when the design priors are aligned with realistic pharmacokinetic information. These results suggest that empirical Bayesian methods are not necessary to achieve robust performance or efficient study designs.

Structured and Scalable AI for High-Dimensional Biological Data: Beyond Lasso Under Correlation
Vinamrata Sharma, Indiana University Bloomington

High-dimensional biological data pose fundamental challenges for artificial intelligence due to strong feature correlations, limited sample sizes, and the need to balance predictive and explanatory power. Traditional sparsity-inducing methods, such as Lasso, often fail under high correlation, leading to unstable feature selection and reduced interpretability. In this work, we present a scalable optimization framework based on Boolean convex programming to address these limitations. Our approach reformulates subset selection using a tight convex relaxation with theoretical guarantees, enabling efficient optimization via projected quasi-Newton methods. We further extend this framework to structured sparsity settings, including group, overlapping, and graph-based regularization. Empirical evaluations on genome-wide association studies (GWAS) demonstrate improved scalability, robustness to correlation, and enhanced feature selection accuracy compared to existing methods. This work provides a unified pathway from classical sparsity models to modern structured AI methods for biological data analysis.

Scalable Bayesian Inference of Optimal Informer Sets for Early Drug Discovery
Alexander J. Fagan, University of Wisconsin - Madison

Informer-based ranking is a statistical framework that leverages historical bioactivity data from large compound libraries across known targets to predict compound activity against a novel target by screening only a small subset of “\informer\" compounds. Bayes Optimal Informer Sets (BOISE) provide a principled approach to selecting these informer sets to maximize the hit rate of subsequent screenings. While highly effective existing methods rely on heuristic sequential selection limiting their scalability to high-throughput settings and million-compound libraries common in industrial drug discovery. To address this scalability issue, we introduce miBOISE, a BOISE variant that optimizes a mutual information-based objective enabling analytic and computationally tractable informer selection. The runtime complexity of the miBOISE model grows linearly with library size and is independent of informer set size.

Bayesian Approach to Composite Strategy for Handling Intercurrent Events for Continuous Outcomes
Kaylee Pascarella, Baylor University

The composite variable strategy handles intercurrent events in clinical trials by incorporating the event directly into the endpoint definition. While straightforward for binary outcomes, implementing this strategy for continuous endpoints presents challenges: investigators must assign a numeric "failure value" to subjects experiencing ICEs, yet standardized guidance is largely absent. Recent proposals recommend defining the failure value relative to a distributional feature of the control population (e.g., the 10th percentile) for clinical interpretability. However, because this value must be estimated from trial data, uncertainty in the failure value must be propagated to the treatment effect estimate to ensure valid inference. We propose two Bayesian approaches that naturally integrate this uncertainty: a two-stage estimator that separates failure definition from treatment effect estimation, and a parametric joint estimator that simultaneously models outcomes and intercurrent event occurrence. Simulations based on asthma trials with FEV1 endpoints demonstrate that, with diffuse priors, both Bayesian methods achieve coverage probabilities comparable to frequentist benchmarks.