ICSA Midwest

NIC-ASA

ICSA Midwest Chapter and NIC-ASA Joint Fall Meeting

October 2-3, 2025

Astellas Pharma Global Development Inc.
2375 Waterview Dr,
Northbrook, IL 60062

This meeting is proudly sponsored by

Table of Contents

Conference Organization Committees

Core Committee Members

Ziqian Geng (Chair), AbbVie
Lei Shu (Co-chair), AbbVie
Dina Elsouda, Astellas
Hui Zhang, Northwestern University
Jennifer Yen, Abbott Laboratories
Jie Ma, Astellas
Jiewei Zeng, AbbVie
John Schoenfelder, AbbVie (retired)
Jung Wook Park, Astellas
Kentaro Takeda, Astellas
Lihui Zhao, Northwestern University
Mandy Jin, AbbVie
Ran Liu, AbbVie
Rishita Nuthethi, Astellas
Saurabh Mukhopadhyay
Xiaohong Huang, AbbVie
Xiaotian Chen, AbbVie
Yang Yang, AbbVie
Yibo Wang, AbbVie
Yiran Bonnie Hu, AbbVie

Subcommittee	Members
Keynote Speaker	Ziqian Geng, Lei Shu
Short Courses & Scientific Program	Mandy Jin, Kentaro Takeda, Xiaohong Huang, Xiaotian Chen
Student Poster & Roundtable	Lihui Zhao, Hui Zhang
Program Book	Yibo Wang, John Schoenfelder
Treasurer	Yang Yang, Dina Elsouda
Registration	Yang Yang, Dina Elsouda, Jiewei Zeng, Rishita Nuthethi
Meeting Rooms	Jung Wook Park, Dina Elsouda

Program Outline

Short Courses on Oct 2, 2025

7:30 am – 8:30 am	Check-in, Registration & Breakfast
8:30 am – 11:30 am	**Short Course 1 (SM.301)** Adaptive Phase 2/3 Designs: Statistical Considerations and Beyond, by Dr. Cong Chen (Merck; virtual)
11:30 – 1:00 pm	**Lunch**
1:00 pm – 4:00 pm	**Short Course 2 (SM.301)** Unleashing the Power of Machine Learning and Deep Learning to Accelerate Clinical Development, by Dr. Yunzhao Xing, Dr. Sheng Zhong, and Dr. Li Wang (AbbVie; in-person)
4:15 pm – 6:00 pm	**Reception (Skygazer Cafe)** Social event with free cocktail and snacks for each registrant

Fall Conference on Oct 3, 2025

7:30 am – 8:30 am	Check-in, Registration & Breakfast
8:30 am – 8:45 am	**Opening Remarks (SM.301)** by Rui (Sammi) Tang, Head, Quantitative Sciences and Evidence Generation, Astellas
8:45 am – 9:30 am	**Keynote Address (SM.301)** **Statistics and Statistician in the AI Era: Deepen the Roots, Widen the Branches** by Xun Chen, VP, Data & Statistical Sciences, AbbVie
9:30 am – 9:45 am	Break / Networking
9:00 am – 4:00 pm	**Student Poster Exhibition (SM.304 + Hallway)** Organizers: Lihui Zhao and Hui Zhang (Northwestern University)
9:45 am - 10:45 am	**Scientific Sessions (in parallel)** • Session 1: Advanced Bayesian Trial Designs and Analyses **(SM.301)** Organizer & Chair: Kentaro Takeda (Astellas) • Session 2: Statistical Considerations for Rare Disease and/or Cell & Gene Therapy Development **(SM.302)** Organizer & Chair: Yusuke Yamaguchi (Astellas)
10:45 am – 11:00 pm	Break / Networking
11:00 am - 12:00 pm	**Scientific Sessions (in parallel)** • Session 3: Innovative Study Designs **(SM.301)** Organizer & Chair: Qi Yan & Lei Shu & Hao Wang (AbbVie) • Session 4: Recent Development in Predictive Modeling and Causal Inference **(SM.302)** Organizer & Chair: Lihui Zhao & Hui Zhang (Northwestern University)

Fall Conference on Oct 3, 2025 (cont'd)

12:00 pm – 1:30 pm	**Lunch + Student Roundtable (SM.303)** Roundtable panel: Jennifer Yen (Abbott), Xiaohong Huang (AbbVie), Kentaro Takeda (Astellas), and Lihui Zhao (Northwestern University)
1:30 pm – 2:30 pm	**Scientific Sessions (in parallel)** • Session 5: Accelerate Drug Development Through Master Protocol Designs **(SM.301)** Organizer & Chair: Edy Yao (AbbVie) • Session 6: Leveraging Large Language Models and Multimodal Machine Learning for Precision Medicine and Targeted Therapies **(SM.302)** Organizer & Chair: Miles Xi (AbbVie)
2:30 pm – 3:00 pm	**Student Poster Q&A Session (SM.304 + Hallway)** Lihui Zhao (Northwestern University), Hui Zhang (Northwestern University), Dave Zhao (UIUC), Mandy Jin (AbbVie), Kentaro Takeda (Astellas)
3:00 pm – 4:00 pm	**Scientific Session** • Session 7: Integrating AI/ML in Clinical Drug Development **(SM.301 + SM.302)** Organizer & Chair: Jia Jia, Mandy Jin & Hui Zheng (AbbVie)
4:00 pm – 4:15 pm	**Student Awards (SM.301)** Lihui Zhao (Northwestern University)
4:15 pm – 4:30 pm	**Closing Remarks (SM.301)** Ziqian Geng (AbbVie) & Lei Shu (AbbVie)

Short Courses

Short Course 1: Adaptive Phase 2/3 Designs: Statistical Considerations and Beyond

Time: 8:30 am – 11:30 am, Oct 2, 2025

Abstract: Following the tremendous success of immune checkpoint inhibitors and other innovative drugs, the last few years have witnessed an explosive growth in number of oncology trials. While the expectation is high for the new drugs or vaccines under development, it is unrealistic to expect all of them to have the same success, especially given the improved standard-of-care. It is imperative to apply cost-effective design strategies to early-to-late transition, whereas the 2-in-1 design strategy plays a central role in decision-making at the Phase 2/3 program level. A comprehensive decision matrix will be used to determine the speed of proceeding to Phase 2/3 after signal detection, and common issues such as how to set the GNG criteria and how to account for the possible predictive biomarker effect will be addressed.

Under FDA's Project Optimus, adaptive Phase 2/3 designs with dose selection become increasingly important. The impact of dose selection on Type I error is fundamentally driven by the probability of pick-the-winner under the null hypothesis. The growing risk of picking a "loser" necessitates ensuring consistency between Phase 2 and Phase 3 results, making it both a regulatory review issue and a sponsor concern. Both considerations will be incorporated into a data-dependent hybrid design to bridge operationally and inferentially adaptive designs, broadening and deepening our understanding of the adaptive Phase 2/3 design.
Attendees of this short course will not only learn technical skills that can be immediately applied in practice, but also gain an enhanced perspective of strategic decision-making in oncology drug-development.

Instructor: Dr. Cong Chen, Merck & Co., Inc

Dr. Cong Chen is Scientific AVP in Early and Mid-stage Oncology Statistics at Merck & Co., Inc., providing fit-for-purpose decision-making strategies and novel statistical approaches for oncology clinical development programs, and supporting oncology external collaborations, competitive intelligence, and high-profile due diligence projects. Prior to taking the role, he led the statistical support for the development of Keytruda (pembrolizumab) and played a key role in accelerating its regulatory approvals.

He is an elected Fellow of the American Statistical Association (2016), an Associate Editor of *Statistics in Biopharmaceutical Research*, a member of the Cancer Clinical Research Editorial Board and a leader of the DIA Innovative Design Working Group. He has published over 100 papers and 10 book chapters on innovative design and analysis methods of clinical trials. He has also given multiple short courses on this subject at statistical conferences and is a frequent speaker and panelist at major clinical conferences, focusing on statistical design strategies and regulatory policy for oncology drug development.

Short Course 2: Unleashing the Power of Machine Learning and Deep Learning to Accelerate Clinical Development

Time: 1:00 pm – 4:00 pm, Oct 2, 2025

Abstract: With the rapid advancement of machine learning (ML) and deep learning (DL) methodology in the last decade, the performances of prediction tasks in many computer science fields (e.g., natural language processing) have been greatly improved. However, the impact of ML/DL in the field of pharmaceutical development has been relatively limited. Hence, we would like to propose a short course to motivate and encourage the use of ML/DL in pharmaceutical development. The course starts with an overview of ML/DL methodology evolution over time and the related key concepts (e.g., back-propagation, hyperparameter tuning, etc.). Then the latest developments in image processing and natural language processing are introduced, together with their novel applications in pharmaceutical development from our recent projects and submitted papers. In terms of the course outline, the materials of the course are divided into three sections:

I. General ML/DL methodology:
II. Image processing and applications: deep convolutional neural networks (DCNN), object detection and segmentation, Region-based CNN (R-CNN), "You Only Look Once" (YOLO), and applications (e.g. psoriasis area and severity prediction)
III. Natural language processing and applications: word embeddings (word2vec), recurrent neural networks and language models, self-attention and transformers, pre-train and fine-tune paradigm and applications (e.g., adverse drug event prediction)

Instructors: Dr. Yunzhao Xing, Dr. Sheng Zhong, and Dr. Li Wang, AbbVie

Yunzhao Xing is an associate director of Statistical Innovation at AbbVie, holding a PhD in Material Science from the University of North Carolina at Chapel Hill and a background in Physics. Prior to AbbVie, he served as a senior scientist at Halliburton, focusing on sensor modeling and simulation. Since joining AbbVie in 2018, Yunzhao has led numerous successful projects in machine learning, deep learning, and image processing. His skill set encompasses web scraping, simulation modeling, and interactive web application development, making him a pivotal contributor to AbbVie's Statistical Innovation Group. Yunzhao is recognized for his commitment to pushing the boundaries of statistical innovation.

Dr. Sheng Zhong is a Director of Statistics at AbbVie Inc. He received his Ph.D. in Statistics from the University of Chicago. At AbbVie, he has led multiple innovative predictive modeling projects across different fields such as clinical trial enrollment duration forecasting, virtual controls based on targeted learning in single-arm trials, and predictive clinical safety monitoring based on structured and text data. His recent works have led to multiple publications and manuscripts. Before joining AbbVie in 2016, Dr. Zhong worked at a big data analytics start-up for heavy machine equipment maintenance, where his work led to 3 US patents.

Li Wang, PhD, is currently Senior Director and Head of Statistical Innovation group at AbbVie. Li is leading Design Advisory which provides strategic and quantitative consulting to all Development teams in all Therapeutic Areas to facilitate innovative thinking and innovative design evaluation. Li leads Clinical Trial Innovation capability in AbbVie to drive Machine Learning and Advanced Analytics research and application in Development and Li is co-leading ASA Biopharma Section Scientific Working Group for “Statistical Perspectives on AI/ML in Pharmaceutical Product Development”. Li also has extensive success in filing and approvals leading to three blockbuster drugs Eliquis, Rinvoq, and Onglyza, improving the lives of millions of patients everyday.

Opening Remarks

Time: 8:30 am – 8:45 am, Oct 3, 2025

Speaker: Rui (Sammi) Tang, Head, Quantitative Sciences and Evidence Generation, Astellas

Dr. Rui (Sammi) Tang is a seasoned drug developer and innovative pharmaceutical leader who has contributed to the successful development and approval of numerous therapies—bringing medicines from research to market that now reach millions of patients every day. With a proven track record of building high-performing teams and driving scientific and operational innovation, she delivers data-driven solutions that accelerate drug development and improve global health outcomes.

As Senior Vice President and Global Head of Quantitative Sciences and Evidence Generation (QSEG) at Astellas Pharmaceuticals, Dr. Tang leads the company's global data and evidence strategy across quantitative analytics, epidemiology, real-world evidence (RWE), biostatistics, programming, medical writing, scientific communication, data systems & enablement, and data management. She is at the forefront of applying Generative AI in regulatory and clinical documentation, AI/ML-powered analytics, and external data to optimize study design and development efficiency.

Dr. Tang serves on the Executive Committee for Data Science & AI at the American Statistical Association (ASA) and is co-founder of DahShu, a global nonprofit advancing data science research and education with over 5,000 members.

Dr. Tang holds a PhD in Statistical Genetics from Michigan Technological University and an Executive MBA from MIT Sloan. She is also an Adjunct Professor at Yale University School of Public Health. She is widely recognized for combining scientific depth with strategic leadership to deliver transformative therapies that improve lives worldwide.

Keynote Speaker

Title: Statistics and Statistician in the AI Era: Deepen the Roots, Widen the Branches

Time: 8:45 am – 9:30 am, Oct 3, 2025

Speaker: Xun Chen, VP, Data & Statistical Sciences, AbbVie Inc.

Xun Chen serves as the Vice President of Data and Statistical Sciences. Bringing her vast experience in clinical sciences and statistical methodology, Xun leads the DSS organization, elevating the critical capabilities and business impact that DSS delivers across AbbVie.

Prior to AbbVie, Xun held roles at Sanofi Pharmaceuticals, where she spearheaded the global Biostatistics and Programing department in successfully bringing a number of innovative treatment solutions to patients worldwide. These include renowned brands like Dupixent, Praluent, Kevzara, Sarclisa, Xenpozyme, Nexviazym, and Altuviiio, as well as the latest major breakthroughsubmissions for treatments in SPMS and Hemophilia. Her leadership in establishing the dedicated Statistical Innovation Hub (SIH) and Analytic Reporting Tool (ART) laid the foundation for modernizing statistical and programming practices within the organization. Before heading the Biostatistics and Programming Department, Xun also successfully led the establishment of a full spectrum clinical sciences and operation platform for Sanofi in China. Xun is a graduate of Columbia University with a Ph.D. in biostatistics.

Throughout her leadership journey, Xun has been dedicated to cultivating a culture of continuous learning and collaboration to promote organic teamwork and drive sustainable collective achievements. Through her role as the President of International Chinese Statistical Association (2023-2025), Xun is committed to expanding her influence and contributions in enhancing data and statistical scienceeducation, research and leadership to meet the evolving needs of society.

Scientific Sessions

Session 1: Advanced Bayesian Trial Designs and Analyses

Time: 9:45 am -10:45 am, Oct 3, 2025

Organizer: Kentaro Takeda (Astellas)

Chair: Kentaro Takeda (Astellas)

Session Abstract: The Bayesian framework is ideal for adaptive learning and flexible decision-making in drug development. Recent FDA guidances on Adaptive Designs and Complex Innovative Trial Designs both mention that Bayesian adaptive designs and analyses are among the most promising approaches to improving the efficiency of drug development. This session will focus on advanced Bayesian clinical trial designs and analysis for drug development.

1. **Title:** BOP2-FE: Bayesian optimal phase II design with futility and efficacy stopping boundaries
   Speaker: Atsuki Hashimoto (Astellas)

   Abstract: The primary purpose of an oncology single-arm trial is to evaluate the effectiveness of anticancer agents and make a go / no-go decision while maintaining patient safety. We propose a flexible Bayesian optimal Phase II design with futility and efficacy stopping boundaries for single-arm clinical trials, named the BOP2-FE design. The proposed BOP2-FE design allows for early stopping of efficacy when the observed antitumor effect is sufficiently higher than the null hypothesis value in the interim looks and retains the benefits of the original BOP2 design, such as explicitly controlling the type I error rate while maximizing power, accommodating different types of endpoint, flexible number of interim looks, and stopping boundaries calculated before the start of the trial. Simulation studies show that the BOP2-FE design reduces the total sample size under the alternative hypothesis while strictly controlling the type I error rate and providing a similar statistical power to the original BOP2 design and a higher statistical power than another existing design.

2. **Title:** Probability of Success for Establishing Noninferiority Across Multiple Visits: Extension of Covariate-Adjusted Bayesian Hierarchical Modeling Framework
   Speaker: Yujie Zhao (Abbvie)

   Abstract: Decision-making is crucial throughout the drug development process, especially at the stage of completion of a proof-of-concept (POC) or Phase II study. In order to determine whether to move forward to a subsequent larger scale confirmatory phase III study, understanding that the uncertainty about the underlying treatment effect and the probability of success (POS) in the phase III study is of the utmost importance. In this research, we proposed and investigated a Bayesian covariate-adjusted hierarchical modeling approach leveraging historical data with longitudinal information to quantify the POS of the pivotal confirmatory trial. Although historical data borrowing methods are widely used and known for the advantages in alleviating recruitment and ethical challenges as well as improving trial efficiency, its integration with longitudinal information across multiple visits and prediction of future trial POS pose methodological challenges. This research not only provides a comprehensive modeling method, but also demonstrates how the proposed model can be used in a Go/No-Go decision-making framework via a glaucoma eye care program example. For the approval of new drugs targeting glaucoma, regulatory agencies typically require a pivotal phase III trial to demonstrate non-inferiority compared to active control. This involves meeting both statistical and clinical margins across multiple visits simultaneously. Simulations were performed to evaluate the key factors that affect the operating characteristics, such as between-trial heterogeneity, subject-level variance and between-visit correlation. The proposed decision-making framework can also be applied to studies in other therapeutical areas with similar settings.

3. **Title:** A constrained hierarchical Bayesian model considering latent biomarker subgroups for time-to-event endpoints in randomized phase II trials
   Speaker: Yifei Huang (Astellas)

Abstract: In randomized Phase III oncology trials, the long-term time-to-event endpoint is the most relevant outcome for participants and regulators. However, in Phase II trials, the short-term binary outcome of tumor response is often used as a surrogate endpoint to evaluate the treatment benefit. This may lead to a high failure rate in phase III trials, as the tumor response may not reflect the actual survival benefit. Moreover, many oncology trials collect biomarker data, especially those that may predict clinical outcomes and identify participants who are more likely to respond to the experimental treatment. Therefore, there is a growing need for a biomarker-based design to enrich the trial by selecting participants whose biomarker levels exceed certain thresholds. This paper proposes a constrained hierarchical Bayesian model that considers latent biomarker subgroups (CHBM-LS) for long-term time-to-event endpoints in Phase II randomized trials. CHBM-LS aggregates the biomarker populations into latent subgroups and accounts for the heterogeneity of treatment effects across biomarker levels in each model. We compare our proposed design with other approaches and show the benefits of CHBM-LS in improving the accuracy of hazard ratio estimates and increasing the power to detect true effects while maintaining control over the type I error rate.

Session 2: Statistical Considerations for Rare Disease and/or Cell & Gene Therapy Development

Time: 9:45 am -10:45 am, Oct 3, 2025

Organizer: Yusuke Yamaguchi (Astellas), Cong Han (Astellas)

Chair: Yusuke Yamaguchi (Astellas)

Session Abstract: Cell & gene therapy (CGT), which can consist of therapies made up of live cells or genetic material that are injected, implanted, or grafted, is a fast-growing and rapidly evolving field across multiple therapeutic areas. Rare or even ultra-rare diseases are often targeted by CGT products, although rare disease drug development also include other treatment modalities. CGT and rare disease drug development are uniquely challenging due to issues such as small patient population that restricts trial design options (e.g., randomized controlled trials may not always be feasible), and heterogeneous natural history that is not well understood. As such, traditional and well-

established trial designs may not be applicable for rare disease clinical investigations. The objective of this session is to discuss statistical considerations related to rare disease and/or CGT drug development, and presenters will share their efforts to overcome these challenges, which would encourage us to further implement innovative trial designs and more efficient statistical analyses.

1. **Title:** Challenges and considerations in dose finding for cell & gene therapy development
   Speaker: Frank Shen (Bristol Myers Squibb)

   Abstract: Dose finding is one of the most important tasks in clinical development, and cell therapy and gene therapy developments are no exception. Traditional methods to support dose finding for small molecules and immunotherapies still have applicability, but cell and gene therapy have many unique challenges for dose finding studies, such as manufacturing challenges, a limited number of doses that can be explored and small sample size dosage variability, unique PK or cellular kinetics, small sample size for rare genetic diseases and additional cofounding factors, and the pressures that come from single dose treatments that are common in the area. We review past development programs that have managed to overcome these issues and provide some advice on what new dose finding methodologies might best fit the unique challenges that go with cell and gene therapy.

2. **Title:** Bayesian quantitative decision-making for two endpoints in rare disease drug development
   Speaker: Haoran Hu (University of Pittsburgh)

   Abstract: A unified methodology to implement the Bayesian quantitative decision-making (QDM) has been established in the context of rare disease drug development, which aimed at building criteria for making definitive Go/No-Go decisions specifically in small-sized study with or without control data. We extend the Bayesian QDM methodologies to two endpoints, which is often encountered as a practical challenge but is not well-documented. A Go/No-Go decision making based on multiple endpoints may precede a pivotal study that could have a single primary endpoint, co-primary endpoints, or multiple primary

endpoints. A bivariate extension capturing the correlation structure between endpoints has a potential to enhance efficiency of the QDM, probably contributing to the reduction of probabilities of inconclusive decision. The present talk begins with a general description of the two-endpoint Bayesian QDM framework, and then proceeds to a detailed illustration of a specific setting.

3. **Title:** Optimizing gene therapy trial design in a rare neuropathy: A case study using a pharmacometrics-informed simulation framework
   Speaker: Alex Sverdlov (Novartis)

   Abstract: The development of disease-modifying gene therapies for rare neurological diseases (RNDs) is hampered by challenges such as small patient populations and disease heterogeneity. To address this, we present a pharmacometrics-informed clinical scenario evaluation framework (CSE-PMx) for systematically comparing and optimizing trial designs through in-silico simulation. We showcase the application of this framework to a hypothetical randomized trial of a one-time gene therapy for Autosomal-Recessive Spastic Ataxia of Charlevoix-Saguenay (ARSACS), leveraging a validated nonlinear disease progression model to generate realistic longitudinal data. Our simulation results demonstrate that a model-based analysis using a nonlinear mixed-effects model (NLMEM) provides substantially greater statistical power to detect a treatment effect compared to conventional analysis methods. This case study illustrates how the CSE-PMx framework can provide a quantitative, evidence-based rationale for selecting efficient trial designs, thereby de-risking and accelerating the clinical development of gene therapies for RNDs.

Session 3: Innovative Clinical Trial Designs

Time: 11:00 am -12:00 pm, Oct 3, 2025

Organizer: Qi Yan(AbbVie), Lei Shu (AbbVie)

Chair: Hao Wang (AbbVie)

Session Abstract: Innovative clinical trial designs are crucial because they increase efficiency, accelerate the delivery of new treatments,

address increasingly complex scientific questions in drug development and strengthen trial results, providing more robust data for regulatory and reimbursement decisions. These advanced designs leverage technologies and novel methodologies such as adaption and master protocols. This session covers the recent advances including seamless Phase 2/3 trials for oncology that integrate early dose selection and confirmatory endpoints while maintaining statistical rigor; adaptive design strategies that enable efficient evaluation of combination therapies; and evidence-based approaches for stratification and randomization in partial blinding platform trials.

1. **Title:** Group-Sequential Testing After Early-Endpoint-Based Dose Selection in Seamless Phase 2/3 Oncology Trials
   Speaker: Haiming Zhou (Daiichi Sankyo)

   Abstract: Oncology development is shifting from “more-is-better” dosing to optimization under FDA Project Optimus. Seamless Phase 2/3 designs are attractive for linking dose exploration with confirmatory evaluation, but when interim dose selection uses early endpoints while the primary endpoint is overall survival (OS), standard independent-increment group-sequential (GS) methods can lose validity and inflate Type I error. We present an error-controlled GS framework tailored to this setting. Stage 1 randomizes multiple doses versus control and selects a dose using early efficacy together with PK/PD and safety. Stage 2 enrolls a new cohort to compare the selected dose with control on OS, allowing multiple interim looks and a final analysis. To maintain strong family-wise error control, we combine (i) closed testing across elementary/intersection hypotheses, (ii) multiplicity adjustment (Simes or Dunnett), and (iii) inverse-normal p-value combination with pre-specified weights based on expected events. Using a disjoint-subjects (patient-wise separation) construction, we derive the joint multivariate normal distribution of the combined statistics across looks and apply α-spending to compute rejection boundaries. This preserves validity even if ORR and OS are perfectly correlated. Extensive simulations demonstrate reduced sample size and trial duration versus traditional operational designs, while maintaining strong Type I error control and high power.

2. **Title:** Statistical approaches to speed up development for combination therapies in oncology
   Speaker: Jun Zhao (Antengene)

   Abstract: Drug developers face significant challenges in developing combination therapies in oncology, which include the complexity of defining the optimal dose level and schedule of multiple agents, the demonstration of the contribution of each component, and the management of multi-dose multi-agent trials. These hurdles contribute to increased trial durations, added trial complexity, and enlarging trial sample sizes, which may hinder patients from accessing new treatments and increase development costs. Here we outline how advanced statistical methods can accelerate the development pathway for combination therapies in oncology by enhancing trial efficiency and using data more effectively. Multiple statistical methods, including innovative trial designs and advanced modeling techniques, can be considered: adaptive trial designs, which use accumulating trial data to make pre-specified, mid-trial modifications, can reduce the time and resources spent; trial designs incorporating real-world and historical data potentially reduce sample sizes; basket or umbrella trial designs can streamline drug development by evaluating multiple therapies or combinations simultaneously under a single protocol, which are particularly useful for biomarker-driven strategies on different tumor types or patient subsets; Bayesian methods or a Bayesian framework are particularly suitable for combination therapies in oncology due to their ability to leverage external information to make more efficient decisions. In summary, the application of advanced statistical methods in design and modeling is essential for the efficient and accelerated development of combination therapies in oncology. The approaches facilitate a more efficient process for identifying synergistic drug combinations, establishing optimal dosing strategy, and providing treatment efficacy, especially in biomarker-defined patient subgroups. These considerations ultimately can lead to faster delivery of more effective therapies to cancer patients.

3. **Title:** Two-Stage Randomization: Optimizing Stratification Strategies for Partial Blinding Platform Trials
   Speaker: Yuchen Yao (AbbVie)

Abstract: Partial blinding platform trials provide an efficient framework for the simultaneous assessment of multiple interventions within a single study while preserving blinding. Although platform trials offer the advantage of sharing concurrent controls across substudies, they also increase the complexity of balancing comparative populations and accommodating intervention-specific eligibility criteria. In therapeutic areas with established baseline prognostic factors, block randomization with stratification remains the standard practice to ensure population balance and valid comparative analysis. Nevertheless, practical recommendations for optimal stratification strategies in partial blinding platform trials remain limited and demand further investigation. We employed simulation to evaluate various stratification strategies for list-based block randomization approaches in a two-stage partial blinding platform trial with three concurrent substudies. Strategies examined included: (1) no stratification, (2) stratification in the first stage only, (3) stratification in the second stage only, and (4) stratification in both stages. These approaches were benchmarked against one-stage randomization methods, both with and without stratification. The simulation demonstrated that the two-stage randomization with stratification in both stages yielded superior balance across stratification factors and greater stability in treatment effect size estimation, achieving performance comparable to one-stage randomization with stratification. Based on these results, we recommend the routine adoption of two-stage randomization with stratification in both stages for partial blinding platform trials. This evidence-based strategy supports robust and reliable analyses and facilitates practical implementation for partial blinding platform trials.

Session 4: Recent Development in Predictive Modeling and Causal Inference

Time: 11:00 am -12:00 pm, Oct 3, 2025

Organizer: Lihui Zhao and Hui Zhang (Northwestern University)

Chair: Lihui Zhao (Northwestern University)

Session Abstract: In this session, three experts will present their recent work in predictive modeling and causal inference. Using the electronic health records (EHRs) from the INSIGHT Clinical Research

Network, the first speaker (Dr. Wenna Xi) will present predictive modeling for suicide attempts among Black children and youth presenting to the emergency room with mental health concerns. The second speaker (Dr. Sihai Dave Zhao) will present a computational method based on causal representation learning to infer how perturbations would affect spatial transcription patterns in tissue by leveraging single-cell Perturb-seq experiments. The third speaker (Dr. Ming Wang) generalizes the usual propensity score (PS)-based causal inference techniques focusing on time-invariant treatments to the longitudinal setting, and will present the dynamic propensity trajectory (DPT) framework and DPT-based matching (DPTM) techniques, which ensure both time-invariant and time-varying covariate balance up to treatment initiation.

1. **Title**: Predicting suicide attempts among Black children and youth presenting to the emergency room with mental health concerns: An EHR-based study
   Speaker: Wenna Xi (Northwestern University)

   Abstract:

   Objectives: Suicide among Black youth has become a public health concern. Many individuals who die by suicide had a history of suicide attempts (SAs); therefore, predicting future SAs within healthcare systems provides a valuable opportunity for early intervention. The goal of this study was to develop a machine learning model to predict SA among Black children and youth presenting to the emergency room (ER) with mental health (MH) concerns.

   Methods: This study was conducted at the encounter level using electronic health records (EHRs) from the INSIGHT Clinical Research Network, which includes data from 7 top academic medical centers in New York City. The study cohort included Black children and youth aged 10-24.9 years who presented to the ER with a MH-related diagnosis (n= 20,886). The outcome was time to SA, and potential predictors were demographics (age, sex, ethnicity, insurance type), diagnoses (grouped by Clinical Classifications Software Refined [CCSR] categories in the past year/month), and healthcare utilization (all-cause and MH-related, by setting [ED, IP], in the past year/month). The dataset was split into training and test sets in a 75:25 ratio. Random

survival forests and LASSO Cox regression models were trained using five-fold cross-validation on the training set, and the model performance was evaluated on the test set.

Results: The final model (LASSO) achieved a c-index of 0.82 on the test set. Top risk factors were sickle cell trait/anemia in the past year (HR=3.63), nausea and vomiting in the past year (HR=1.62), low back pain in the past year (HR=1.69), malaise and fatigue in the past month (HR=2.17), intent of injury, assault in the past year (HR=2.09), depressive disorders in the past year (HR=1.44), essential hypertension in the past year (HR=1.53), other specified and unspecified skin disorders in the past year (HR=1.59), suicidal ideation/attempt/intentional self-harm in the past year (HR=1.55), female sex (HR=1.33).

Conclusions: The model, developed using data exclusively from Black children and youth, identified well-known risk factors (e.g., depression, past suicidality, and female sex), as well as conditions more specific to the Black population (e.g., sickle cell trait/anemia) that have not been reported in models developed for the general population, suggesting that in addition to general risk factors, Black youth with certain health conditions require extra attention for suicide prevention.

2. **Title:** Causal effects of gene perturbations across intact tissue

   Speaker: Sihai Dave Zhao (University of Illinois Urbana-Champaign)

   Abstract: A critical task in preclinical drug development is to understand how genomic perturbations affect a tissue's gene expression pattern. Direct measurement is difficult, as genome-wide profiling technologies are typically destructive. Causal inference methods can be applied, but most large-scale perturbation experiments are performed on dissociated single cells rather than tissues because transcriptomics on intact tissue is still prohibitively expensive. We describe a computational method based on causal representation learning to infer how perturbations would affect spatial transcription patterns in tissue by leveraging single-cell Perturb-seq experiments.

3. **Title**: Dynamic Propensity Trajectory Modeling and Matching with Time-Dependent Covariates for Causal Inference
   Speaker: Ming Wang (Case Western Reserve University)

   Abstract: In observational studies, propensity score (PS)-based causal inference techniques are commonly used to address selection bias in treatment assignment. However, most existing PS research focuses on time-invariant treatments within a cross-sectional design, with limited attention given to propensity score methods in a longitudinal context, particularly when treatment initiation varies and the focus is on time-to-event outcomes. To fill this gap, we introduce the dynamic propensity trajectory (DPT) framework and DPT-based matching (DPTM) techniques, which ensure both time-invariant and time-varying covariate balance up to treatment initiation. In the post-matching analysis, we evaluate the causal effects of treatment on event outcomes following treatment initiation. Extensive simulation studies are conducted to assess the empirical performance of our methods compared to existing ones. Additionally, we apply the proposed methods to the Chronic Renal Insufficiency Cohort (CRIC) study to assess the effects of antihypertensive medications on reducing cardiovascular disease risk in patients with chronic kidney disease.

Session 5: Accelerating Drug Development Through Novel Platform Designs

Time: 1:30 pm – 2:30 pm, Oct 3, 2025

Organizer: Edy Yao (AbbVie)

Chair: Edy Yao (AbbVie)

Session Abstract: Platform trial designs allow for the simultaneous evaluation of multiple interventions within an overarching trial structure, which optimizes patient recruitment through shared control arm, enables adaptive modifications on the investigational arms based on interim results, and ultimately shortens time to market and reduce resources burden. This session will advocate for broader adoption of platform designs in both industry-led and academic research settings. Challenging topics in platform design such as optimization of patient allocation, integration of longitudinal biomarkers, and sample size re-estimation will be featured in this session. Attendees will gain insights

into the statistical considerations required to successfully implement these modern trial frameworks that are transforming the drug development landscape.

1. **Title:** Statistical Considerations for Evaluating Combination Therapies with Shared Comparator Arms in A Randomized Platform Trial
   Speaker: Yibo Wang (AbbVie)

   Abstract: Clinical trials in rare diseases often face significant challenges due to low prevalence, making it difficult to recruit sufficient patients for adequate statistical power. This constraint calls for innovative design strategies that can maximize the efficiency under limited total sample sizes. Platform trial designs offer a promising solution by evaluating multiple investigational drugs simultaneously against a shared reference arm, thereby reducing the overall sample size required. In this presentation, we investigate a platform trial framework under a fixed total sample size, reflecting practical constraints in rare disease research. A key practical consideration is the staggered availability of investigational drugs — not all investigational drugs may be ready to be evaluated concurrently, and some may be dropped early from the platform due to futility or external factors. To address these dynamics, we develop and evaluate design strategies that determine when to add or drop arms, and more critically, how to optimize patient allocation across treatment arms over time to maximize statistical power for treatment comparisons. Our findings provide practical guidance for designing more informative and resource-efficient platform trials under real-world operational limitations.

2. **Title:** Bayesian Biomarker-Assisted Platform Design for Dose Ranging in Multi-Agent Multi-Dose Trials
   Speaker: Ruitao Lin (MD Anderson Cancer Center, Virtual)

   Abstract: Assessing the long-term benefits of new treatments can be expensive and time-consuming, particularly in disease areas with unmet medical needs. While platform trials enable the evaluation of multiple interventions simultaneously, they currently cannot assess studies involving multiple agents and doses or utilize longitudinal biomarkers in decision-making.

We propose a Bayesian biomarker-assisted platform design that offers a unified framework for evaluating multiple investigational agents and their doses in a multi-stage, randomized controlled trial. The design streamlines the drug evaluation process and decreases development costs by including proof-of-concept, futility and superiority monitoring, and dose optimization in a single trial, while avoiding over-allocating patients to a shared placebo or active control arm. To facilitate making real-time interim group sequential decisions, temporarily unobserved long-term responses are estimated from longitudinal biomarker measurements. Design parameters and the maximum sample size are fine-tuned to achieve good frequentist properties. The proposed design is illustrated by a trial of three targeted agents for systemic lupus erythematosus (SLE), evaluated by their 24-week response rates. Extensive simulations show that the proposed design compares favorably to several conventional platform designs.

3. **Title:** Using Surrogate Information to Improve Confirmatory Platform Trial with Sample Size Re-estimation
   Speaker: Liwen Wu (Takeda, Virtual)

   Abstract: Platform design which allows simultaneous exploration of multiple arms with a common control is becoming essential for efficient drug development. However, one of the critical challenges for confirmatory platform trials is immature data for interim decisions, particularly for the treatment arm selection and sample size determination with limited data available. We use a modified conditional power (CP) for both treatment arm selection and sample size determination at interim analysis for the proposed platform trial. The modified CP uses the available data from both primary and surrogate endpoints. We also demonstrate the application in a case study of a lung cancer trial.

Session 6: Leveraging Large Language Models and Multimodal Machine Learning for Precision Medicine and Targeted Therapies

Time: 1:30 pm – 2:30 pm, Oct 3, 2025

Organizer: Miles Xi (AbbVie)

Chair: Miles Xi (AbbVie)

Session Abstract: Recent advances in computational methods are reshaping biomedical research and healthcare by enabling new strategies for data integration, disease detection, and model generalization. This session highlights three complementary perspectives: (1) using multimodal signals such as language, eye-tracking, and facial cues for early dementia detection; (2) applying adversarial regularization to mitigate site-specific effects and improve model transportability; and (3) leveraging large language models to enhance interoperability across fragmented health data systems. Collectively, these presentations showcase how modern analytic techniques can reduce barriers in data use, unlock richer indicators of disease, and deliver more reliable and generalizable results for biomedical research and clinical practice.

1. **Title:** Leverage LLMs to Enhance Health Data Interoperability
   Speaker: Dr. Yikuan Li (George Mason University)

   Abstract: Recent advances in large language models (LLMs) have opened new opportunities to overcome longstanding barriers in health data interoperability. By transforming unstructured clinical narratives into standardized formats like FHIR, LLMs can streamline the extraction, integration, and harmonization of real-world data across disparate systems. For the pharmaceutical industry, this shift enables more efficient and scalable access to high-quality, interoperable datasets critical for clinical trial optimization, drug safety surveillance, and real-world evidence generation. This talk will explore how LLMs are being leveraged to bridge data silos, accelerate data curation workflows, and enhance the value of both structured and unstructured health data in research settings.

2. **Title:** Adversarial Regularization Techniques for Effective Removal of Site-Specific Effects
   Speaker: Dr. Reuben Retnam (Takeda)

   Abstract: Classification models trained on multi-site patient data often suffer from site-specific biases due to differences in data collection or class imbalance across sites. Such biases can lead models to learn site-dependent artifacts, limiting generalizability to new populations. This talk presents an adversarial learning framework designed to mitigate site-specific effects while

maintaining strong predictive performance. Results demonstrate improved out-of-sample generalization across cohorts and highlight the flexibility of this framework across multiple model types. The approach offers a practical strategy for reducing bias and enhancing model robustness in biomedical research pipelines.

3. **Title:** Multimodal Signals for Dementia Detection: Language, Eye-Tracking, and Facial Cues
   Speaker: Dr. Hyeju Jang (Indiana University Indianapolis, Virtual)

   Abstract: This talk will present insights from recent research on detecting dementia using signals across three modalities: language, eye-tracking, and facial expressions. The key findings from the studies will highlight how each modality contributes to identifying cognitive decline and how combining them improves performance. The talk will focus on high-level takeaways, practical implications, and next steps toward building accessible, AI-powered screening tools.

Session 7: Integrating AI/ML in Clinical Drug Development and Other Fields

Time: 3:00 pm – 4:00 pm, Oct 3, 2025

Organizer: Jia Jia (AbbVie), Mandy Jin (AbbVie)

Chair: Hui Zheng (AbbVie)

Session Abstract: This session explores cutting-edge advances in integrating Artificial Intelligence (AI) and Machine Learning (ML) within clinical drug development and related fields. The first presentation demonstrates how Large Language Models and ML techniques are transforming causal inference and Bayesian decision-making in clinical trials by structuring and harmonizing diverse data sources to enhance statistical robustness and interim decision frameworks. The second presentation addresses the persistent challenge of missing data in longitudinal randomized clinical trials, detailing targeted learning strategies under MAR and MNAR assumptions to ensure reliable treatment effect estimation. The third presentation introduces a novel generative modeling approach for hypergraphs—Denoising Diffused Embeddings—which tackles high-dimensionality and interpretability

challenges, with demonstrated benefits in electronic health record analysis.

Collectively, the session illustrates the power and potential of advanced AI/ML methodologies in accelerating drug development, improving data reliability, and unlocking new insights in healthcare and beyond.

1. **Title:** Leveraging Large Language Models and Machine Learning to Strengthen Causal Inference and Bayesian Decision-Making in Clinical Trials
   Speaker: Yunxiao He (FocusKPI inc.)

   Abstract: Large language models (LLMs) offer new opportunities to strengthen causal inference and decision-making in clinical trials. By converting unstructured sources—such as physician notes, patient narratives, and imaging reports—into structured covariates and latent features (e.g., embeddings), LLMs enable more precise information extraction and enrichment. They also facilitate data harmonization by standardizing variable definitions and reconciling heterogeneous coding schemes across external datasets, and supporting covariate discovery and confounder identification through systematic review of literature, protocols, and observational data. Together, these capabilities reduce manual burden, increase comparability, and improve the robustness of statistical models for synthetic control and external data borrowing. When integrated into Bayesian go/no-go decision frameworks, LLM-augmented evidence contributes to stronger prior specification, more accurate likelihoods, and reduced posterior uncertainty, enabling more reliable interim decisions on efficacy or futility. Incorporating such enriched data inevitably increases model complexity, creating opportunities to leverage machine learning and deep learning methods for high-dimensional representation learning and scalable inference. The convergence of LLMs, Bayesian statistics, and modern ML approaches offers a pathway toward faster, more rigorous, and more generalizable evidence generation in clinical development. Some use cases will be used to illustrate the capabilities of the various techniques involved.

2. **Title:** Handling Missing Data in Longitudinal Randomized Clinical Trials Within the Framework of Targeted Learning Under MAR and MNAR
Speaker: Mandy Jin (AbbVie)

Abstract: Missing data is a common issue in longitudinal randomized clinical trials (RCTs), where the effect of treatment is estimated by a pre-specified statistical model. The Targeted Learning framework, through targeted maximum likelihood estimation (TMLE), harnesses the strengths of machine learning with Super Learner (SL) to reduce bias and improve efficiency in RCT settings while preserving valid causal inference. In this research project, we propose methods for handling missing data in longitudinal RCTs within the framework of targeted learning through longitudinal maximum likelihood estimation (LTMLE), with missing data mechanisms MAR and missing not at random (MNAR). The methods are applied to a real public RCT dataset.

3. **Title:** Denoising Diffused Embeddings: a Generative Approach for Hypergraphs
Speaker: Gongjun Xu (University of Michigan, Virtual)

Abstract: Hypergraph data, which capture multi-way interactions among entities, are becoming increasingly prevalent in the big data eta. Generating new hyperlinks from an observed, usually high-dimensional hypergraph is an important yet challenging task with diverse applications, such as electronic health record analysis and biological research. This task is fraught with several challenges. The discrete nature of hyperlinks renders many existing generative models inapplicable. Additionally, powerful machine learning-based generative models often operate as black boxes, providing limited interpretability. Key structural characteristics of hypergraphs, including node degree heterogeneity and hyperlink sparsity, further complicate the modeling process and must be carefully addressed. To tackle these challenges, we propose Denoising Diffused Embeddings (DDE), a general generative model architecture for hypergraphs. DDE exploits potential low-rank structures in high-dimensional hypergraphs and adopts the state-of-the-art diffusion model framework. Theoretically, we show that when true embeddings are accessible, DDE exactly reduces the task of generating new high-dimensional hyperlinks to generating new low-dimensional

embeddings. Moreover, we analyze the implications of using estimated embeddings in DDE, revealing how hypergraph properties--such as dimensionality, node degree heterogeneity, and hyperlink sparsity--impact its generative performance. Simulation studies demonstrate the superiority of DDE over existing methods, in terms of both computational efficiency and generative accuracy. Furthermore, an application to a symptom co-occurrence hypergraph derived from electronic medical records uncovers interesting findings and highlights the advantages of DDE.

Student Poster Session

Chak Kwong (Tommy) Cheng, University of Illinois Chicago
Weighted Bayesian Bootstrap for Function Regression

Chenze Li, Ohio State University
Transfer Learning for Random Forest: Statistical Error Bounds and an Application to Multi-Hospital EHR Data

Feiyang Deng, University of Michigan
Discrete Survival Knowledge Distillation for Competing Risks Analysis

Hengde Ouyang, University of Illinois Chicago
Mixed-Effects Location Scale (MELS) Models: An Application Using Skin Cancer Tissue Microarray (TMA) Data

Jiafeng Zhu, Northwestern University
Calibrating Single-Arm Studies via External Controls: Resolving Non-Identifiability to Estimate Treatment Effect Modifiers for Continuous Outcomes

Ke Xie, Washington University in Saint Louis
A Permutation-Based Pruning Method for Controlling Type I Error in the Virtual Twins Procedure

Lingxuan Kong, University of Michigan
Bayesian Variable Selection on Small Sample Trial Data via Adaptive Posterior Posterior-Informed Shrinkage Prior

Mengqi Lin, University of Michigan
Controlling the False Discovery Proportion in Observational Studies with Hidden Bias

Ngoc Duong, Northwestern University
Aligning Observational and RCT Data to Predict Long-term Survival Outcomes Using Stacked Survival Models

Nidhi Pai, University of Minnesota

Quantifying Physical Activity Intervention Effects via Functional Regression

Pengbo Wan, Michigan State University
Total Events Avoided via GPS Matching: Evaluating the Causal Impact of the EPA PM2.5 Standard on Infant Mortality

Priyanka Banik, Northern Illinois University
Trees vs. Splines: A Comparative Analysis of BART and BMARS

Sam Lam, University of Illinois Urbana-Champaign
Targeted Maximum Likelihood Estimation using Real-World Outcomes to Predict Combination Therapy Efficacy in Clinical Trials

Scott Zuo, Northwestern University
Clustering-Augmented IPW for Robust Causal Effect Estimation in Observational Studies

Souradipto Dastidar, University of Minnesota Twin Cities
Leveraging External Data for Equitable Opioid Clinical Trials: A Bayesian Optimal Enrichment Framework

Xingran Chen, University of Michigan
A Unified Framework for Inference with General Missingness Patterns and Machine Learning Imputation

Xinyu Li, University of Wisconsin-Madison
Split–Bootstrap Tests for Poisson Latent-Space Network Models

Yiming Shi, Washington University in Saint Louis
Iterative Reference Selection for Normalizing Compositional Microbiome Data

Yunyi Wang, University of Texas Health Science Center at Houston
Event Projection for Time-to-event Data Using Bayesian Method: A Comparative Study

Zhichen Xu, Washington University in Saint Louis
Subgroup Identification via Interaction Tree and Mixed Model for Repeated Measures with Application to Alzheimer's Disease