Invited Data Science Speakers

Panel Discussion: Executive Insights: AI/ML Transforming the Clinical Research Landscape

Dr. Rick Landin

Brief Bio: Dr. Richard Landin, President and CEO, has over 25 years of drug development experience ranging from preclinical to post-commercialization. His therapeutic areas of expertise include precision medicine, oncology, asthma and allergy, migraine, insomnia, depression, general anxiety disorder, multiple sclerosis, lupus, and rheumatoid arthritis. Dr. Landin has established a history of building, managing and leading biometrics organizations. In addition, he has served and led cross functional, multi-corporate teams, leading to regulatory submission and approval of multiple products. Dr. Landin holds a Ph.D. is Statistics from Texas A&M University. He has advanced statistical theory for the advancement of drug development, with successes including the following: development of the Theory of Selective Score Inflation to redesign depression studies, the creation of innovative endpoints for the measure of Sleep Maintenance, innovative data presentation of QTc data, and most recently serving as a catalyst in the development of a precision medicine toolbox based on advances in applied math/machine learning.

Steve Rosenberg

Steve's career in life science and healthcare spans more than 40 years. He has led the development and deployment of cloud-based solutions with a focus on driving more integrated approaches to patient-centric trial management. Today, he continues to leverage this deep experience to drive uMotif's hypergrowth.

Steve was most recently Senior Vice President and General Manager of Oracle Health Sciences. Before that, Steve was at the forefront of advancing the adoption of electronic data capture and was the visionary behind the integrated clinical technology suite introduced by Phase Forward which was acquired by Oracle in 2010.

Steve has been named twice to the prestigious PharmaVoice 100 list. He recently served on the Board of Anju Software and currently sits on Prix Galien's Digital Media and USA Digital Health Advisory boards. He also served a six-year term on CDISC's Board.

Steve and his wife Pam founded the non-profit The One By One Project to meet the urgent needs of people in greater Boston who would otherwise fall through the cracks. In his spare time, Steve enjoys summers on Cape Cod with his family.

Dr. Ming Tan

Brief Bio: Dr. Tan is a professor and the chair of the Department of Biostatistics, Bioinformatics, and Biomathematics at Georgetown University and its Lombardi Comprehensive Cancer Center with university tenure. His research focuses on innovative designs, robust causal inference methods for clinical trials and RWD, and statistical learning and ML/AI methods for discovering and evaluating drug combinations and targets. He has served on multiple government and industry panels, including the FDA Advisory Committee, Mock FDA ACs and DMCs. He came to Georgetown in 2012 after tenures at the University of Maryland School of Medicine and the Marlene and Stewart Greenebaum Comprehensive Cancer Center, St. Jude Children's Research

Hospital, and The Cleveland Clinic. He received his Ph.D. in Statistics from Purdue University, West Lafayette, Indiana.

DS session 2:
AI/Machine Learning-Driven Innovations for Smarter Clinical Trials

Dr. Brian Hobbs

Title: Applying ML to Discern Predictive from Prognostic Biomarkers and Simple Solutions for Design

Abstract: Distinguishing predictive biomarkers from prognostic biomarkers is crucial for advancing precision oncology medicine. However, traditional statistical methods often fail to account for complex interactions within the tumor microenvironment (TME) and patient heterogeneity, leading to suboptimal trial designs and biomarker inferences.

This presentation integrates machine learning (ML) and Bayesian approaches to address these challenges, drawing from recent studies on thoracic cancers, non-small cell lung cancer (NSCLC), and radiation-induced immunosuppression. A patient similarity-embedded Bayesian model, utilizing power priors and localized prognostic patterns, outperforms conventional regression and ML in identifying subgroups based on anticancer immunity markers in NSCLC, adjusted for clinical factors [1]. Extending this, survival ML models (e.g., random survival forests) delineate prognostic TME subtypes in NSCLC using PD-L1 expression and CD3 counts, revealing high-risk profiles like high PD-L1/low CD3 with reduced 5-year survival [2].

A Bayesian counterfactual ML framework identifies patient profiles where proton beam therapy mitigates lymphopenia compared to intensity-modulated radiation therapy in esophageal cancer [3]. These methods leverage embeddings, synthetic data augmentation, and causal inference to discern predictive effects while adjusting for prognostic confounders.

The presentation concludes with practical recommendations for biomarker-guided trial designs. Using optimal efficiency predictive probability monitoring, we propose randomized phase II designs (e.g., enrichment, stratified, or pooled control arms) that incorporate futility stopping for multiple biomarker subpopulations, reducing sample sizes by 23-64% compared to traditional single-arm trials while enhancing comparative efficacy insights [4].

References

[1] **Yu D, Huang M, Kane MJ, Hobbs BP**. A patient similarity-embedded Bayesian approach to prognostic biomarker inference with application to thoracic cancer immunity. J R Stat Soc Ser C Appl Stat. 2025;00:1-24. doi:10.1093/jrsssc/qlaf001

[2] **Yu D, Kane MJ, Koay EJ, Wistuba II, Hobbs BP**. Machine learning identifies prognostic subtypes of the tumor microenvironment of NSCLC. Sci Rep. 2024;14(1):15004. doi:10.1038/s41598-024-64977-7

[3] **Yu D, Kane MJ, Chen Y, Lin SH, Mohan R, Hobbs BP**. Bayesian Counterfactual Machine Learning Individualizes Radiation Modality Selection to Mitigate Immunosuppression. JCO Clin Cancer Inform. 2025;(In Press).

[4] **Zabor EC, Kaizer AM, Pennell NA, Hobbs BP**. Optimal predictive probability designs for randomized biomarker-guided oncology trials. Front Oncol. 2022;12:955056. doi:10.3389/fonc.2022.955056

Brief Bio: Dr. Hobbs completed a doctoral degree in biostatistics at the University of Minnesota and then joined The University of Texas MD Anderson to pursue a Postdoctoral Fellowship in 2010. In 2011 he joined the faculty at MD Anderson Cancer Center as a tenure-track Assistant Professor. In Houston, Dr. Hobbs was a faculty member at The University of Texas Graduate School of Biomedical Sciences and adjunct faculty member at Rice University in the Department of Statistics. He was promoted to Associate Professor with tenure in 2017, and then recruited to join Cleveland Clinic to establish a new Section of Cancer Biostatistics. Dr. Hobbs held academic appointments in Cleveland Clinic's Taussig Cancer Institute and Lerner Research Institute's Department of Quantitative Health Sciences and served as co-Director of the Biostatistics and Bioinformatics Core for the Case Comprehensive Cancer Center. Additionally, he lectured in Cleveland Clinic's Lerner College of Medicine. Dr. Hobbs re-joined The University of Texas System at Dell Medical School in August 2020 as a tenured Associate Professor. After over 15 years in academia, Dr. Hobbs has joined Telperian Inc in the role of Chief Scientific Officer in 2024.

Dr. Hobbs served as a regular reviewer for many journals, including Journal of Clinical Oncology, and has advised several companies. Industry leaders have adopted Dr. Hobbs's methodological innovations, including an oncology trial which was one of the first five proposals accepted into the FDA's Complex Innovative Design Program (CID). In 2016, Dr. Hobbs was selected by The University of Minnesota for the Emerging Leader Award, an honor bestowed on alumni on the basis of impactful contributions within 10 years of graduating from one of The School of Public Health's 20 programs. Recognized as an expert in clinical oncology research methodology, in 2017 Dr. Hobbs was invited to lead the publication of National Cancer Institute's Clinical Trials Design Task Force providing consensus recommendations for first-in-human cancer drug trials that use seamless designs. In 2020, he was invited to contribute to an article for Nature Reviews Clinical Oncology describing the current state of tumor agnostic trials. In 2021, Dr. Hobbs was invited to review the landscape of basket trials in the Journal of Clinical Oncology. His research to advance diagnostic and therapeutic strategies is described by over 150 academic publications (https://scholar.google.com/citations?hl=en&user=SaB00PgAAAAJ&view_op=list_works).

Dr. Michael Kane/Dr.Arvind Rao

Title: Integrating Transformer Models to Simplify Clinical Trial Design and Resource Efficiency

Abstract: FClinical trial design and execution have seen increasing sophistication starting with Bayesian adaptive designs in 1990 through new classes of "learn as you go" (LAGO) trial designs. However, these types of trials tend to require extra resources to implement and more enrollees to fit parameters while evaluating efficacy. In this talk we explore how transformer models, including foundational models provide opportunities for simplifying trial planning and design through the integration of information that could have otherwise evade trialists or need

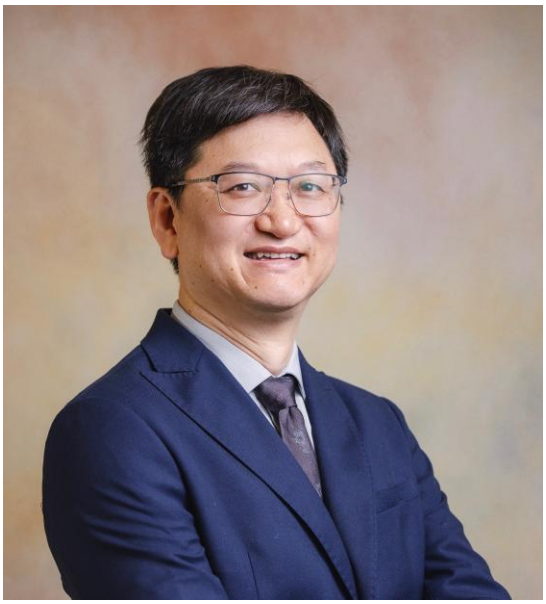
to be learned during the trial. These models, which can be used to extract and integrate information over a wide variety of data sources, provide the opportunity to better characterize patients, understand underlying mechanisms of disease, and create studies based on hypotheses of therapies functionalized by their mechanism of action. In this talk we provide two or three specific proof of concept use cases in the area of radiomics and hypothesis generation illustrating potential applications in trial planning and design.

Brief Bio:

Michael Kane is a Co-Founder and Advisor at Telperian Inc. as well as an Associate Professor at the University of Texas MD Anderson Cancer Center. His current work focuses on methods for integrating multimodal data in supervised learning problems and the incorporation of those results by foundational models for knowledge and hypothesis generation.

Arvind Rao is a Professor at the University of Michigan Ann Arbor, with prior roles at the University of Texas MD Anderson Cancer Center and Carnegie Mellon University. His research focuses on cancer bioinformatics, AI applications in healthcare including radiology and pathology, and spatial biology.

Dr.Cheng Su

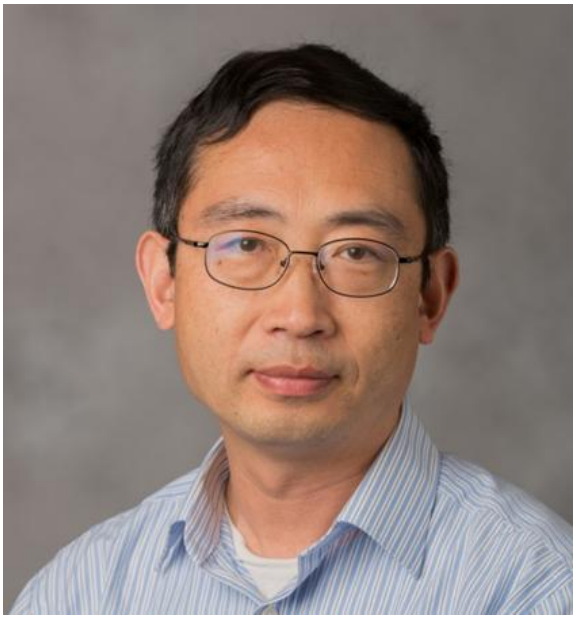


Title: Unlocking AI Potential in RBQM: Today and Tomorrow

Abstract: Risk-Based Quality Management (RBQM) is central to ensuring efficient, high-quality clinical trials, yet many RBQM activities remain manual, fragmented, and resource-intensive. This presentation explores how emerging capabilities in Generative Artificial Intelligence (GAI) and agentic frameworks could shape the future of RBQM.

Brief Bio: Dr. Cheng Su is an Executive Director of Data Science and Head of Data Quality & Innovation at BioMarin Pharmaceutical. Cheng is passionate about applying statistical thinking and advanced analytics to improve drug research and development. He has extensive expertise in risk-based quality management (RBQM), development of analysis systems, predictive modeling, preclinical research, genomics, and biomarkers. In his current role, Cheng leads AI-driven initiatives to transform clinical development, including solutions for risk assessment, protocol deviation detection, and clinical study build.

Dr. Shining Deng



Title: Advancing Tumor Response Assessment: Integrating AI-Assisted RECIST in Clinical Trials

Abstract: For more than two decades, RECIST has been the gold standard for objectively assessing treatment response in solid tumors, relying on manual measurements of tumor size from radiographic images. While widely adopted, this process can be time-consuming and subject to reader variability. Recent advances in artificial intelligence (AI) present an opportunity to transform this paradigm by automating lesion measurements and enabling comprehensive evaluation across the entire body—not just target lesions. AI-assisted RECIST promises faster, more consistent, and potentially more informative assessments, which could improve decision-making in clinical development. Our team has been actively investigating AI-based radiographic approaches, including AI-assisted RECIST, in multiple clinical trials spanning early- and late-stage oncology programs. In this presentation, we will share practical experiences, methodological challenges, and key statistical considerations encountered when integrating AI-assisted RECIST into clinical studies, as well as discuss implications for regulatory acceptance and future applications in oncology drug development.

Brief Bio: Shibing Deng is a biomarker statistician at Pfizer, La Jolla, CA. He leads a group of statisticians to support translational biomarker research and development in oncology. Shibing joined Pfizer after receiving his PhD in biostatistics from UNC Chapel Hill in 2004. Over the years, Shibing has provided statistical support for preclinical research, early and late phase clinical development at Pfizer. Prior to joining Pfizer, he worked as a statistician at UNC Chapel Hill, SAS Institute and Pharsight Corp in Cary NC. Shibing has over 100 peer reviewed publications and his research interests include statistical methods in genomics, preclinical and translational biomarkers.

DSA session 3:
Technology applications from vendors implementing AI solutions

Matt Purri, PhD

Brief Bio: *Dr. Matthew Purri is the co-founder and Chief Technology Officer of Octozi. He earned his PhD in Artificial Intelligence from Rutgers University, specializing in computer vision and machine learning. During his studies, he held research roles at Waymo, Amazon, and Clarifai, and authored multiple papers presented at top AI conferences including CVPR, ECCV, and NeurIPS. He later developed advanced natural language processing models at Altana AI. As an undergraduate, he founded an autonomous drone startup aimed at improving pesticide dispersion to combat the Zika virus.*

Title of Talk: ***Octozi: Agentic AI for More Efficient Medical Data Review***

Abstract: Clinical trial data cleaning represents a critical bottleneck in drug development, with manual review processes struggling to manage exponentially increasing data volumes and complexity. This paper presents Octozi, an artificial intelligence-assisted platform that combines large language models with domain-specific heuristics to transform medical data review. In a controlled experimental study with experienced medical reviewers (n=10), we demonstrate that AI assistance increased data cleaning throughput by 6.03-fold while simultaneously decreasing cleaning errors from 54.67% to 8.48% (a 6.44-fold improvement). Crucially, the system reduced false positive queries by 15.48-fold, minimizing unnecessary site burden. Economic analysis of a representative Phase III oncology trial reveals potential

cost savings of $5.1 million, primarily driven by accelerated database lock timelines (5-day reduction saving $4.4M), improved medical review efficiency ($420K savings), and reduced query management burden ($288K savings). These improvements were consistent across reviewers regardless of experience level, suggesting broad applicability. Our findings indicate that AI-assisted approaches can address fundamental inefficiencies in clinical trial operations, potentially accelerating drug development timelines such as database lock by 33% while maintaining regulatory compliance and significantly reducing operational costs. This work establishes a framework for integrating AI into safety-critical clinical workflows and demonstrates the transformative potential of human-AI collaboration in pharmaceutical clinical trials.

Henry Liu, Ph.D

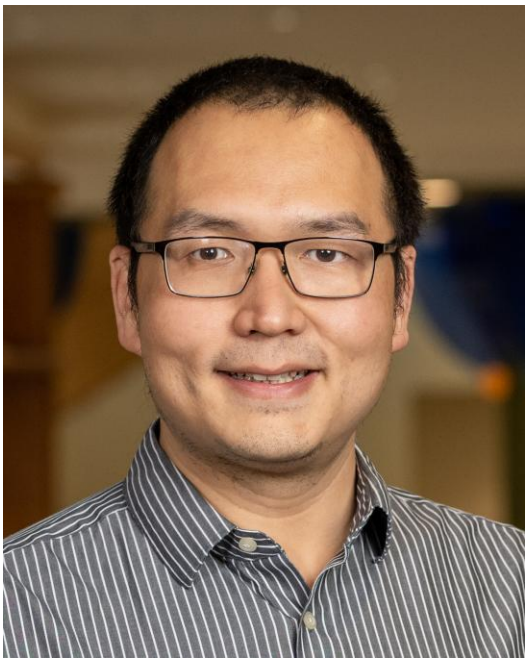
Brief Bio:
Dr. Liu is the CEO of Janus Data Intelligence, an AI solutions company under Q2BI Corporation. Q2BI specializes in delivering biometrics and AI-driven solutions for clinical trials, serving leading pharmaceutical companies worldwide. He holds a Ph.D. from Peking University and brings over 12 years of experience in AI and clinical trial data science. At Q2BI, he leads the development of AI-powered solutions that are transforming clinical trial data processing and analysis.

Title: AI Agents in Clinical Trial Data Processing, Current and Future

Abstract: Clinical trial data processing is facing a growing mismatch. Data volumes and complexity keep rising, while quality expectations, traceability requirements, and delivery timelines remain uncompromising. This talk explores how agentic AI built on large language models can move biometrics workflows beyond simple question answering and toward structured task execution across tools and data. Using concrete application patterns in clinical trial settings, we discuss interactive data exploration, automated programming for analysis outputs, contextual linking across domains such as safety events and concomitant medications, and evidence retrieval that supports interpretation and review. We will also examine the hard boundary of today's agents, including why they can complete independent tasks but still struggle to reach production grade correctness without validation and human oversight. Finally, we introduce the Agent Skills approach for step by step context loading and auditable workflows, and we close with a practical framework for human AI collaboration that separates what can be delegated now, what may become reliable next, and what expert judgment is likely to remain uniquely human in safety critical work.

DSA session 4:
From Complexity to Clarity: AI methods for high-dimensional data

Dr Jian Kang



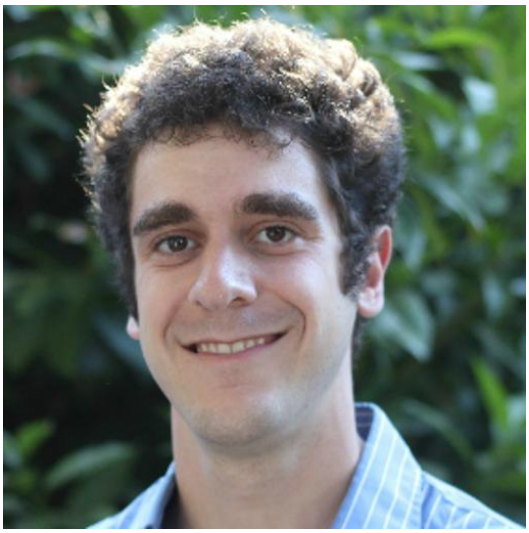
Title: Split Conformal Prediction for Uncertainty Quantification in Brain-Computer Interfaces

Abstract: Brain-computer interfaces (BCIs) convert noisy and complex neural signals into digital commands, enabling hands-free device interaction across assistive, clinical, and consumer applications. We propose an efficient uncertainty quantification method for EEG BCI systems, focusing on the widely studied P300 speller, using split conformal prediction. Our framework constructs confidence sets at multiple levels (stimulus, half-sequence, and character), enabling reliable early stopping and abstention in online free-typing tasks. It adapts to varying signal-to-noise ratio (SNR) across participants while maintaining guaranteed prediction accuracy. Extensive simulations and large-scale EEG datasets validate the method, demonstrating consistent improvements in system robustness and BCI-utility. This work contributes to

developing fast, accurate, and reliable BCI systems. This is joint work with Bangyao Zhao, Yixin Wang and Jane Huggins.

Brief Bio: Dr. Jian Kang is a Professor and Associate Chair for research in the Department of Biostatistics at the University of Michigan, Ann Arbor. His main research interests are developing statistical methods and theory for large-scale complex biomedical data analysis with focuses on Bayesian methods, statistical artificial intelligence and machine learning with applications to imaging and precision medicine.

Dr Sam Gross



Title: Training cancer classifiers on high dimensional data with reliable performance estimates

Abstract: The Machine Learning team at GRAIL conducts research on how to train cancer classifiers like the one that powers the Galleri™ multi-cancer early detection test. These classifiers take in high dimensional input data reflecting the methylation states observed on cell free DNA (cfDNA) and use that to predict cancer status. In this talk, Samuel will discuss some of the computational and modeling considerations that allows GRAIL to train such models in a way that produces reliable estimates of the performance those models could achieve in a real world application.

Brief Bio: Samuel Mesirov (né Gross) is the Director of Machine Learning at GRAIL. He has a PhD in Statistics from Stanford University where he studied data fusion in high dimensional modeling problems. In addition to his academic background, Samuel has over a decade of experience in industry building and leading teams devoted to solving impactful modeling problems in healthcare and medicine.

Dr Jeff Palmer

Title: Role of statistics in driving translational development using diverse data sources (wearables, omics, labs)

Abstract: This talk will focus on the role of statistics and data science in translational clinical development and Model-Informed Drug Development (MIDD). MIDD is now widely accepted as an integral part of clinical development, with the majority of global regulatory authorities recognizing its importance through guidance documents and publications.The Quantitative Medicine Center of Excellence (QM CoE) within FDA (CDER) has identified key areas of active MIDD interest including, but not limited to, biomarker endpoint development, clinical trial simulation, and AI/ML applications.I will cover novel statistical approaches for biomarker analyses, with an emphasis on high-dimensional data (transcriptomics, proteomics, etc). I will also cover current and future AI/ML applications in MIDD.

Brief Bio: Jeff Palmer is currently Executive Director and head of Translational Clinical Sciences Statistics at Pfizer, leading a team of statisticians, bioinformaticians and computational biologists driving statistical strategy and innovation in translational clinical development across the full portfolio. He also co-chairs the enterprise-wide Model-Informed Drug Development (MIDD) committee and community, advancing model-based techniques to ensure robust quantitative decision making across the development portfolio. In a previous role at Pfizer, he was a statistics group head leading early clinical development in rare diseases. Jeff and his team provided the statistical strategy and support for all pre-clinical and early rare disease clinical development programs, spanning 30+ indications across metabolic, hematology, cardiology, neurology and endocrine, and across multiple therapeutic modalities. Prior to Pfizer he had worked for over ten years with various other pharma and consulting companies supporting mainly rare diseases, oncology, and neurology. Jeff is active in the local statistics community, serving as an elected councilmember of the New England Statistical Society (NESS), and serving on the steering committee of the New England Rare Disease Statistics (NERDS) conference. Jeff received his MS in statistics from the University of Chicago and conducted his doctoral research in statistics at Carnegie Mellon University.

Dr Gregory Alexander

Title: Statistical considerations for generating confirmatory evidence from high dimensional data

Brief Bio: Gregory Alexander, PhD is the Director for Biostatistics Programs within the FDA/CDRH. Dr. Alexander's current work in medical device regulation is preceded by over 25 years working in industry, applying statistical methodology for translational research, biomarker discovery, development of invitro assays and predictive algorithms. At GRAIL, LLC, a subsidiary of Illumina, his work focused on the development and validation of multi-cancer early detection diagnostic tests. Prior to that, at Genomic Health (acquired by Exact Sciences) his work included the development and validation of several LDT's which employed technologies such as multiplex PCR and IHC to aid disease prognosis, as well as NGS for molecular profiling of tumors from blood samples. At CareDx (formally XDx), his work centered on development of multi-gene expression assays used in non-invasive monitoring for acute rejection of organ transplants. At Cytokinetics, Inc., Dr. Alexander developed algorithms used in high-content image-based screening of small molecules for the potential treatment of cancer, cardiomyopathies, fungal and musculoskeletal diseases.

DSA session 5:
Mobile Health in Clinical Trials: Statistical, Regulatory, and Implementation Challenges

Wessam Sonbol, CEO Delve Health

Integrating Wearable Technologies into Clinical Trials: Best Practices for Regulatory and Statistical Considerations

Wearable technologies are increasingly incorporated into clinical trials to capture continuous, real-world physiological data that extend beyond episodic clinic-based assessments. While these data streams offer new opportunities to characterize pharmacodynamics, functional outcomes, and safety signals, they also introduce statistical and regulatory challenges that must be addressed to meet U.S. Food and Drug Administration (FDA) expectations for endpoint reliability, interpretability, and auditability.

This presentation focuses on best practices for integrating wearable-derived measures into clinical trials in a manner consistent with FDA guidance on digital health technologies (DHTs), estimand framework principles (ICH E9 [R1]), and expectations for data integrity. We will discuss statistical considerations for defining fit-for-purpose digital endpoints, including pre-specification of estimands, alignment of wearable-derived variables with meaningful clinical concepts, and strategies for handling missingness and noncompliance inherent to passive data collection.

Key analytical challenges will be explored, including transformation of high-frequency sensor data into analysis-ready endpoints, management of within-subject correlation, and control of multiplicity when wearable-derived measures are evaluated alongside traditional clinical endpoints. Case examples will illustrate how wearables can support characterization of pharmacodynamic effects—such as changes in activity, heart rate variability, or sleep metrics—as well as early detection of safety signals, including cardiopulmonary or autonomic changes, in both therapeutic and device studies.

Regulatory considerations will be emphasized throughout, including FDA expectations for endpoint justification, traceability from raw sensor data to derived variables, and documentation required to support reproducibility and regulatory review. The session will also address how wearable data may contribute to assessments of treatment burden and quality of care, particularly when used as supportive or exploratory endpoints in confirmatory trials.

This presentation provides biostatisticians with a practical framework for designing, analyzing, and defending wearable-derived data in clinical trials, with an emphasis on statistical rigor, regulatory alignment, and readiness for FDA review.