2026 UMBC-Stanford Workshop on Clinical Trials and Regulatory Science

From Statistical Rigor to AI-enabled Decision-making

University of Maryland Baltimore County (UMBC)
September 18-19, 2026

UMBC, (Sat at ITE Building,104, and Fri at ILSB 223)
1000 Hilltop Circle, Baltimore, MD 21250

Objectives

The series of the *UMBC-Stanford Workshop on Clinical Trial and Regulatory Science* aims to bring together regulators, academic researchers, and industry professionals to discuss prominent issues of common interest, to share and exchange information and experiences, and to improve close collaborations to promote biomedical innovation in modern regulatory science.

Target audience

Regulatory scientists, statisticians, health policy professionals, clinicians, biopharmaceutical product development professionals, and other relevant research scientists.

Organizing committees

Faculty from Stanford University and UMBC, leaders from industry.

Sponsors & Supporting Organizations

- UMBC, Dean's Office of College of Natural and Mathematical Sciences, and Department of Mathematics and Statistics
- Stanford University, Center for Innovative Study Design.
- International Society for Biopharmaceutical Statistics (ISBS)
- Merck & Co. Inc. (If the Merck conference grant application gets approved by Sep)
- M-CERSI, University of Maryland Center for Excellence of Regulatory Science and Innovation.

Registration

Conference website: https://www2.umbc.edu/circ/hosting/1st_draft/index.html. Online registration is required for everyone. Regular registration for **Sat**. workshop: $200 (before or on Aug.30), $300 (after Aug. 30); government employees and academia: $75. For short course at **Friday** afternoon, registration is $100. For all **graduate students** in statistics/biostatistics with valid ID, the registration fee is **FREE** for Sat workshop and Fri short course **until Aug. 30**. The free student registrations need to be done early for food and drink preparations. After Aug. 30, student registrations are increased to $40. Dual registration for workshop and short course: $250 (before or on Aug. 30), $350 (after Aug. 30). Credit card is acceptable online, and check or cash payment is acceptable onsite (Sep.18-19).

Lodging

When making reservation, please use the booking link provided under conference website: https://www.hilton.com/en/book/reservation/deeplink/?ctyhocn=BWIBADT&departuredate=2026-12-31&corporateCode=2656730

Double Tree BWI - (Hilton hotel chain)

890 Elkridge Landing Road
Linthicum, MD 21090
1-800-810-0271
Free airport shuttle to BWI: Yes

Free WIFI: Yes
Breakfast: For purchase
Distance to UMBC campus: 8 miles
UMBC rate (standard room): $99 per night

Parking information (TBD)

For more Information

Please contact Dr. Yi Huang, Department of Mathematics and Statistics, University of Maryland Baltimore County (UMBC), 1000 Hilltop Circle, Baltimore, MD 21250. E-Mail: yihuang@umbc.edu.

2026 UMBC-Stanford Workshop on Clinical Trials and Regulatory Science

– From Statistical Rigor to AI-enabled Decision-making

Scientific Program

Friday Sept. 18 Program

Registration: ILSB 223 in Interdisciplinary Life Science Building, open after 1:30pm

Welcome Speech by (Department Chair, Dr. Andrei Draganescu)
2:00 pm - 2:05 pm, ILSB 223

Lai Memorial Short Course

Causal Generalist Medical AI

Instructor: Dr. Hongtu Zhu,
Kenan Distinguished Professor at the University of North Carolina at Chapel Hill,
JASA Editor, the COPSS 2025 Snedecor Award winner.
Moderator: Dr. Tian Lu, Stanford University
Venue: ILSB 223
Time: 2:10 pm – 3:30 pm, Short course session I
3:30 pm – 3:40 pm, Coffee Break
3:40 pm – 5:00 pm, Short course session II

Conference Mixer
ILSB 223, 5:00 pm – 6:00pm,
Networking and conference mixer with catered food.
Highly encourage students and junior researchers to participate.

Evening Banquet (invitation only, 6:30pm – 9:00pm)

Saturday Sept. 19 Program

(ITE Building 104/ 102, Lecture Halls)

Poster Session

Poster: Lobby of ITE building, open: 9:00am – 5:00pm
(For poster presenters, please setup your poster during 8:00am – 9:00am at Sat, Sept. 19)
(We will try to provide poster stand and board.)

Saturday Sept. 19 Program

(ITE Building 104/102, Lecture Halls)

Registration: Lobby of ITE building, open after 8:00 am

8:00 am – 8:30 am: Continental Breakfast

8:30 am – 9:00 am: Opening and Welcome Remarks (ITE 104)

Chair: Dr. Yi Huang, UMBC

- Provost or President, UMBC (10min) – Karl stainer, or provost or president.
- Dr. Ying Lu (5min): Dr. T. L. Lai's lifetime achievement, Inaugural Lai Memorial Short Course series
- Dr. Jie Chen (5min): Dr. Joe Heyse's lifetime achievement, Inaugural Joe Heyse Memorial Short Course series
- Dr. Yi Huang: UMBC-Stanford Workshop history and traditions, conference notice and announcement, two memorial short course series in alternating years, and two student poster awards named after them.

9:00 am – 9:50 am: Keynote Speech (ITE 104)

Chair: Dr. Yi Huang (or someone from FDA), UMBC

Evolving Role of Statisticians in Drug Development in the Artificial Intelligence and Digital Health Era

Dr. Mark Levenson, Director, Office of Biostatistics, Center for Drug Evaluation and Research, FDA

9:50 am – 10:10 am: Coffee Break

10:10 am – 11:00 am: Keynote Speech (ITE 104)

Chair: Dr. Jie Chen, Taimei Intelligence Biopharma R&D

Topic/ title (TBD)

Dr. Lisa C. Lupinacci, SVP, Merck Research Laboratories

11:00 am – 12:00 pm: Panel discussion (ITE 104)

Chair: Drs. Jie Chen and Tian Lu, Taimei Intelligence Biopharma R&D and Stanford University

AI/ML: From Statistical Rigor to AI-enabled Decision-making

Panelists: (approximate 6 panelists)

Dr. Lisa C. Lupinacci, Merck Research Laboratories
Dr. Mark Levenson, FDA
Dr. Hongtu Zhu, UNC
Dr. Mike Baiocchi, Stanford University
One or two more panelists here

12:00 am – 12:20 am: Conference Group Photo

12:20 pm – 1:40 pm: Lunch Break

Catered lunch at the lobby of the ITE building

1:40 pm – 2:40 pm: Keynote Speech (ITE 104)

Chair: Dr. Ying Lu, Stanford University

Title: What are we doing with synthetic data?

Dr. Mike Baiocchi, Stanford University.

2:40 – 4:10 pm Invite Session

Session 1 – Room ITE 104
Title: AI Models and Tools for Clinical Trials
Organizer: Dr. Ying Lu, Stanford University
Chair: UMBC colleague

Title : Evidence-based Clinical Trial Design: When the Power of AI Meets Statistical Rigor
Dr. Will Ma, President and CEO, HopeAI

Title : Navigating the AI Journey: Practical Steps for Pharma's Clinical Research Transformation
Dr. Gary Ma, Taimei Technology

Title : Unifying Clinical Trials, Literature, and Real-World Data: An AI Platform for Clinical Evidence Synthesis and Management
Dr. Hui Wang, Founder of SageAI

Title : The AI Revolution: Bubble or Breakthrough for Drug Development (not Discovery)?
Yuan Ji, CSO & Co-founder of Bayesoft Inc

Session 2 -- Room 102
Title: Advanced methods for Clinical Trial design and Data synthesis
Organizer: Dr. Yi Huang, UMBC
Chair: UMBC faculty

Title : TBD
Dr. Willam Wang, Merck Research Laboratories

Title : Improving power and early decision-making in clinical trial information-adaptive design through integrated covariate and auxiliary outcome information.
Dr. Chixiang Chen, School of Medicine, UMB

Title : TBD
Dr. Martin Kline, CDER, FDA

Title : One-class support vector machines integrated Bayesian approaches for external control borrowing in clinical trials
Dr. Yi Huang, UMBC

4:10 – 4:30 pm Coffee Break

4:30 – 6:00 pm Invited Session

Session 3 – Room ITE 104
Memorial Session for Dr. Tze Lai
Organizers: Dr. Ying Lu, Stanford University
Chair: Dr. Zhaohai Li, George Washington U

Title : Adaptive Design, Stochastic Approximation, and Stochastic Optimization
Professor Ying Kuen Cheung, Columbia University

Title : Challenge in the analysis of large scale data
Professor Zhezhen Jin, Columbia University

Title : TBD
Professor Feifang Hu, George Washington University

Title : Matching Hearts, Improving Lives: Interpretable Modeling, Simulation-Based Heart Transplant Matching, and New Tools for Informative Missingness and Evolving Clinical Data with Shifts
Professor Jiayang Sun, George Mason University

Session 4 – Room 102
Innovative statistical methodologies to solve complex clinical problems
Organizers: Dr. Ling Shen, CareDx Inc.
Chair: Dr. Ling Shen, CareDx Inc

Title : Bias reduction in g-computation for covariate adjustment in randomized clinical trials
Dr. Haitao Chu, Pfizer

Title : Multi-state model framework to characterize patient journey and time-dependent covariates
Dr. Jing Shi, CareDx Inc

Title : Inverse probability weighting to adjust diagnostic performance from observational data
Ms. Sijia Wang, CareDx Inc

Title : Inverse Probability of Treatment Weighting: A Simple and Effective Approach to Covariate Adjustment for Survival Endpoints in Randomized Clinical Trials
Dr. Zhiwei Zhang, Gilead Sciences

Conference Reception
Lobby of ITE building, 6:00 – 7:00 pm

Closing Remarks from Dr. Yi Huang and Special singer - Mrs. Letitia Lai. (10min)

Abstracts

(Abstracts are listed in the alphabetic order of speaker's last name and first name.)

Title: What are we doing with synthetic data?

- **Mike Baiocchi, PhD, Stanford University**

The generation of realistic-looking data enjoying similar statistical properties to some target real-world distribution is known as *synthetic data generation*. To some data scientists, the augmentation of datasets using synthetically generated data is an alluring proposition: in the best case, it generates realistic data *in silico* at a fraction of the cost of authentic data which may be found *in vivo* or *in vitro*. To traditional (bio)statisticians, synthetic data raises serious skepticism. In this talk, we clarify the types, origins, and potential use cases of synthetic data. We also make it clear how limited its applications are, and why "traditional" (bio)statisticians tend to not be as excited about it. Further, we will offer an empirical method for determining if synthetic data are useful in a given application.

Title: Adaptive Design, Stochastic Approximation, and Stochastic Optimization
Ying Kuen Cheung, PhD., Columbia University

In their seminal 1979 Annals paper, Lai and Robbins established the equivalence between a parametric adaptive design driven by maximum likelihood estimation and a nonparametric stochastic approximation scheme. This elegant result has motivated a rich literature on recursive maximum likelihood estimation. It also offers a principled lens for model-based designs in dose-finding clinical trials — where the search for a target dose is a one-dimensional root-finding problem, exactly the setting stochastic approximation was built for. In this talk, we revisit the Lai–Robbins equivalence and trace its relevance to modern trial design. Viewing model-based methods such as the continual reassessment method through the stochastic approximation lens clarifies why they work — and when they may fail to converge to the intended target. We then turn to drug combinations, where the target dose becomes a contour rather than a point. In this context, we will discuss how Lai's work on stochastic optimization motivates further development of Bayesian decision-theoretic designs.

Title: Bias reduction in g-computation for covariate adjustment in randomized clinical trials

- **Haitao Chu, MD, PhD, Senior Director, Pfizer**

G-computation is a powerful covariate adjustment method for analyzing randomized clinical trials. It typically relies on fitting canonical generalized linear models. However, this could be problematic when the sample size or event number is small relative to the number of covariates. Common issues include the underestimation of the variance and the potential nonexistence of maximum likelihood estimators. Bias reduction methods are commonly employed to address these issues, including Firth correction, which guarantees the existence of corresponding estimates. Yet, their application within g-computation remains underexplored. In this article, we analyze the asymptotic bias of g-computation estimators and propose a novel bias-reduction method that improves both estimation and inference. Our approach performs bias correction via generalized Oaxaca-Blinder estimators, and thus the resulting estimators are guaranteed to be bounded. The proposed debiased estimators use slightly modified versions of maximum likelihood or Firth correction estimators for nuisance parameters. We also introduce a simple small-sample bias adjustment for variance estimation to improve finite-sample inference validity. Through extensive

simulations, we demonstrate that our method offers superior finite-sample performance, effectively addressing the bias-efficiency tradeoff. Finally, we illustrate its practical utility by reanalyzing a completed randomized clinical trial. In this example, our method improves precision in a small subgroup analysis for which the standard method fails to fit the regression model. This is a joint work with Drs. Xin Zhang and Lin Liu.

Title: The AI Revolution: Bubble or Breakthrough for Drug Development (not Discovery)?

- **Yuan Ji, CSO & Co-founder of Bayesoft Inc**

AI is transforming drug development. At Bayesoft, we have seen substantial gains in efficiency across statistical programming, evidence synthesis, scientific reasoning, trial design generation, and code development. But the same technology that makes AI powerful also creates its greatest weakness: generative AI is inherently probabilistic. It can produce different answers to the same question, and sometimes confidently generate outputs that are wrong.vThat is a manageable risk in brainstorming. It is a serious problem in regulated drug development, where analyses, decision rules, and computational results must be reproducible, traceable, and exact. The solution is not to make everything AI-driven. It is to use probabilistic AI where intelligence, synthesis, and creativity add value—and enforce deterministic computation wherever uncertainty is unacceptable. The real breakthrough will come from systems that know the difference. Whether today's AI boom becomes a bubble or a lasting revolution may ultimately depend on our ability to harness its intelligence without surrendering scientific and computational rigor. And this would require human input and governance, which in term demands society to continue foster human learning and education.

Title: Challenge in the analysis of large scale data

- **Zhezhen Jin, PhD, Columbia University**

Analysis of large-scale data is challenging due to data storage and computational complexity. When analyzing large-scale data, subsampling methods and divide-and-conquer procedures are appealing, because they ease the computational burden, while preserving the validity of inferences. In this talk, several challenges and issues will be reviewed and a perturbation subsampling approach will be presented based on independent and identically distributed stochastic weights for analyzing large-scale data. The method can be justified based on optimizing convex objective functions by establishing the asymptotic consistency and normality of the resulting estimators. The method simultaneously provides consistent point and variance estimators. We demonstrate the finite-sample performance of the proposed method using simulation studies and a real-data analysis.

Title: Evolving Role of Statisticians in Drug Development in the Artificial Intelligence and Digital Health Eras

- **Mark Levenson, PhD, Center for Drug Evaluation and Research, FDA**

Artificial intelligence (AI) and machine learning (ML) are reshaping drug development and regulatory science. This presentation examines the evolving role of statisticians across the pharmaceutical product life cycle. The presentation discusses challenges to the application of AI, which statisticians are uniquely qualified to address. The presentation addresses evaluation of AI tools for regulatory decision-making through a risk-based credibility assessment framework, illustrated with case examples. Emerging intersections between AI, digital health technologies, decentralized clinical trials, and real-world evidence are also explored.

Title: Navigating the AI Journey: Practical Steps for Pharma's Clinical Research Transforma-tion

- **Gary Ma, PhD, Taimei Technology**

The integration of AI into clinical research is no longer a distant vision but an operational imperative for pharmaceutical companies. However, the path to successful transformation is often fraught with challenges. This presentation offers a pragmatic guide for navigating this journey. We will break down the essential steps, from building a robust data foundation and upskilling teams to implementing AI solutions in specific trial scenarios. By focusing on practical challenges and effective pathways, this session aims to equip attendees with the knowledge and framework needed to successfully implement and scale AI initiatives, ultimately enhancing the quality and speed of bringing new therapies to patients.

Title: Evidence-based Clinical Trial Design: When the Power of AI Meets Statistical Rigor

- **Will Ma, PhD,** Founder & CEO

Clinical trial design decisions — from endpoint selection to sample size estimation to comparator arm construction — are only as good as the evidence informing them. Yet in practice, development teams often face fragmented literature, incomplete data, and limited access to patient-level information from prior studies, leading to designs built more on convention than on evidence. This talk presents a framework where AI and statistical rigor converge to enable truly evidence-based clinical trial design. The approach, developed at HopeAI, integrates three core capabilities: (1) systematic, human-curated synthesis of clinical outcome data across indications through the PURE Evidence platform; (2) generation of synthetic individual patient data (SynthIPD) from published aggregate results, enabling trial-level simulations grounded in real-world evidence; and (3) AI Statistician that work alongside biostatisticians to translate synthesized evidence into actionable design parameters. We illustrate this framework through oncology case studies, demonstrating how AI-curated evidence and synthetic data can inform adaptive designs, optimize randomization strategies, and strengthen regulatory submissions by grounding design assumptions in a more complete evidence base. Critically, each AI-generated output is validated through established statistical methods — ensuring that the power of AI amplifies, rather than replaces, statistical judgment. When AI meets statistical rigor, the result is clinical trial designs that are both innovative and defensible.

Title: Multi-state model framework to characterize patient journey and time-dependent covariates

- **Jing Shi, PhD, CareDx**

Multi-State Model (MSM) is a class of flexible models including survival and competing risk models. It can accommodate any time-dependent discrete events. We demonstrated the utility of MSM using longitudinal data from KOAR data. Schematic presentations of MSMs analytic framework were used to illustrate clinical questions as concrete analytic problems. Hazard regression methods can be engineered to implement MSMs. Extended Cox regression was used to estimate the Hazard Ratio (HR).

Title: Matching Hearts, Improving Lives: Interpretable Modeling, Simulation-Based Heart Transplant Matching, and New Tools for Informative Missingness and Evolving Clinical Data with Shifts

- **Jiayang Sun, PhD., Bernard Dunn Eminent Scholar, and Chair, George Mason University**

This talk is dedicated to Professor Tze Leung Lai. Heart disease remains a leading cause of death in the United States. Heart transplantation can substantially improve survival for patients with end-stage heart

failure, yet many donor hearts are discarded. This talk highlights the value of integrated statistical thinking and interdisciplinary collaboration through four research projects: (1) interpretable machine learning models for post-heart-transplant survival, (2) simulation-based approaches to improve donor-recipient matching, (3) DACr for handling informative missing data, and (4) prediction under distributional shift through selection-bias modeling (SBMM). Joint with Wei Dai, Zixiang Xu, Jie Xu, Lucas Zhu, and Jason Goldberg.

Title: Unifying Clinical Trials, Literature, and Real-World Data: An AI Platform for Clinical Evidence Synthesis and Management

- **Hui Wang, PhD, Lumbrita**

Clinical evidence is generated across diverse data sources spanning clinical trials, published literature, and real-world evidence. These sources remain scattered across disconnected systems and are integrated largely by hand. This fragmentation makes evidence synthesis slow, costly, and difficult to reproduce or audit — whether the goal is supporting a pre-market submission or meeting continuous post-market obligations. We present SageMed Bridge, an integrated AI platform that unifies these heterogeneous data sources and their analytics within a single architecture and automates the full evidence workflow: data capture, management, integration, analysis, and synthesis. This single, configurable platform follows FDA-recommended human-centered design, combining autonomous agentic workflows with deterministic analysis, human approval gates at each critical step, and a complete audit trail to produce reproducible and verifiable outputs. In this presentation, we illustrate the platform's ability to collect, link, and perform rigorous statistical analysis across multiple data sources.

Title: Inverse probability weighting to adjust diagnostic performance from observational data

- **Sijia Wang, MS, CareDx**

Diagnostic accuracy is commonly evaluated by comparing an index test with a definitive reference standard. In observational studies, however, reference-standard verification is often performed selectively rather than systematically. When the probability of verification depends on the index-test result or clinical characteristics associated with disease status, complete-case area under the curve (AUC) analyses among verified observations may not represent diagnostic performance in the target population.
Several approaches can address selective verification, including outcome-modeling, multiple-imputation, and likelihood-based methods. These alternatives commonly require modeling the disease outcomes missing among unverified observations, and some require simultaneous specification of the disease and verification mechanisms. Inverse probability weighting provides a transparent alternative: observations with verified outcomes are weighted by the inverse of their estimated probability of verification, thereby reconstructing the covariate distribution of the target population under conditional exchangeability, positivity, and correct specification of the verification model.

We present a practical framework for inverse probability weighted estimation of AUCs. The framework emphasizes selection-mechanism-informed covariate specification, flexible modeling of nonlinear relationships, stabilized weights, positivity and overlap assessment, covariate-balance diagnostics, effective sample size, and principled management of influential weights. For longitudinal data with repeated diagnostic tests, patient-level bootstrap resampling is used, with the verification model, stabilization factor, truncation threshold, and weighted AUC re-estimated within each bootstrap sample to propagate uncertainty from the complete estimation procedure. Alternative implementations of weight-truncation

rules provide sensitivity analyses, while clinically defined rejection episodes prevent serial positive biopsies from being treated as independent diagnostic events.

Compared with complete-case analysis, this framework explicitly addresses the selection process producing the verified sample while remaining compatible with empirical AUC estimation. Its implementation and practical considerations are illustrated using an observational study of donor-derived cell-free DNA and biopsy-confirmed rejection after pediatric heart transplantation.

Title: Inverse Probability of Treatment Weighting: A Simple and Effective Approach to Covariate Adjustment for Survival Endpoints in Randomized Clinical Trials

- **Zhiwei Zhang, PhD., Senior Director, Gilead Sciences**

Covariate adjustment aims to improve the statistical efficiency of randomized trials by incorporating information from baseline covariates. Popular methods for covariate adjustment include analysis of covariance for continuous endpoints and standardized logistic regression for binary endpoints. For survival endpoints, while some covariate adjustment methods have been developed, they are not commonly used in practice for various reasons, including high demands for theoretical and methodological sophistication as well as computational skills. In this article, we point out that inverse probability of treatment weighting (IPTW) is a simple and effective approach to covariate adjustment for survival endpoints. We provide theoretical results that justify its use in conjunction with the Cox model and the Kaplan-Meier analysis, as well as numerical results that demonstrate its effectiveness in finite samples of small to moderate sizes. Compared to other covariate adjustment methods for survival endpoints, IPTW has several major advantages including simplicity, generality and ease of implementation, and thus holds great promise as a practical approach to covariate adjustment for survival endpoints.

Lai Memorial Short Course - Causal Generalist Medical AI

- Hongtu Zhu, PhD, Kenan Distinguished Professor, University of North Carolina at Chapel Hill

The rapid evolution of flexible and reusable artificial intelligence (AI) models is reshaping modern medical science. In this lecture, I will introduce Causal Generalist Medical AI (Causal GMAI) - a new paradigm that integrates causal inference with generalist AI models to enhance interpretability, robustness, and generalizability in medical decision-making. Causal GMAI leverages self-supervised, semi-supervised, and supervised learning across diverse multimodal data sources, including medical imaging, electronic health records, clinical trials, laboratory measurements, genomics, knowledge graphs, and clinical text. This unified framework enables a single model to perform a broad range of clinical and translational tasks with minimal task-specific supervision. By explicitly incorporating causal reasoning, Causal GMAI moves beyond purely predictive modeling to infer underlying causal mechanisms. This capability improves diagnostic accuracy, supports more reliable treatment recommendations, and advances personalized and precision medicine. The lecture will highlight the methodological foundations, illustrative applications, and future opportunities for building trustworthy, clinically actionable AI systems.