Short Course of the 2026 Workshop on Biostatistics and Bioinformatics

Title: Empowering Your Study Through Federated Learning: PDA and the New MOSAiC Framework

Dean's Meeting Center, 26th Floor, 25 Park Place, Georgia State University, Atlanta, GA 30303

1:00pm-5:00pm, May 8 (Friday)

Yong Chen, Ph.D., Professor of Biostatistics and Founding Director of the Center for Health AI and Synthesis of Evidence (CHASE) at the University of Pennsylvania

Abstract:

The rapid expansion of digital health data has transformed medical research, providing unprecedented opportunities to generate evidence from large-scale and multi-institutional electronic health records (EHR). Yet fully leveraging these data remains challenging: protecting patient privacy, managing high-dimensional and heterogeneous information, and enabling meaningful collaboration across institutions all pose substantial hurdles.

To address these challenges, our team has developed **Privacy-preserving Distributed Algorithms** (PDA) (https://pdamethods.org/) —a suite of federated learning tools that enable rigorous multi-site analyses without sharing individual

patient data. PDA supports a broad range of statistical and machine learning tasks, spanning association studies, causal inference, clustering, and counterfactual analyses. PDA has been successfully deployed across data-centric networks such as OHDSI, PCORnet, the International Agency for Research on Cancer, and the NIH RECOVER Initiative, demonstrating its utility in pharmacoepidemiology, early disease prediction, subphenotyping, and hospital performance evaluation.

In this short course, I will also introduce for the first time our new framework MOSAiC (Multi-site One-Shot Aggregation of Compressed Risk Functions)—a unified paradigm for modern distributed research networks. Multi-site studies increasingly underpin biomedical discovery, yet inference across sites remains constrained by privacy regulations, data heterogeneity, sparse events in smaller centers, and the practical burden of multi-round communication. MOSAiC reframes federated learning as a mathematical problem of compressing and aggregating local risk functions, leveraging advances in tensor networks, a state-of-the-art technique for high-dimensional function approximation.

MOSAiC achieves four properties previously unattainable in federated algorithms beyond linear models: one-shot communication, lossless recovery of pooled-data estimates, inclusion of all sites regardless of size or event sparsity, and analytic submodel exploration without recontacting partners. I will illustrate MOSAiC's validity and efficiency through applications in drug relabeling, drug repurposing, and post-market safety surveillance.

Together, PDA and MOSAiC provide a powerful foundation for enabling privacy-preserving, scalable, and scientifically robust multi-institutional research.

About the instructor:

Yong Chen is a Professor of Biostatistics and Founding Director of the Center for Health AI and Synthesis of Evidence (CHASE) at the University of Pennsylvania. He also directs the Penn Computing, Inference, and Learning (PennCIL) lab, focusing on evidence synthesis, machine learning/AI, and clinical evidence generation.

Dr. Chen serves as an Associate Editor for the *Journal of the American Statistical Association* (JASA) and *The Annals of Applied Statistics* (AoAS), a Statistical Consultant for *New England Journal of Medicine*-AI, and is a Commissioner on the Lancet Commission on Rare Disease.

During the pandemic, as Biostatistics Core Director for the RECOVER COVID Initiative, he led data studies on post-COVID conditions using information from over 9 million pediatric patients across 40 health systems. This work generated some of the first timely real-world evidence on the effectiveness and safety of COVID vaccines, as well as insights into the impacts of long COVID.

Dr. Chen is one of twenty Commissioners serving internationally on the Lancet Commission on Rare Diseases. He is also a Statistical Editor for the Annals of Internal Medicine, a Statistical Consultant for New England Journal of Medicine-AI, and an Associate Editor for both the JASA-ACS and The Annals of Applied Statistics. Dr. Chen has authored over 300 peer-reviewed papers in statistics and medical informatics. He is an elected Fellow of the American Statistical Association, and the American College of Medical Informatics.