

Department of Statistics Seminar Series



Bingxin Zhao
**Department of Statistics and Data
Science**
University of Pennsylvania

Resampling-based pseudo-training in genomic predictions

Wednesday, October 22nd, 2025

12:10-1:30 PM

SEC 117

Zoom Meeting: Meeting ID: 983 2864 8232

Password: 665123

<https://rutgers.zoom.us/j/98328648232?pwd=gcHifQ9OouKZnUY0EHLe8sjNXaV1oj.1>

Light refreshments will be served in Hill 452, 11:40am

Abstract:

In this talk, I will present a resampling-based pseudo-training framework for genomic prediction that enables model development using only summary-level data. We show that generating pseudo-training and validation statistics from summary results achieves asymptotic equivalence to conventional training while avoiding the need for individual-level datasets. Simulations and real data applications suggest that pseudo-training performs comparably to standard approaches with large datasets and substantially better when tuning data are limited. We highlight two platforms built on this framework: PennPRS (<https://pennprs.org/>), a cloud-based computing infrastructure supporting large-scale, no-code polygenic risk score training with purely summary data resources, and GCB-Hub (<https://www.gcbhub.org/>), which applies pseudo-training to proteome-wide association studies for protein-disease mapping and drug discovery. Together, these advances demonstrate how resampling-based pseudo-training methods can broaden accessibility, scalability, and impact of genomic prediction across diverse biomedical research settings.

Bio:

Dr. Bingxin Zhao is an Assistant Professor in the Department of Statistics and Data Science at the Wharton School, University of Pennsylvania, with a secondary appointment in Department of Medicine, Perelman School of Medicine. His research focuses broadly on statistics, AI in science and medicine, and inter-organ connections such as heart-brain and eye-brain links (<https://www.bingxinzhao.com/>).