Innovations in Clinical AI and Precision Medicine: Experimental Design, Bias Reduction, and Advanced Analytical Methods

National Cheng Kung University
December 18-21, 2024

Syllabus

Instructors: Yu Shyr, Ph.D.

Chair, Department of Biostatistics

Director, Vanderbilt Center for Quantitative Sciences

Harold L. Moses Chair in Cancer Research

Professor of Biostatistics, Biomedical Informatics, and Health Policy

Vanderbilt University Medical Center

E-mail: yu.shyr@vanderbilt.edu

Chih-Yuan Hsu, Ph.D.

Research Assistant Professor

Department of Biostatistics

Vanderbilt University Medical Center

E-mail: chih-yuan.hsu@vumc.org

The key concepts in using clinical AI for precision medicine focus on prevention and treatment strategies that incorporate individual molecular profiles and clinical information. Technologies such as single-cell next-generation sequencing (NGS), spatial transcriptomics, liquid biopsy for circulating tumor DNA (ctDNA) analysis, microbiomics, radiomics, and other high-throughput assays have gained tremendous popularity in recent years due to their ability to generate large volumes of data quickly and cost-effectively. The rise of big data has significantly advanced the goals of precision medicine. However, many challenges remain across the continuum of data capture and utilization—from analyzing high-throughput biomarkers to fully leveraging electronic health records (EHR), ultimately aiming for clinical decision-making guided by a patient's genome.

In this course, I will explore the evolving landscape of applied mathematical and statistical data science, addressing novel study designs, bias reduction, and integration of various Omics data. Additionally, I'll discuss the necessity for statisticians, biologists, and clinicians to shift their perspectives to keep pace with the rapid growth of information technology, machine learning, and the AI revolution. These areas provide significant opportunities for our profession to enhance its role in precision medicine research.

I. Design and Conduct of Clinical AI-Based Biomedical Research

This section will cover the theoretical and practical challenges in designing and conducting clinical AI-based biomedical research, including real-world studies and high-dimensional experiments. Topics will include identifying **types of bias** in clinical AI studies, defining **primary objectives**, **selecting design strategies and strengthening features**, determining **precision analysis** and **sample sizes**, implementing **matching techniques**, assessing **baseline imbalances**, and handling **missing data**.

II. Analysis of Biomedical Research

We will explore analytical methods appropriate for various designs, including **novel Bayesian clinical trials**, **cohort**, **case-control**, **cross-sectional**, and **repeated measures designs**. The course will cover statistical approaches for **continuous, categorical, and time-to-event endpoints**, with a focus on understanding procedures and their applications in data analysis. **Common errors** in analyzing biomedical data will be highlighted, along with strategies to avoid them. We will also discuss **statistical modeling approaches** using real-world data, supplemented by published examples.

This course will delve into applying **advanced matching methods** and **missing data analysis** techniques in biomedical studies. Furthermore, ensemble learning methods such as **boosting, bagging, and stacking** will be introduced, demonstrating how these techniques can aggregate models to enhance prediction accuracy in real-world biomedical data. Like **causal inference** methods, these advanced techniques extend beyond traditional statistical models and may require careful explanation for non-quantitative researchers.

III Topics:

Date	Topic
Dec 18 (Wednesday) 6:00 PM – 9:00 PM (Lab: 8:00 PM – 9:00 PM)	Overview of clinical AI based biomedical research Type of biases Study design Sample size determination and precision analysis Lab (1) Markov Chain Monte Carlo
Dec 19 (Thursday) 6:00 PM – 9:00 PM (Lab: 8:00 PM – 9:00 PM)	Statistical methods in biomedical research I Univariate data analysis (parametric and non-parametric tests) Regression (linear and non-linear) analysis for continuous, binary, count, and ordered data Lab (2) Generalized linear regression, Missing value imputation, and Variable selection
Dec 20 (Friday) 6:00 PM – 9:00 PM (Lab: 8:00 PM – 9:00 PM)	Statistical methods in biomedical research II Mixed effect model, GEE model, Cox model Propensity score matching (PSM) and inverse probability of treatment weighting (IPTW) Lab (3) Mixed-effects model and GEE
Dec 21 (Saturday) 9:00 AM – 12:00 PM 1:00 PM – 4:00 PM (Lab: 3:00 PM – 4:00 PM)	Missing data analysis Multivariate model building strategy Machine learning methods in high-dimensional data analysis Ridge regression, Lasso, Elastic net model Boosting, bagging, and stacking, aggregate multiple models Lab (4) Application to real data examples

IV. Software

1. PS Sample Size and Power Calculations (developed by William Dupont, free download).
2. R

V. Grade

Lab report (35%); Final (65%)