APRIL 26
2024

2024 BIOSTATISTICS PRACTICUM/APEX SYMPOSIUM

COLUMBIA

MAILMAN SCHOOL
OF PUBLIC HEALTH

BIOSTATISTICS

Join us in celebrating the Master's class of 2024!

Please scan the QR code to see the full student titles and abstracts

SCHEDULE

TIME	EVENT	LOCATION
12:00 - 12:45	Welcome from Dr. Berhane & Keynote Speaker	Hammer 401
1:00 - 2:00	Session 1	Rooms HSC 107, 108A/B, 109A/B, 207, 209A/B, 210
2:00 - 3:30	Coffee/Tea Break	
2:30 - 3:30	Session 2	
3:45 - 4:45	Session 3	
5:00 - 7:00	Graduation Celebration	Bard Ballroom

KEYNOTE SPEAKER

Nadiya Pavlishyn (Class of 2019)

Nadiya Pavlishyn is a 2019 graduate of the Mailman School of Public Health, Biostatistics MS, Theory & Methods track. She applied to Mailman from Stony Brook University as an Applied Math & Statistics and Economics double major. However, before she joined our MS program, she was also a proud alum of our department's BEST program in 2015. She has worked as a Research Data Associate at NYU Dept of INternal Medicine, as the Research Coodinator & Statistician at CUMC Department of Radiology, and as a Senior Healthcare Data Scientist at Aetion. Currently, she is the Associate Director of Data Science at Aetion. Come to hear about Nadiya's time at Mailman, her professional journey, and her words of advice for our graduating class as they embark on their own professional pursuits!

ROOM TOPICS

SESSION 1

Topic	Room
Bayesian Statistics	LL207
Cancer Research	LL108A
Causal Inference & Environmental Health	LL210
Clinical Trial	LL108B
Data Analysis	LL109A
Health Policy	LL107
Machine Learning	LL109B
Statistical Genetics	LL209B
Study Design	LL209A

SESSION 2

Topic	Room
Causal Inference	LL207
COVID-19 Research	LL209A
Data Analysis	LL210
Environmental Health Research	LL109A
Health Policy	LL108B
Longitudinal Data Analysis	LL108A
Machine Learning	LL109B
Topics in Mental Health	LL107
Observational Study	LL209B

SESSION 3

Topic	Room
Cardiovascular Death	LL108B
COVID-19 Research	LL109B
Topics in Data Visualizations	LL109A
Topics in Epidemiology	LL107
Machine Learning and Dimension Reduction	LL108A
Social Behavior Research	LL209A
Statistical Genetics	LL209B
Survival Analysis	LL207
Transfer Learning	LL210

SESSION 1

HSC 207 *Bayesian Statistics* Host: Ryan Wei	**Zirui Zhang** Comprehensive metabolomic and proteomic data analysis in Gulf War Illness with exercise tolerance tests **Baoyi Feng** Impact of collaborative care models on depression screening and treatment in primary care settings across New York State **Charles Chen** Comparison of Different Setups for A Novel Bayesian Framework for Model Contamination in Serial Dilution Immunoassay **Zekai Jin** Mediation Analysis of Brain Connectivity and Apathy under Hierarchical Drift Diffusion Modeling
HSC 108A *Cancer Research* Host: Madison Stoms	**Jiazhen Liu** APEX Internship Report: Full Process of RNA-SEQ Data Analysis **Jiong Ma** Challenges and Strategies in Evaluating CAR-T Therapy: Navigating the Discrepancy Between Progression-Free and Overall Survival Outcomes **Yifei Zhao** DNAGmPT: a Transformer Model Predicting DNA Methylation from Sequence and Gene Expression **Yimin Chen** Assessing the Predictive Value of Pathologic Complete Response for Survival in Triple-Negative Breast Cancer Cases
HSC 210 *Causal Inference and Environmental Health* Host: Ting-Hsuan Chang	**Ziyue Yang** Impact of heat alert on different population in US **Lingke (Lincole) Jiang** Sample Size Calculations for High-Dimensional Mediation Analysis **Jasmine Niu** Simulation study of estimating causal effects of continuous exposures **Hongpu Min** How the causal effects of NWS heat alerts vary across characteristics of communities
HSC 108B *Clinical Trials* Host: Anja Shahu	**Yi Huang** Enhancing Clinical Site Selection: Developing an internal tool for Statistical Analysis in the Drug Review Process **Emily Potts** Are Safety Lead-In Phase II Clinical Trials Really Safe? **Xiyuan (Angel) Ji** A Multi-center, Single-Arm, Phase II Clinical Trial of Combination Therapy in Patients with Malignant Peripheral Nerve Sheath Tumors **Shengzhi Luo** The impact of Metabolic Fatty Liver Disease(MFLD) in the physical examination
HSC 109A *Data Analysis* Host: Guadalupe Antonio Lopez	**Emmanuel Latio** Merging of Departmental of Datasets to Analyze Student Trajectories **Pei Liu** Moderation effect of persistent homology-based functional connectivity on cognitive decline in Alzheimer's Disease **Tiffany Zhou** Exploring the Impact of Enhanced Skin-to-Skin Contact Through a Novel Medical Device on Maternal and Infant Health Outcomes **Feng Yan** g.bkmr R command for automated implementation of a BKMR approach for time-varying environmental mixtures

SESSION 1

HSC 107 *Health Policy* Host: Safiya Sirota	**Yiting Zhang** Validity and Reliability of an Evidence Metric for Public Policy Standards **Haoyang Wang** Predicting ICU Patients' Mortality Outcome with Biomarkers' Data by Applying Biological Aging Algorithms **Ruyi Chen** Analyzing Prescription Trends Across Demographics and States: A Biostatistical Approach to Pharmaceutical Marketing Strategies **Isabella Patino** Analyzing Salary and Job Requirement Disparities between Government and Private Sector Public Health Occupations: Utilization of Lightcast Job Postings
HSC 109B *Machine Learning* Host: Serena Wang	**Shun Xie** User Churn Model and the Exploration of the Effect of Quarantine Status on Churn Likelihood **Yuze Yuan** Using Real Words Data to Predict High-Demand Clothing Materials: A Strategy for Healthier and Efficient Inventory Management **Yuanhao Zhang** A Data-Driven Approach to Predict Hot Clothing Design in the Post-COVID Era **Shaohan Chen** Study and Implementation of Kernel Ridge Regression
HSC 208B *Statistical Genetics* Host: Yuqi Miao	**Anjing Liu** Molecular Quantitative Trait Loci Discovery Using Quantile Regression **Haochen Sun** Detecting and correcting discrepancies in summary statistics for fine-mapping studies **Zining Qi** Refined Approaches for Missing Data Imputation Enhance Quantitative Trait Loci Discovery in Multi-Omics Analysis **Yvonne Chen** Analysis of Spatial Architecture in Actinic Keratoses
HSC 209A *Study Design* Host: Wenbo Fei	**Yichen Lyu** Macro for Win Ratio **Jiaoyang Li** What Statistical Designs for Rare Disease Trial proposals improve Clinical Trial Readiness for Regulatory Approval? A Pilot and Feasibility Study **Yuchen Hua** Prospective analysis of pancreatic cancer and meat intake based on 15 cohorts **Junyan Zhu** DeepMed: whether gender disparities exist in income when education levels are the same

SESSION 2

HSC 207 *Causal Inference* Host: Ryan Wei	**Ziqing Wang** R implementation of a multistate approach for stochastic interventions on a time-to-event mediator in the presence of competing risks **Yuchen Zhang** Simulation Study in Mediation analysis with time-varying mediators and time-to-event outcomes accounting for competing risks **Riya Bhilegaonkar** Causal Inference Program Evaluation of Pediatric Transitions of Care (TOC) **Binyue Hu** Addressing Bias in Mendelian Randomization: Introducing the Penalized Inverse-Variance Weighted Estimator with Application to Obesity-Related Exposures and Type 2 Diabetes Risk
HSC 209A *COVID-19 Research* Host: Wenbo Fei	**Zozo Chunyu** Mendelian Randomization Analysis to Investigate the correlation of Covid-19 and Alzheimer's Disease in European Populations **John Cheng** Estimating the Effects of Long COVID on Employment Loss **Wenjun Mo** Healthcare Access and COVID-19 Outcomes in New York City **Yuan Yao** Analysis of the relationship between Covid-19 Prevention and Control and Depression in college students ——A case study of undergraduates at Shanghai Fudan University
HSC 210 *Data Analysis* Host: Ting-Hsuan Chang	**Ruihan Zhang** Analyzing Pharmaceutical Production Yield Losses: A Biostatistical Approach to Enhancing Efficiency in Pharmaceutical Manufacturing **Meng Fang** Analyzing Evacuation Dynamics in Louisiana During Hurricane Ida: A Comprehensive Zipcode-Level Study **Shuting Kang** COVID-19 Vaccine Claims Cleaning Program and Missing Dose Investigation **Sarah Tsang** Compiling Data for Government Grants
HSC 109A *Environmental Health Research* Host: Guadalupe Antonio Lopez	**Jennifer Osei** Heat Exposure on Health Outcomes in Ghana: Redefining & Comparing Climate Heath Stress Index Exposure Measures for Atmospheric Heat through Comprehensive Model Optimization **Tharina Messeroux** Exploring Housing Satisfaction Levels in Public Housing: Preliminary Analysis for Smoke-Free Policy Compliance **Yining Chen** Identifying Changes in Human Mobility and Movement Patterns Linked to California Wildfires **Eileen Ramirez del Rio** Neurobehavioral Developmental Outcomes in Young Children Born to Ecuadorian Women with Exposures to Ethylenethiourea (ETU)

SESSION 2

HSC 108B *Health Policy* Host: Anja Shahu	**Jingchen Chai** Medicare Data Insights: Strategic Approaches to Enhancing Drug and Vaccine Safety **Shodai Inose** Health System Strengthening in South Sudan: Understanding Post-Independence National Trends in Utilization of Maternal and Child Health Services **Nirali Patel** Racial and Ethnic Disparities in Postpartum Readmissions at New York State Hospitals **Angsi Shi** LIGHTSPEED" - How Project "Lightspeed" Changed Our Approach to Imminent Public Health Crisis
HSC 108A *Longitudinal Data Analysis* Host: Madison Stoms	**Mingyi Du** Predicting CHD from Metabolic Components Using Dynamic Bayesian Network in longitudinal cohorts **Xicheng Xie** Analyzing trends in humid heat stress in New York State: A Bayesian spatiotemporal analysis **Jiawen Zhao** Examining the Impact of a Sleep Intervention on Nurses' Mental Health Status (Stress and Depression Level): A Longitudinal Analysis using GLMM and GEE **Chenyao Ni** Longitudinal Profiles of Suicidal Ideation in Older Adults with Depression: Associations with Depressive Symptoms, Cognitive Deficits, and Clinical Measures
HSC 109B *Machine Learning* Host: Serena Wang	**Hongru Tang** Inference in Linear Regression with Informative Sampling: A Comparison of Methods **Noah Zhou** Predictive Modeling of Diabetes Risk Using Biostatistical and Data Science Methods **Jiahe Deng** Discriminating Malignant Breast Cancer from Benign: Insights from FNA Image Analysis **Yang Yi** Identifying multiple sclerosis subtypes using unsupervised machine learning and clinical data
HSC 107 *Topics in Mental Health* Host: Safiya Sirota	**Qingyue Zhuo** The Impact of Functional Connectivity Measures on ADHD Comorbidity Profiles in Children **Haotian (Matthew) Ma** Correlation between cognition, functional impairment, and neuropsychiatric symptoms **Shangsi Lin** Exploring the Impact of Oversampling Techniques on the Performance of Postpartum Depression Prediction Models and the Mediating Role of Substance Usage: An Investigation Based on New York State Patient Data **Jiayi Shi** Predictive Modeling of Childhood Anxiety Disorders Using Task-Based fMRI: Insights from the ABCD Study
HSC 209B *Observational Study* Host: Yuqi Miao	**Sophie Chen** The role of neural flexibility in cognitive aging **Landi Guo** Analysis of Persistent Homology-Based Functional Connectivity and Cognitive Function Across the Lifespan **Yuhuan Lin** Modeling Piecewise Relationships in Laboratory Data Using Linear Mixed Effects Model **William Anderson** Uranium exposure and kidney tubule biomarker levels in the Multi-Ethnic Study of Atherosclerosis (MESA)

SESSION 3

HSC 108B *Cardiovascular Disease* Host: Anja Shahu	**Yucheng Li** Early Onset and Progressive Cardiac Remodeling in Pediatric Sickle Cell Disease- A Longitudinal Echocardiographic Study **Zizhao Lin** Association of public water arsenic exposure with incident fatal and non-fatal CVD **Qilin Zheng** Heart Failure Prediction **Xuesen Zhao** Investigating the genomic interactions between Alzheimer's Disease and cardiovascular/cerebrovascular risk factor in ROSMAP cohort
HSC 109B *COVID-19 Research* Host: Serena Wang	**Wenjia Zhu** Cytokine and Chemokines Dysregulation in Long COVID Patients **Yahui Zhou** The Paxlovid Rebound Study **Zuoqiao Cui** Estimating the Causal effect of Cigarette Smoking on COVID-19 severity: A Comprehensive Mendelian Randomization Analysis **Xin Ren** Predicting Mutations in SARS-CoV-2 with a Deep Learning Model
HSC 109A *Topics in Data Visualization* Host: Guadalupe Antonio Lopez	**Chao Gao** Enhancing Clinical Trial Data Review with Interactive Visualization Through R/R Shiny **Fengdi Zhang** Clinical Data Management and Visualization in Medical Device Industry **Jingya Yu** Exploring the Pharmacokinetics of ERAS-007: A Comprehensive Analysis and Visualization **Yiying Wu** Establishing a Public Health Infrastructure in Vernon, CA: A Comparative Demographic Analysis and GIS Evaluation of Healthcare Accessibility
HSC 107 *Topics in Epidemiology* Host: Safiya Sirota	**Tingyi Li** Risk factors of diabetes **Tvisha Devavarapu** Meta-Analysis of Resistance to first-line Anti-Tuberculosis Drugs within Central Nervous System Tuberculosis Cases **Anzhuo Xie** Stroke Data and Underlying factors Analysis **Olivia Schulist** Migration and Nutrition Epidemiology in Mexico and the United States
HSC 108A *Machine Learning and Dimension Reduction* Host: Madison Stoms	**Qinzhen Sun** Comparative Analysis of Penalized Regression Models with High-dimensional Imaging Biomarkers in a Large-scale Alzheimer's Disease Study **Runze Cui** Regularized Generalized Method of Moments **Ruiji Pan** Using Machine Learning Methods to Predict Customer Segmentation **Zhaoqiany Xiong** Investigating the Model Performance in Multi-Pollutant Mixtures with Complex Confounders

SESSION 3

HSC 209A *Social Behavior Research* Host: Wenbo Fei	**Melike Aksoy** Marital Status and Sedentary Behaviors in Older Adults **Jingyi Tian** Analysis of Influencing Factors of Online Public Opinion Dissemination on Public Health Emergencies: Based on Weibo Platform **Xiangyi He** Optimizing School-Based Caries Prevention: A Comparative Study of Simple vs. Complex Regimens in Underserved Communities **Farizah Rob** Social relationships age the immune system in early midlife: Evidence from the National Longitudinal Study of Adolescent to Adult Health
HSC 209B *Statistical Genetics* Host: Yuqi Miao	**Tianshu Liu** Evaluation of Eigen Higher Criticism, Eigen Berk-Jones, and Omnibus Tests on Multi-Trait GWAS Summary Statistics **Yixuan Jiao** Assumption Validation and Comparison across Different Spatially Variable Genes Identification Methods **Yunxi Yang** Robust multi-source data integration for enhanced ADRD risk prediction for underrepresented Populations **Jingyi Yao** A computational method for the detection of cell type-specific gene-gene correlation in multi-sample single-cell RNA-seq data
HSC 207 *Survival Analysis* Host: Ryan Wei	**Kyle Schichl** Characterizing Apathy in ADNI using Joint Latent Class Linear Mixed Models **Youlan Shen** A Comparative Analysis of Cox Hazards and Generalized Linear Models in Evaluating the Impact of PM2.5 Exposure on Mortality Rates **Wenjing Yang** Circadian polygenetic risk scores determining hyperglycemia in multi-ethnic study of atherosclerosis **Haoyu Tian** Study On the Early Prediction of Myopia in Children and Adolescents by the interaction between environment and Genes
HSC 210 *Transfer Learning* Host: Ting-Hsuan Chang	**Zhengwei Song** Enhancing Breast Cancer Risk Prediction in Hispanic Women Through Transfer Learning **Hongjie Liu** Transfer Learning of Conditional Average Treatment Effect **Zijian Xu** Transfer Causal Learning with Instrumental Variables **Jiajun Tao** Transfer learning in Cox model

2024 BIOSTATISTICS PRACTICUM/APEX SYMPOSIUM

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

BIOSTATISTICS