2022 ICSA APPLIED STATISTICS SYMPOSIUM

Gainesville, Florida
June 19-22, 2022

International Chinese Statistical Association
泛華統計協會

Contents

Welcome

2022 ICSA Applied Statistics Symposium
June 19-22, 2022
Gainesville, Florida, USA

On behalf of the organizing committee, we welcome you to the campus of the University of Florida. We are thankful to the ICSA executive committee for selecting Gainesville, FL, to be the venue for the 2022 Applied Statistics Symposium. We are super excited to have the unique distinction of hosting this conference face to face since the beginning of the pandemic. We plan to follow all safety protocols to create a safe environment for our participants.

The program committee has worked diligently to bring you an exciting program which consists of 3 plenary and 2 special invited lectures, plus 70 invited sessions on the theme of "Statistical Innovations in the Era of Artificial Intelligence and Data Science". In addition, there are 36 posters, 6 short courses and 5 oral presentations by student paper award recipients, and an after-dinner talk. In addition to the academic components of the conference, the local organizing committee has organized several social events including the opening mixture, entertainment programs, local outings, and a fabulous banquette.

We hope you will enjoy the next three and half days on campus. Besides attending numerous technical sessions, do take advantage of various local attractions Gainesville has to offer including a historic downtown, and various state and city parks with extensive nature trails. UF displays a very impressive campus listed on the National Register of Historic Places. You will be able to view beautiful brick construction buildings that showcase traditional Gothic architecture. A number of natural springs are within driving distance and so are three major metropolis and numerous beaches on both sides of the state. Feel free to contact the local organizers or the student volunteers if you need any assistance.

We thank all individuals and entities who contribute to the success of this event, notably, the deans of our two colleges, chair of department of biostatistics at UF, the staff members, the student volunteers, all committee members, our sponsors, notably the National Science Foundation for supporting the students and junior researchers attending the conference, chairs of neighboring statistics departments for encouraging their students to attend, and so on. Last but not least, we thank all the participants. Without your involvement, the event could not be a success. Go Gators!

Samuel Wu and Somnath Datta
Co-chairs, 2022 ICSA Applied Statistics Symposium Organizing Committee

Program Overview

ICSA 2022 Applied Statistics Symposium (Program and Activity Overview)

Theme: Statistical Innovations in the Era of Artificial Intelligence and Data Science

Date	Time	Event	Location
Sunday, June 19	7:30-18:30	Registration	HPNP Lobby
	8:30-12:30	Short Courses SC01, SC02, SC03, SC04	See Program Book
	12:30-13:30	Lunch	
	13:30-17:30	Short Courses SC02, SC03, SC05, SC06	See Program Book
	18:30-22:00	**Welcome Reception and Mixer**	Hilton Hotel
Monday, June 20	7:30-18:30	Registration	
	8:00-8:30	**Welcome and Opening Remarks**	
	8:30-9:30	**Plenary Keynote Talk 1**	
	9:30-10:00	Coffee Break	
	10:00-11:40	Invited Sessions 1A to 1H	
	11:40-13:00	Lunch	
	13:00-14:40	Invited Sessions 2B to 2H	
	14:34-15:00	Coffee Break	
	15:00-16:40	Invited Sessions 3A to 3H	
	16:40-17:00	Coffee Break	
	17:00-18:40	Invited Sessions 4A to 4H	
	19:00-20:00	**Poster Session**	
Tuesday, June 21	7:30-18:30	Registration	
	8:30-9:30	**Plenary Keynote Talk 2**	
	9:30-10:00	Coffee Break	
	10:00-11:40	Invited Sessions 5A to 5H	
	11:40-13:00	Lunch	
	13:00-14:40	**Special invited Session**	
	14:40-15:00	Coffee Break	
	15:00-16:40	Invited Sessions 6A to 6H	
	16:40-17:00	Coffee Break	
	17:00-18:40	Invited Sessions 7A to 7H	
	19:00-22:00	**Conference Banquet**	
Wednesday, June 22	**8:30-9:30**	**Plenary Keynote Talk 3**	
	9:30- 10:00	Coffee Break	
	10:00-11:40	Invited Sessions 8A to 8H	
	11:40-13:00	Lunch	
	13:00-14:40	Invited Sessions 9A to 9H	
	14:40	Adjournment	

Session	HPNP Room
Plenary	1404
A	G312
B	G112
C	G101
D	G103
E	G114
F	G301
G	1101
H	1102

Keynote Lecture 1

Keynote Speaker

David O. Siegmund, Ph.D., who holds the John D. and Sigrid Banks Chair at Stanford University, Stanford, CA, is a statistician who is comfortable in both the airy heights of theory and the practicalities of real-world applications. He works at the interface between probability and statistics, applying the tools he develops to topics as diverse as the design of medical clinical trials and mapping the locations of genes that are involved in specific physiological traits. His work has earned him several awards, including a Guggenheim Fellowship in 1974, the Humboldt Prize in 1980, and membership in the American Academy of Arts and Sciences in 1994. In 2002 he was elected to the National Academy of Sciences. His Inaugural Article, published in this issue of PNAS, reviews recent methodological developments in quantitative trait locus mapping and addresses the problem of mapping with selected, rather than random, samples.

Location and Time: HPNP Auditorium (1404), June 20 (Monday), 8:30 am – 9:30 am

Organizer: ICSA special lecture committee

Keynote Host: Samuel Wu, Ph.D., University of Florida

Title: Change detection, estimation, and segmentation

Abstract: I will first discuss the maximum score statistic to detect and estimate via confidence regions change-points in the level, slope, or other property of a Gaussian process and to segment the process when there appear to be multiple changes. Sequential detection is also considered. Examples involving temperature variations, levels of atmospheric greenhouse gases, temporal incidence of hate crimes, suicide rates, incidence of Covid-19, and excess deaths during the Covid-19 pandemic illustrate the general theory.

I will describe research in progress for spatio-temporal processes, where the spatial features can be either (A) unstructured vectors of observations or (B) random fields where changes of interest are geometrically clustered. Examples include low and (perhaps sparse) high dimensional cases.

I also mention the special problems posed by temporal and/or spatial dependence. Failure to account for correlations can lead to inflated false positive rates. while the change-points themselves can lead to upwardly biased estimates of correlations that result in loss of power.

Aspects of this research involve collaboration with Fang Xiao, Li Jian, Liu Yi, Nancy Zhang, Benjamin Yakir and Li (Charlie) Xia.

Keynote Lecture 2

Keynote Speaker

Jianqing Fan, Ph.D., is a statistician, financial econometrician, and data scientist. He is Frederick L. Moore'18 Professor of Finance, Professor of Statistics, and Professor of Operations Research and Financial Engineering at the Princeton University where he chaired the department from 2012 to 2015. He is the winner of The 2000 COPSS Presidents' Award, Morningside Gold Medal for Applied Mathematics (2007), Guggenheim Fellow (2009), Pao-Lu Hsu Prize (2013) and Guy Medal in Silver (2014).

Location and Time: HPNP Auditorium (1404), June 21 (Tuesday), 8:30 am – 9:30 am

Organizer: ICSA special lecture committee

Keynote Host: Somnath Datta, Ph.D., University of Florida

Title: Measuring housing activeness from multi-source big data and machine learning

Abstract: Measuring timely high-resolution socioeconomic outcomes is critical for policy-making and evaluation, but hard to reliably obtain. With the help of machine learning and cheaply available data such as social media and nightlight, it is now possible to predict such indices in fine granularity. This paper demonstrates an adaptive way to measure the time trend and spatial distribution of housing activeness with the help of multiple easily-accessible datasets. We first identified the regional activeness status at the individual level from energy consumption data and then matched it with nightlight and land use data geographically. Then, we introduce the principle of robustification via truncation and factor-adjusted regularization methods for prediction (FarmPredict) to deal with two important stylized features in big data. The heterogeneity of big data is mitigated through the use of the government land planning data. Farm-Predict effectively lifts the prediction space and solves the colinearity problem in high-dimensional data. It is applicable to all machine learning algorithms. FarmPredict allows us to extend the regional results to the city level, with a 75% out-of-sample explanation of the spatial and timeliness variation in the housing usage. FarmPredict is not only a model but an analytical framework of machine learning on high-dimensional data, showing broad potential applications to other social science problems. Since energy is indispensable for life, our method is highly transferable with the requirement of only public and accessible data. Our paper demonstrates the power of machine learning in understanding socioeconomic outcomes when the census and survey data is costly or unavailable.

(Joint work with Yang Zhou, Lirong Xue, Zhengyu Shi, Libo Wu)

Special Invited Talks

Special Invited Speaker

Xihong Lin, Ph.D., is Professor and former Chair of the Department of Biostatistics, Coordinating Director of the Program in Quantitative Genomics at the Harvard T. H. Chan School of Public Health, and Professor of the Department of Statistics at the Faculty of Arts and Sciences of Harvard University, and Associate Member of the Broad Institute of Harvard and MIT.

Dr. Lin is an elected member of the National Academy of Medicine. She received the 2002 Mortimer Spiegelman Award from the American Public Health Association, and the 2006 Committee of Presidents of Statistical Societies (COPSS) Presidents' Award and the 2017 COPSS FN David Award. She is an elected fellow of American Statistical Association (ASA), Institute of Mathematical Statistics, and International Statistical Institute.

Dr. Lin's research interests lie in development and application of statistical and computational methods for analysis of massive data from genome, exposome and phenome, and scalable statistical inference and learning for big genomic, epidemiological and health data.

Location and Time: HPNP Auditorium (1404), June 21 (Tuesday), 13:00 pm – 13:45 pm

Organizer: Somnath Datta, Ph.D., University of Florida

Keynote Host: Ji-Hyun Lee, DrPH, University of Florida

Title: Lessons learned from the COVID-19 pandemic: a statistician's reflection

Abstract: In this article, I will discuss my experience as a statistician involved in COVID-19 research in multiple capacities in the last two years, especially in the early phase of the pandemic. I will reflect on the challenges and the lessons I have learned in pandemic research regarding data collection and access, epidemic modeling and data analysis, open science and real time dissemination of research findings, implementation science, media and public communication, and partnerships between academia, government, industry and civil society. I will also make several recommendations on preparing for the next stage of the pandemic and for future pandemics.

Special Invited Talks

Special Invited Speaker

Nilanjan Chatterjee, Ph.D., received his Bachelor's and Master's degree from the Indian Statistical Institute, Calcutta and a Ph.D. in statistics from the University of Washington, Seattle in 1999. He served as Chief of the Biostatistics Branch of the Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute (NCI) for almost eight years. Dr. Chatterjee now serves as a Bloomberg Distinguished Professor at the Johns Hopkins University with joint appointments at the Bloomberg School of Public Health (Biostatistics) and the School of Medicine (Oncology). He remains a Special Volunteer with DCEG.

Dr. Chatterjee's research focuses on a diverse set of quantitative issues that arise in design, analysis, interpretation and public health translation of modern molecular and genetic epidemiologic studies.

Location and Time: HPNP Auditorium (1404), June 21 (Tuesday), 13:45 pm – 14:30 pm

Organizer: Somnath Datta, Ph.D., University of Florida

Keynote Host: Ji-Hyun Lee, DrPH, University of Florida

Title: Predictive model building through integration of information across disparate data sources and summary-statistics

Abstract: Model building based on classical statistical methods, as well as modern machine learning techniques, typically requires availability of a single adequately large dataset, or multiple harmonized datasets across a group of similar studies. In the future, however, development of complex models incorporating a variety of factors from different domains will require integration of information from disparate data sources, which, individually may have information only on subsets of the explanatory variables of interest. Moreover, information from some studies may only be available through pre-computed summary-statistics, generated under certain forms of ”reduced” models. In this talk, I will describe some of our recent efforts towards developing statistical methods for model building through data integration under a semiparametric generalized meta-analysis framework. I will illustrate the unique opportunity data integration methods provide through an application involving the development of a COVID-19 mortality risk calculator through integration of information across diverse datasets.

Special Invited Talks

Banquet Speaker

Lee-Jen Wei, Ph.D., was graduated from Fu Jen Catholic University's Mathematics Department in 1970. He obtained his PhD from the University of Wisconsin–Madison in 1975. He has been a tenured Professor of Biostatistics at Harvard University since 1991 and was the co-director of the Bioinformatics Core at the Harvard School of Public Health from 2003 to 2007. From 2003 to 2004, he served as the acting chair of the Department of Biostatistics at Harvard University. Under his supervision, the department successfully converted the doctor of science degree program in biostatistics (a professional degree) to a conventional (art and sciences) Ph.D. program at the Harvard Graduate School. This was an important accomplishment since the department had tried this conversion for more than 20 years without success.

Professor Wei has developed and published a number of novel quantitative methods for analyzing data from experimental and observational studies. Specifically, he has published many papers on monitoring drug and device safety and related topics. The resulting procedures have been utilized for various drug and device regulatory evaluations involving safety issues. His extensive experience in quantitative science for making inferences about the drug and device safety is readily applicable to the general industry product safety issues.

Location and Time: Ben Hill Griffin Stadium Champions Club (121 Gale Lemerand Drive), June 21 (Tuesday), 20:00 pm – 20:45 pm

Organizer: ICSA special lecture committee

Keynote Lecture Host: Samuel Wu, Ph.D., University of Florida

Title: Lost in translation

Abstract: One of the main goals of conducting a clinical, comparative study is to obtain robust, clinically interpretable treatment effect estimates with respect to harm-benefit perspectives at the patient's level via efficient and reliable quantitative procedures. To accomplish this goal, it is important to know how to effectively translate new developments in basic data science research into clinical research and practice. Unfortunately, some commonly used statistical procedures are not translational. That is, results of the analysis may be misinterpreted or difficult to comprehend. A notorious example is use of the p-value for clinical decision making, which is not an appropriate quantifier for assessing the clinical utility of a new therapy or strategy. In this talk, we will discuss several translational problems and present possible remedies.

Keynote Lecture 3

Keynote Speaker

Susan Murphy, Ph.D., is the Mallinckrodt Professor of Statistics and of Computer Science, Radcliffe Alumnae Professor at the Radcliffe Institute, Harvard University. Her research focuses on improving sequential, individualized, decision making in health, in particular on clinical trial design and data analysis to inform the development of mobile health treatment policies. Susan is a Fellow of the Institute of Mathematical Statistics, a Fellow of the College on Problems in Drug Dependence, a former editor of the Annals of Statistics, a member of the US National Academy of Medicine and a 2013 MacArthur Fellow.

Location and Time: HPNP Auditorium (1404), June 22 (Wednesday), 8:30 am – 9:30 am

Organizer: ICSA special lecture committee

Keynote Lecture Host: Guogen Shan, Ph.D., University of Florida

Title: Inference for longitudinal data after adaptive sampling

Abstract: Adaptive sampling methods, such as reinforcement learning (RL) and bandit algorithms, are increasingly used for the real-time personalization of interventions in digital applications like mobile health and education. As a result, there is a need to be able to use the resulting adaptively collected user data to address a variety of inferential questions, including questions about timevarying causal effects. However, current methods for statistical inference on such data (a) make strong assumptions regarding the environment dynamics, e.g., assume the longitudinal data follows a Markovian process, or (b) require data to be collected with one adaptive sampling algorithm per user, which excludes algorithms that learn to select actions using data collected from multiple users. These are major obstacles preventing the use of adaptive sampling algorithms more widely in practice. In this work, we proved statistical inference for the common Z-estimator based on adaptively sampled data. The inference is valid even when observations are non-stationary and highly dependent over time, and (b) allow the online adaptive sampling algorithm to learn using the data of all users. Furthermore, our inference method is robust to miss-specification of the reward models used by the adaptive sampling algorithm. This work is motivated by our work in designing the Oralytics oral health clinical trial in which an RL adaptive sampling algorithm will be used to select treatments, yet valid statistical inference is essential for conducting primary data analyses after the trial is over.

SC01: Causal Inference with R

Location and Time: HPNP G114, Sun, June 19, 8:30 - 12:30
Length: Half-day
Instructors: Prof. Babette Brumback (University of Florida)

Abstract: One of the primary motivations for clinical trials and observational studies of humans is to infer cause and effect. Disentangling causation from confounding is of utmost importance. Causal Inference with R explains and relates different methods of confounding adjustment in terms of potential outcomes and graphical models, including standardization, doubly robust estimation, difference-in-differences estimation, and instrumental variables estimation. Several real data examples, simulation studies, and analyses using R motivate the methods throughout. The course assumes familiarity with basic statistics and probability, regression, and R. The course will be taught with a blend of lecture and worked examples.

Teaching Plan:
First part:
Introduction – 15 minutes
Potential Outcomes and Effect Measures – 30 minutes
Causal Directed Acyclic Graphs – 1 hr

15 minute break

Second part:
Standardization and Doubly Robust Estimation – 1 hr
Difference-in-Differences Estimation – 30 minutes
Instrumental Variables Estimation – 30 minutes

About the Instructors: Babette A. Brumback, Ph.D. is Professor in the Department of Biostatistics at the University of Florida; she won the department's Outstanding Teacher Award for 2020-2021. A Fellow of the American Statistical Association, she has researched and applied methods for causal inference since 1998, specializing in methods for time-dependent confounding, complex survey samples and clustered data.

SC02: Leveraging Real-World Data in Clinical Trial Design and Analysis

Location and Time: HPNP G103, Sun, June 19, 8:30 - 17:30
Length: Full-day
Instructor: Dr. Chenguang Wang (Regeneron Pharmaceuticals, Inc.)

Abstract: The amount of real-world data (RWD) collected from sources other than protocol-driven clinical studies is increasing ultra-rapidly. The clinical evidence that can be derived from analysis of these RWD is considered as real-world evidence (RWE) that can complement the knowledge derived from traditional well-controlled clinical trials. Leveraging RWE can potentially save time and cost of the investigational study and improve the efficiency of regulatory decision-making. Incorporating RWD in regulatory decision-making demands much more than "mixing" RWD with investigational clinical trial data. The RWD has to undergo appropriate analysis for deriving the right RWE. Moreover, such analysis has to be integrated with the design and analysis of the investigational study for regulatory decision-making. The standard clinical trial toolbox does not offer ready solutions for incorporating RWD. In this course, the instructor(s) will cover a series of methods they have developed for leveraging real-world data in clinical trial design and analysis. Their work has been recognized by the FDA and received The FDA CDRH Excellence in Scientific Research Award and The FDA Scientific Achievement Award.

Teaching Plan: In Part I of the course, we introduce a new method for proposing performance goals—numerical target values pertaining to effectiveness or safety endpoints in single-arm medical device clinical studies—by leveraging RWE. The method applies entropy balancing to address possible patient dissimilarities between the study's target patient population and existing real-world patients, and can take into account operation differences between clinical studies and real-world clinical practice.

In Part II of the course, we introduce a method that extends the Bayesian power prior approach for a single-arm study to leverage external RWD. The method uses propensity score methodology to pre-select a subset of RWD patients that are similar to those in the current study in terms of covariates, and to stratify the selected patients together with those in the current study into more homogeneous strata. The power prior approach is then applied in each stratum to obtain stratum-specific posterior distributions, which are combined to complete the Bayesian inference for the parameters of interest.

In Part III of the course, we introduce several extensions of the PS-integrated method in Part II. These extensions include 1) a frequentist PS-integrated composite likelihood approach for incorporating RWE in single-arm clinical studies; 2) leveraging multiple RWD sources in single-arm medical device clinical studies; 3) leveraging RWD for the evaluation of diagnostic tests for low prevalence diseases; 4) augmenting both arms of a randomized controlled trial by leveraging RWD; and 5) PS-integrated approach for survival analysis.

In Part IV of the course, we describe an R package, psrwe, that implements a PS-integrated power prior (PSPP) method, a PS-integrated compos- ite likelihood (PSCL) method, and a PS-integrated weighted Kaplan-Meier estimation (PSKM) method for the methods in Parts II and III. Illustrative examples are provided to demonstrate each of the approaches.

In Part V of the course, we introduce a propensity score-based Bayesian non-parametric Dirichlet process mixture model that summarizes subject-level information from randomized and RWD to draw inference on the causal treatment

effect in exploratory analysis.

About the Instructor: Dr. Chenguang Wang is a Senior Director and the Head of Statistical Innovation at Regeneron. Previously, Dr. Wang was an Associate Professor with Johns Hopkins University and an FDA Mathematical Statistician. Dr. Wang has extensive experience in clinical trial design and analysis in the regulatory setting. Dr. Wang holds B.S. and M.S. degrees in Computer Science and has abundant experience developing statistical software.

SC03: Marginal Models in Analysis of Correlated Binary Data with Time-Dependent Covariates

Location and Time: HPNP G301, Sun, June 19, 8:30 - 17:30
Length: Full-day
Instructors: Prof. Jeffrey Wilson (Arizona State University); Prof. Din Chen (Arizona State University)

Abstract: This workshop is based on the book: "Marginal Models in Analysis of Correlated Binary Data with Time Dependent Covariates" co-authored by Drs. Jeffrey R. Wilson, Elsa Vazquez-Arreola, and (Din) Ding-Geng Chen, published by Springer in 2020, which is the first book to systematically introduce marginal models to analyze correlated binary data with time-dependent covariates in clinical trials and observational studies using R and SAS. This workshop provides a thorough presentation of correlated binary data with time-dependent covariate. It gives a detailed step-by-step illustration of their implementation using R and SAS. Longitudinal data or contain correlated data due to the repeated measurements on the same subject. The changing values usually consist of time-dependent covariates and their association with the outcomes present different sources of correlation. Most methods used to analyze longitudinal data would average the effects of time-dependent covariates on outcomes over time and provide a single regression coefficient per time-dependent covariate. Such an approach prevents analysts and researchers the opportunity to following the changing impact of time-dependent covariates on the outcomes. The workshop addresses such issues through the use of partitioned regression coefficients. We further use examples of correlated data with time-dependent covariate on obesity from the Add Health study and cognitive impairment diagnosis in the National Alzheimer's Coordination Center.

Teaching Plan: Morning Session (8:30am to 12:30pm):
1. Fundamentals of estimation of regression coefficients in cross-sectional data
    a. Review of the estimation of regression models
    b. Generalized estimating equation (GEE) and generalized linear mixed models
    c. Generalized Method of Moments estimates;
2. Presentation on data with time-dependent covariates and discussion on the partitioned matrix.

Afternoon Session (1:30pm to 4:30pm):

3. Present correlated data with time-dependent covariates. Illustrate longitudinal data and the analysis using linear mixed models for continuous endpoints, generalized linear mixed model and GEE for categorical endpoints.

4. Bayesian analysis in this partitioned data matrix using MCMC is applied.

About the Instructors: Dr. Jeffrey Wilson is a Professor of Statistics and Biostatistics at Arizona State University. Dr. Wilson's research experience includes grants as PI and co-PI from the NIH, NSF, USDA, Arizona Department of Health Services, and the Arizona Disease Research Commission. He is presently the Statistics Associate Editor for The Journal of Minimally Invasive Gynecology and a former Chair of the Editorial Board of the American Journal of Public Health. He has published more than 85 articles in leading journals such as Statistics in Medicine, American Journal of Public Health, Journal of Royal Statistics Society, Computational Statistics, and Australian Journal of Statistics, among others. He has consulted with pharmaceutical companies and hospitals while representing them before the FDA and other federal government healthcare agencies. He has taught specialized Biostatistics classes at Mayo Clinic. He has led similar courses for Phoenix Children's Hospital, Barrow Neurological Center, St. Joseph's Hospital, and Banner Hospital. He is the former Director of the School of Health Management and Policy He is a former Director and co-Director of the Biostatistics Core in the NIH Center for Alzheimer at Arizona State University.

Dr. (Din) Ding-Geng Chen is now the executive director and professor in biostatistics at College of Health Solutions, Arizona State University. He was the Wallace H. Kuralt distinguished professor in Biostatistics at University of North Carolina-Chapel Hill, a professor in biostatistics at the University of Rochester Medical Center, the Karl E. Peace endowed eminent scholar chair and professor in biostatistics from the Jiann-Ping Hsu College of Public Health at the Georgia Southern University. Dr. Chen is an elected fellow of the American Statistical Association (ASA), an elected member of the International Statistics Institute (ISI), and a senior expert consultant for biopharmaceuticals and government agencies with extensive expertise in clinical trial biostatistics. Dr. Chen has more than 200 referred professional publications and co-authored/co-edited 33 books on biostatistics clinical trials, biopharmaceutical statistics, interval-censored survival data analysis, meta-analysis, public health statistics, statistical causal inferences; statistical methods in big-data sciences and Monte-Carlo simulation-based statistical modeling. Dr. Chen has been invited nationally and internationally to give short courses at various scientific conferences.

SC04: Statistical methods for analyzing transmission and control of infectious diseases

Location and Time: HPNP G312, Sun, June 19, 8:30 - 12:30
Length: Half-day
Instructors: Dr. Ira Longini (University of Florida); Dr. Yang Yang (University of Florida); Dr. Matt Hitchings (University of Florida)

Abstract: Application of statistical inference methods to infectious disease data is a key tool in understanding transmissibility of pathogens and the effectiveness of interventions. In this half-day course, we will learn about different sources of data that arise from passive surveillance, active case finding and clinical studies, and methods for inferring key parameters from such data. The types of data sources to be covered include epidemic curve data, household-based observational data, and data arising from serosurveillance studies. We will also cover common computational algorithms for statistical inference and a few software packages that implement these algorithms. In addition, we will briefly introduce several advances in modeling frameworks to address challenges arising from the pandemic of COVID-19. Upon completion of this course, participants will recognize the various types of infectious disease data, common models designed to analyze these data, key parameters of epidemiological importance including intervention efficacies, and promising research directions in the field of infectious disease modeling.

Teaching Plan: The course will be divided into three sessions each of 70min, with two 15-min breaks.

First session: History of infectious disease modeling; types of infectious disease data (case numbers, serology, household data including time of symptom onset) and the underlying hierarchy of information; Overview of transmission parameters of epidemiological importance such as the basic reproductive number, final attack rate, and secondary attack rate; Different measures of vaccine efficacies and effectiveness of vaccination programs.

Second session: Detail on classic models that are fitted to epidemic curve data, final size models with fixed and random infectious periods for close contact groups (e.g., households), discrete-time chain binomial models and continuous-time survival models for sequential data of symptom onsets or laboratory confirmations among close contact groups, statistical inference from serosurveillance data, and agent-based models.

Third session: Computational methods (EM and Monte Carlo EM algorithms, traditional MCMC, Approximate Bayesian Computing, Particle Filtering, and Hamiltonian Monte Carlo). We will introduce a few R packages (e.g. surveillance, transtat, serosolver) and show some data examples; recent advances in statistical transmission models to address challenges the a rose during the pandemic of COVID-19 (e.g., presymptomatic and asymptomatic infectiousness, under-testing, delayed reporting, etc.).

About the Instructors: Dr. Ira Longini is a professor of biostatistics in the College of Public Health and Health professions as well as Emerging Pathogens Institute at the University of Florida. He works on the mathematical modeling, stochastic processes and biostatistics applied to epidemiological infectious disease problems. He has specialized in the mathematical and statistical theory of epidemics–a process that involves constructing and analyzing mathematical models of disease transmission, disease progression and the analysis of infectious disease data based on these models. In addition, he works extensively in the design and analysis of vaccine and infectious disease prevention trials and observational studies.

Dr. Yang Yang is an associate professor of biostatistics in the College of Public Health and Health professions as well as Emerging Pathogens Institute at the University of Florida. His research focuses on statistical methods for disease transmission dynamics, efficacy evaluation, missing data and surveillance bias. He also works on ecological modeling and genetic association for clinical outcomes.

Dr. Matt Hitchings is an Assistant Professor in the Department of Biostatistics at the University of Florida. His primary focus is evaluating the effectiveness of interventions against infectious disease, through clinical trials, observational studies, and development and application of mathematical models. Recently he has been conducting observational studies of vaccine effectiveness using passive surveillance data in Brazil, and developing a framework for analysis of serological data for pathogens including SARS-CoV-2 and dengue virus.

SC05: Spatial analysis with Gaussian Markov random fields

Location and Time: HPNP G114, Sun, June 19, 13:30 - 17:30
Length: Half-day
Instructors: Dr. Debashis Mondal (Washington University)

Abstract: Gaussian Markov random fields have been applied with much success to account for discrete spatial variation in both lattice and areal unit data. Applications include astronomy, agriculture, computer vision, climate studies, epidemiology, image analysis, geology and other areas of environmental science. Lattice-based Gaussian Markov random fields are extremely adaptable to swift and uncomplicated statistical computations and provide ways to develop complex and hierarchical models through local specifications, and, for these reasons, have contributed to considerable success in the analysis of spatial data. This short course gives an introduction to spatial models based on Gaussian Markov random fields. The course covers statistical computation for spatial linear mixed models, particularly, residual maximum likelihood (REML) estimation and kriging or

prediction. The course also presents statistical computation for general spatial mixed models using Markov Chain Monte Carlo (MCMC) sampling methods. Practicum sessions will introduce various R codes with applications from environmental sciences and geographical epidemiology.
The course will end with a summary of the topics and ideas covered and a list of further resources.

Teaching Plan: Lecture 1: Introduction to spatial statistics, Gaussian Markov random fields, conditionals and intrinsic autoregressions.

Lecture 2: Spatial mixed models, REML, kriging, h-likelihood and MCMC computations.

Break

Lecture 3 and 4: Statistical calculations using R-codes. Applications from environmental sciences and geographical epidemiology.

Summary and further resources.

About the Instructors: Debashis Mondal, PhD, is an associate professor in the Department of Mathematics and Statistics at Washington University in St Louis. Mondal's research interests include spatial statistics; computational science and machine learning; and applications in environmental sciences, ecology, including microbial ecology, and geographical epidemiology. Mondal won an NSF CAREER Award in 2013 and the International Indian Statistical Association's Young Researcher Award in 2015. He is also an elected member of the International Statistical Institute. Mondal earned his doctorate in statistics at the University of Washington, Seattle.

SC06: Bayesian Computational Tools for Clinical Data

Location and Time: HPNP G312, Sun, June 19, 13:30 - 17:30
Length: Half-day
Instructor: Prof. Sujit Ghosh (North Carolina State University); Dr. Amy Shi (AstraZeneca Pharmaceutical)

Abstract: The Bayesian paradigm provides a structured and practical way of expressing complicated models through a sequence of simple conditional distributions making them useful for simple to complex data structures required to address multiple phases of clinical trials, particularly for those that involves different types of data irregularities (missing values, censored data, etc.). Over the recent years there have been tremendous efforts on developing Bayesian analytics for leveraging data from sources outside of prospectively designed study, referred to as external data such as various Real-World-Data (RWD) sources, historical clinical data, and data from multiple trials within a grand hierarchical structure. Thus, development of appropriate statistical models and related inference are warranted that are not only based on solid theoretical guarantees but also making sure that such complex models are estimable and interpretable in practical settings for modern clinical trials. Thus, one of the main goals of the proposed short course is to present the modern analytical tools that are easily accessible to practitioners by providing a glimpse of theoretical backgrounds supplemented by many practical examples derived from real case studies. This will be accomplished by illustrating numerous real-data examples (using software demos) ranging from two-arm trials to more complex hierarchical models that involves handling data irregularities commonly faced by practitioners.

Teaching Plan: The first part of the short course will begin with a brief overview of Bayesian machine learning (BML) methods for randomized controlled trials (RCTs) using various study designs including sample size determination methods. In particular, it will showcase the use of Bayesian posterior predictive methods for properly handling missing and censored data, a feature that are not readily employed my routine ML methods. The second part of the course will involve more realistic and complex models that have recently emerged in the modern era used by pharmaceutical industries and regulatory agencies, and then showcase the use of modern BML methods through various real case studies. Throughout the tutorial practical applications and worked-out examples will be emphasized without getting into the theoretical underpinnings of the methods, but relevant literature will be provided for those wishing to learn more in-depth notions of BML tools. The concepts and methods discussed will be demonstrated using the popular software packages (R and SAS) developed by the presenters, but those are implementable by any other software capable of coding Markov Chain Monte Carlo (MCMC) methods.

The two-parts of the course will consist of the following topics:

Part I - Introduction to Bayesian Methods for Clinical Trials

1. Basics of Bayesian Methods for RCTs (20min)
2. Predictive Distributions and Sample Size Determination (20min)
3. Computational Methods using Monte Carlo Methods (35min)
4. Primer on Bayesian Software (via R, Stan and SAS) (30min)

(15min break)

Part II – BML methods with real-data examples

1. Bayesian regression models using 'brms' R package (35min)
2. GLMs and Multi-level models PROC BGLIMM (40min)
3. Penalized regression models with data irregularities (30min)
4. Q&As and additional demos on demand (15min)

About the Instructor: Professor Sujit Kumar Ghosh has

over 25 years of experience in conducting, applying, evaluating and documenting statistical analysis of biomedical and environmental data. Prof. Ghosh is actively involved in teaching, supervising and mentoring graduate students at the doctoral and master levels. He has supervised over 40 doctoral graduate students and published over 125 peer-reviewed journal articles in various areas of statistics with applications in biomedical and environmental sciences, econometrics and engineering. He has recently co-authored a book (with Dr. Reich) titled "Bayesian Statistical Methods," which is being used as a textbook at several universities. Prof. Ghosh has delivered over 180 invited lectures, seminars at national and international meetings. He has also delivered several short courses and served as short-term visiting professor at several institutions in various countries. Prof. Ghosh received the International Indian Statistical Association (IISA) Young Investigator Award in 2008; was elected a Fellow of the American Statistical Association (ASA) in 2009; was elected as the President of the NC Chapter of ASA in 2013 and also elected as the President of the IISA in 2017.

Dr. Amy Shi is currently a Statistical Science Associate Director at AstraZeneca Pharmaceutical in the Late CVRM (Cardiovascular Renal Metabolism) group. Much of her work involves with taking part in clinical trials as a statistician and researching for innovative statistical methods. Before joining AstraZeneca, she was a Principal Research Statistician Developer in the Bayesian Modeling Group at SAS from 2010 to 2021. Her job responsibility was to enhance the Bayesian capabilities of SAS software, with a focus on generalized linear mixed models, multilevel hierarchical settings, variable selection, choice modeling, and machine learning. She developed a couple of SAS Bayesian procedures (PROC BCHOICE and PROC BGLIMM) and many functional packages. Dr. Shi has a MS in Statistics from the Michigan State University and a Ph.D. in Biostatistics from the University of North Carolina at Chapel Hill.

Scientific Program (Mon, Jun. 20 - Wed, Jun. 22)

Welcome and Opening Remarks: Mon, June 20 8:00-8:30 (EDT)

Session W : Welcome and Opening Remarks
Location: HPNP Auditorium (1404)
Organizer: Symposium Organizing Committee.
Chair: Peihua Qiu, Ph.D., University of Florida.

8:00-8:05 Welcome - Dr. Peihua Qiu, Chair of Department of Biostatistics

8:05-8:10 Welcome - Dr. Michael Perri, Dean of College of Public Health and Health Professions

8:10-8:20 Welcome - Dr. Zhezhen Jin, President of International Chinese Statistical Association

8:20-8:25 Welcome - Somnath Datta, Co-chair of organizing committee

8:25-8:30 Opening Remarks - Ji-Hyun Lee, Chair of local committee

Plenary Keynote Talk 1: Mon, June 20 8:30-9:30 (EDT)

Session P1 : Plenary Keynote Talk 1
Location: HPNP Auditorium (1404)
Organizer: ICSA Special Lecture Committee.
Chair: Samuel Wu, Ph.D., University of Florida.

8:30-9:30 Change detection, estimation, and segmentation
David O. Siegmund. Stanford University

Sessions 1A-1H: Mon, June 20 10:00-11:40 (EDT)

Session 1A : Causal Inference And Its Applications
Location: HPNP G312
Organizer: Xinping Cui, University of California, Riverside, Esra Kurum, University of California, Riverside.
Chair: Xinping Cui, University of California, Riverside.

10:00-10:25 A causal approach to functional mediation analysis with application to a smoking cessation intervention
Donna Coffman. Temple University

10:25-10:50 Estimating the Average Treatment Effect in Randomized Clinical Trials with All-or-None Compliance
Zhiwei Zhang. NIH/NCI

10:50-11:15 Survey Weighting Strategies In Causal Mediation Analysis
Haoyu Zhou. Temple University

11:15-11:40 Discussion: Causal Inference and its Applications
Esra Kurum. University of California, Riverside

Session 1B : Latent Variable Models In The Data Science Era
Location: HPNP G112
Organizer: Yuqi Gu, Columbia University, Gongjun Xu, University of Michigan.
Chair: Yuqi Gu, Columbia University.

10:00-10:25 Identifiable Deep Generative Models via Sparse Decoding
♦*Gemma Moran*[1]*, Dhanya Sridhar*[2]*, Yixin Wang*[3] *and David Blei*[1]. [1]Columbia University [2]Mila and Universite de Montreal [3]University of Michigan

10:25-10:50 Population-Level Balance in Signed Networks
♦*Weijing Tang and Ji Zhu*. University of Michigan

10:50-11:15 Likelihood estimation of sparse topic distributions in topic models and its applications to Wasserstein document distance calculations
♦*Xin Bing, Florentina Bunea, Marten Wegkamp and Seth-Strimas Mackey*. Cornell University

11:15-11:40 High-dimensional principle component analysis with heterogeneous missingness
♦*Ziwei Zhu*[1]*, Tengyao Wang*[2] *and Richard Samworth*[3]. [1]University of Michigan, Ann Arbor [2]London School of Economics [3]University of Cambridge

Session 1C : Some Advances In Statistical Machine Learning
Location: HPNP G101
Organizer: Taps Maiti, Michigan State University.
Chair: Vojtech Kejzlar, Skidmore College.

10:00-10:25 Structurally Sparse Bayesian Neural Networks: Spike and Slab Shrinkage Priors
Sanket Jantre, ♦*Shrijita Bhattacharya and Tapabrata Maiti*. Michigan State University

10:25-10:50 An Adaptive Stochastic Approximation Algorithm for Randomized Decision GAN
Faming Liang. Purdue University

10:50-11:15 Volcano and valley prior with adhesive shrinkage for high dimentional data
Liangliang Zhang. case western reserve university

11:15-11:40 Information-preserving Bayesian models for efficient and robust learning
Sandeep Madireddy. Argonne NationalLaboratory

Session 1D : Machine Learning/Artificial Intelligence In Biomedical Research With 'big' Data
Location: HPNP G103
Organizer: Xiang-Yang Lou, University of Florida/Department of Biostatistics, Qing Lu, University of Florida/Department of Biostatistics.
Chair: Xiang-Yang Lou, University of Florida/Department of Biostatistics.

10:00-10:25 AI for Regulatory Science
Weida Tong. FDA

10:25-10:50 Causal networks for drug discovery

♦*Tao Xu*[1]*, Shicheng Guo*[2]*, Jinyung Zhao*[1] *and Momiao Xiong*[3]. [1]University of Florida [2]University of Wisconsin-Madison [3]University of Texas Health Science Center at Houston

10:50-11:15 New Toolkits for Disease Network Biology

Jake Chen. UAB Informatics Institute

11:15-11:40 Achieving Differential Privacy with Matrix Masking in Big Data

Aidong Ding[1], ♦*Samuel Wu*[2]*, Guanhong Miao*[2] *and Shigang Chen*[2]. [1]Northeastern University [2]University of Florida

Session 1E : Statistical Challenges And Advances In Complex Data Analysis

Location: HPNP G114

Organizer: Yichuan Zhao, Georgia State University.

Chair: Yichuan Zhao, Georgia State University.

10:00-10:25 Nontraditional Statistical Methods based on Wasserstein Distances and Conformal Prediction Set

Xiaoming Huo. Georgia Institute of Technology

10:25-10:50 Bayesian Spatially Varying Weight Neural Networks with the Soft-Thresholded Gaussian Process Prior

Jian Kang. University of Michigan

10:50-11:15 Some Recent Advances on the analysis of Interval-Censored Case-cohort Failure Time Data

(Tony) Jianguo Sun. University of Missouri

11:15-11:40 An Efficient Method for Clustering Multivariate Longitudinal Data

Junyi Zhou[1], ♦*Ying Zhang*[2] *and Wanzhu Tu*[3]. [1]Amgen Inc [2]UNMC [3]Indiana University

Session 1F : Statistical Methods And Applications For Analyzing Real-World Data

Location: HPNP G301

Organizer: Kelly Zou, Viatris.

Chair: Ying Lu, Stanford University.

10:00-10:25 WeightP2V: a flexible risk prediction framework with patient representation weighted by medical concepts

Jia Guo and ♦*Shuang Wang*. Columbia University

10:25-10:50 Efficient Algorithms and Implementation of a Semiparametric Joint Model for Longitudinal and Competing Risks Data: With Applications to Massive Biobank Data

Shanpeng Li[1]*, Ning Li*[1]*, Hong Wang*[2]*, Jin Zhou*[1]*, Hua Zhou*[1] *and* ♦*Gang Li*[1]. [1]UCLA [2]Central South University

10:50-11:15 A statistical quality assessment method for longitudinal observations in electronic health record data with an application to the VA million veteran program

Hui Wang[1]*, Ilana Belitskaya-Levy*[1]*, Fan Wu*[1]*, Jennifer Lee*[2]*, Mei-Chiung Shih*[1]*, Philip Tsao*[2] *and* ♦*Ying Lu*. [1]Department of Veterans Affairs, Palo Alto, CA, USA [2]Stanford University

11:15-11:40 Floor Discussion.

Session 1G : Recent Advances In Survival And Recurrent Events Analysis For Complex Data Structures

Location: HPNP 1101

Organizer: Dongdong Li, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute.

Chair: Dongdong Li, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute.

10:00-10:25 Structured variable selection in Cox model with time-dependent covariates

♦*Guanbo Wang*[1]*, Yi Yang*[1]*, Mirelle Schnitzer*[2]*, Tom Chen*[3]*, Rui Wang*[3] *and Robert Platt*[1]. [1]McGill University [2]University of Montreal [3]Harvard University

10:25-10:50 Robust Estimation for Recurrent Event Analysis in the Presence of Informative Event Censoring

♦*Tom Chen*[1]*, Rui Wang*[1] *and Victor Degruttola*[2]. [1]Harvard Pilgrim Health Care and Harvard Medical School [2]Harvard School of Public Health

10:50-11:15 Variance Estimation for Cox Model When Using Propensity Score Weighting

♦*Di Shu*[1]*, Jessica G Young*[2]*, Sengwee Toh*[2] *and Rui Wang*[2]. [1]University of Pennsylvania [2]Harvard University

11:15-11:40 Statistical Analysis of Recurrent Events from Administrative Databases

Yi Xiong. Fred Hutchinson Cancer Center

Session 1H : Statistical Inference For Two-Phase Studies With Outcome-Dependent Sampling

Location: HPNP 1102

Organizer: Natalie DelRocco, University of Florida Department of Biostatistics.

Chair: Adam Ding, Northeastern University Department of Mathematics.

10:00-10:25 Robust methods for Two-Phase Studies under generalized linear models

♦*Jacob Maronge*[1]*, Jonathan Schildcrout*[2] *and Paul Rathouz*[3]. [1]University of Texas MD Anderson Cancer Center [2]Vanderbilt University Medical Center [3]Dell Medical School at the University of Texas at Austin

10:25-10:50 Epidemiological Study Designs for Quantitative Longitudinal Data

♦*Jonathan Schildcrout, Chiara Digravio and Ran Tao*. VUMC

10:50-11:15 Statistical Methods for Selective Biomarker Testing in Two-Phase Studies

♦*Natalie Delrocco*[1]*, Adam Ding*[2] *and Samuel Wu*[1]. [1]University of Florida [2]Northeastern University

11:15-11:40 Design and Analysis Strategies with "Secondary" Use Data

Sarah Lotspeich. UNC

Sessions 2B-2H: Mon, June 20 13:00-14:40 (EDT)

Session 2B : Advanced Research In Bio-Molecular And Imaging Data By Our Young Researchers

Location: HPNP G112
Organizer: Susmita Datta, Department of Biostatistics, University of Florida.
Chair: Zhigang Li, Department of Biostatistics, University of Florida.

13:00-13:25 Outcome-guided Bayesian Clustering for Disease Subtype Discovery Using High-dimensional Transcriptomic Data
Lingsong Meng and ♦*Zhiguang Huo*. Department of Biostatistics, University of Florida

13:25-13:50 Double soft-thresholded multigroup model for vector-valued image regression with application to DTI imaging
♦*Arkaprava Roy*[1] *and Zhou Lan*[2]. [1]University of Florida [2]Yale University

13:50-14:15 Joint analysis and visualization of DNA methylation and nucleosome occupancy in single-molecule and single-cell data
Rhonda Bacher. University of Florida

14:15-14:40 Unity in diversity: Commonalities in these three different data analytical techniques
Arkaprava Roy, Rhonda L Bacher, Zhiguang Huo and ♦*Susmita Datta*. University of Florida

Session 2C : Emerging Topics In Statistical Learning For Biomedical Data

Location: HPNP G101
Organizer: Li-Xuan Qin, Memorial Sloan Kettering Cancer Center.
Chair: Carrie Wright, Johns Hopkins University.

13:00-13:25 A Semiparametric Approach to Developing Well-calibrated Models for Predicting Binary Outcomes
♦*Yaqi Cao*[1]*, Ying Yang*[2] *and Jinbo Chen*[1]. [1]University of Pennsylvania [2]Tsinghua University

13:25-13:50 How does data preprocessing impact statistical learning in microRNA studies?
Li-Xuan Qin. MSKCC

13:50-14:15 A Bayesian Reinforcement Learning Approach for Optimizing Combination Antiretroviral Therapy in People with HIV
♦*Yanxun Xu*[1]*, Wei Jin*[1]*, Yang Ni*[2] *and Leah Rubin*[1]. [1]Johns Hopkins University [2]Texas A&M University

14:15-14:40 HID machine: A Random Forest-based High Order Interaction Discovery Method for High-Dimensional Genomic Data
♦*Min Lu, Yifan Sha and Xi Chen*. University of Miami

Session 2D : Statistics In Biosciences (Sibs): Real World Challenges And Recent Methodological Developments

Location: HPNP G103
Organizer: X. Joan Hu, Department of Statistics and Actuarial Science Simon Fraser University, Hongzhe Lee, University of Pennsylvania.
Chair: Hongkai Ji, Johns Hopkins Bloomberg School of Public Health.

13:00-13:25 Multi-sample single-cell RNA-seq data analysis and visualization - methods, software, and benchmark
♦*Hongkai Ji*[1]*, Boyang Zhang*[1]*, Wenpin Hou*[1]*, Zhicheng Ji*[2]*, Zeyu Chen*[3]*, E John Wherry*[3] *and Stephanie Hicks*. [1]Johns Hopkins Bloomberg School of Public Health [2]Duke University School of Medicine [3]University of Pennsylvania Perelman School of Medicine

13:25-13:50 An efficient segmentation algorithm to estimate sleep duration from actigraphy data
Jonggyu Baek[1]*,* ♦*Margaret Banker*[2]*, Erica Jensen*[2]*, Xichen She*[2]*, Karen Peterson*[2]*, Andrew Pitchford*[3] *and* ♦*Peter Song*. [1]University of Massachusetts Medical School [2]University of Michigan [3]Iowa State University

13:50-14:15 Semiparametric estimation for length-biased interval-censored data with a cure fraction
Pao-Sheng Shen[1]*,* ♦*Yingwei Peng*[2]*, Hsin-Jen Chen*[3] *and Chyong-Mei Chen*[3]. [1]Tunghai University [2]Queen's University [3]National Yang Ming Chiao Tung University

14:15-14:40 Floor Discussion.

Session 2E : Some Recent Methods For Sequential Monitoring Of Complex Data

Location: HPNP G114
Organizer: Peihua Qiu, University of Florida.
Chair: Peihua Qiu, University of Florida.

13:00-13:25 A Robust Dynamic Screening System By Estimation of the Longitudinal Data Distribution
♦*Lu You*[1] *and Peihua Qiu*[2]. [1]University of South Florida [2]University of Florida

13:25-13:50 Transparent Sequential Learning for Statistical Process Control
Peihua Qiu. Founding Chair, Department of Biostatistics

13:50-14:15 Statistical Quality Control Using Image Intelligence: A Sparse Learning Approach
Yicheng Kang. Bentley University

14:15-14:40 Adaptive Process Monitoring Using Covariate Information
♦*Kai Yang*[1] *and Peihua Qiu*[2]. [1]Medical College of Wisconsin [2]University of Florida

Session 2F : Big Data, Machine Learning And Graphical Methods

Location: HPNP G301
Organizer: Kelly Zou, Viatris.
Chair: Yuqi Gu, Columbia University.

13:00-13:25 A Latent State Space Model for Learning Brain Dynamics for Mental Disorders
Yuanjia Wang. Columbia University

13:25-13:50 Clinical practice management of primary open-angle glaucoma in the United States: An analysis of real-world evidence
Joseph Imperato[1]*, Kelly Zou*[2]*, Jim Li*[2] *and* ♦*Tarek Hassan*[3]. [1]IQVIA [2]Medical Analytics and Real-World Evidence, Viatris Inc [3]Global Therapeutic Area Lead, Ophthalmology, Viatris Inc

13:50-14:15 Bayesian Pyramids: Identifiable Multilayer Discrete Latent Structure Models for Discrete Data

♦*Yuqi Gu*[1] *and David Dunson*[2]. [1]Columbia University [2]Duke University

14:15-14:40 Role of AI/ML and Big Data Analytics in Drug and Digital Medicine Development

Peter Zhang. Otsuka Pharmaceuticals (US)

Session 2G : Recent Development In Survival Analysis In Clinical Trials

Location: HPNP 1101
Organizer: Tianmeng Lyu, Novartis, Dong Xi, Gilead Sciences.
Chair: Dong Xi, Gilead Sciences.

13:00-13:25 On the Use of Restricted Mean Survival Time in Time-to-Event Data Analysis

Lihui Zhao. Northwestern University

13:25-13:50 From Logic-respecting Efficacy Estimands to Logic-ensuring Analysis Principle for Time-to-event Endpoint in Randomized Clinical Trials with Subgroups

Yi Liu[1]*, Miao Yang*[1]*, Siyoen Kil*[2]*, Jiang Li*[3]*, Shoubhik Mondal*[4]*, Hong Tian*[3]*, Liwei Wang,* ♦*Yue Shentu*[5] *and Godwin Yung*[6]. [1]Nektar Therapeutics [2]LSK [3]Beigene [4]AstraZeneca [5]Daiichi Sankyo Inc. [6]Genentech

13:50-14:15 A MCP-Mod approach to designing and analyzing survival trials with potential non-proportional hazards

♦*Xiaodong Luo, Yuan Sun and Zhixing Xu*. Sanofi

14:15-14:40 Bayesian inference for a principal stratum estimand on recurrent events truncated by death

♦*Tianmeng Lyu, Björn Bornkamp, Guenther Mueller-Velten and Heinz Schmidli*. Novartis

Session 2H : Challenges And Recent Developments In Multi-Outcome Analysis

Location: HPNP 1102
Organizer: Ming Wang, Penn State College of Medicine.
Chair: Ming Wang, Penn State College of Medicine.

13:00-13:25 Alternative multivariate endpoints and related statistical models for clinical trials in Alzheimer disease

Guoqiao Wang. Division of Biostatistics, Washington University in St Louis

13:25-13:50 Joint multivariate copula-frailty modeling of multiple-type recurrent events and the terminal event

Menglu Liang and ♦*Ming Wang*. Penn State College of Medicine

13:50-14:15 Knowledge-guided Bayesian Factor Analysis for Integrative Analysis of Multi-Omics Data

Qiyiwen Zhang, Changgee Chang and ♦*Qi Long*. University of Pennsylvania

14:15-14:40 Synergistic Self-learning Approach to Establishing Individualized Treatment Rules from Multiple Benefit Outcomes in a Calcium Supplementation Trial

♦*Yiwang Zhou*[1] *and Peter Song*[2]. [1]Department of Biostatistics, St. Jude Children's Research Hospital [2]Department of Biostatistics, University of Michigan

Sessions 3A-3H: Mon, June 20, 15:00-16:40 (EDT)

Session 3A : Recent Advances In Statistical Methods For Causal Inference And Personalized Medicine

Location: HPNP G312
Organizer: Ming Wang, Penn State College of Medicine.
Chair: Ming Wang, Penn State College of Medicine.

15:00-15:25 Evaluating different methods for estimating optimal treatment based on observational data

Qian Xu, Qi Zheng and ♦*Maiying Kong*. University of Louisville

15:25-15:50 Evidence factors from multiple, possibly invalid, instrumental variables

Anqi Zhao[1]*,* ♦*Youjin Lee*[2]*, Dylan Small*[3] *and Bikram Karmakar*[4]. [1]National University of Singapore [2]Brown University [3]University of Pennsylvania [4]University of Florida

15:50-16:15 Estimation of marginal treatment effect on binary outcome with multiple robustness and information borrow from secondary outcomes

♦*Chixiang Chen*[1]*, Shuo Chen*[1]*, Qi Long*[2]*, Sudeshna Das*[3] *and Ming Wang*[4]. [1]University of Maryland, School of Medicine [2]University of Pennsylvania [3]Harvard Medical School [4]Pennsylvania State University

16:15-16:40 Estimation Of Marginal Treatment Effect On Binary Outcome With Multiple Robustness And Information Borrow From Secondary Outcomes

Xiasuan Cai[1]*, Xinru Wang*[1]*, Justin Baker*[2]*, Jukka-Pekka Onnela*[3] *and* ♦*Linda Valeri*[1]. [1]Columbia University [2]McLean Hospital [3]Harvard University

Session 3B : New Advances In High-Dimensional Data Analysis

Location: HPNP G112
Organizer: Arkaprava Roy, University of Florida.
Chair: Arkaprava Roy, University of Florida.

15:00-15:25 On Statistical Inference with High Dimensional Sparse CCA

♦*Nilanjana Laha*[1]*, Nathan Huey*[1]*, Brent Coull*[2] *and Raharshi Mukherjee*[1]. [1]Harvard University [2]Harvard Mukherjee

15:25-15:50 A Graphical Lasso model for Hermitian matrices to detect global time-lagged Teleconnections

♦*Indranil Sahoo*[1]*, Joseph Guinness*[2] *and Brian J. Reich*[3]. [1]Virginia Commonwealth University [2]Cornell University [3]North Carolina State University

15:50-16:15 Multilayer Adjusted Cluster Point Process Model: Application to Microbial Biofilm Image Data Analysis

♦*Suman Majumder*[1]*, Brent Coull*[1]*, Jessica Markwelch*[2]*, Floyd Dewhirst*[3]*, Jacqueline Starr*[4] *and Kyu Ha Lee*[1]. [1]Harvard T.H. Chan School of Public Health [2]Marine Biological Laboratories [3]Forsyth Institute [4]Brigham and Women's Hospital

16:15-16:40 Correlated Wishart Matrices Classification via an Expectation-Maximization Composite Likelihood-Based Algorithm
Zhou Lan. Yale University

Session 3C : Machine Learning And Deep Learning Methods For Complex And Big Data
Location: HPNP G101
Organizer: Yichuan Zhao, Georgia State University.
Chair: Yichuan Zhao, Georgia State University.

15:00-15:25 Generative models for diabetic retinopathy
Lingsong Zhang. Purdue University

15:25-15:50 Divide and conquer approaches for nonparametric regression and variable selection
Sapuni Chandrasena and ♦*Rong Liu*. University of Toledo

15:50-16:15 A Bayesian Semi-supervised Approach to Keyword Extraction with Only Positive and Unlabeled Data
Guanshen Wang[1], ♦*Yichen Cheng*[2], *Yusen Xia*[2], *Qiang Lin*[3] *and Xinlei Wang*[1]. [1]Southern Methodist University [2]Georgia State University [3]University of Science and Technology of China

16:15-16:40 Deep learning approaches for predicting virus-host interactions and drug response
Zhongming Zhao. University of Texas Health Science Center at Houston

Session 3D : Advance In Statistical Methods For Complex Data
Location: HPNP G103
Organizer: Dehan Kong, University of Toronto.
Chair: Dehan Kong, University of Toronto.

15:00-15:25 Predicting long-term breast cancer risk with mammogram imaging data
♦*Shu Jiang*[1], *Jiguo Cao*[2], *Bernard Rosner*[3] *and Graham Colditz*[1]. [1]Washington university school of medicine [2]Simon fraser university [3]Harvard School of medicine

15:25-15:50 Fighting Noise with Noise: Causal Inference with Many Candidate Instruments
♦*Xinyi Zhang, Linbo Wang, Stanislav Volgushev and Dehan Kong*. University of Toronto

15:50-16:15 Smooth nonparametric dynamic prediction for competing risks via deep learning
Zhiyang Zhou. University of Manitoba

16:15-16:40 Distributed Cox Proportional Hazards Model Using Summary-level Information
♦*Dongdong Li*[1], *Wenbin Lu*[2], *Di Shu*[3], *Sengwee Toh*[1] *and Rui Wang*[4]. [1]Harvard Medical School [2]North Carolina State University [3]University of Pennsylvania Perelman School of Medicine [4]Harvard Medical School and Harvard T.H. Chan School of Public Health

Session 3E : Recent Advancement In Statistical Learning Methods For High-Dimensional Biomedical Data
Location: HPNP G114
Organizer: Zhiguang Huo, Department of Biostatistics, University of Florida.
Chair: Zhiguang Huo, Department of Biostatistics, University of Florida.

15:00-15:25 On p-value combination of independent and frequent signals: asymptotic efficiency and Fisher ensemble
♦*Yusi Fang*[1], *Chung Chang*[2] *and George Tseng*[1]. [1]Biostatistics, University of Pittsburgh [2]Applied Math, National Sun Yat-sen University

15:25-15:50 Improve Health Equality for Polygenic Risk Score (PRS) by Joint Penalized Regression of GWAS Summary Statistics from Two Ancestries
♦*Peng Liu*[1], *Max G'sell*[1], *Bernie Delvin*[2] *and Kathryn Roeder*[1]. [1]Carnegie Mellon University [2]University of Pittsburgh

15:50-16:15 High-dimension to high-dimension screening for detecting genome-wide epigenetic regulators of gene expression
Hongjie Ke[1], *Zhao Ren*[2], *Shuo Chen*[1], *George Tseng*[2], *Jianfei Qi*[1] *and* ♦*Tianzhou Ma*[1]. [1]University of Maryland [2]University of Pittsburgh

16:15-16:40 The mediating role of neuroimaging data in age-related cognitive decline
♦*Hwiyoung Lee and Shuo Chen*. University of Maryland, Baltimore

Session 3F : Advanced Statistical Learning Methods For Dynamic Systems
Location: HPNP G301
Organizer: Rongjie Liu, Florida State University.
Chair: Rongjie Liu, Florida State University.

15:00-15:25 A Computing Algorithm for Parameter Estimation of Ultra-high Dimensional VAR Model
Hongyu Miao. Florida State University

15:25-15:50 Generalized Ordinary Differential Equation (GODE) Model and Its Link to Deep Learning
Hulin Wu. University of Texas Health Science Center at Houston

15:50-16:15 Nonparametric Bayesian Q-learning for adjusting partial compliance in multi-stage randomized trials
♦*Indrabati Bhattacharya, Brent Johnson and Ashkan Ertefaie*. University of Rochester

16:15-16:40 Dynamic Topological Data Analysis for Brain Networks
Moo Chung'. University of Wisconsin-Madison

Session 3G : Geometric Statistics In Medical Image Computing
Location: HPNP 1101
Organizer: Hani Doss, University of Florida.
Chair: Hani Doss, University of Florida.

15:00-15:25 Statistical Analysis of Shape Networks
♦*Anuj Srivastava, Xiaoyang Guo, Aditi Basu Bal and Tom Needham*. Florida State University

15:25-15:50 Feature Gradient Flow for Interpretation of Deep Learning Models
P. Thomas Fletcher. University of Virginia

15:50-16:15 Nested Homogeneous Spaces: Construction, Learning and Applications
Baba Vemuri. University of Florida

16:15-16:40 Integrated Construction of Multimodal Atlases with Structural Connectomes in the Space of Riemannian Metrics
Sarang Joshi. University of Utah

Session 3H: The Jiann-Ping Hsu Invited Session on Biostatistical and Regulatory Sciences
Location: HPNP 1102
Organizer: Lili Yu and Karl Peace, JPH College of Public Health, Georgia Southern University.
Chair: Lili Yu and Karl Peace, JPH College of Public Health, Georgia Southern University.

15:00-15:25 Covariate-Balancing-Aware Interpretable Deep Learning Models for Treatment Effect Estimation
♦*Kan Chen, Qishuo Yin and Qi Long*. University of Pennsylvania

15:25-15:50 How to Implement the "One Patient, One Vote" Principle under the Framework of Estimand?
♦*Naitee Ting*. Boehringer ingelheim

15:50-16:15 Cox Model for Weibull Survival Data
Mario Keko, ♦*Marwan Alsharman, Djhenne Dalmacy, Lili Yu*. Georgia Southern University

16:15-16:40 An Application of the Cure Model to A Cardiovascular Clinical Trial
♦*Varadan Sevilimedu, S Ma, P Hartigan, TC Kyriakides*. Memorial Sloan Kettering Cancer Center

Sessions 4A-4H: Mon, June 20, 17:00-18:40 (EDT)

Session 4A : Recent Developments For Causal Inference: Theory, Method, And Application (This session is co-sponsored by the Caucus for Women in Statistics (CWS))
Location: HPNP G312
Organizer: Guanyu Hu, University of Missouri.
Chair: Guanyu Hu, University of Missouri.

17:00-17:25 Calibrated Optimal Decision Making with Multiple Data Sources and Limited Outcome
♦*Hengrui Cai, Wenbin Lu and Rui Song*. North Carolina State University

17:25-17:50 A Focusing Framework for Testing Bi-Directional Causal Effects with GWAS Summary Data
Ting Ye. University of Washington

17:50-18:15 Sensitivity Analysis of Individual Treatment Effects: A Robust Conformal Inference Approach
Ying Jin[1], ♦*Zhimei Ren*[2] *and Emmanuel Candès*[1]. [1]Stanford University [2]University of Chicago

18:15-18:40 Causal inference of time-varying effects in non-stationary time series using mobile health data
♦*Xiaoxuan Cai*[1], *Jukka-Pekka Onnela*[2], *Justin Baker*[3], *Habib Rahimi-Eichi*[3] *and Linda Valeri*[1]. [1]Columbia University [2]Harvard University [3]McLean Hospital

Session 4B : High-Dimensional Statistical Inference For Big Complicated Data
Location: HPNP G112
Organizer: Gongjun Xu, Department of Statistics, University of Michigan, Yinqiu He, Data Science Institute, Columbia University.
Chair: Yinqiu He, Data Science Institute, Columbia University.

17:00-17:25 Anti-Concentration of Suprema of Gaussian Processes with Applications to High-Dimensional CLTs
Alexander Giessing. University of Washington

17:25-17:50 Multiple-Splitting Projection Test for High-Dimensional Mean Vectors
Wanjun Liu[1], ♦*Xiufan Yu*[2] *and Runze Li*[3]. [1]LinkedIn Corporation [2]University of Notre Dame [3]Penn State University

17:50-18:15 Two-sample hypothesis testing of multiple-network data
♦*Yinqiu He*[1], *Xuming He*[2], *Ji Zhu*[2] *and Gongjun Xu*[2]. [1]Columbia University [2]University of Michigan

18:15-18:40 Doubly Debiased Lasso: High-Dimensional Inference under Hidden Confounding
♦*Zijian Guo*[1], *Domagoj Cevid*[2] *and Peter Buhlmann*[2]. [1]Rutgers [2]ETH, Zurich

Session 4C : New Fronts In Joint Modeling And Machine Learning
Location: HPNP G101
Organizer: Zhigang Li, University of Florida.
Chair: Lihui Zhao, Northwestern University.

17:00-17:25 Joint modeling for longitudinal and interval censored survival data
Ding-Geng Chen. Arizona State University

17:25-17:50 Heterogeneous Data Integration And The Predictive Ability of Cancer Survival Models
Yi Guo. Health Outcomes & Biomedical Informatics, University of Florida

17:50-18:15 Regression Analysis of Mixed Panel-Count Data with Application to Cancer Studies
♦*Yimei Li*[1], *Liang Zhu*[2], *Lei Liu*[3] *and Leslie Robison*[4]. [1]St Jude Children's Research Hospital [2]Eisai [3]Washington University [4]St. Jude Children's Research Hospital

18:15-18:40 Joint modeling in presence of informative censoring in palliative care studies
♦*Quran Wu*[1], *Michael Daniels*[2], *Areej Jawahri*[3], *Marie Bakitas*[4] *and Zhigang Li*[1]. [1]Department of Biostatistics, University of Florida [2]Department of Statistics, University of Florida [3]Department of Oncology, Massachusetts General Hospital [4]School of Nursing, University of Alabama at Birmingham

Session 4D : Knowledge-Guided Machine Learning And Statistical Modeling In Longitudinal Studies With Survival Endpoints
Location: HPNP G103
Organizer: Colin Wu, National Heart, Lung and Blood Institute, Xin Tian, National Heart, Lung and Blood Institute.
Chair: Xin Tian, National Heart, Lung and Blood Institute.

17:00-17:25 Design and Analysis of a Multi-Platform Trial of Patients Hospitalized for COVID-19
♦*Eric Leifer*[1]*, Lucy Kornblith, Jeffrey Berger, Lana Castellucci, Michael Farkouh, Ewan Goligher, Patrick Lawler and Scott Berry*. [1]NIH/NHLI

17:25-17:50 Knowledge-Guided Model Building and Estimation with Time-to-Event Outcomes and Longitudinal Covariates
♦*Colin O. Wu*[1]*, Xiaoyang Ma and Xin Tian*. [1]Division of Intramural Research

17:50-18:15 Dynamic Risk Prediction Triggered by Intermediate Events Using Survival Tree Ensembles
♦*Yifei Sun*[1]*, Sy Han Chiou*[2]*, Colin Wu*[3]*, Meghan Mcgarry*[4] *and Chiung-Yu Huang*[4]. [1]Columbia Univeristy [2]University of Texas at Dallas [3]National Heart, Lung, and Blood Institute [4]University of California San Francisco

18:15-18:40 Dealing With Competing Risks in Clinical Trials
James Troendle. NIH

Session 4E : Robust Information Integration From Multiple Studies In Clinical And Biomedical Research

Location: HPNP G114
Organizer: Ming Wang, Pennsylvania State University, Chixiang Chen, University of Maryland.
Chair: Chixiang Chen, University of Maryland.

17:00-17:25 On multi-site collaboration, data sharing, and analytic strategy in medical research
♦*Jing Huang*[1]*, Rui Duan*[2] *and Yong Chen*[1]. [1]University of Pennsylvania [2]Harvard University

17:25-17:50 Integrating summary information from many external studies with heterogeneous populations
Peisong Han. University of Michigan

17:50-18:15 Data Integration Methods Targeting Underrepresented Populations in Precision Medicine
Rui Duan. Harvard University

18:15-18:40 Integrated Analysis of Randomized Clinical Trials with Real-World Data
♦*Xiaofei Wang*[1]*, Dasom Lee*[2] *and Shu Yang*[2]. [1]Duke University [2]NC State University

Session 4F : Statistical Innovation In Complex And High Dimensional Data

Location: HPNP G301
Organizer: Jiaying Weng, Bentley University.
Chair: Zi Ye, Lehigh University.

17:00-17:25 Change detection in certain random intensity-driven point processes through repeated testing
Moinak Bhaduri. Bentley University

17:25-17:50 A nonparametric multi-sample test for high-dimensional compositional data with applications to the human microbiome
Qingyang Zhang. University of Arkansas

17:50-18:15 Minimum discrepancy approach for dimension reduction by filtered feature
Pei Wang. Miami University

18:15-18:40 Nonparametric Mixture Model: Application in Contaminated Trials
Zi Ye. Lehigh University

Session 4G : Enhance Decision Making In Early Oncology Studies To Expedite Drug Development

Location: HPNP 1101
Organizer: Gaohong Dong, BeiGene.
Chair: Kathy Zhang, BeiGene.

17:00-17:25 A Bayesian hierarchical monitoring design for phase II cancer clinical trials: Incorporating information on response duration
♦*Jian Wang*[1]*, Jing Ning*[1]*, Junsheng Ma*[1]*, Chunyan Cai*[2] *and Naval Daver*[1]. [1]The University of Texas MD Anderson Cancer Center [2]Marketplace Data Science, Uber

17:25-17:50 Bayesian Interim Monitoring for Faster Decision-Making in Early Phase Trials
Victoria Chang, Kathy Zhang and ♦*Gaohong Dong*. BeiGene

17:50-18:15 Discussant: Ying Lu.

18:15-18:40 Floor Discussion.

Session 4H : Design And Analysis Of Computer Experiments

Location: HPNP 1102
Organizer: Abhyuday Mandal, University of Georgia.
Chair: Ting Zhang, University of Georgia.

17:00-17:25 Modeling and Active Learning for Experiments with Quantitative-Sequence Factors
Abhyuday Mandal. University of Georgia

17:25-17:50 Lioness Algorithm for Finding Optimal Design of Experiments
♦*Hongzhi Wang, Qian Xiao and Abhyuday Mandal*. University of Georgia

17:50-18:15 A Simulation Optimization Approach for Sequential Accelerated Life Testing via Approximate Bayesian Inference
Ye Chen[1]*,* ♦*Qiong Zhang*[2]*, Mingyang Li*[3] *and Wenjun Cai*[4]. [1]Virginia CommonwealthUniversity [2]Clemson University [3]USF [4]Virginia Tech

18:15-18:40 Optimal Crossover Designs for Generalized Linear Models
♦*Jeevan Jankar*[1]*, Abhyuday Mandal and Jie Yang*[2]. [1]University of Georgia [2]University of Georgia

Plenary Keynote Talk 2: Tue, June 21, 8:30-9:30 (EDT)

Session P2 : Plenary Keynote Talk 2

Location: HPNP Auditorium (1404)
Organizer: ICSA Special Lecture Committee.
Chair: Somnath Datta, Ph.D., University of Florida.

8:30-9:30 Measuring housing activeness from multi-source big data and machine learning
Jianqing Fan. Princeton University

Sessions 5A-5H: Tue, June 21, 10:00-11:40 (EDT)

Session 5A : Statistical Methodologies In Causal Inference With Application In Drug Development

Location: HPNP G312
Organizer: Jiarui Lu, Novartis Pharmaceuticals Corporation, Dong Xi, Gilead Sciences.
Chair: Tianmeng Lyu, Novartis Pharmaceuticals Corporation.

10:00-10:25 Time and Causality: Learning Causal Structures from Longitudinal Data
Siyi Deng[1], ♦*Jiarui Lu*[2] *and Dong Xi*[3]. [1]Cornell University [2]Novartis pharmaceuticals corporation [3]Gilead Sciences

10:25-10:50 Minimax optimal subgroup identification
♦*Matteo Bonvini*[1], *Edward H. Kennedy*[1] *and Luke J. Keele*[2]. [1]Carnegie Mellon University [2]University of Pennsylvania

10:50-11:15 A Bayesian Machine Learning Approach for Estimating Heterogeneous Survivor Causal Effects: Applications to a Critical Care Trial
♦*Xinyuan Chen*[1], *Michael O. Harhay*[2], *Guangyu Tong*[3] *and Fan Li*[3]. [1]Mississippi State University [2]University of Pennsylvania [3]Yale University

11:15-11:40 Application of the causal inference in estimands for a principal stratum in clinical trials
Yongming Qu. Eli Lilly and Company

Session 5B : Recent Developments Of Dimension Reduction In Integrating Big And Complex Data

Location: HPNP G112
Organizer: Zhihua Su, University of Florida.
Chair: Zhihua Su, University of Florida.

10:00-10:25 Nonlinear envelope model
♦*Bing Li*[1], *Zhihua Su*[2] *and Dennis Cook*[3]. [1]Penn State University [2]University of Florida [3]University of Minnesota

10:25-10:50 Asymptotic distribution for partial least square prediction when the number of sample is small
♦*Liliana Forzani*[1] *and R. Dennis Cook*[2]. [1]Universidad Nacional del Litoral [2]University of Minnesota

10:50-11:15 A unified framework to high dimensional sufficient dimension reduction
♦*Shanshan Ding*[1], *Wei Qian*[1] *and Lan Wang*[2]. [1]University of Delaware [2]University of Miami

11:15-11:40 Envelope-based Partial Least Squares with Application to Cytokine-based Biomarker Analysis for COVID-19
♦*Yeonhee Park*[1], *Zhihua Su*[2] *and Dongjun Chung*[3]. [1]University of Wisconsin [2]University of Florida [3]Ohio State University

Session 5C : Precision Digital Health Care Via Machine Learning (This session is co-sponsored by the Statistical Learning and Data Science (SLDS) Section of ASA)

Location: HPNP G101
Organizer: Glen Wright Colopy, LifeBell AI / ASA SL&DS Section Program Chair.
Chair: Samaneh Nasiri, Harvard Medical School.

10:00-10:25 Designing Reinforcement Learning Algorithms for Digital Interventions: Pre-implementation Guidelines
♦*Anna L. Trella*[1], *Kelly W. Zhang*[1], *Inbal Nahum-Shani*[2], *Vivek Shetty*[3], *Finale Doshi-Velez*[1] *and Susan A. Murphy*[1]. [1]Harvard University [2]University of Michigan [3]University of California, Los Angelas

10:25-10:50 Oblique random survival forests version 2.0: faster and more interpretable
♦*Byron Jaeger and Nicholas Pajewski*. Wake Forest School of Medicine

10:50-11:15 Going Beyond Spike-and-slab: L1-ball Sparsity Prior With Applications On Image Data Analysis
♦*Leo Duan and Maoran Xu*. University of Florida

11:15-11:40 Floor Discussion.

Session 5D : Statistical Methods For Complex And High Dimensional Data

Location: HPNP G103
Organizer: Xueying Tang, University of Arizona.
Chair: Xueying Tang, University of Arizona.

10:00-10:25 Consistent and scalable Bayesian joint variable and graph selection for disease diagnosis leveraging functional brain network
♦*Xuan Cao*[1] *and Kyoungjae Lee*[2]. [1]University of Cincinnati [2]Sungkyunkwan University

10:25-10:50 Bayesian mixture models, non-local prior formulations and MCMC algorithms
Jairo Alberto Fuquenepatino. UC Davis

10:50-11:15 Two-component Gibbs samplers: Convergence rate and asymptotic variance
♦*Qian Qin*[1] *and Galin Jones*[2]. [1]University of Minneota [2]University of Minnesota

11:15-11:40 Efficient Algorithms and Theory for High-Dimensional Bayesian Varying Coefficient Models
Ray Bai. University of South Carolina

Session 5E : Modern Streaming Data Analysis: Change-Point Problems And Applications

Location: HPNP G114
Organizer: Jie Chen, Augusta University, Yajun Mei, Georgia Institute of Technology.
Chair: Ruizhi Zhang, University of Nebraska- Lincoln.

10:00-10:25 Detection of multiple change points in multiple profiles
♦*Jie Chen*[1] *and Shirong Deng*[2]. [1]Augusta University [2]Wuhan University

10:25-10:50 Changepoint Analysis of Hourly Sky-cloudiness Conditions in Canada
Mo Li[1], ♦*Qiqi Lu*[1] *and Xiaolan Wang*[2]. [1]Virginia Commonwealth University [2]Environment and Climate Change Canada

10:50-11:15 Learning under concept drift
Yuekai Sun. University of Michigan

11:15-11:40 Inference for Gaussian Multiple Change-point Model via Bayesian Information Criterion
♦*Yue Niu*[1]*, Ning Hao*[1] *and Han Xiao*[2]. [1]University of Arizona [2]Rutgers University

Session 5F : Emerging Development In The Analysis Of Data With Complex Features
Location: HPNP G301
Organizer: Wenqing He, University of Western Ontario.
Chair: Wenqing He, University of Western Ontario.

10:00-10:25 Feature Screening with Large Scale and High Dimensional Survival Data
Grace Yi[1]*,* ♦*Wenqing He*[1] *and Raymond Carroll*[2]. [1]University of Western Ontario [2]Texas A&M University, University of Technology Sydney

10:25-10:50 Analysis of the Cox Model with Longitudinal Covariates with Measurement Errors and Partly Interval Censored Failure Times, with Application to an AIDS Clinical Trial
♦*Yanqing Sun*[1]*, Qingning Zhou*[1] *and Peter Gilbert*[2]. [1]University of North Carolina at Charlotte [2]Fred Hutchinson Cancer Research Center and University of Washington

10:50-11:15 Learning Optimal Dynamic Treatment Regimens Subject to Stagewise Risk Control
Mochuan Liu[1]*, Yuanjia Wang*[2]*, Haoda Fu*[3] *and* ♦*Donglin Zeng*[1]. [1]University of North Carolina [2]Columbia University [3]Eli Lilly and Company

11:15-11:40 A new Bayesian method for handling covariate measurement error and detection limit in regression models
♦*Muhire Kwizera*[1]*, Roderick Little*[2]*, Matthew Perzanowski*[3] *and Qixuan Chen*[1]. [1]Department of Biostatistics, Columbia University [2]Department of Biostatistics, University of Michigan [3]Department of Environmental Health Sciences, Columbia University

Session 5G : Statistical Leadership In Drug Development In The New Era Of Data Science
Location: HPNP 1101
Organizer: Yijie Zhou, Vertex Pharmaceuticals, Jun Zhao, Astellas Pharma.
Chair: Jun Zhao, Astellas Pharma.

10:00-10:25 Opportunities and Challenges of Using Real-world Data for Signal Identification and Evidence Generation to Inform Study Design and Scientific Questions in Medical Research
Yiyue Lou. Vertex Pharmaceuticals

10:25-10:50 Empowering Real-World Evidence Generation in Rare Conditions: Collaborative data initiatives
♦*Jia Zhong, James Signorovitch and Eric Wu*. Analysis Group

10:50-11:15 Assessing Mediational Processes using Joint Longitudinal Models in the Framework of Individual Measurement Occasions
♦*Jin Liu*[1]*, Robert Perera*[2] *and Yijie Zhou*[1]. [1]Vertex Pharmaceuticals [2]Virginia Commonwealth University

11:15-11:40 Discussion
Yijie Zhou. Vertex

Session 5H: Student Paper Competition Winners
Location: HPNP 1102
Organizer: Organizing Committee.
Chair: Organizing Committee.

10:00-10:25 Sensitivity Analysis under the f-Sensitivity Models: A Distributionally Robust Optimization Viewpoint
♦*Ying Jin*[1]*, Zhimei Ren*[2] *and Zhengyuan Zhou*[3]. [1]Stanford University [2]University of Chicago [3]New York University

10:25-10:50 Fast Distributed Principal Component Analysis for Large-Scale Federated Data
♦*Shuting Shen, Junwei Lu and Xihong Lin*. Harvard University

10:50-11:15 High-Dimensional Dynamic Process Monitoring By PCA-Based Sequential Learning
♦*Xiulin Xie and Peihua Qiu*. University of Florida

11:15-11:40 Supervised Learning of Physical Activity Features from Functional Accelerometer Data
♦*Margaret Banker and Peter X.K. Song*. University of Michigan

Special Invited Talks: Tue, June 21, 13:00-14:30 (EDT)

Session S1 : Special Invited Talks
Location: HPNP Auditorium (1404)
Organizer: Somnath Datta, Ph.D., University of Florida.
Chair: Ji-Hyun Lee, Ph.D., University of Florida.

13:00-13:45 Lessons Learned from the COVID-19 Pandemic: A Statistician's Reflection
Xihong Lin. Harvard University

13:45-14:30 Predictive model building through integration of information across disparate data sources and summary-statistics
Nilanjan Chatterjee. Johns Hopkins University

Sessions 6A-6H: Tue, June 21, 15:00-16:40 (EDT)

Session 6A : Recent Advances In Mendelian Randomization
Location: HPNP G312
Organizer: Chong Wu, Florida State University.
Chair: Chong Wu, Florida State University.

15:00-15:25 Inference of nonlinear causal effects with GWAS summary data
♦*Ben Dai*[1]*, Chunlin Li*[2]*, Haoran Xue*[2]*, Wei Pan*[2] *and Xiaotong Shen*[2]. [1]The Chinese University of Hong Kong [2]The University of Minnesota

15:25-15:50 Causal analysis with rerandomization estimators (CARE)
♦*Chong Wu*[1] *and Jingshen Wang*[2]. [1]FLORIDA STATE UNIVERSITY [2]University of California, Berkeley

15:50-16:15 Breaking the Winner's Curse in Mendelian Randomization: Rerandomized Inverse Variance Weighted Estimator
Xinwei Ma[1]*,* ♦*Jingshen Wang*[2] *and Chong Wu*[3]. [1]UC San Diego [2]UC Berkeley [3]Florida State University

16:15-16:40 Constrained maximum likelihood-based Mendelian randomization robust to both correlated and uncorrelated pleiotropic effects

♦*Haoran Xue*[1]*, Xiaotong Shen*[2] *and Wei Pan*[1]. [1]Division of Biostatistics, School of Public Health, University of Minnesota [2]School of Statistics, University of Minnesota

Session 6B : Recent Advances In Dimension Reduction Techniques

Location: HPNP G112

Organizer: Dipankar Bandyopadhyay, Virginia Commonwealth University.

Chair: Shanshan Ding, University of Delaware.

15:00-15:25 Significance testing for canonical correlation analysis in high dimensions

Ian Mckeague[1] *and* ♦*Xin Zhang*[2]. [1]Columbia University [2]Florida State University

15:25-15:50 Dimension Reduction Forests: Local Variable Importance using Structured Random Forests

♦*Joshua Loyal*[1]*, Ruoqing Zhu*[1]*, Yifan Cui*[2] *and Xin Zhang*[3]. [1]University of Illinois at Urbana-Champaign [2]National University of Singapore [3]Florida State University

15:50-16:15 Envelope model for function-on-function linear regression

♦*Zhihua Su*[1]*, Bing Li*[2] *and Dennis Cook*[3]. [1]University of Florida [2]Pennsylvania State University [3]University of Minnesota

16:15-16:40 Floor Discussion.

Session 6C : Statistical Methods For Assessing Genomic Heterogeneity

Location: HPNP G101

Organizer: Yuchao Jiang, University of North Carolina at Chapel Hill.

Chair: Yuchao Jiang, University of North Carolina at Chapel Hill.

15:00-15:25 Robust Statistical Inference for Cell Type Deconvolution

♦*Jingshu Wang and Dongyue Xie*. University of Chicago

15:25-15:50 Single-cell eco-evolutionary dynamics of intratumor heterogeneity

Meghan Ferrall-Fairbanks. University of Florida

15:50-16:15 Neural Network Models for Sequence-Based TCR and HLA Association Prediction

♦*Si Liu, Phil Bradley and Wei Sun*. Fred Hutchinson Cancer Center

16:15-16:40 A statistical framework for cell-type-specific transcriptomics-wide association studies with an application to breast cancer

Xiaoyu Song. Icahn School of Medicine at Mount Sinai

Session 6D : Novel Statistical Modeling And Computing Methods For Complex Data

Location: HPNP G103

Organizer: Victor Hugo Lachos Davila, University of Connecticut.

Chair: Victor Hugo Lachos Davila, University of Connecticut.

15:00-15:25 New Bounded response models for target variables

Jorge Bazan. USP

15:25-15:50 Penalized complexity priors for the skewness parameter of power links

♦*Jose Ordonez*[1]*, Marcos Prates*[2]*, Jorge Bazan*[3] *and Victor Lachos*[4]. [1]Federal University of Bahia [2]Federal University of Minas Gerais [3]ICMC - USP [4]University of Conneticut

15:50-16:15 Linear Mixed-effects Models For Censored Data With Serial Correlation Errors Using The Multivariate Student's T-distribution

♦*Kelin Zhong*[1]*, Rommy C. Olivari*[2]*, Aldo M. Garay*[2] *and Victor H. Lachos*[3]. [1]Department of Statistics, UConn [2]Department of Statistics, Federal University of Pernambuco [3]Department of Statistics, University of Connecticut

16:15-16:40 Floor Discussion.

Session 6E : Modern Streaming Data Analysis: Detection And Identification

Location: HPNP G114

Organizer: Jie Chen, Augusta University, Ruizhi Zhang, University of Nebraska – Lincoln.

Chair: Ruizhai Zhang, University of Nebraska – Lincoln.

15:00-15:25 Low-Rank Robust Subspace Tensor Clustering for Metro Passenger Flow Modeling

Nurretin Sergin, Jiuyun Hu and ♦*Hao Yan*. Arizona State University

15:25-15:50 Optimal Parallel Sequential Change Detection under Generalized Performance Measures

Zexian Lu[1]*, Yunxiao Chen*[2] *and* ♦*Xiaoou Li*[1]. [1]University of Minnesota [2]London School of Economics and Political Sciences

15:50-16:15 Differentially private approaches for streaming data analysis

Wanrong Zhang. Harvard University

16:15-16:40 Active sequential change-point detection under sampling control

Yajun Mei. Georgia Institute of Technology

Session 6F : Deep Learning With Application And Uncertainty Quantification

Location: HPNP G301

Organizer: Xinping Cui, University of California, Riverside.

Chair: Xinping Cui, University of California, Riverside.

15:00-15:25 Random walk with restart with graph embedded neural network to inform potential targets

♦*Yushi Liu, Bochao Jia and Rick Higgs*. Eli Lilly

15:25-15:50 Learning interactions in Reaction Diffusion Equation with Deep Learning

Sichen Chen[1]*,* ♦*Nicolas Brunel*[2]*, Xin Yang*[3] *and Xinping Cui*[1]. [1]Department of Statistics, University of California, Riverside [2]Laboratoire de Mathématiques et Modélisation d'Evry, ENSIIE [3]Department of Mathematics , University of California, Riverside

15:50-16:15 An optimal transport approach for selecting a representative subsample

Ping Ma. University of Georgia

16:15-16:40 Distribution-free uncertainty quantification for classification
♦*Sasha Podkopaev and Aaditya Ramdas*. Carnegie Mellon University

Session 6G : Recent Advances In Clinical Trial Design And Practice
Location: HPNP 1101
Organizer: Shu Wang, University of Florida, Chung-Chou (Joyce) Chang, University of Pittsburgh.
Chair: Chung-Chou (Joyce) Chang, University of Pittsburgh.

15:00-15:25 A hybrid efficacy/effectivenss estimand for binary composite endpoints in clinical trials
♦*Xingyuan Li and Nathan Morris*. Eli Lilly and Company

15:25-15:50 Bayesian adaptive model selection design for optimal biological dose finding in phase I/II clinical trials
Ruitao Lin. The University of Texas MD Anderson Cancer Center

15:50-16:15 A Simulation Study Evaluating Phase I Clinical Trial Designs for Combinational Agents
♦*Shu Wang, Elias Sayour and Ji-Hyun Lee*. University of Florida

16:15-16:40 Bayesian Response Adaptive Randomization Design with A Composite Endpoint of Mortality and Morbidity
♦*Zhongying Xu and Chung-Chou Chang*. University of Pittsburgh

Session 6H : New Developments In Modern Nonparametric Statistics And The Applications
Location: HPNP 1102
Organizer: Yichuan Zhao, Georgia State University.
Chair: Yichuan Zhao, Georgia State University.

15:00-15:25 Doubly robust U-statistic with applications
♦*Ao Yuan, Anqi Yin and Ming Tan*. Georgetown University

15:25-15:50 Joint Semiparametric Models for Case-Cohort Designs
Weibin Zhong[1] *and* ♦*Guoqing Diao*[2]. [1]Bristol Myers Squibb [2]George Washington University

15:50-16:15 Novel empirical likelihood inference for the mean difference with right-censored data
Kangni Alemdjrodo[1] *and* ♦*Yichuan Zhao*[2]. [1]Purdue University [2]Georgia State University

16:15-16:40 Asymptotic Normality of Gini Correlation in High Dimension with Applications to the K-sample Problem
♦*Yongli Sang*[1] *and Xin Dang*[2]. [1]University of Louisiana at Lafayette [2]University of Mississippi

Sessions 7A-7H: Tue, June 21, 17:00-18:40 (EDT)

Session 7A : Novel Statistical Methods For -Omic Data Analysis
Location: HPNP G312
Organizer: Xiaoyu Song, Icahn School of Medicine at Mount Sinai.
Chair: Xiaoyu Song, Icahn School of Medicine at Mount Sinai.

17:00-17:25 LongStrain: An integrated strain-level analytic pipeline utilizing longitudinal metagenomics data
Boyan Zhou and ♦*Huilin Li*. New York University

17:25-17:50 An all-in-one statistical framework that simulates realistic single-cell omics data and infers cell heterogeneity structure
Jingyi Jessica Li. UCLA

17:50-18:15 Data-Type Weighted Multi-Omics Spectral Clustering for Disease Subtyping
♦*Peifeng Ruan and Hongyu Zhao*. Yale University

18:15-18:40 Deep Learning Methods for Retinal Imaging Genetics
Wei Chen. University of Pittsburgh

Session 7B : Modern Time Series And Network Methods In Data Science.
Location: HPNP G112
Organizer: Xinping Cui, University of California, Riverside, Ping Ma, University of Georgia.
Chair: Ping Ma, University of Georgia.

17:00-17:25 Collaborative Spectral Clustering in Attributed Networks
Pengsheng Ji. Univ. of Georgia

17:25-17:50 High Quantile Regression for Tail Dependent Time Series
Ting Zhang. University of Georgia

17:50-18:15 Dimension Reduction in Time Series Under the Presence of Conditional Heteroscedasticity
Murilo Dasilva, ♦*T. N. Sriram and Yuan Ke*. University of Georgia

18:15-18:40 Multiple autocovariance changepoints problems in high-dimensional time series
Yuan Ke. University of Georgia

Session 7C : Innovative Approach Of Hidden Markov Model
Location: HPNP G101
Organizer: Hyoyoung Choo-Wosoba, National Cancer Institute, Center for Cancer Research, Biostatistics and Data Management Section.
Chair: Paul Albert, Branch Chief Senior Investigator.

17:00-17:25 Bayesian Semiparametric Hidden Markov Tensor Partition Models for Longitudinal Data with Local Variable Selection
Giorgio Paulon, Peter Mueller and ♦*Abhra Sarkar*. UT-Austin

17:25-17:50 Non-Standard Applications of Hidden Markov Models in the Biosciences
♦*Jordan Aron*[1]*, Matthew O. Gribble*[2]*, Li C. Cheung*[3]*, and Paul Albert*[3]. [1]University of Minnesota [2]University of Alabama at Birmingham School of Public Health [3]National Cancer Center

17:50-18:15 A hidden Markov model approach for a joinpoint trend analysis
♦*Hyoyoung Choo-Wosoba, Philip Rosenburg and Paul Albert*. National Cancer Institute

18:15-18:40 Discussant: Paul Albert.

Session 7D : Statistical Advances And Applications In Analyzing Large Scale & Multi-Omic Single-Cell Data
Location: HPNP G103
Organizer: Rhonda Bacher, University of Florida; Department of Biostatistics.
Chair: Rhonda Bacher, University of Florida; Department of Biostatistics.

17:00-17:25 iscTrack, a semi-supervised algorithm and interactive single-cell tool to track emerging transcriptional states in serial samples
Jiannong Li, Scott Cukras, Sathya Sriramareddy, Keiran Smalley, Xiaoqing Yu and ♦*Ann Chen*. Moffitt Cancer Center

17:25-17:50 Deep learning methods for cell type identification and gene expression imputation
Sijie Yao, Xiaoqing Yu and ♦*Xuefeng Wang*. Moffitt Cancer Center

17:50-18:15 Nonparametric Interrogation of Transcriptional Regulation in Single-Cell RNA and Chromatin Accessibility Multiomic Data
Yuchao Jiang. UNC Chapel Hill

18:15-18:40 A statistical framework for scRNA-seq data modeling: simulation and applications
♦*Guoshuai Cai*[1]*, Xizhi Luo*[1]*, Fei Qin*[1] *and Feifei Xiao*[2]. [1]University of South Carolina [2]University of Florida

Session 7E : Modern Streaming Data Analysis: Process Monitoring
Location: HPNP G114
Organizer: Jie Chen, Augusta University, Yajun Mei, Georgia Institute of Technology.
Chair: Yajun Mei, Georgia Institute of Technology.

17:00-17:25 Fault Classification for High-dimensional Data Streams: A Directional Diagnostic Framework Based on Multiple Hypothesis Testing
Dongdong Xiang. East China Normal University

17:25-17:50 Adversarially Robust Sequential Hypothesis Testing
Shuchen Cao[1]*,* ♦*Ruizhi Zhang*[1] *and Shaofeng Zou*[2]. [1]University of Nebraska-Lincoln [2]University at Buffalo, The State University of New York

17:50-18:15 Recent advances in quality and industrial analytics
Fugee Tsung. HKUST

18:15-18:40 Asymptotic Optimality Theory for Active Quickest Detection with Two Affected Streams
Qunzhi Xu. Georgia Institute of Technology

Session 7F : Discriminant And Cluster Analysis For Complex Data
Location: HPNP G301
Organizer: Xin (Henry) Zhang, Florida State University.
Chair: Guanyu Hu, University of Missouri.

17:00-17:25 Conditional probability tensor decompositions for multivariate categorical response regression
♦*Aaron Molstad*[1] *and Xin Zhang*[2]. [1]University of Florida [2]Florida State University

17:25-17:50 Quadratic Discriminant Analysis by Projection
Ruiyang Wu and ♦*Ning Hao*. University of Arizona

17:50-18:15 A Doubly-Enhanced EM Algorithm for Model-Based Tensor Clustering
♦*Qing Mai, Xin Zhang, Yuqing Pan and Kai Deng*. Florida State University

18:15-18:40 Stochastic Low-rank Tensor Bandits for Multi-dimensional Online Decision Making
Will Wei Sun. Purdue University

Session 7G : Design And Analysis In Vaccine Development And Its Challenges
Location: HPNP 1101
Organizer: Bo Fu, Sanofi, Jun Zhao, Astellas.
Chair: Bo Fu, Sanofi.

17:00-17:25 Assessing the Role of Antibody in Vaccine Protection
Dean Follmann. NIH

17:25-17:50 Sensitivity Analysis for Evaluating Principal Surrogate Endpoints Relaxing the Equal Early Clinical Risk Assumption
♦*Ying Huang, Yingying Zhuang and Peter Gilbert*. Fred Hutchinson Cancer Research Center

17:50-18:15 Durability of Covid-19 Vaccines
Yu Gu. University of North Carolina

18:15-18:40 Statistical Consideration for Accelerated COVID-19 Vaccine Clinical Development in the Pandemic
James Zhou. HHS/ASPR/BARDA

Session 7H : Methods For Inference On Variable Importance Using Machine Learning (This session is co-sponsored by the Statistical Learning and Data Science (SLDS) Section of ASA)
Location: HPNP 1102
Organizer: Brian Williamson, Kaiser Permanente Washington Health Research Institute.
Chair: Brian Williamson, Kaiser Permanente Washington Health Research Institute.

17:00-17:25 Inference for model-agnostic variable importance
♦*Brian Williamson*[1]*, Susan Shortreed*[1]*, Peter Gilbert*[2]*, Noah Simon*[3] *and Marco Carone*[3]. [1]Kaiser Permanente Washington Health Research Institute [2]Fred Hutchinson Cancer Research Center [3]University of Washington

17:25-17:50 Variable importance measure for spatial machine learning models with application to air pollution exposure prediction
♦*Si Cheng, Ali Shojaie, Lianne Sheppard and Adam Szpiro*. University of Washington

17:50-18:15 Floodgate: inference for model-free variable importance
♦*Lu Zhang and Lucas Janson*. Harvard University

18:15-18:40 Regularization on Ensembles of Tree and Variable importance
♦*Siyu Zhou and Lucas Mentch*. University of Pittsburgh

Banquet Talk: Tue, June 21, 20:00-20:45 (EDT)

Session S2 : Banquet Talk

Location: Ben Hill Griffin Stadium Champions Club (121 Gale Lemerand Drive)
Organizer: ICSA Special Lecture Committee.
Chair: Samuel Wu, Ph.D., University of Florida.

20:00-20:45 Lost in translation
Lee-Jen Wei. Harvard University

Plenary Keynote Talk 3: Wed, June 22, 8:30-9:30 (EDT)

Session P3 : Plenary Keynote Talk 3

Location: HPNP Auditorium (1404)
Organizer: ICSA Special Lecture Committee.
Chair: Guogen Shan, Ph.D., University of Florida.

8:30-9:30 Inference for longitudinal data after adaptive sampling
Susan Murphy. Harvard University

Sessions 8A-8H: Wed, June 22, 10:00-11:40 (EDT)

Session 8A : Ultra-High Dimensional Variable Selection And Zero-Inflated Negative Binomial Spatial And Temporal Regression

Location: HPNP G312
Organizer: Hsin-Hsiung Huang, University of Central Florida.
Chair: Hsin-Hsiung Huang, University of Central Florida.

10:00-10:25 Spatiotemproal Zero-Inflated Bayesian Negative Binomial Regression Using Nearest Neighbor Gaussian Process and Polya-Gamma Mixtures
♦*Qing He and Hsin-Hsiung Huang*. University of Central Florida

10:25-10:50 An Exchangeable Prior on Partitions for Clustering
♦*Charles Harrison, Qing He and Hsin-Hsiung Huang*. University of Central Florida

10:50-11:15 Multi-Omics Integrative Analysis for Incomplete Data Using Weighted p-value Adjustment Approaches
Wenda Zhang[1], *Joshua Habiger*[2], *Hsin-Hsiung Huang*[3] *and* ♦*Yen-Yi Ho*[1]. [1]University of South Carolina [2]Oklahoma State University [3]University of Central Florida

11:15-11:40 Sparse Bayesian Matrix-variate Regression with High-dimensional Binary Response Data
♦*Hsin-Hsiung Huang*[1] *and Shao-Shuan Wang*[2]. [1]University of Central Florida [2]National Central University

Session 8B : Recent Developments In Functional Data Analysis

Location: HPNP G112
Organizer: Gang Li, UCLA.
Chair: Gang Li, UCLA.

10:00-10:25 Multilevel Modeling of Spatially Nested Functional Data: Spatiotemporal Patterns of Hospitalization Rates in the U.S. Dialysis Population
Yihao Li[1], *Danh Nguyen*[2], *Sudipto Banerjee*[1], *Connie Rhee*[2], *Kamyar Kalantar-Zadeh*[2], *Esra Kurum*[3] *and* ♦*Damla Senturk*. [1]UCLA [2]UCI [3]UCRiverside

10:25-10:50 Online Estimation for Functional Data
Fang Yao. Beijing University

10:50-11:15 Functional ANOVA for High-Dimensional Spectral Analysis
♦*Robert Krafty*[1], *Marie Tuft*[2], *Fabio Ferrarelli*[3], *Ori Rosen*[4] *and Zeda Li*[5]. [1]Emory University [2]Sandia National Laboratory [3]University of Pittsburgh [4]University of Texas [5]Baruch College, City University of New York

11:15-11:40 Factor-augmented model for functional data
Yuan Gao[1], *Han Lin Shang*[2] *and* ♦*Yanrong Yang*[1]. [1]The Australian National University [2]Macquarie University

Session 8C : Recent Advances In Robust Statistical Models For Censored And Missing Data

Location: HPNP G101
Organizer: Vicror Hugo Lachos Davila, University of Connecticut.
Chair: Jorge Luis Bazan Guzman, University of Sao Paulo.

10:00-10:25 Censored autoregressive regression models with Student-t innovations
Katherine Andreina Loorvaleriano[1], ♦*Fernanda Langschumacher*[2], *Christian Galarza*[3] *and Larissa Avilamatos*[1]. [1]University of Campinas [2]Ohio State University [3]Escuela Superior Politécnica del Litoral

10:25-10:50 Lasso regularization for censored skew-t regression and high dimensional predictors
Victor Hugo Lachos. University of Connecticut

10:50-11:15 Extending multivariate Student's-t semiparametric mixed models for longitudinal data with censored responses and heavy tails
Thalita Mattos[1], *Victor Hugo Lachos*[2], *Luis Mauricio Castro*[3] *and* ♦*Larissa Matos*[1]. [1]Universidade Estadual de Campinas [2]University of Connecticut [3]Pontificia Universidad Católica de Chile

11:15-11:40 Floor Discussion.

Session 8D : Recent Advances In Latent Variable Analysis

Location: HPNP G103
Organizer: Gongjun Xu, University of Michigan.
Chair: Xiangbin Meng, Northeast Normal University.

10:00-10:25 A Note on Statistical Inference for Noisy Incomplete Binary Matrix
Gongjun Xu. University of Michigan

10:25-10:50 VEMIRT: A Variational EM Algorithm-based Shiny App for High-dimensional IRT Applications
♦*Chun Wang*[1], *Gongjun Xu*[2], *Chenchen Ma*[2], *Ruoyi Zhu*[1] *and Jiaying Xiao*[1]. [1]University of Washington [2]University of Michigan

10:50-11:15 A random effect hidden Markov model for process data
Xueying Tang. University of Arizona

11:15-11:40 Tree-informed Bayesian multi-source domain adaptation
♦*Zhenke Wu*[1]*, Zehang Li*[2]*, Irena Chen*[1] *and Mengbing Li*[1]. [1]University of Michigan, Ann Arbor [2]University of California, Santa Cruz

Session 8E : New Advances In Microbiome Related Data Analysis
Location: HPNP G114
Organizer: Zhigang Li, University of Florida.
Chair: Zhigang Li, University of Florida.

10:00-10:25 LinDA: Linear Models for Differential Abundance Analysis of Microbiome Compositional Data
♦*Jun Chen*[1] *and Xianyang Zhang*[2]. [1]Mayo Clinic [2]Texas A&M University

10:25-10:50 Identifying Microbial Interaction Networks Based on Irregularly Spaced Longitudinal Data
♦*Jiang Gui*[1]*, Jie Zhou*[1]*, Weston Viles*[2] *and Annie Hoen*[1]. [1]Dartmouth College [2]University of Southern Maine

10:50-11:15 Synergy Regression of Microbiome and Metabolome Data
Yue Wang. Arizona State University

11:15-11:40 A Novel Causal Mediation Analysis Approach for Zero-Inflated Count Mediators
♦*Meilin Jiang*[1]*, Seonjoo Lee*[2]*, A. James O'malley*[3]*, Yaakov Stern*[2] *and Zhigang Li*[1]. [1]University of Florida [2]Columbia University [3]Geisel School of Medicine at Dartmouth

Session 8F : Statistical Computation Of Big Data With Biomedical Applications
Location: HPNP G301
Organizer: Sharmistha Guha, Texas A&M University, Statistics.
Chair: Sharmistha Guha, Texas A&M University, Statistics.

10:00-10:25 Bayesian data compression
♦*Rajarshi Guhaniyogi*[1] *and Aaron Scheffler*[2]. [1]Texas A & M University [2]UC San Francisco

10:25-10:50 A 'Divide-and-Conquer' AECM Algorithm for Large non-Gaussian Longitudinal Data
♦*Reuben Retnam*[1]*, Sanvesh Srivastava*[2] *and Dipankar Bandyopadhyay*[1]. [1]Virginia Commonwealth University [2]University of Iowa

10:50-11:15 Bayesian Generalized Sparse Symmetric Tensor-on-Vector Regression
♦*Sharmistha Guha and Rajarshi Guhaniyogi*. Texas A&M University

11:15-11:40 Ordinal Causal Discovery
♦*Yang Ni and Bani Mallick*. Texas A&M University

Session 8G : Recent Development In Innovative Clinical Trial Designs
Location: HPNP 1101
Organizer: Dong Xi, Gilead Sciences.
Chair: Jiarui Lu, Novartis.

10:00-10:25 Graphical representation of the Hochberg procedure and other equally weighted tests
♦*Dong Xi*[1] *and Frank Bretz*[2]. [1]Gilead Sciences [2]Novartis

10:25-10:50 A unified framework for weighted parametric group sequential design (WPGSD)
Keaven Anderson, Zifang Guo, Jing Zhao and ♦*Linda Sun*. Merck & Co., Inc.

10:50-11:15 Statistical Interactions in a Clinical Trial
Naitee Ting. Boehringer Ingelheim Pharmaceuticals, Inc.

11:15-11:40 Deep historical borrowing framework in confirmatory clinical trials with multiple endpoints
Tianyu Zhan. AbbVie

Session 8H : Some Popular Applications In Data Integration
Location: HPNP 1102
Organizer: Gauri Datta, University of Georgia and U.S. Census Bureau.
Chair: Gauri Datta, University of Georgia and U.S. Census Bureau.

10:00-10:25 Multivariate Global-Local Priors for Small Area Estimation
Tamal Ghosh[1]*,* ♦*Malay Ghosh*[2] *and Jerry Maples*[3]. [1]Citibank, Tampa [2]University of Florida [3]United States Census Bureau

10:25-10:50 Pseudo-Bayesian Small Area Estimation
♦*Juhyung Lee*[1]*, Gauri Datta*[2] *and Jiacheng Li*[2]. [1]University of Florida [2]University of Georgia

10:50-11:15 Incorporating heterogeneous offsets in hierarchical disease mapping
♦*Emily Peterson and Lance Waller*. Emory University

11:15-11:40 Floor Discussion.

Sessions 9A-9H: Wed, June 22, 13:00-14:40 (EDT)

Session 9A : Bayesian Calibration Of Computer Models
Location: HPNP G312
Organizer: Vojtech Kejzlar, Skidmore College, Department of Mathematics and Statistics.
Chair: Tapabrata Maiti, Michigan State University, Department of Statistics and Probability.

13:00-13:25 On estimating photometric redshift of galaxies by augmenting observation with simulation
Arindam Fadikar. Argonne National Laboratory

13:25-13:50 A theoretical framework of the scaled Gaussian stochastic process in prediction and calibration
♦*Mengyang Gu*[1]*, Fangzheng Xie*[2] *and Long Wang*[3]. [1]University of California, Santa Barbara [2]Indiana University Bloomington [3]Johns Hopkins University

13:50-14:15 Bayesian Calibration and Model Mixing
Matthew Pratola. Dept. of Statistics, The Ohio State University

14:15-14:40 An efficient approach for computer model calibration with variational Bayesian inference
♦*Vojtech Kejzlar*[1] *and Taps Maiti*[2]. [1]Skidmore College [2]Michigan State University

Session 9B : Novel Developments For Functional Data Analysis
Location: HPNP G112
Organizer: Raymond Wong, Texas A&M University.
Chair: Muxuan Liang, University of Florida.

13:00-13:25 Adaptive Frequency Band Analysis for Functional Time Series
♦*Pramita Bagchi*[1] *and Scott Bruce*[2]. [1]George Mason University [2]Texas A&M University

13:25-13:50 Sliced Elastic Distance for Climate Model Validation
Robert Garrett[1], ♦*Trevor Harris*[2] *and Bo Li*[1]. [1]University of Illinois at Urbana-Champaign [2]Texas A&M University

13:50-14:15 MARGARITA: Marginal-Product Basis Representation for Multi-dimensional Functional Data Analysis
William Consagra, Arun Venkataraman and ♦*Xing Qiu*. University of Rochester

14:15-14:40 Floor Discussion.

Session 9C : Statistical Methods For High Dimensional Microbiome Data
Location: HPNP G101
Organizer: Somnath Datta, University of Florida, Subha Guha, University of Florida.
Chair: Subha Guha, University of Florida.

13:00-13:25 What Can We Learn About the Bias of Microbiome Studies from Analyzing Data from Mock Communities
Mo Li[1], ♦*Glen Satten*[2], *Ni Zhao*[1], *Angel Rivera*[3] *and Robert Tyx*[3]. [1]Johns Hopkins University [2]Emory University [3]CDC

13:25-13:50 Nonparametric Bayesian approaches for identifying differentially abundant genera between multiple groups in microbiome data
Archie Sachdeva, Somnath Datta and ♦*Subharup Guha*. University of Florida

13:50-14:15 Deep ensemble learning over the microbial phylogenetic tree (DeepEn-Phy)
♦*Wodan Ling*[1], *Youran Qi*[2], *Xing Hua*[1] *and Michael Wu*[1]. [1]Fred Hutchinson Cancer Center [2]Amazon

14:15-14:40 IFAA: Robust association identification and Inference For Absolute Abundance in microbiome analyses
Zhigang Li. University of Florida

Session 9D : Recent Advancements In Statistical Data Integration
Location: HPNP G103
Organizer: Jin Jin, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health.
Chair: Jin Jin, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health.

13:00-13:25 Meta Clustering for Collaborative Learning
♦*Chenglong Ye*[1], *Reza Ghanadan*[2] *and Jie Ding*[3]. [1]University of Kentucky [2]Google [3]University of Minnesota

13:25-13:50 Joint integrative analysis of dependent data sources
♦*Emily Hector*[1] *and Peter Song*[2]. [1]North Carolina State University [2]University of Michigan

13:50-14:15 Data Integration Via Analysis of Subspaces
♦*Jack Prothero*[1], *Meilei Jiang*[2], *Quoc Tran-Dinh*[3], *Jan Hannig*[3] *and J.s. Marron*[3]. [1]National Institute of Standards and Technology [2]Meta [3]UNC Chapel Hill

14:15-14:40 Synthetic-data-based transfer learning approach for multi-site risk prediction
♦*Tian Gu and Rui Duan*. Department of Biostatistics, Harvard T.H. Chan School of Public Health

Session 9E : Modern Business Statistical Analysis
Location: HPNP G114
Organizer: Aidong Adam Ding, Northeastern University, Shaobo Li, The University of Kansas.
Chair: Shaobo Li, The University of Kansas.

13:00-13:25 Penalized quantile regression
♦*Ben Sherwood and Shaobo Li*. University of Kansas

13:25-13:50 On the use of Minimum Penalties in Multivariate Regression
♦*Brad Price*[1] *and Ben Sherwood*[2]. [1]West Virginia University [2]University of Kansas

13:50-14:15 Joint Modeling of Playing Time and Purchase Propensity in Massively Multiplayer Online Role Playing Games Using Crossed Random Effects
Trambak Banerjee. University of Kansas

14:15-14:40 Measuring goodness-of-fit for bankruptcy prediction and its application to U.S. and Polish data
♦*Xiaorui Zhu and Dungang Liu*. University of Cincinnati

Session 9F : Application And Theory Of Statistical Test And Evaluation
Location: HPNP G301
Organizer: Aidong Adam Ding, Northeastern University.
Chair: Aidong Adam Ding, Northeastern University.

13:00-13:25 Statistical Evaluation of Deep Learning-based Side-channel Analysis
Aidong Ding. Northeastern University

13:25-13:50 Improved Meta-Analysis of ROC curves
♦*Buddika Peiris*[1] *and Shuang Yang*[2]. [1]Worceter Polytechnic Institute [2]Worcester Polytechnic Institute

13:50-14:15 Signal-noise ratio of genetic associations and statistical power of SNP-set tests
Hong Zhang[1], *Ming Liu*[2], *Jiashun Jin*[3] *and* ♦*Zheyang Wu*[2]. [1]Merck Research Laboratories [2]WPI [3]Carnegie Mellon University

14:15-14:40 BEAUTY Powered BEAST
Kai Zhang[1], ♦*Zhigen Zhao*[2] *and Wen Zhou*[3]. [1]UNC [2]Temple [3]Colorado State University

Session 9G : Statistical Challenges In Clinical Trials For Alzheimer Disease
Location: HPNP 1101
Organizer: Guoqiao Wang, Washington University in St. Louis, Changxing Ma, University at Buffalo.
Chair: Yan Li, Washington University in St. Louis.

13:00-13:25 Dose change and statistical power in the Aducanumab trial
Guogen Shan. University of Florida

13:25-13:50 A More Efficient Outcome for Alzheimer Disease Research: the Item Response Theory Based Score for the Clinical Dementia Rating (CDRR)

♦*Yan Li, Guoqiao Wang, Chengjie Xiong, Krista L Moulder and John C Morris*. Washington University in St. Louis

13:50-14:15 Floor Discussion.

Session 9H : Statistics Education In The Era Of Ai And Data Science

Location: HPNP 1102

Organizer: Steven Foti, University of Florida.

Chair: Steven Foti, University of Florida.

13:00-13:25 Interactive Graphics: A Bridge from Coding to Programming

Adam Loy. Carleton College

13:25-13:50 Case studies to community engagement: bringing hands-on data science experiences to the classroom

♦*Carrie Wright*[1]*, Stephanie Hicks*[1]*, Ava Hoffman*[1]*, Michael Rosenblum*[1]*, Michael Breshock*[1]*, Qier Meng*[1]*, Margaret Taub, Leah Jager*[1]*, Tyler Derreth*[1] *and Mindi Levin*[1]. [1]Johns Hopkins Bloomberg School of Public Health

13:50-14:15 Constructing a Modern Data Visualization Course: Topics, Reflections, and Feedback

Steven Foti. University of Florida

14:15-14:40 Foundations for NLP-assisted formative assessment feedback for short-answer tasks in large-enrollment classes

Susan Lloyd, ♦*Matthew Beckman, Dennis Pearl, Rebecca Passoneau, Zhaohui Li and Zekun Wang*. Penn State University

No.	Last Name	First Name	Title
1	Anyaso-Samuel	Samuel	Adjusting for informative cluster size in pseudo-value based regression approaches with clustered time to event data
2	Chakraborty	Nilanjana	A Bayesian framework for sparse estimation in High Dimensional Mixed Frequency Vector Autoregressive Models
3	Chan	Lap Sum	DrFARM: Identification and inference for pleiotropic variants in GWAS
4	Daw	Ranadeep	REDS: Random Ensemble Deep Spatial prediction
5	Dilma	Eleni	Class Distance model for community detection
6	Fang	Yusi	On p-value combination of independent and frequent signals: asymptotic efficiency and Fisher ensemble
7	Ge	Lin	Tailoring Capture-Recapture Methods to Estimate Registry-Based Case Counts Based on Error-Prone Diagnostic Signals
8	Hampton	Hayden	Deep Belief Network Anomaly Detection using Least Square Support Vector Methods
9	Han	Qiyu	Statistical Inference for Low Rank Matrix Regression with Adaptively Collected Data
10	Kang	Huining	A mixture model approach for identifying genes whose isoform abundances are associated with survival outcome
11	Kang	Tong	Analyzing Dental Fluorosis Data using a Novel Bayesian Model for Clustered Longitudinal Outcomes with an Inflated Category
12	Adhikary	Avizit Chandra	PARD: Patient-specific Abnormal Region Detection in Alzheimer's Disease Studies
13	Li	Wenhao	A Comparison of Two Approaches to Dynamic Prediction: Joint Modeling and Landmark Modeling
14	Lindberg	David	A Bayesian Nonparametric Approach to an HIV Assessment Survey with Missing Data and Skip Conditions
15	Liu	Jinyuan	A Distance-based Semiparametric Regression Framework for Between-subject attributes: Applications to High-dimensional Sequences of Microbiome and Wearables
16	Lu	Nicholas	Analyzing the Impact of Different Countries' Approaches to the COVID-19 Pandemic on Their Cumulative Infection Curves By Using Nonparametric Density Regression and Clustering Methods
17	Mao	Siqi	BERT based Financial Sentiment Index Enhanced (BERTFSIE) Models for Financial Markets Forecast
18	Yang	Yuting	Regression Analysis of a Future State Entry Time Distribution Conditional on a Past State Occupation in a Progressive Multistate Model
19	Roy	Samrat	A Regularized High Dimension Low Tubal-Rank Tensor Regression
20	Saha	Sudipto	A Constrained Bayesian Multiscale Spatial Model using the Truncated Normal Distribution
21	Samanta	Srijata	A generalized likelihood based Bayesian approach for scalable joint regression and covariance selection in high dimensions
22	Tan	Xiaoqing	Leveraging Models from Heterogeneous Data Sources to Improve Personalized Treatment Effect Estimation

23	Wang	Wei	Multivariate Survival Analysis in Big Data: A Divide-and-Combine Approach
24	Wang	Xing	Extreme and Inference for Tail Gini Functionals with Applications in Tail Analysis of Systemic Risk
25	Wang	Hongwei	Clinical Trials with External Control: Beyond Propensity Score Matching
26	Xie	Xiulin	Control Charts For Dynamic Process Monitoring With An Application To Air Pollution Surveillance
27	Yu	Mengxin	Are Latent Factor Regression and Sparse Regression Adequate?
28	Yue	Xiaowei	Physics-Constrained Bayesian Optimization
29	Zhang	Lu	StarTrek: Combinatorial Variable Selection with False Discovery Rate Control
30	Zhong	Weibin	Application of Two-step GONOGO Criteria and Model-based Design for Dose Finding Based on Efficacy
31	Zhou	Doudou	RISE: Rank in Similarity Graph Edge-Count Two-Sample Test
32	Zou	Jian	CGMM: an algorithm for constrained model-based clustering

Index of Authors