SHORT COURSES

	Short Course Registration Fees					
	By January 15			After January 15		
	Half Day	Second Half Day	Full Day	Half Day	Second Half Day	Full Day
Member	$250	$200	$350	$275	$225	$375
Non-Member	$325	$290	$425	$350	$315	$450

Sunday, March 22, 2020

SC 1. Implementing Bayesian Adaptive Designs: From Theory to Practice

Full Day | 8:00 am – 5:00 pm

Ying Yuan, University of Texas MD Anderson Cancer Center

J. Jack Lee, University of Texas MD Anderson Cancer Center

Description: As a statistical framework, a Bayesian approach is intuitive, logical, coherent, elegant, and adaptive in nature. It is uniquely suitable for the design and analysis of clinical trials. The learning curve of Bayesian methods, however, is steep and the complexity of Bayesian computation can be intimidating. To overcome these hurdles, this short course is designed to provide an overview of Bayesian theory and its application to adaptive clinical trials. The emphasis is on implementing such designs by turning theory into practice. Easy-to-use Shiny applications and downloadable standalone programs will be introduced to facilitate the study design, conduct, and analysis of Bayesian adaptive methods. The main application areas include adaptive dose finding, adaptive toxicity and efficacy evaluation, posterior probability and predictive probability for interim monitoring of study endpoints, outcome-adaptive randomization, hierarchical models, adaptive biomarker identification and validation, multi-arm, multi-stage designs, and platform designs, etc. Bayesian adaptive designs allow flexibility in clinical trial conduct, increase study efficiency, enhance clinical trial ethics by treating more patients with more effective treatments, increase the overall success rate for drug development and can still preserve frequentist operating characteristics by controlling type I and type II error rates. Lessons learned from real trial examples and practical considerations for conducting adaptive designs and will be given.

SC 2. Practical solutions for working with electronic health records data

Full Day | 8:00 am – 5:00 pm

Rebecca Hubbard, University of Pennsylvania

Description: The widespread adoption of electronic health records (EHR) as a means of documenting medical care has created a vast resource for the study of health conditions, interventions, and outcomes in the general population. Using EHR data for research facilitates the efficient creation of large research databases, execution of pragmatic clinical trials, and study of rare diseases. Despite these advantages, there are many challenges for research conducted using EHR data. To make valid inference, statisticians must be aware of data generation, capture, and availability issues and utilize appropriate study designs and statistical analysis methods to account for these issues.

This short course will introduce participants to the basic structure of EHR data and analytic approaches to working with these data through a combination of lecture and hands-on exercises in R. The first part of the course will cover issues related to the structure and quality of EHR data, including data types and methods for extracting variables of interest; sources of missing data; error in covariates and outcomes extracted from EHR data; and data capture considerations such as informative visit processes and medical records coding procedures. Participants will have the opportunity to explore a synthetic EHR-derived data set to gain familiarity with the structure of EHR data and data exploration and visualization tools for identifying data quality issues. In the second half of the course, we will discuss statistical methods to mitigate some of the data quality issues arising in EHR, including missing data and error in EHR-derived covariates and outcomes. R code will be provided for implementation of the presented methods, and hands-on exercises will be used to compare results of alternative approaches.

This short course is of interest to researchers without prior experience working with EHR data as well as more experienced individuals interested in learning practical solutions to some common analytic challenges. The overarching objective of this course is to provide participants with an introduction to the structure and content of EHR data as well as a set of practical tools to investigate and analyze this rich data resource.

SC 3. Design and Analysis of Sequential, Multiple Assignment, Randomized Trials for small and large samples

Full Day | 8:00 am – 5:00 pm

Kelley Kidwell, University of Michigan

Thomas Braun, University of Michigan

Roy Tamura, University of South Florida

Description: Sequential, multiple assignment, randomized trials (SMARTs) have been implemented in oncology, drug abuse, ADHD, obesity, depression, insomnia, autism, and smoking cessation, among other areas. A SMART is a multi-stage trial design that allows for individuals to be randomized at two or more stages based on intermediate outcomes. SMART design has primarily been focused on informing the construction of dynamic treatment regimens (DTRs) or adaptive interventions. DTRs are evidence-based treatment guidelines where treatment can be altered over time based on the individual. Most SMARTs are conducted in large samples and analyzed using frequentist methods to explore potential delayed effects and treatment interactions over time to estimate and compare DTRs. More recently, Bayesian and frequentist methods have been developed to apply the SMART design in rare diseases, or more generally, small samples to find the best overall treatment sharing information across stages. Thus, a SMART design can also be used to strengthen inference on the best single treatment. The Bayesian methods developed to analyze SMART data in small samples may also be extended to find the most effective DTRs. This short course will introduce SMART design for both large and small samples. Case studies will be used as examples and R code will be provided for practice.

SC 4. Programming with hierarchical statistical models: Using the BUGS-compatible NIMBLE system for MCMC and more

Half Day | 8:00 am – 12:00 pm

Christopher Paciorek, University of California, Berkeley

Description: NIMBLE (r-nimble.org) is a system for fitting and programming with hierarchical models in R that builds on the BUGS language for declaring models. NIMBLE provides analysts with a flexible system for using MCMC, sequential Monte Carlo, MCEM, and other techniques on user-specified models. It provides developers and methodologists with the ability to write algorithms in an R-like syntax that can be easily disseminated to users. C++ versions of models and algorithms are created for speed, but these are manipulated from R without any need for analysts or algorithm developers to program in C++. While analysts can use NIMBLE as a drop-in replacement for WinBUGS or JAGS, NIMBLE provides greatly enhanced functionality in a number of ways.

This hands-on tutorial will first show how to specify a hierarchical statistical model using BUGS syntax and fit that model using MCMC. Participants will learn how to customize the MCMC for better performance (choosing samplers and blocking schemes) and how to specify one's own statistical distributions and functions to extend the syntax of BUGS. We will demonstrate the use of NIMBLE for biostatistical methods such as semiparametric random effects models and clustering models using Bayesian nonparametric techniques. We will also demonstrate the use of NIMBLE's built-in reversible jump MCMC for variable selection and the use of NIMBLE's CAR-based spatial models.

SC 5. Multivariate meta-analysis methods

Half Day | 1:00 pm – 5:00 pm

Haitao Chu, University of Minnesota Twin Cities

Yong Chen, University of Pennsylvania

Description: Comparative effectiveness research aims to inform health care decisions concerning the benefits and risks of different prevention strategies, diagnostic instruments and treatment options. A meta-analysis is a statistical method that combines results of multiple independent studies to improve statistical power and to reduce certain biases compared to individual studies. Meta-analysis also has the capacity to contrast results from different studies and identify patterns and sources of disagreement among those results. The increasing number of prevention strategies, assessment instruments and treatment options for a given disease condition, as well as the rapid escalation in costs, have generated a need to simultaneously compare multiple options in clinical practice using innovative and rigorous multivariate meta-analysis methods.

This short course, co-taught by Drs. Chu and Chen who have collaborated on this topic for more than a decade, will focus on most recent developments for multivariate meta-analysis methods. This short course will offer a comprehensive overview of new approaches, modeling, and applications on multivariate meta-analysis. Specifically, this short course will discuss the contrast-based and the arm-based network meta-analysis methods for multiple treatment comparisons; network meta-analysis methods for multiple diagnostic tests; multivariate extension of network meta-analysis; and multivariate meta-analysis methods estimating complier average causal effect in randomized clinical trials with noncompliance.

Case studies will be used to illustrate the principles and statistical methods introduced in this course. R codes with real examples will also be provided. This application oriented short course should be of interest to researchers who would apply up-to-date multivariate meta-analysis methods and who are interested in developing novel methods for multivariate meta-analysis. We anticipate that it will be well-received by an interdisciplinary scientific community, and play an important role in improving the rigor and broadening the applications of multivariate meta-analysis.

SC 6. Statistical Network Analysis with Applications to Biology

Half Day | 8:00 am – 12:00 pm

Ali Shojaie, University of Washington

George Michailidis, University of Florida

Description: Networks and network analysis methods are increasingly used by biomedical scientists and computational biologists to glean insight into cellular functions and mechanisms of disease propagation and initiation. While many approaches have been recently proposed, statistical and machine learning tools commonly play a key role in such analyses. This course provides a practical introduction to statistical network analysis methods for biological application. This short course will cover the following classes of methods: (i) statistical methods for network-structured data analysis; (ii) inference methods for undirected networks. The course will primarily focus on methods that are widely used in biological applications and, in particular, in the analysis of -omics data, as well as recent developments in statistical machine learning. Throughout, the emphasis will be on practical applications of network analysis methods, as well as their limitations, including validation of results and tools for reproducible research. Case studies using publicly available -omics data will be used to describe various statistical network analysis methods.

SC 7. Trial Design and Analysis Using Multisource Exchangeability Models

Half Day | 1:00 pm – 5:00 pm

Joseph Koopmeiners, University of Minnesota

Brian Hobbs, Cleveland Clinic

Alex Kaizer, University of Colorado

Description: Modern biomedical applications often call statisticians to estimate the effect of a treatment or intervention in sub-groups defined by demographic, genetic, or other participant information. This results in increasingly smaller sample sizes, which reduces power. Hierarchical modeling allows sub-group specific effects to be "shrunk" together, thus borrowing strength and increasing precision. However, standard hierarchical approaches are limited because they lack the flexibility to model complex relationships between sub-groups, where some sub-groups are exchangeable, while others are not. In this short course, we discuss trial design using multi-source exchangeability models (MEMs), which provide a flexible approach to estimating sub-group-specific effects, while accounting for complex relationships between subgroups. We provide an overview of the methodology and a comparison with standard hierarchical modeling approaches. We then discuss multi-source modeling in the context of trial design, focusing specifically on platform and basket trial designs, illustrating the advantage of multi-source trial designs vs. standard designs. The ability to incorporate other adaptive elements, such as adaptive randomization, will also be discussed. Much of the course will be illustrated via the basket package in R.

TUTORIALS

	Tutorial Registration Fees	
	By January 15	After January 15
Member	$75	$85
Non-Member	$85	$95
Student	$40	$50

Monday, March 23 - Tuesday, March 24, 2020

T 1. Statistical methods for geometric functional data

Monday, March 23 | 8:30 am – 10:15 am

Karthik Bharath, University of Nottingham, UK

Sebastian Kurtek, The Ohio State University

Description: How can one quantify variation in Hippocampal shapes obtained from MRI images as 2D curves? How does one model intra-tumour heterogeneity using samples of pixel densities? Answers to such questions on functional data with rich geometric structure require methods that are at a nascent developmental stage, and are typically not part of the standard functional data toolbox.

In this tutorial, we shall introduce some modern statistical and computational tools for handling such functional data objects. The first part of the tutorial will focus on the representation of such data and computation of descriptive summaries such averages and PCA, with numerous references to existing works and computing resources. The focus then moves to understanding the challenges involved in developing regression models involving such data objects. The last part of the tutorial will present an overview of the current state-of-the-art, and suggest future directions of research with a view towards inference.

T 2. Disease Risk Modeling and Visualization using R

Monday, March 23 | 10:30 am - 12:15 pm

Paula Moraga, University of Bath, UK

Description: Disease risk models are essential to inform public health and policy. These models can be used to quantify disease burden, understand geographic and temporal patterns, identify risk factors, and measure inequalities. In this tutorial we will learn how to estimate disease risk and quantify risk factors using areal and geostatistical data. We will also create interactive maps of disease risk and risk factors, and introduce presentation options such as interactive dashboards. We will work through two disease mapping examples using data of malaria in The Gambia and cancer in Pennsylvania, USA. We will cover the following topics:

- Model disease risk in different settings,
- Manipulate and transform point, areal and raster data using spatial packages,
- Retrieve high resolution spatially referenced environmental data using the raster package,
- Fit and interpret spatial models using Integrated Nested Laplace Approximations (INLA) (http://www.r-inla.org/),
- Map disease risk and risk factors using leaflet (https://rstudio.github.io/leaflet/) and ggplot2 (https://ggplot2.tidyverse.org/),

The tutorial examples will focus on health applications, but the approaches covered are also applicable to other fields that use georeferenced data including epidemiology, ecology or demography. We will provide clear descriptions of the R code for data importing, manipulation, modeling and visualization, as well as the interpretation of the results. The tutorial materials are drawn from the book 'Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny' by Paula Moraga (2019, Chapman & Hall/CRC Biostatistics Series).

T 3. Integration of Genetics and Imaging Data in Scientific Studies

Monday, March 23 | 1:45 pm - 3:30 pm

Debashis Ghosh, Colorado School of Public Health

Description: In this tutorial, we will discuss issues and approaches in the consideration of combining genetics and imaging data in biological and biomedical studies. A variety of motivating examples will be described. A common life-cycle pipeline for analytics will be discussed, along with some emergent lessons that have been learned through the literature. I will also focus on the types of questions that typically asked with these data sources and the roles of regression modelling and machine learning in these contexts.

T 4. Causal Inference Using the R TWANG Package for Mediation and Continuous Exposures

Monday, March 23 | 3:45 pm - 5:30 pm

Donna Coffman, Temple University

Description: When randomized experiments are infeasible, analysts must rely on observational data in which treatment (or exposure) is not randomly assigned (e.g., in health policy research or when determining the effects of environmental exposures). In addition, knowing the mechanisms or pathways through which a treatment works requires causal inference methods because the mediator is not randomly assigned. This tutorial aims to promote the use of causal inference methods for mediation and continuous exposures using the R twang package. The twang package recently was expanded to handle mediation and continuous exposures. We will first introduce causal mediation using the potential outcomes framework and weighting methods for estimating the causal mediation effects. We then will illustrate the implementation of gradient (or generalized) boosting models (GBM) for estimating the weights using the R twang package. Next, we will introduce the generalized propensity score (GPS) for continuous exposures. We will illustrate the implementation of GBM for estimating the GPS using the R twang package. The tutorial will provide relevant statistical background knowledge of mediation, the GPS, GBM, and weighting but will focus on implementation rather than statistical theory. Attendees should have some familiarity with propensity score analysis (e.g., for binary treatments/ exposures) and regression models, but knowledge of causal mediation, GPS, and GBM is not necessary. Attendees will be provided with the R code.

T 5. Fundamentals of difference-in-differences studies

Tuesday, March 24 | 1:45 pm - 3:30 pm

Laura A. Hatfield, Harvard Medical School

Bret Zeldow, Harvard Medical School

Description: A popular design in policy and economics research, difference-in-differences contrasts a treated group's pre- to post-intervention change in outcomes to an untreated comparison group's change in outcomes over the same period. The difference between the changes in the treated and comparison groups may be interpreted as the causal effect of the intervention if one assumes that the comparison group's change is a good proxy for the treated group's counterfactual change if it had not been treated. In this tutorial, we review the fundamentals of difference-in-differences studies, including key causal assumptions and ways to assess their plausibility, selection of a good comparison group, matching and regression techniques, statistical inference, and robustness checks.

T 6. R package development

Tuesday, March 24 | 3:45 pm - 5:30 pm

John Muschelli, Johns Hopkins University

Description: The jump from R programming with scripts to packages can be quite large. We hope to answer some of the basic questions of getting you started with package development answering the questions of: How do you create a basic R package? What are some R package best practices? How do I know if I can install this package? How do I depend on other packages? The tutorial will go through a simple 2-function package and describe resources to use after the course, including the R Package Development YouTube series: https://www.youtube.com/watch?v=79s3z0gIuFU&list=PLk3B5c8iCV-T4LM0mwEyWIunIunLyEjqM&index=1

ROUNDTABLES

Registration is required. Roundtable Registration Fee: $45
Monday, March 23 12:15-1:30 p.m.

R 1. Statistical positions in government

Paul Albert, National Cancer Institute

Description: The federal government provides exciting career opportunities for biostatisticians. There are positions ranging from mathematical statisticians, postdoctoral fellows, and tenure-track investigators. We will discuss these different types of positions, including the different types of work and the citizen requirements. We will discuss how to locate positions and the application/interview process. Focus will be on positions at the National Institutes of Health and the Food and Drug Administration where most government biostatisticians work.

R 2. How Can We Improve Biostatistical Reviewing for Medical Journals?

Cynthia Garvan, University of Florida

Description: The scientific community is justifiably concerned about both the rigor and reproducibility of medical research. From Statistics Done Wrong (Reinhart, 2015) to findings from a recent National Academies of Sciences, Engineering, and Medicine workshop convened to address questions about the reproducibility of scientific research, lack of statistics education has been identified as a major culprit in the generation of poor science. Beyond a lack of statistics education for researchers, a lack of education for biostatistical reviewers is problematic. In this roundtable we will discuss steps needed to improve this vital contribution of the biostatistician to advance medical research.

R 3. Early career mentoring: What do I do now?

Lance A. Waller, Emory University

Description: A career in the field of Biostatistics can be rewarding but also a challenge to navigate early in one's career. Some parts of the field seem to be changing quickly, others seem to stay the same. Departments and research groups grow and shrink, scientific (and funding!) priorities shift with new technology, new discoveries, and new approaches. In this roundtable, we will consider multiple issues involved with beginning a career in Biostatistics. We will discuss the different "currencies of success" associated with careers in academic, industry, and government organizations. We will discuss communication skills, funding strategies, collaboration skills, and opportunities to contribute to the field in multiple ways. Please feel free to bring questions (or send them to the facilitator beforehand) to allow the discussion to address your needs as well as these guidelines.

R 4. Publish or Perish in Biostatistics

Geert Molenberghs, Hasselt University and KU Leuven, Belgium

Description: Like statistics and biostatistics itself, publishing in biostatistics journals is in full transition: from paper to also electronic to electronic only; what about open access? What about reproducibility and, relatedly, scientific integrity?; all of this against the background of privacy protection. Do we publish in a journal owned by a commercial publisher, in a society-owned journal, in a cooperative journal, or perhaps in no journal at all? Do we prefer a statistics or a data science journal – or is this a false dichotomy? What is the relative status of theory, theorems, methodology, modeling, data analysis, and simulations? Apart from being an author, what are the relative advantages and drawbacks of acting as referee or Associate Editor? Should we give weight to impact factors or are they ignorable?

R 5. Understanding the NIH Grant Review Process

Scarlett L. Bellamy, Drexel University

Description: Have you ever wondered what it's like to be member of an NIH study section? Have you ever wondered about the review process for grants that you have submitted or plan to submit? In this roundtable we will discuss the NIH review process, from the perspective of a current member of Biostatistical Methods and Research Design (BMRD) Study Section. Attendees should leave the discussion with a better understanding the grant review process to better inform how they might prepare future grants or as they consider service on future study sections.

R 6. Data Science Programs

Joel Greenhouse, Carnegie Mellon University

Description: Academic and online data science programs are popping up everywhere. Employers now post positions for data scientist and rarely for statisticians or data analyst. If statistical thinking is the bedrock of data science, how can we insure that statistics and good statistical thinking play a proper role in the training of the next generation of statistical scientist? What has your experience been with the emergence of data science at your University or your place of employment. These, as well as other participant generated questions will be the source of discussion for this roundtable.

R 7. Being a Biostatistician in a Medical Center

Bryan E. Shepherd, Vanderbilt University Medical Center

Description: Statisticians are in great demand in medical centers. This can be both exciting and daunting. We will discuss strategies for flourishing in a medical center, from gaining respect among medical collaborators, to identifying and pursuing interesting research projects, to protecting one's time.

R 8. How to navigate collaborative research

Andrea B. Troxel, NYU School of Medicine

Description: We will discuss best practices for working with collaborators to develop grant proposals, guidelines for effort allocation for both faculty and staff, and timelines for grant preparation. We will also discuss common roadblocks that arise, and offer tips for troubleshooting challenging situations.

R 9. Running a Statistical Consulting Business

Alicia Y. Toledano, Biostatistics Consulting, LLC

Description: Running your own consultancy has many benefits, such as choosing your clients and projects, setting your own hours, and possibly working from home. This roundtable will focus on meeting challenges and carrying out responsibilities associated with those benefits. We will discuss making decisions related to: incorporation, using an attorney to review contracts, accounting, insurance, SOPs including for quality control, and having subcontractors and/or employees. Based on time and attendees' interests, we may also discuss one or more of: 1) Deciding what projects to undertake, with respect to areas of statistical expertise and 2) project type, such as short- or long-term; papers, grants, and/or FDA submissions; 3) How to get clients; 4) Working with clients that are not local; and 5) Ensuring your continued professional development statistically, and in soft skills like working as part of an interdisciplinary team. Come with questions and/or suggestions!