SHORT COURSES

	RECEIVED BY **FEB 1ST**			RECEIVED AFTER **FEB 1ST**		
	HALF DAY	2ND HALF DAY	FULL DAY	HALF DAY	2ND HALF DAY	FULL DAY
MEMBER	**$225**	**$190**	**$325**	**$250**	**$215**	**$350**
NONMEMBER	**$275**	**$240**	**$375**	**$300**	**$265**	**$400**

SUNDAY, MARCH 25, 2018

SC1. Bayesian Adaptive Clinical Trials

Full Day | 8:00 am – 5:00 pm

Scott Berry, Berry Consultants

Description: This course focuses on the innovative uses of Bayesian adaptive clinical trials. The application of Bayesian statistics to flexible adaptive trials is becoming a powerful innovation within the clinical trials industry. This is evidenced by the currently proposed FDA work in PDUFA VI:

> "To facilitate the advancement and use of complex adaptive, Bayesian, and other novel clinical trial designs, FDA will conduct the following activities during PDUFA VI: a. FDA will develop the staff capacity to enable processes to facilitate appropriate use of these types of methods. This staff will support the computationally intensive review work necessary to evaluate complex adaptive, Bayesian, and other novel clinical trial designs, with a particular focus on clinical trial designs for which simulations are necessary to evaluate the operating characteristics."

This demonstrates the increased desire and demand for FDA to be able to evaluate Bayesian Adaptive clinical trials. The course will focus on the specific strengths of the Bayesian approach and present numerous examples of trials in phase I, II, and III in the clinical trials space. Examples from a wide variety of therapeutic areas such as cardiovascular, neurology, pain, oncology, and rare disease will be discussed. The course will present examples from drugs and devices.

The course will discuss the Bayesian approach with the strengths and weaknesses it brings embedded in trial design, the software needs for using the Bayesian approach and for building adaptive trials, and the receptivity of these approaches to multiple stakeholders.

SC2. Machine Learning for Biomarker Discovery

Full Day | 8:00 am – 5:00 pm

Noah Simon, University of Washington

Description: We will present a number of supervised learning methods that can be applied to Biomedical Big Data: In particular we will cover penalized approaches to regression and classification; as well as support vector machines, and tree-based methods.

We will consider the analysis of "high-dimensional" data sets from genomics, transcriptomics, metabolomics, proteomics, and other fields. These data are typically characterized by a huge number of molecular measurements (such as genes) and a relatively small number of samples (such as patients). In addition, we will discuss the use of these tools in the development of prognostic and predictive biomarkers. Throughout the course, we will focus on common pitfalls in the supervised analysis of Biomedical Big Data and how to avoid them.

This course assumes some previous exposure to linear regression and statistical hypothesis testing.

SC3. An Introduction to the Joint Modeling of Longitudinal & Survival Data, with Applications in R

Full Day | 8:00 am – 5:00 pm

Dmitri Rizopoulos, Erasmus University Medical Center

Description: In follow-up studies different types of outcomes are typically collected for each subject. These include longitudinally measured responses (e.g., biomarkers), and the time until an event of interest occurs (e.g., death, dropout). Often these outcomes are separately analyzed, but in many occasions it is of scientific interest to study their association. This type of research question has given rise in the class of joint models for longitudinal and time-to-event data. These models constitute an attractive paradigm for the analysis of follow-up data that is mainly applicable in two settings: First, when focus is on a survival outcome and we wish to account for the effect of endogenous time-dependents covariates measured with error, and second, when focus is on the longitudinal outcome and we wish to correct for non-random dropout.

This full-day course is aimed at applied researchers and graduate students, and will provide a comprehensive introduction into this modeling framework. We will explain when these models should be used in practice, which are the key assumptions behind them, and how they can be utilized to extract relevant information from the data. Emphasis is given on applications, and after the end of the course participants will be able to define appropriate joint models to answer their questions of interest.

This course assumes knowledge of basic statistical concepts, such as standard statistical inference using maximum likelihood, and regression models. In addition, basic knowledge of R would be beneficial but is not required. Participants are required to bring their own laptop with the battery fully charged. Before the course instructions will be sent for installing the required software.

The primary target audience includes statisticians working in applied environments where hierarchical modelling and survival analysis are key issues; this would include biostatisticians working in the pharmaceutical industry, regulatory agencies, or academic centers.

SC4. Neuroimaging Analysis within R

Half Day | 8:00 am – 12:00 pm

John Muschelli, Johns Hopkins Bloomberg School of Public Health

Kristin Linn, University of Pennsylvania

Description: In the neuroimaging community, there is a diverse and large set of software tools currently being used by analysts and researchers. There have been great strides in standardizing the syntax for multiple pieces of software such as nypipe in Python, yet many of these languages do not have the statistical sophistication demanded to solve cutting-edge neuroimaging problems. R is a programming language that has the state of the art statistical tools that are relevant to imaging analysis. Already, a number of neuroimaging researchers use R as their primary language, and we believe this community will grow rapidly in the future.

With the rapid and increasing number of open-access neuroimaging datasets, such as the Alzheimer's Disease Neuroinitiative Initiative (ADNI) and the Human Connectome Project (HCP), there is a void for an analysis framework that (1) is reproducible and can deal with high-dimensional data structures, (2) is open-access and accessible to a large community, and (3) provides the best environment to perform fast and advanced statistical methods needed for such complex data. The R programming language satisfies 1,2, and 3. In this tutorial, we will provide tutorials on how to use R for structural magnetic resonance imaging (MRI) analysis. We will show how to perform entire image analysis in R, from the scans in raw image format to the statistical analysis after image preprocessing, with an emphasis on reproducibility by using a single programming language. This course will use a real multiple sclerosis dataset and will show the steps of going from the raw image files to performing multiple sclerosis lesion classification with a number of classifiers

entirely in R. In this hands-on tutorial, attendees will be given instructions for setup and data before the course, so that they are able to follow along and perform the analysis during the tutorial.

The topics to be cover in the course are as follows:

- **(a)** Introduction to the Statistical Software R (JM)
- **(b)** Read and Write Images (JM)
- **(c)** Visualization (JM and JPF)
- **(d)** Inhomogeneity Correction (JPF)
- **(e)** Brain Extraction (JM)
- **(f)** Image Segmentation (JPF)
- **(g)** Coregistration Within and Between MRI Studies (JM)
- **(h)** Intensity Normalization (JPF)
- **(i)** Harmonization of multi-site datasets (JPF)

Only within recent years has it become possible to perform entire image analysis completely and reproducibly in R. New R packages such as ANTsR (http://stnava.github.io/ANTsR/) and fslr (https://cran.r-project.org/web/packages/fslr/index.html) have made this possible. R is a powerful and open source statistical software that many members of ENAR already use for post-processing statistical analysis; yet members may be unaware of the new and powerful potential of R for image preprocessing, allowing for the creation of a streamline reproducible pipeline for entire image analysis. This tutorial is designed to educate and instruct in a step-by-step manner how to perform an entire image analysis in R, and should be a useful exercise for researchers interested in any type of imaging data, not just MRI. In addition, if time permits, we will show how we can use several statistical methods from existing R packages to attack the crucial problem of harmonizing datasets coming from different imaging sites.

SC5. Causal Inference Using Structural Nested Mean Models

Half Day | 8:00 am – 12:00 pm

Judith Lok, Harvard T.H. Chan School of Public Health

Description: Structural nested models allow investigators to estimate the causal effect of a time-varying treatment in the presence of confounding mechanisms. When a time-varying treatment is repeatedly adapted to evolving prognostic factors and the value of the prognostic factors is affected by prior treatment, the effect of the treatment cannot simply be estimated by conditioning on the patient characteristics. This treatment-confounder feedback is common in observational studies. The course will describe g-estimation of Structural Nested Mean Models (SNMMs) using single-robust and doubly robust estimators. I will present methods for estimating and testing the parameters of SNMMs. I will also cover coarse Structural Nested Mean Models, and illustrate these methods with an application to the effect of antiretroviral treatment in HIV-positive patients.

SC6. Reproducible Research with R & RStudio

Half Day | 1:00 pm – 5:00 pm

Keith A. Baggerly, The University of Texas M.D. Anderson Cancer Center

Description: In recent years, it has become apparent that sizeable numbers of papers in the biomedical literature are neither replicable (we can't get similar results with new samples) nor reproducible (we can't even get the reported results starting from the same data). Some problematic cases have even affected patient care.

In partial response to the situation, the National Institutes of Health (NIH) rolled out a "Rigor and Reproducibility Initiative" in 2016, placing more emphasis on how these issues are addressed in studies they fund. Journals are similarly beginning to ask for more documentation.

But, just as the pressure to work more reproducibly is increasing, tools which make it easier to do so are being introduced, lowering barriers to entry.

In this short course, we will:

- discuss some examples motivating the shift to RR
- survey the simple nature of the most common problems
- discuss organizing data as projects
- use RStudio, knitr, and R Markdown to illustrate
- the use of literate programming to interleave
- text describing the analyses with the code
- producing the results
- use RStudio, devtools, and roxygen2 to construct
- a basic R package
- survey other commonly used tools and give pointers
- how they might be used and where to learn more

This course presumes some working knowledge of R. Attendees are requested to bring laptops with recent versions of R and RStudio installed, as well as the R packages knitr, R Markdown, devtools, roxygen2, and RTools (this last is for Windows PCs; it's required to compile R packages).

Suggested references include:

- Reproducible Research with R and RStudio (2e) by Christopher Gandrud
- R Packages by Hadley Wickham

SC7. Survival for Precision Medicine

Half Day | 1:00 pm – 5:00 pm

Lu Tian, Stanford University School of Medicine
Haoda Fu, Eli Lilly

Description: The idea of precision medicine is to use patient's specific characteristics, such as genetic make-up, biomarker profile, clinical history, environmental exposure, etc., to guide clinical decision making for effective prevention and treatment. Recent advances in high-throughput and information technologies can easily and robustly generate a large amount data to characterize individual patients, offering extraordinary opportunities to develop and promote precision medicine in daily clinical practice. The development of such a smart and targeted strategy needs to be empirically data-driven and the corresponding challenges in the statistical front are huge, mainly because analyzing heterogeneous treatment effect/associations, i.e., interaction, is much more difficult than analyzing the homogeneous counterparts, i.e., main effect.

The goal of this course to introduce recently developed statistical and machine learning techniques for precision medicine. Most of the course will focus on how to optimally assign treatment to patients according to his or her personal characteristics in the context of two-arm randomized clinical trial. However, we will also discuss extensions to multi-arm trials and observational studies. Specifically, we will cover the following topics: the casual inference framework for personalized treatment effect based on counterfactual outcomes; estimation of optimal treatment selection rule including subgroup analysis, tree-based method, regression modeling, modified covariates approach, Q-learning and outcome weighted learning; and validation and statistical inference of the optimal treatment selection strategy. We will also discuss the computational perspective of the aforementioned methods including dimension reduction via regularization and applications of machine learning methods. We will provide examples for how to construct and evaluate estimated optimal treatment regimes in R.

There is no requirement for prior exposure to precision medicine or machine learning methods, but graduate-level statistical knowledge of basic methods such as regression, interactions, and analysis of randomized clinical trials is expected.

TUTORIALS

	RECEIVED BY **FEB 1ST**	RECEIVED AFTER **FEB 1ST**
MEMBER	**$75**	**$85**
NONMEMBER	**$85**	**$95**
STUDENT	**$40**	**$50**

MONDAY, MARCH 26, 2018 – TUESDAY, MARCH 27, 2018

T1. Micro-randomized Trials for Constructing Mobile Health Interventions

Monday, March 26 | 10:30 am – 12:15 pm

Susan Murphy, Harvard University

Description: Mobile devices along with wearable sensors make possible the ability to deliver treatments anytime and anywhere. These mobile interventions are being employed across a variety of health fields, including to support HIV medication adherence, encourage physical activity and healthier eating as well as to support recovery in addictions. The treatments in the mobile intervention are often time-varying and might be delivered many times across 100s or 1000s or more time points. In this tutorial we discuss a type of factorial trial design, namely the (stratified) micro-randomized trial, for use in constructing mobile interventions. We discuss primary hypotheses for these factorial designs and how to determine the sample size so as to test these hypotheses with a given power. We will also review secondary analyses that can be used to estimate and test interaction effects between time-varying context (e.g., location, current stress classification, time of day, mood and ambient noise) and time-varying treatments. These discussions will be made concrete by using three real-life micro-randomized trials (namely in physical activity, smoking cessation and encouraging adherence) to clarify ideas and analyses.

T2. Design & Analysis of Medical Studies Using Electronic Health Records Data

Monday, March 26 | 1:45 pm – 3:30 pm

Rebecca Hubbard, Perelman School of Medicine, University of Pennsylvania

Yong Chen, Perelman School of Medicine, 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 routine clinical practice. 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.

In this tutorial, we will discuss topics related to the design and analysis of research studies using EHR data. In the first part of the tutorial we 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. In the second half of the tutorial, we will discuss statistical methods that mitigate some of these

issues, including missing data and error in EHR-derived covariates and outcomes. We will also discuss cutting-edge methods developed to address unique challenges in the EHR context such as privacy-preserving computation in the context of distributed research networks. The overarching objective of this tutorial is to provide participants with an introduction to the structure and content of EHR data as well as a set of appropriate tools to investigate and analyze this rich data resource.

T3. Overview of the FDA Draft Guidance on Multiple Endpoints

Monday, March 26 | 3:45 pm – 5:30 pm

Jeff Maca, QuintilesIMS, Inc.

Description: The FDA released a draft guidance document for industry "Multiple Endpoints in Clinical Trials" in January, 2017. This draft guidance aims to share statistical methods and the agencies thinking on issues which can occur in studies which have multiple endpoints. The guidance document also shares methods for handling the issues which arise in such studies. This tutorial will give an overview of the content of the guidance document, with examples of the common multiplicity methodologies, as well as an introduction to the graphical methodology (Bretz, 2009)

T4. Fast & Easy RNA-seq Computational Workflow Using Bioconductor

Tuesday, March 27 | 8:30 am – 10:15 am

Michael Love, University of North Carolina Gillings School of Public Health

Description: Here we walk through an end-to-end gene-level RNA-seq differential expression workflow using Bioconductor packages and other open source software. We will start from the FASTQ files, show how to generate gene quantifications using Salmon and to import these data into R. We perform exploratory data analysis (EDA) for quality assessment and to explore the relationship between samples, perform differential gene expression analysis using different statistical packages, and visually explore the results. Additionally, we will point to (but not cover) new methods and workflows being developed for single-cell analysis.

T5. Multi-Modal Imaging: Promises & Pitfalls

Tuesday, March 27 | 1:45 pm – 3:30 pm

Martin Lindquist, Johns Hopkins Bloomberg School of Public Health

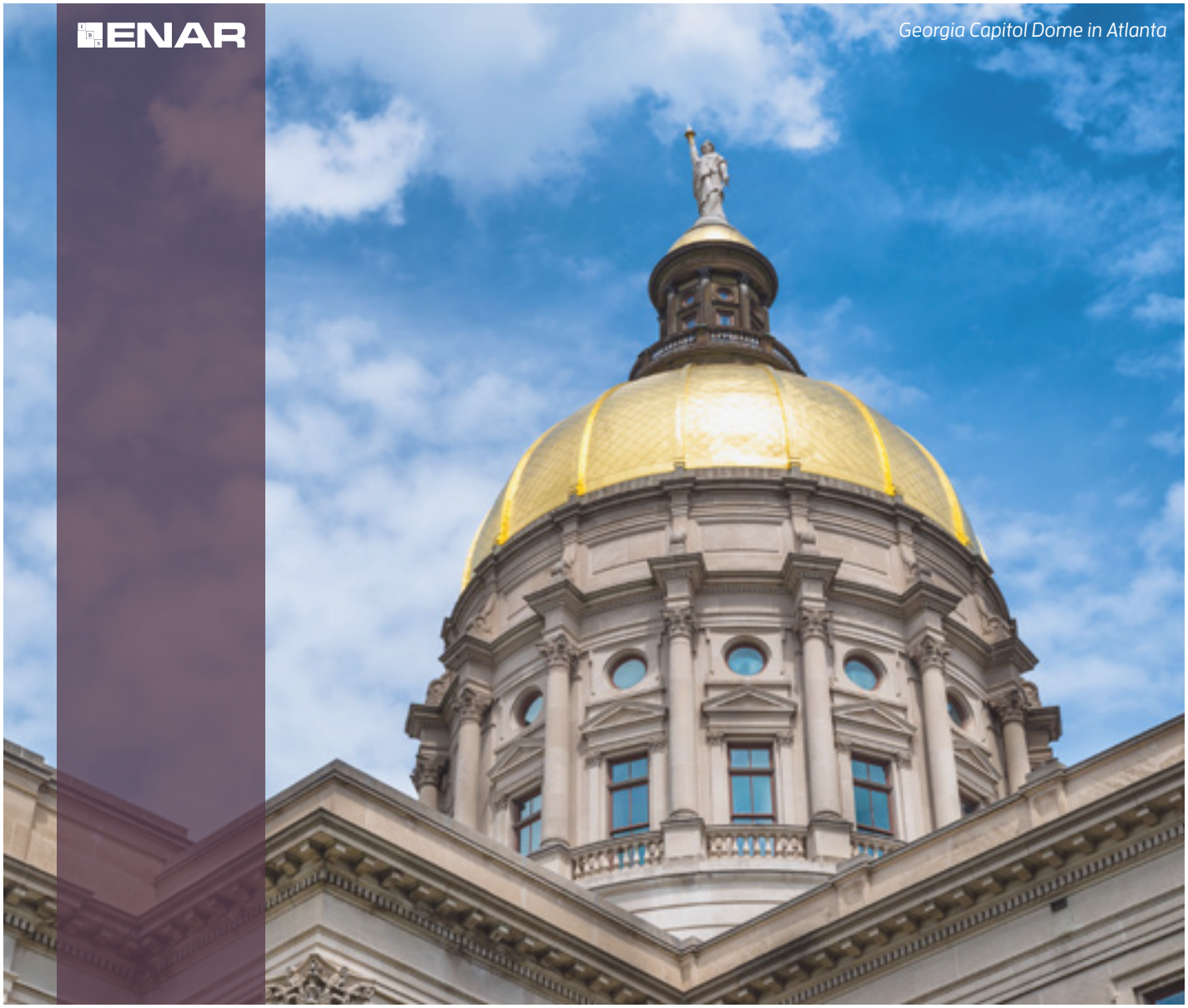
Description: All methods used for human neuroimaging have their own limitations and strengths. Therefore the current trend is toward increasingly interdisciplinary approaches that use multiple methodologies to overcome some of the limitations of each method used in isolation. For example, fMRI data are increasingly combined with EEG and MEG data to improve temporal precision, among other benefits. Neuroimaging data is also being combined with transcranial magnetic stimulation to integrate the powerful ability of neuroimaging to observe brain activity with the ability afforded by TMS to manipulate brain function and examine causal effects. In addition, integrating genetics with brain imaging is seen as a way to study how a particular subset of polymorphisms may affect functional brain activity. Each of these multi-modal approaches promise to be important topics of future research, and to fully realize their promise, novel statistical techniques will be needed. Here we discuss these promises, as well as challenges that need to be addressed.

T6. Integrative Analyses of High-throughput Multi-platform Genomics Data

Tuesday, March 27 | 3:45 pm – 5:30 pm

Veera Baladandayuthapani, The University of Texas M.D. Anderson Cancer Center

Description: Due to rapid technological advances, various types of genomic, epigenomic, transcriptomic, and proteomic data with different sizes, formats, and structures have become available. Each of these distinct data types provides a different, partly independent and complementary, high-resolution view of various biological processes. Modeling and inference in such studies is challenging,



Georgia Capitol Dome in Atlanta

not only due to high dimensionality, but also due to presence of structured dependencies (e.g. pathway/regulatory mechanisms, serial and spatial correlations etc.). Integrative analyses of these multi-domain data combined with patients' clinical outcomes can help us understand the complex biological processes that characterize a disease, as well as how these processes relate to the eventual progression and development of a disease. This tutorial will cover integrative statistical and computational frameworks that acknowledge and exploit these inherent complex structural relationships for both biomarker discovery and clinical prediction to aid translational medicine. These approaches will be illustrated using several biomedical case examples, especially in oncology.

ROUNDTABLES

REGISTRATION IS REQUIRED | FEE: $40

MONDAY, MARCH 26 | 12:15 pm – 1:30 pm

R1. Obtaining Grant Funding for Your Research

Michelle Dunn, Data Collaboratory

SOLD OUT

Description: For some academic positions, research funding is nice; for others, it is essential. If finding and obtaining grants is important to your career, this roundtable is for you. We will talk about finding biostatistics grant funding opportunities from NIH as well as other sources. You will come away with tips that will save precious time.

R2. Effective Interdisciplinary Collaborations

Reneé Moore, Emory University Rollins School of Public Health

SOLD OUT

Description: Many of us, statisticians, find that the greatest joy in our work is making meaningful contributions to the application of our choice (e.g. medicine, public health, business) as a part of interdisciplinary research teams. Being an effective member of these teams is analogous to sewing together the relevant pieces to make a beautiful quilt. To be an effective member one, obviously, must have strong analytic skills and be able to communicate the statistical methodology, but also one must have strong interpersonal skills and develop a collaborative approach to address the needs of the interdisciplinary team. We will discuss key elements for effectively enhancing interdisciplinary experiences, successful leadership as part of a research team, and how to address common challenges that arise in this environment. In addition, we will discuss the incorporation of training students and interns while weaving together the pieces to be an effective member of interdisciplinary collaborations.

R3. Wearable & Implantable Technology

Ciprian Crainiceanu, Johns Hopkins University Bloomberg School of Public Health

SOLD OUT

Description: The round table will focus on the many aspects of wearable and implantable technology measurements and their application to health studies. This is in response to the explosion of studies utilizing various types of devices including accelerometers, heart and glucose monitors, GPS, and ecological momentary assessment (EMA) devices. We will discuss recent developments in the area, potential for funding applications, and strategies for using WIT to reduce bias and measurement error in observational studies and clinical trials.

R4. Practical Issues in Clinical Trial Design & Analysis

Peter Thall, The University of Texas M.D. Anderson Cancer Center

SOLD OUT

Description: This roundtable will focus on methods for dealing with practical considerations that arise in the process of clinical trial design, conduct, and analysis. Depending on the attendees' interests, we will discuss a variety of recent developments, including designs discussed in the 2016 book 'Bayesian Designs for Phase I-II Clinical Trials' by Yuan, Nguyen and Thall, as well as new utility-based designs for randomized trials and personalized medicine.

R5. Career Opportunities for Statisticians at the CDC

Timothy Green, Centers for Disease Control and Prevention

SOLD OUT

Description: The Centers for Disease Control and Prevention (CDC) employs more than 400 mathematical and health statisticians across six coordinating offices, the Center for Global Health, and the National Institute for Occupational Safety and Health. CDC statisticians serve a critical role in supporting the Agency's mission through the compilation, analysis, and interpretation of statistical information used to inform public health activities. Using

a multidisciplinary approach to investigate public health problems, statisticians serve as technical experts in the development of study designs, data collection systems, software and analytic methodologies. In this roundtable, we will discuss different roles of statisticians at CDC. We will examine opportunities at various career stages and discuss potential career trajectories within the Agency.

SOLD OUT

R6. Grant Proposal 101 for Statisticians

Debashis Ghosh, Colorado School of Public Health

Description: In the current funding climate for grant-sponsoring organizations such as NIH and NSF, research grants of all types have become more competitive than ever before. This roundtable will focus on strategies and resources for researchers to be able to submit competitive research methodology grants for these mechanisms. In addition, review processes for various grant mechanisms will be described.

SOLD OUT

R7. Statistics and Connections: Opportunities and Challenges for Statistics for Studying how Neurons, Organisms and People Communicate

Brian Caffo, Johns Hopkins University Bloomberg School of Public Health

Description: In this roundtable we will discuss the development of statistical tools for studying connections. Connectivity dominates brain studies from functional neuroimaging studies to the microscopic study of individual neurons. Connectivity similarly dominates our study behavior and human connections of all sorts. Several parallel efforts address the statistical problems of connectivity including: random graphs, graphical models and covariance modeling. In this roundtable we discuss the emerging field of connectomics and both the unique and common aspects of different applied connectivity settings. We will further discuss statistical challenges and ways that statisticians can have a large impact on this field.

SOLD OUT

R8. Modeling and Inference for Spatial-Temporal BIG DATA

Sudipto Banerjee, University of California at Los Angeles

Description: With the growing capabilities of Geographic Information Systems (GIS) and user-friendly software, statisticians today routinely encounter geographically referenced data containing observations from a large number of spatial locations and time points. Over the last decade, a variety of models have been proposed for researchers to better understand the complex nature of spatial and temporal variability. However, their computational complexity renders such models unfeasible for massive data sets. This roundtable will discuss and offer solutions to some of the challenges facing spatial statisticians and GIS analysts with regard to analyzing massive amounts of spatial data. Recent research and developments will be discussed and emphasis will be accorded to practical implementation, the state of current software and tools, and the different types of inferential questions being posed to spatial statisticians today.

SOLD OUT

R9. Identifying the Non-Identifiable: Non-Progressive Cancers in Screened Populations

Yu Shen, The University of Texas M.D. Anderson Cancer Center

Ruth Etzioni, Fred Hutchinson Cancer Research Center

Description: Over the past decades, some cancer screening programs such mammography in breast cancer and colonoscopy in colon cancer, among others, have been considered powerful weapons to prolong survival and increase cure. At the same time, however, the debate about the potential harm due to overdiagnosis has also attracted attention. Though several cancer screening studies have been conducted for breast, lung, colon and prostate, estimation of the cancer natural history and overdiagnosis can be extremely challenge due to various biases and an unobservable disease process in the screening cohort. In this roundtable lunch, we will discuss study designs, data sources, methodological challenges and opportunities in the area of cancer screening with a particular focus on leveraging available data to learn about heterogeneity of the underlying disease progression process.

SOLD OUT