| | REGISTRATION RECEIVED | | | | | |
|---|---|---|---|---|---|---|
| | BY **FEB 1** | | | AFTER **FEB 1** | | |
| | 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** |



## SC1: Regression Modeling Strategies

**FULL DAY: 8:00 am – 5:00 pm**

**Frank E Harrell Jr.**
*Vanderbilt University School of Medicine*

**Description:** All standard regression models have assumptions that must be verified for the model to have power to test hypotheses and for it to be able to predict accurately. Of the principal assumptions (linearity, additivity, distributional), this course will emphasize methods for assessing and satisfying the first two. Practical but powerful tools are presented for validating model assumptions and presenting model results. This course provides methods for estimating the shape of the relationship between predictors and response using the widely applicable method of augmenting the design matrix using restricted cubic splines. Even when assumptions are satisfied, overfitting can ruin a model's predictive ability for future observations. Methods for data reduction will be introduced to deal with the common case where the number of potential predictors is large in comparison with the number of observations. Methods of model validation (bootstrap and cross-validation) will be covered, as will auxiliary topics such as modeling interaction surfaces, efficiently utilizing partial covariable data by using multiple imputation, variable selection, overly initial observations, collinearity, and shrinkage. The methods covered will apply to almost any regression model, including ordinary least squares, logistic regression models, quantile regression, longitudinal data analysis, and survival models.

## SC2: Data Science for Statisticians

**FULL DAY: 8:00 am – 5:00 pm**

**Amelia McNamara**
*Smith College Visiting Assistant Professor of Statistical & Data Sciences, MassMutual Faculty Fellow*

**Description:** As statistics becomes more computational and the term 'data science' is gaining traction, it is clear there are skills statisticians need to stay current. This short course will get you up to speed on many of the recent developments in this field. While there are many possible tools to use for data science, we will focus on the R ecosystem. Topics covered will include:

- Data visualization (ggplot2)
- The tidyverse (dplyr, broom, tidyr)
- APIs and web scraping (rvest, httr)
- Version control (git and GitHub)
- Reproducible research (RMarkdown and Project TIER)
- Finding help (StackOverflow, google, twitter)
- Interactivity (shiny and leaflet)

Participants should bring their own laptop with R and RStudio installed.

## SC3: Getting SMART about Dynamic Treatment Regimens

**FULL DAY: 8:00 am – 5:00 pm**

**Kelley Kidwell**
*University of Michigan*

**Description:** The effective treatment and management of a wide variety of health disorders often requires individualized, sequential decision-making. To do this, each patient's treatment is dynamically adapted over time based on the patient's history and changing disease state. Dynamic Treatment Regimens operationalize individualized decision making using a sequence of decision rules that specify whether, how, for whom, or when to alter the dose, type, or delivery of pharmacological, behavioral, and/or psychosocial treatments. Recently, there has been a surge of clinical and methodological interest in developing and evaluating Dynamic Treatment Regimens in clinical trials. For example, there is great interest in the use of sequential multiple assignment randomized trials (or SMART, a type of multi-stage randomized trial design) to build Dynamic Treatment Regimens. This non-technical workshop will provide a brief introduction to Dynamic Treatment Regimens and describe how SMART designs can be used to develop high-quality Dynamic Treatment Regimens. A variety of case studies will be presented to illustrate these methods.

**Condensed outline of the course:**

- Dynamic treatment regimens
- Sequential multiple assignment randomized trial (SMART) designs
- SMART design principles
- SMART case studies

## SC4: Statistical Methods for Brain Network Analysis Using Multimodal Imaging Data

**HALF DAY: 8:00 am – 12:00 pm**

**Ying Guo**
*The Rollins School of Public Health of Emory University*
**Phebe B. Kemmer**
*The Rollins School of Public Health of Emory University*

**Description:** Neuroimaging technology has played a central role in advancing scientific understanding of normal brain function in humans and in investigating the neural basis for major psychiatric and neurological disorders. In recent years, the study of human brain based on multimodal imaging technologies has become one of the new frontiers of neuroscience research. It provides the opportunity to combine the strengths of individual modalities to investigate the brain architecture from both functional and structural perspectives. For example, functional magnetic resonance imaging (fMRI) measures the hemodynamic response related to neural activity in the brain dynamically. Diffusion tensor imaging (DTI) can provide information on structural connectivity among brain networks. It has become increasingly clear that the most effective research approaches should utilize multimodal fusion to capitalize on the strength of each modality.

This course aims to provide an introduction to the analytical techniques and statistical methods for investigating functional and structural brain networks using fMRI and DTI data. The course will start with a discussion on some basic principles of data acquisition and preprocessing steps for fMRI and DTI and highlight characteristics of the data that are particularly relevant to statistical modeling. We will then proceed to cover the analytical procedures in multimodal connectivity analysis including: data structure and management, brain network node specification, statistical methods for network estimation and visualization of functional and structural networks. We will also introduce and demonstrate popular software for imaging processing, statistical analysis and data visualization for both fMRI and DTI data.

## SC5: Artificial Intelligence, Machine Learning, and Precision Medicine

**HALF DAY: 8:00 am – 12:00 pm**

**Haoda Fu**
*Eli Lilly and Company, Indiana University School of Medicine*
**Yufeng Liu**
*University of North Carolina at Chapel Hill*

**Description:** This half day short course will provide an overview of statistical machine learning, and artificial intelligence

techniques with applications to the precision medicine, in particular to deriving optimal individualized treatment strategies for personalized medicine. This short course will cover both treatment selection and treatment transition. The treatment selection framework is based on outcome weighted classification. We will cover logistic regression, support vector machine (SVM), – learning, robust SVM, and angle based classifiers for multi-category learning, and we will show how to modify these classification methods into outcome weighted learning algorithms for personalized medicine. The second part of short course will also cover the treatment transition. We will provide an introduction on reinforcement learning techniques. Algorithms, including dynamic programming for Markov Decision Process, temporal difference learning, SARSA, Q-Learning algorithms, actor-critic methods, will be covered. We will discuss on how to use these methods for developing optimal treatment transition strategies. The techniques discussed will be demonstrated in R.

**Intended audience:** This course is intended for graduate students who have some knowledge of statistics and want to be introduced to statistical machine learning, or practitioners who would like to apply statistical machine learning techniques to their problems on personalized medicine and other biomedical applications.

**Prerequisite:** Participants should be familiar with linear regression and statistical hypothesis testing, as well as some familiarity with R or another programming language.

## SC6: Statistical Evaluation of Medical Risk Prediction Models in the Presence of Competing Risks

**HALF DAY: 1:00 pm – 5:00 pm**

**Michael W. Kattan**
*Cleveland Clinic*
**Thomas A. Gerds**
*University of Copenhagen*

**Description:** Risk predictions are useful to inform patients and to guide medical decision making and the basis for precision medicine. The basis of a risk prediction model typically use variables such as age, gender, smoking status, disease status and history, blood tests and genetic markers. A suitable regression model combines these variables into a prediction. Prediction performance describes how well a model will do on future patients. Unfortunately, this is never knowable. The best we can do is to simulate the model being applied to future patients by repeatedly splitting the data set into training and validation part.

In this course we address the survival setting with competing risks. For a given time horizon there are three groups: subjects with event, subjects with a competing risk, and subjects event-free. There also is a possibility that subjects are not followed until the horizon (right censored). We discuss descriptive tools (calibration plots, retrospective risk re-classification) and inference based on the performance measures Brier score and time-dependent AUC adapted to the situation with right censored data and competing risks.

Attendees should have a basic understanding of regression models for survival analysis and experience using the R language. This short course is based on R.

| | REGISTRATION RECEIVED | |
|---|---|---|
| | BY **FEB 1** | AFTER **FEB 1** |
| MEMBER | **$75** | **$85** |
| NONMEMBER | **$85** | **$95** |
| STUDENT | **$40** | **$50** |



## T1: Open Source and Interactive Scientific Computing with R/Jupyter

**Monday, March 13: 8:30 am – 10:15 am**

**Andrea Foulkes**
*Mount Holyoke College*

**Description:** The simple and powerful collaborative notebook environment of iPython has only very recently become available for integration with R applications with the development of the open Jupyter Project (**https://jupyter.org/**). The Jupyter Notebook is an interactive computational environment that integrates code, rich text, mathematical formulations and visual representations of data. The protocol offers a seamless framework for creating and implementing reproducible research, providing the tools for documenting processes of data management, analysis and visualization. In this tutorial, you will learn the basic functionalities of Jupyter, starting from creating new code and markdown cells, to generating reports and slideshows, and ultimately to sharing your notebooks with the broader scientific community. We will also explore the opportunities for using Jupyter both for collaborative scientific computing and as an effective teaching environment for a broad range of data science and statistics courses while engaging in reproducible research practice.

## T2: Statistical Analysis of RNA Sequencing Data

**Monday, March 13: 10:30 am – 12:15 pm**

**Mingyao Li**
*University of Pennsylvania*

**Description:** RNA sequencing (RNA-seq) allows an unbiased survey of the entire transcriptome in a high-throughput manner. It has rapidly replaced microarrays as the major platform for transcriptomics studies. In this tutorial, I will cover several topics related to RNA-seq data analysis, including read alignment, quality control, gene expression quantification, and analysis of differential expression and alternative splicing.
I will also briefly introduce some statistical issues in single-cell RNA-seq data analysis.

## T3: Data-driven Modelling Techniques for Medical Decision Making

**Tuesday, March 14: 8:30 am - 10:15 am**

**Madhu Mazumdar**
*Mount Sinai Health System*
*Tisch Cancer Institute*
**Liangyuan Hu**
*Mount Sinai Health System*

**Description:** Focused clinical questions have generated considerable data from both randomized trials and observational studies, providing a wealth of opportunities for research in clinical decision making, healthcare policy evaluation and

evidence-based practice. Properly formulating the research question and using appropriate statistical methods to solve the question is critical. Often times simple linear regression models are not sufficient, especially in dealing with multi-stage designs and monitoring therapy in management of chronic diseases. This tutorial will review the basics of several data-driven statistical modelling techniques, and demonstrate applications of the methods to longitudinal data sets in cancer. The participant should leave the tutorial with an ability to visualize patterns and issues observed in the data; properly identify complications associated with the data; propose solutions to 'correct' these complications and implement software to conduct statistical analysis. Topics covered include: issues popularly observed in databases collected in actual clinical practice; statistical methods to address the data issues; data-driven modelling techniques to conduct comparative effectiveness analysis; evidence-based optimal medical decision making (baseline or sequential); R software for visualizing data patterns and conducting statistical analyses.

## T4: Statistical Issues in Comparative Effectiveness Research

**Tuesday, March 14: 1:45 pm – 3:30 pm**

**Sharon-Lise Normand**
*Harvard Medical School*

**Description:** Comparative Effectiveness Research (CER) refers to a body of research that generates and synthesizes evidence on the comparative benefits and harms of alternative interventions to prevent, diagnose, treat, and monitor clinical conditions, or to improve the delivery of health care. The evidence from CER is intended to support clinical and policy decision making at both the individual and the population level. While the growth of massive health care data sources has given rise to new opportunities for CER, several statistical challenges have also emerged. This tutorial will provide an overview of the types of research questions addressed by CER, review the main statistical methodology currently utilized, and highlight areas where new methodology is required. Inferential issues in the "big data" context are identified. Examples from cardiology and mental illness will illustrate methodological issues.

## T5: NIH Perspectives on Peer Review

**Tuesday, March 14: 3:45 pm – 5:30 pm**

**Michael S. Lauer**
*National Institutes of Health*
**Sally A. Amero**
*National Institutes of Health*

**Description:** The NIH peer review process forms the cornerstone of the NIH extramural research mission. The NIH currently handles approximately 80,000 applications and engages approximately 25,000 reviewers per year. Our goal is to assure, in a manner free from inappropriate influences, unbiased robust evaluations of NIH grant applications by a large body of scientific experts. This tutorial will cover the nuts and bolts of this process, and will explore current issues and initiatives related to peer review, such as rigor and reproducibility.

**REGISTRATION IS REQUIRED: $40**

**ROUNDTABLES MONDAY, MARCH 13** 12:15 pm – 1:30 pm

## R1: Publishing Without Perishing: Strategies for Success in Publishing in Biostatistical Journals

**Marie Davidian**
*North Carolina State University*

SOLD OUT

**Description:** Contributing to the advance of our discipline through publication of articles in peer-reviewed journals is a fundamental expectation for junior and not so-junior biostatistical researchers alike. Success publishing one's work ensures that it will be widely disseminated to researchers and practitioners who stand to benefit. In addition, funding agencies and academic institutions place considerable importance on a successful record of publication. Accordingly, understanding the peer review and editorial processes of top journals and mastering the art of writing an effective journal article are keys to success in publishing. How does one determine the best outlet for one's work? What are the essential elements of a successful journal article? How does one maximize the chance of acceptance? What strategies can ensure that a published paper is read and cited? How does one make optimal use of limited space and additional supplementary material in conveying the message? What are the roles of the editor, associate editor, and referees? What considerations do editors use when evaluating a paper? This roundtable will provide a forum for candid discussion of these and other questions.

## R2: Some Thoughts on How Statisticians Can Become Leaders

**Jesse A. Berlin**
*Johnson & Johnson*

SOLD OUT

**Description:** What does it take for a statistician to become a leader in an organization? The obvious direction is to work toward becoming a department chair, leading other statisticians. That requires skills that go beyond simply being a really good statistician. There are notable examples of statisticians who have earned leadership roles in academic settings as deans and, in industry, in roles that are much broader than leading statisticians. In this roundtable, we'll discuss some of the principles to keep in mind if you're interested in advancement. The main take-home message is that it's not just IQ, but emotional intelligence and communication skills that get recognition.

## R3: The Joy of Study Section!

**Leslie McClure**
*Drexel University*

SOLD OUT

**Description:** There is much discussion in the scientific community about validity and reproducibility, bringing a stronger desire to involve statisticians in the entire scientific process. One aspect of this is the involvement of statisticians in reviewing grants that are non-statistical, such as clinical research or basic science research proposals. However, there is a lack of training for statisticians about how to participate in review panels that are primarily comprised of non-statisticians. In this roundtable, we will discuss how the process of grant reviews works, tips for providing statistical reviews for non-statistical research and how to get involved in non-statistical review panels. In addition, we will discuss how involvement in reviewing non-statistical grants can enhance the biostatistician's own career.

## R4: Strategies for Success in Collaborative Grant Writing

**Rebecca Hubbard**
*University of Pennsylvania*

SOLD OUT

**Description:** Collaborative grant writing is a necessary component of most careers in academic biostatistics and can be rewarding or vexing (or both). Forging strategies for effective and efficient collaboration can mean the difference

between frustration and productivity. Successful collaborative grant writing can also be the entry point into positive and fruitful scientific relationships. In this session we will discuss a range of considerations relevant to collaborative grant writing including strategies for building positive collaborative relationships, strategies for dealing with collaborations that have gone awry, tips and tricks for effective writing, and approaches to furthering methodological research through participation in collaborative grants.

## R5: A Statistician in the Age of Data Science

**Nick Chamandy**
*Lyft*

SOLD OUT

**Description:** In this roundtable I will attempt to unshroud the enigma of 'Data Science' by offering several definitions up for scrutiny. I will argue that the lack of a standardized definition has contributed to angst among the Statistics community (and among data scientists!). This will hopefully shed some light on the ostensibly complex relationship between Statistics and Data Science. I will also reflect what I have learned in my nine years as a statistician in the tech industry, including the challenges of building and managing a data science team.

## R6: Biostatistical Tools for Clinical Neuroimaging

**Ciprian Crainceanu**
*Johns Hopkins Bloomberg School of Public Health*

**Description:** Neuroimaging is increasingly popular in statistics, though little research has been dedicated to the brain affected by various pathologies. Cancer, stroke, multiple sclerosis, Alzheimer's disease, and traumatic brain injury produce major changes in the brain. These changes make it difficult to use standard neuroimaging tools for analysis and require new methods and a deeper understanding of the specific pathology. We will discuss the various problems that remain open, including bias field correction, lesion segmentation, lesion tracking, intensity normalization, and spatial registration. We also will discuss new tools in R that combine the strength of existent neuroimaging tools with a familiar computational environment. Free online courses such as Introduction to Neurohacking with R will be introduced, while NeuroconductoR, a new software platform for Neuroimaging R programming will be introduced and discussed.

## R7: Introduction to Causal Inference

**Eric Tchetgen Tchetgen**
*Harvard T.H. Chan School of Public Health*

SOLD OUT

**Description:** This roundtable discussion reviews the foundations of causal inference and introduces a systematic methodology for defining, estimating, and testing causal claims in experimental and observational studies. We will discuss the general causal framework as well as discuss various practical and applied settings where causal inference is useful. We will discuss the three primary causal questions of interest, namely, questions regarding (1) the effects of potential interventions, (2) counterfactual probabilities and (3) mediation.

## R8: Statistical Evaluation of Drug Safety Data

**Amy Xia**
*Amgen*

SOLD OUT

**Description:** There has been growing awareness of the importance of the statistical evaluation of drug safety data, both in the pre-marketing and post-marketing settings. Careful and comprehensive approaches are warranted in safety evaluation. This roundtable will discuss some key issues and emerging statistical methodologic developments in safety evaluation. Specifically, we discuss the following topics: prospective program-level safety planning, evaluation and reporting; the impact of adverse event grouping on statistical analysis; the applications of Bayesian methods in safety signal detection and safety monitoring; meta-analysis for analyzing safety data; and safety graphics. In addition, we cover aspects related to benefit-risk assessment.