**SHORT COURSES**

**Implementing Bayesian Adaptive Designs: From Theory to Practice**

**SC 1.**

**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 will be given.

**Design and Analysis of Sequential, Multiple Assignment, Randomized Trials for small and large samples**

**SC 3.**

**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.
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.

**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...
TUTORIALS

**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 intratumour 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 as 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/),

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.

**Tutorial Registration Fees**

<table>
<thead>
<tr>
<th></th>
<th>By January 15</th>
<th>After January 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Member</td>
<td>$75</td>
<td>$85</td>
</tr>
<tr>
<td>Non-Member</td>
<td>$85</td>
<td>$95</td>
</tr>
<tr>
<td>Student</td>
<td>$40</td>
<td>$50</td>
</tr>
</tbody>
</table>

**Monday, March 23 - Tuesday, March 24, 2020**

---

T 1. Statistical methods for geometric functional data

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 intratumour 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 as 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

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/),

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

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

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.
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.

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=79c3z0gluFU&list=PLk3B5c8lCV-T4LM0mweYIuunlunLyEqjM&index=1
Should we give weight to impact factors or are they ignorable? Apart from being an author, what are the relative status of theory, theorems, methodology, modeling, data analysis, science journal – or is this a false dichotomy? What is the relative electronic only; what about open access? What about reproducibility biostatistics journals is in full transition: from paper to also electronic to like statistics and biostatistics itself, publishing in Description: your needs as well as these guidelines. to the facilitator beforehand) to allow the discussion to address to the field in multiple ways. Please feel free to bring questions (or send funding strategies, collaboration skills, and opportunities to contribute and government organizations. We will discuss communication skills, “currencies of success” associated with careers in academic, industry, and government organizations. 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. 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. Description: Early career mentoring: What do I do now? 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?