SHORT COURSES

	Short Course Registration Fees					
	By Feb. 1st			After Feb. 1st		
	Half Day	2nd Half Day	Full Day	Half Day	2nd Half Day	Full Day
Member	$250	$200	$350	$275	$225	$375
Non-Member	$325	$290	$425	$350	$315	$450

Short Courses Sunday, March 24, 2019

SC1. Bayesian Inference and Clinical Trial Designs Using Historical Data

Full Day | 8:00 am – 5:00 pm
Ming-Hui Chen, University of Connecticut
Fang Chen, SAS Institute Inc.

Description: Clinical trials are understandably expensive. However, similar trial data are often available from previous studies or experiments. Borrowing information from historical data can potentially help reducing trial cost and providing more accurate estimation while maintaining desirable qualities, such as control type I error and power. This short course provides a comprehensive review of Bayesian methods for borrowing historical information and proper use of these methods in Bayesian clinical trial designs. Several case studies are illustrated using software code that is explained in detail.

The course focuses on using historical data in design areas that include design of non-inferiority clinical trials, design of superiority clinical trials, methods for go/no-go decisions, sequential meta-analysis design, and joint analysis that combines the results from multiple trials. Special topics that are discussed include Monte Carlo simulation, Bayesian sample size determination, analysis of recurrent events, and frailty regression. The examples are shown using SAS, including the SAS macro language and the MCMC procedure, with a strong focus on technical details.

The course also includes an introduction of the Bayesian approach to inference (presented from a biopharmaceutical perspective) and outlines approaches in using and borrowing historical information, including variations of the power prior and meta-analytic-predictive prior.

SC2. Big Data, Data Science and Deep Learning for Statistician

Full Day | 8:00 am – 5:00 pm
Hui Lin, DowDuPont
Ming Li, Amazon

Description: The increasing volume and sophistication of data pose new challenges and needs for data science. There is a pressing need for data scientists who can bring actionable insight from the vast amount of data collected. In the past several years, deep learning has gained traction in many areas, and it becomes an essential tool in data scientist's toolbox. In this course, students will develop a clear understanding of the big data cloud platform, learn basic data manipulation, preprocess and machine learning skills, and understand the motivation and use cases of deep learning through hands-on exercises. We will also cover the "art" part of data science including the general data science project flow, common pitfalls, and soft skills to effectively communicate with business stakeholders. The course is for audiences with a statistical background. This course will prepare statisticians to be successful data scientists and deep learning scientist in various industries and business sectors.

We will use the Databricks community edition cloud platform throughout the training course to cover hands-on sessions including (1) big data platform using Spark through R sparklyr package; (2) introduction to Deep Neural Network, Convolutional Neural Network and Recurrent Neural Networks and their applications; (3) deep learning examples using TensorFlow through R keras package. The primary audiences for this course are (1) statistician in traditional industry sectors such as manufacturing, pharmaceutical, and banking; (2) statistician in government agencies; (3) statistical researchers in universities; (4) graduate students in statistics departments. The prerequisite knowledge is MS level education in statistics and entry level of R knowledge. No software installation is needed in participants' laptop and the cloud platform is easily accessed through web browsers such as Chrome or Firefox with the internet connection.

SC3. Analysis of Medical Cost Data: Statistical and Econometric Tools

Full Day | 8:00 am – 5:00 pm
Lei Liu, Washington University in St. Louis
Tina Shih, University of Texas MD Anderson Cancer Center

Description: Rapid growth in medical costs in the U.S. has been a major policy concern and one of the recurrent themes in presidential debates for decades. Medical cost data are routinely collected in billing records of hospitals and claims of health insurance plans (e.g., Medicare, Medicaid, or commercial insurance) or in national surveys (e.g., Medical Expenditure Panel Study). The wide availability of such data has motivated the development and application of state-of-the-art statistical and econometric methods. The policy relevance of medical cost estimates makes medical cost research extremely important because inaccurate statistical inferences could lead to misguided policy decisions.

This short course will summarize the up-to-date analytical methods for medical cost research. The short course will be co-taught by a biostatistician (Liu) and a health economist (Shih), who have collaborated on this topic for more than a decade. This interdisciplinary collaboration has resulted in multiple grants and numerous papers. Medical cost research has also sparked the interest of other quantitative scientists, leading to the development of a growing number of new analytical methods. Therefore, we think the timing is ripe to deliver a short course to summarize the latest methodological development from our group and other researchers to advance the knowledge in medical cost research.

Our short course will offer a comprehensive compilation of new approaches, modeling, and applications on medical cost analyses. It aims not only to synthesize the disparate literature of this fast growing field, but, in doing so, to foster new methodological development, new perspectives, new questions, and a broader understanding of medical cost research. While we intend to discuss recent methodological development in the analysis of medical cost data, materials will be presented in a way that is understandable to clinical researchers and policy analysts with moderate training in statistics and/or econometrics.

This application oriented short course is of interest to researchers who would apply up-to-date statistical tools to medical cost data. 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 medical cost analysis.

SC4. StatTag for Connecting R, SAS, and Stata to Word: A Practical Approach to Reproducibility

Half Day | 8:00 am – 12:00 pm
Abigail Baldridge, Northwestern University
Luke Rasmussen, Northwestern University

Description: Reproducibility, wherein data analysis and documentation is sufficient so that results can be recomputed or verified, is an increasingly important component of statistical practice. "Weaving" tools such as R Markdown facilitate reproducibility by combining narrative text and analysis code in one plain-text document, but are of limited use when manuscripts or reports must be generated in MS Word (e.g. due to journal requirements or client preference). To address this challenge, we have created

StatTag, a free, open-source program that embeds statistical results from R, SAS, or Stata directly in Microsoft Word. StatTag is available as a Word plugin (Windows) or standalone application (Mac) that links statistical code files to Word documents. From Word, users attach one or more code files to an active document, and use the StatTag interface to "tag" selected statistical output – estimates, tables, or figures. The user instructs StatTag to insert the selected statistical output into the Word document, whereupon StatTag invokes the appropriate statistical software and places the result within the document text. Inserted results can then be updated automatically or on demand, and will retain their linkage to the code even when the document changes hands, is redlined, or the text is copied and pasted elsewhere. The StatTag interface also allows direct user interaction with the code file; users may view, edit and re-run statistical code directly from Word. StatTag improves over other similar software in that it functions directly from Word, and it allows the usage of more than one statistical software and code.

In this short course, we will:

- introduce approaches for reproducible research with focus on data analysis and publication
- introduce StatTag, a reproducible research tool for Word with SAS, Stata and/or R
- lead a hands-on session during which participants will generate an abstract with StatTag in the software version of their choice and update their abstract through a brief peer review
- connect users to the StatTag knowledge base and summarize the information learned

This course is intended for a broad audience; prerequisites are experience preparing documents in Word and conducting analysis in any one of R, SAS, or Stata. In addition to the in-person course, participants will have access to an online course and materials before and after the conference.

SC5. Personalized Medicine: Subgroup Identification in Clinical Trials

Half Day | 8:00 am – 12:00 pm

Ilya Lipkovich, Eli Lilly
Alex Dmitrienko, Mediana, Inc

Description: This short course will provide a description of a broad class of statistical methods dealing with exploratory subgroup analysis in clinical trials as one of the key components of personalized medicine. This includes subgroup search/biomarker discovery methods that can be applied both in early and late-phase clinical trials. Subgroup identification from observational data will not be considered.
We will begin with a broad review of existing approaches to subgroup/biomarker identification in the context of personalized medicine illustrating the key elements of principled data-driven subgroup evaluation and then focus on a recursive partitioning method SIDES (Subgroup Identification Based on Differential Effect Search, Lipkovich et al., 2011) and its extensions SIDEScreen (Lipkovich and Dmitrienko, 2014) and Stochastic SIDEScreen (Lipkovich et al, 2017).

Key elements of SIDES and related methods will be discussed including generation of multiple promising subgroups based on different splitting criteria, evaluation of variable importance (VI), implementing VI-based biomarker screening, and addressing Type I error rate and subgroup effect inflation using resampling
based methods.

Case studies from both early and late clinical development programs will be used to illustrate the principles and statistical methods introduced in this course. A software tool implementing SIDES and related methods will be presented (RSIDES package developed by the authors, http://biopharmnet.com/subgroup-analysis-software/).

SC6. Design of Matched Studies with Improved Internal and External Validity

Half Day | 1:00 pm – 5:00 pm

José R. Zubizarreta, Harvard University

Description: In observational studies of causal effects, matching methods are extensively used to approximate the ideal study that would be conducted if controlled experimentation was possible. In this short course, we will explore recent advancements in matching methods to design matched studies with improved internal and external validity. With these matching methods, we will: (1) directly obtain flexible forms of covariate balance, ranging from mean balance to balance of entire joint distributions, (2) produce self-weighting matched samples that are representative of target populations by design, and (3) handle multiple treatment doses without resorting to a generalization of the propensity score, instead balancing the original covariates. We will discuss extensions to matching with instrumental variables, in discontinuity designs, and for matching before randomization in experiments. The methods discussed build upon recent advancements in computation and optimization for large data sets. We will use the statistical software package 'designmatch' for R.

Participants will gain a clear picture of role of matching for causal inferences, and its pros and cons. They will learn how to construct balanced and representative matched samples, improving on traditional matching methods on the estimated propensity score. The target audience of the workshop is applied researchers with quantitative training and familiarity with traditional regression methods. Facility with R is ideal, but not strictly necessary as well-documented step-by-step code will be provided.

SC7. Smart Simulations with SAS and R

Half Day | 1:00 pm – 5:00 pm

Mehmet Kocak, University of Tennessee Health Science Center

Description: In statistical methodology research and practice, simulations are among the ways to show operating characteristics of the proposed method against the existing methods or alternative approaches. Depending on the response variables of interest in such simulations, univariate or multivariate, iterative or non-iterative, simulation designs must be considered very carefully to produce generalizable and reproducible conclusions regardless of the simulation platform, and this task is much more difficult and under-recognized than typically thought. In this short course, we will introduce simple to more complex simulation designs and the importance of simulation size; we will describe potential pitfalls that may not be easily recognizable and suggest what metadata to be captured for a clear description of the simulation process and results. We plan to carry out examples both in SAS and R to show similarities and differences between the two platforms. In doing so, we will utilize Graphical Analytics techniques, which are indispensable components of statistical learning and practice, and must be made part of any simulation plans as well.

Course participants are highly encouraged to have a personal computer with at least one of SAS or R (and R-studio) installed to practice alongside the instructor as the following modules are being presented:

Module-1: Simulating data for univariate random variables following Gaussian Distribution, Student-t-Distribution, Gamma Distribution and its special cases, Beta Distribution, Binomial Distribution, Poisson Distribution, etc.

Module-2: Simulation designs for one-sample hypothesis testing for continuous, binary, and survival endpoints. In this module, we will also illustrate iterative simulation designs such as Phase-I Dose Escalation Design, and Simon's Two-stage designs.

Module-3: Simulation designs for two- or more-sample hypothesis testing for continuous, binary, and survival endpoints. One of the main focus here will be Empirical Power calculations for Randomized Clinical Trials.

Module-4: Simulation designs for Multivariate random variables and designs that require iterative processing. We will compare and contrast SAS and R as two simulation platforms and discuss ways to improve efficiency in simulation
design and conduct.

TUTORIALS

	Tutorial Registration Fees	
	By Feb. 1st	After Feb. 1st
Member	$75	$85
Non-Member	$85	$95
Student	$40	$50

T1. An Introduction to Causal Effect Estimation with Examples Using SAS® Software

Monday, March 25 | 8:30 am – 10:15 am
Michael Lamm, SAS Institute Inc.

Description: How can you estimate a causal effect from nonrandomized data? As statisticians and data scientists are increasingly tasked with analyzing data that come from observational studies rather than randomized experiments, this is a question of increasing importance. This tutorial provides an overview of methods for estimating causal effects for dichotomous treatments. In particular, it illustrates causal effect estimation by propensity-score-based matching, inverse probability weighting, and doubly robust methods by using examples relevant to the biological and life sciences. The analyses are performed using the PSMATCH and CAUSALTRT procedures in SAS/STAT® software. Also briefly discussed are approaches for constructing and evaluating the underlying models, comparisons of the estimation methods, and the assumptions required for identifying and estimating treatment effects.

T2. Building Effective Data Visualizations with ggplot2

Monday, March 25 | 10:30 am – 12:15 pm
Lucy D'Agostino McGowan, Johns Hopkins Bloomberg School of Public Health

Description: "If you're navigating a dense information jungle, coming across a beautiful graphic or a lovely data visualization, it's a relief. It's like coming across a clearing in the jungle." – David McCandless.

The ability to create polished, factual, and easily-understood data visualizations is a crucial skill for the modern statistician. Visualizations aid with all steps of the data analysis pipeline, from exploratory data analysis to effectively communicating results to a broad audience. This tutorial will first cover best practices in data visualization. We will then dive into a hands on experience building intuitive and elegant graphics using R with the ggplot2 package, a system for creating visualizations based on The Grammar of Graphics.

T3. Meta-Analysis of Clinical Trials: Effects-at-Random or Studies-at-Random?

Monday, March 25 | 1:45 pm – 3:30 pm
Jonathan J. Shuster, University of Florida

Description: Meta-Analysis and Systematic Reviews stand at the top of most "Evidence Pyramids". Virtually all random-effects meta-analyses ever done (classical or Bayes) use the "Effects-at-Random" premise, where the random effect size for each study is drawn from an urn and the population mean of the urn is estimated. The almost never used "Studies-at-Random" instead presumes that the observed studies are a random sample of studies, drawn from a large conceptual urn of studies. The important distinction is that in the "effects-at-random" presumption, there can be no association between the random effect sizes and the study design parameters, which determine study weights. It is impossible to prove beyond a reasonable doubt that no such association exists. The framework for inference in studies-at-random, which estimates the mean outcome in the urn of studies, using the study sample sizes as its weights, offers many advantages over effects-at-random. We cite three here. First, in the target population, the mean of each completed study is known without error. Single-stage cluster sampling methods can easily be applied. Second, studies-at-random, but not effects-at-random, recognize that the study sample sizes are random variables, a source of variation conveniently not considered in effects-at-random. Third, the asymptotic distribution of effects-at-random, but not studies-at-random require either normal assumptions or large samples within studies. Both approaches are asymptotic in the number of studies being combined. Of note, we shall present two eye-opening real situations for effects-at-random, where keeping the point estimates as they were, but cutting the standard errors uniformly in half, cause a highly significant result to become non-significant. This cannot happen to studies-at-random. We shall apply studies-at-random methods to three situations: (1) Low event-rate binomial trials, (2) Trials with quantitative endpoints, and (3) Bland-Altman analysis with repeated measures within subjects. Unlike the classical repeated measures Bland-Altman methods, it is completely robust to the lack of independence within subjects.

T4. Modern Multiple Imputation

Monday, March 25 | 3:45 pm – 5:30 pm
Michael R. Elliott, University of Michigan

Description: In the four decades since it was first proposed, multiple imputation has come to offer a comprehensive and practical solution to the problem of making statistical inference when missing data is present. This tutorial will provide a brief overview of the theoretical background behind multiple imputation, and then discuss a variety of practical implementations beyond the fully model-based setting, including use of chained equations, and predictive mean matching. We will conclude with a review of relevant software packages for creating and analyzing multiply imputed datasets, including SAS, R, and IVEWare.

T5. A Primer on Python for Statistical Programming and Data Science

Tuesday, March 26 | 1:45 pm – 3:30 pm
Christopher Fonnesbeck, Vanderbilt University Medical Center

Description: Though Python is ostensibly a general-purpose programming language, it has quickly become a dominant language for machine learning and data science applications. This is due in part to its fundamental strengths as a high-level language, and in part to the powerful set of third-party packages that comprise the Python "scientific stack". In this hands-on tutorial, we will first cover the fundamentals of Python programming, including data structures, control flow, functions, and classes, with particular attention paid to aspects of the language that is idiomatic. The second part of the course will comprise a survey of Python libraries that are relevant for modern data analysis, particularly in the context of data science and probabilistic programming. These include: NumPy, SciPy, Jupyter, pandas, dask, scikit-learn, PyMC3, matplotlib, Seaborn, and TensorFlow. Demonstrations will be motivated with real-data examples, using Jupyter notebooks to allow for interaction and experimentation.

T6. Analysis of Patient-Reported Outcomes

Tuesday, March 26 | 3:45 pm – 5:30 pm
Joseph Cappelleri, Pfizer Inc
Andrew G. Bushmakin, Pfizer Inc

Description: Patient-reported outcomes are often relevant in studying a variety of diseases and outcomes that cannot be assessed adequately without a patient's evaluation and whose key questions require patient's input on the impact of a disease or a treatment. To be useful to patients, researchers and decision makers, a patient-reported outcome (PRO) must undergo a validation process to support that it measures what it is intended to measure accurately and reliably. In this tutorial, the core topics of validity and reliability of a PRO measure will be discussed. In addition, the specialized topics of clinically important responder and clinically important difference on a PRO measure will be featured. Other analytic areas such as longitudinal data analysis of a PRO measure will be highlighted. Illustrations will be provided through real-life and simulated examples, including simulation-based learning of the methodologies. Material is drawn in part from the book "Patient-Reported Outcomes: Measurement, Implementation and Interpretation" (Cappelleri, Zou, Bushmakin et al. 2013).

ROUNDTABLES

Registration is required. Roundtable Registration Fee: $40.00

Roundtables Monday, March 25, 2019 | 12:15 pm – 1:30 pm

R1. From Working Group to Center: How to Establish and Grow Research Groups

Jason Roy, Rutgers School of Public Health

Description: Research working groups can be great breeding grounds for new ideas and cross-disciplinary collaborations. However, there are challenges in maintaining active participation, given competing demands on people's time. In this roundtable, we will discuss this in the context of our experience growing a causal inference working group and later establishing the Center for Causal Inference. Whether your goal is to have a small working group or to form a research center, you should come away with useful tips to increase participation and productivity.

SOLD OUT

R2. Time Management: Taming Your Inbox

Elizabeth Stuart, Johns Hopkins Bloomberg School of Public Health

Description: Do you have trouble keeping email under control? Do things slip through the cracks? o you worry about not being able to find time for focused and intense work ("important but not urgent"), due to the "urgent but not important" tasks that come up? This roundtable will discuss strategies for time management, especially for dealing with email and the tasks that come with it. Come with your strategies and hear from others about what works for them!

SOLD OUT

R3. Practical Issues in Clinical Trial Design and Analysis for Precision Medicine

Peter F. Thall, M.D. Anderson Cancer Center

Description: This round table discussion will focus on methods for dealing with practical considerations that arise in the process of clinical trial design, conduct, and analysis, with particular attention to newer phase I-II and randomized trial designs that include subgroup-specific decision making. 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.

SOLD OUT

R4. Statistical Leadership

Bhramar Mukherjee, University of Michigan Rogel Cancer Center

Description: In this discussion, I will focus on two types of leadership positions, an outward leadership challenge as a statistician leading a group of non-quantitative but exceptionally talented biomedical researchers and an inward leadership role as a Department chair to lead a Biostatistics department to the next generation. In the era of health big data, we need to lead as both an active doer and a careful thinker. I have benefitted greatly in my own career by simply being at the table where scientific strategies are being defined and scientific discoveries are being made. As a statistician, it is important to be in the room where it happens and truly learn to embrace, adopt and appreciate the diversity in data science and in the society!

SOLD OUT

R5. Strategies for Success in Publishing

Heping Zhang, Yale University

Description: Publishing in peer-reviewed journals is a fundamental expectation for all biostatisticians. Understanding how peer-review and the editorial process works, and what makes an effective journal articles are critical for success in publishing. This roundtable will focus on a range of strategies for successful publication of your research.

SOLD OUT

R6. Submitting your Grant to NIH

Peter Kozel, NIH/NIDDK

Description: Have you ever been confused by the NIH grant system? Want to know tips for working out the appropriate institute and funding opportunities to submit to? Interested in what happens after you submit your application? The objective of this roundtable is to raise an awareness of how the NIH peer review process works, and to discuss some general do's and don'ts of application submission.

SOLD OUT

R7. Analytic Challenges of Administrative Health Data

Rebecca Hubbard, University of Pennsylvania

Description: There is currently enormous enthusiasm for conducting research using data from electronic health records (EHR). However, analyzing EHR data entails many practical challenges. This roundtable will discuss key challenges for the analysis of EHR data including: missing and mismeasured variables, confounding by indication, and informative observation processes. The objective of the roundtable will be to raise awareness about concerns arising in the analysis of EHR data and to share best practices for addressing these challenges.

R8. Tips for Interviewing Well

Joseph C. Cappelleri, Pfizer Inc

Description: This roundtable will focus on ways to interview well and therefore increase the chance of receiving a job offer. Many tips will be provided and discussed. Among them are researching the industry and the institution, clarifying your "selling points" and the reasons you want the job, and preparing for common interview questions. Basic and subtle interviewing skills will be discussed including (among others) how to make a good first impression, to get on the same side as the interviewer, and to empower yourself through thinking positive.

R9. Effective Collaboration and Statistical Leadership—in Drug Development and Beyond

Lei Shen, Eli Lilly and Company

Description: Modern day drug development is highly complex and requires deep technical expertise in many scientific disciplines as well as effective collaboration among teammates with different expertise. The skills to address challenges that inevitably come up along the way are not the monopoly of any single discipline, but a case can be made that statisticians - thanks to our training in uncertainty quantification and ability to think in probabilistic terms - are in a prime position to step up as problem solvers. Somewhat paradoxically, our statistical leadership can be greatly enhanced by not being merely "statisticians who work in drug development", but rather "drug developers who happen to be statisticians". This roundtable gives us the opportunity to discuss and debate the requisite skillsets for us to develop - and conscious effort to make - to be successful in this realm. And are there close parallels in other industries, in government agencies, and in academic research?