ICSA Short Course Information

2022 ICSA Short Course Schedule

Time	Presenters	Title
06/19/2022, AM	Prof. Babette Brumback	Causal Inference with R
	Dr. Ira Longini Dr. Yang Yang Dr. Matt Hitchings	Statistical methods for analyzing transmission and control of infectious diseases
06/19/2022, PM	Dr. Debashis Mondal	Spatial analysis with Gaussian Markov random fields
	Prof. Sujit Ghosh Dr. Amy Shi	Bayesian Computational Tools for Clinical Data
06/19/2022, Whole Day	Dr. Chenguang Wang	Leveraging Real-World Data in Clinical Trial Design and Analysis
	Prof. Jeffrey Wilson Prof. Din Chen	Marginal Models in Analysis of Correlated Binary Data with Time-Dependent Covariates

ICSA Short Course Information

Course Title

Causal Inference with R

Abstract

One of the primary motivations for clinical trials and observational studies of humans is to infer cause and effect. Disentangling causation from confounding is of utmost importance. Causal Inference with R explains and relates different methods of confounding adjustment in terms of potential outcomes and graphical models, including standardization, doubly robust estimation, difference-in-differences estimation, and instrumental variables estimation. Several real data examples, simulation studies, and analyses using R motivate the methods throughout. The course assumes familiarity with basic statistics and probability, regression, and R. The course will be taught with a blend of lecture and worked examples.

Teaching Plan (Half-Day)

First part:

Introduction -- 15 minutes
Potential Outcomes and Effect Measures -- 30 minutes
Causal Directed Acyclic Graphs -- 1 hr
15 minute break

Second part:

Standardization and Doubly Robust Estimation -- 1 hr
Difference-in-Differences Estimation -- 30 minutes
Instrumental Variables Estimation -- 30 minutes

Prerequisites

Familiarity with regression and R

Category

Methodology

Short Course Presenter (contact in bold)

Presenter: **Prof. Babette Brumback** *Email:* brumback@ufl.edu
Affiliation: University of Florida

Biography:

Babette A. Brumback, Ph.D. is Professor in the Department of Biostatistics at the University of Florida; she won the department's Outstanding Teacher Award for 2020-2021. A Fellow of the American Statistical Association, she has researched and applied methods for causal inference since 1998, specializing in methods for time-dependent confounding, complex survey samples and clustered data.

ICSA Short Course Information

Course Title

Leveraging Real-World Data in Clinical Trial Design and Analysis

Abstract

The amount of real-world data (RWD) collected from sources other than protocol-driven clinical studies is increasing ultra-rapidly. The clinical evidence that can be derived from analysis of these RWD is considered as real-world evidence (RWE) that can complement the knowledge derived from traditional well-controlled clinical trials. Leveraging RWE can potentially save time and cost of the investigational study and improve the efficiency of regulatory decision-making.

Incorporating RWD in regulatory decision-making demands much more than "mixing" RWD with investigational clinical trial data. The RWD has to undergo appropriate analysis for deriving the right RWE. Moreover, such analysis has to be integrated with the design and analysis of the investigational study for regulatory decision-making. The standard clinical trial toolbox does not offer ready solutions for incorporating RWD.

In this course, the instructor(s) will cover a series of methods they have developed for leveraging real-world data in clinical trial design and analysis. Their work has been recognized by the FDA and received The FDA CDRH Excellence in Scientific Research Award and The FDA Scientific Achievement Award.

Teaching Plan (One-Day)

In Part I of the course, we introduce a new method for proposing performance goals—numerical target values pertaining to effectiveness or safety endpoints in single-arm medical device clinical studies—by leveraging RWE. The method applies entropy balancing to address possible patient dissimilarities between the study's target patient population and existing real-world patients, and can take into account operation differences between clinical studies and real-world clinical practice.

In Part II of the course, we introduce a method that extends the Bayesian power prior approach for a single-arm study to leverage external RWD. The method uses propensity score methodology to pre-select a subset of RWD patients that are similar to those in the current study in terms of covariates, and to stratify the selected patients together with those in the current study into more homogeneous strata. The power prior approach is then applied in each stratum to obtain stratum-specific posterior distributions, which are combined to complete the Bayesian inference for the parameters of interest.

In Part III of the course, we introduce several extensions of the PS-integrated method in Part II. These extensions include 1) a frequentist PS-integrated composite likelihood approach for incorporating RWE in single-arm clinical studies; 2) leveraging multiple RWD sources in single-arm medical device clinical studies; 3) leveraging RWD for the evaluation of diagnostic tests for low prevalence diseases; 4) augmenting both arms of a randomized controlled trial by leveraging RWD; and 5) PS-integrated approach for survival analysis.

In Part IV of the course, we describe an R package, psrwe, that implements a PS-integrated power prior (PSPP) method, a PS-integrated compos- ite likelihood (PSCL) method, and a PS-integrated weighted Kaplan-Meier estimation (PSKM) method for the methods in Parts II and III. Illustrative examples are provided to demonstrate each of the approaches.

In Part V of the course, we introduce a propensity score-based Bayesian non-parametric Dirichlet process mixture model that summarizes subject-level information from randomized and RWD to draw inference on the causal treatment effect in exploratory analysis.

Prerequisites

Basic understanding about clinical trials; basic knowledge about R

Category

Methodology

Short Course Presenter (contact in bold)

Presenter: **Dr. Chenguang Wang** *Email:* chenguang.wang@regeneron.com

Affiliation: Regeneron Pharmaceuticals, Inc.

Biography:

Dr. Chenguang Wang is a Senior Director and the Head of Statistical Innovation at Regeneron. Previously, Dr. Wang was an Associate Professor with Johns Hopkins University and an FDA Mathematical Statistician. Dr. Wang has extensive experience in clinical trial design and analysis in the regulatory setting. Dr. Wang holds B.S. and M.S. degrees in Computer Science and has abundant experience developing statistical software.

ICSA Short Course Information

Course Title

Marginal Models in Analysis of Correlated Binary Data with Time-Dependent Covariates

Abstract

This workshop is based on the book: "Marginal Models in Analysis of Correlated Binary Data with Time Dependent Covariates" co-authored by Drs. Jeffrey R. Wilson, Elsa Vazquez-Arreola, and (Din) Ding-Geng Chen, published by Springer in 2020, which is the first book to systematically introduce marginal models to analyze correlated binary data with time-dependent covariates in clinical trials and observational studies using R and SAS. This workshop provides a thorough presentation of correlated binary data with time-dependent covariate. It gives a detailed step-by-step illustration of their implementation using R and SAS. Longitudinal data or contain correlated data due to the repeated measurements on the same subject. The changing values usually consist of time-dependent covariates and their association with the outcomes present different sources of correlation. Most methods used to analyze longitudinal data would average the effects of time-dependent covariates on outcomes over time and provide a single regression coefficient per time-dependent covariate. Such an approach prevents analysts and researchers the opportunity to following the changing impact of time-dependent covariates on the outcomes. The workshop addresses such issues through the use of partitioned regression coefficients. We further use examples of correlated data with time-dependent covariate on obesity from the Add Health study and cognitive impairment diagnosis in the National Alzheimer's Coordination Center

Teaching Plan (One-Day)

Morning Session (8:30am to 12:30pm):

1. Fundamentals of estimation of regression coefficients in cross-sectional data
   a. Review of the estimation of regression models
   b. Generalized estimating equation (GEE) and generalized linear mixed models
   c. Generalized Method of Moments estimates;
2. Presentation on data with time-dependent covariates and discussion on the partitioned matrix.

Afternoon Session (1:30pm to 4:30pm):

3. Present correlated data with time-dependent covariates. Illustrate longitudinal data and the analysis using linear mixed models for continuous endpoints, generalized linear mixed model and GEE for categorical endpoints.
4. Bayesian analysis in this partitioned data matrix using MCMC is applied.

Prerequisites

Classical Regression and Generalized linear model

Category

Methodology

Short Course Presenter (contact in bold)

*Presenter 1:*Prof. Jeffrey Wilson *Email:* jeffrey.wilson@asu.edu
Affiliation: Arizona State University
Biography:

Dr. Jeffrey Wilson is a Professor of Statistics and Biostatistics at Arizona State University. Dr. Wilson's research experience includes grants as PI and co-PI from the NIH, NSF, USDA, Arizona Department of Health Services, and the Arizona Disease Research Commission. He is presently the Statistics Associate Editor for The Journal of Minimally Invasive Gynecology and a former Chair of the Editorial Board of the American Journal of Public Health. He has published more than 85 articles in leading journals such as Statistics in Medicine, American Journal of Public Health, Journal of Royal Statistics Society, Computational Statistics, and Australian Journal of Statistics, among others. He has consulted with pharmaceutical companies and hospitals while representing them before the FDA and other federal government healthcare agencies. He has taught specialized Biostatistics classes at Mayo Clinic. He has led similar courses for Phoenix Children's Hospital, Barrow Neurological Center, St. Joseph's Hospital, and Banner Hospital. He is the former Director of the School of Health Management and Policy He is a former Director and co-Director of the Biostatistics Core in the NIH Center for Alzheimer at Arizona State University.

*Presenter 2:***Prof. Din Chen** *Email:* dinchen@asu.edu
Affiliation: Arizona State University
Biography:

Dr. (Din) Ding-Geng Chen is now the executive director and professor in biostatistics at College of Health Solutions, Arizona State University. He was the Wallace H. Kuralt distinguished professor in Biostatistics at University of North Carolina-Chapel Hill, a professor in biostatistics at the University of Rochester Medical Center, the Karl E. Peace endowed eminent scholar chair and professor in biostatistics from the Jiann-Ping Hsu College of Public Health at the Georgia Southern University. Dr. Chen is an elected fellow of the American Statistical Association (ASA), an elected member of the International Statistics Institute (ISI), and a senior expert consultant for biopharmaceuticals and government agencies with extensive expertise in clinical trial biostatistics. Dr. Chen has more than 200 referred professional publications and co-authored/co-edited 33 books on biostatistics clinical trials, biopharmaceutical statistics, interval-censored survival data analysis, meta-analysis, public health statistics, statistical causal inferences; statistical methods in big-data sciences and Monte-Carlo simulation-based statistical modeling. Dr. Chen has been invited nationally and internationally to give short courses at various scientific conferences.

ICSA Short Course Information

Course Title

Statistical methods for analyzing transmission and control of infectious diseases

Abstract

Application of statistical inference methods to infectious disease data is a key tool in understanding transmissibility of pathogens and the effectiveness of interventions. In this half-day course, we will learn about different sources of data that arise from passive surveillance, active case finding and clinical studies, and methods for inferring key parameters from such data. The types of data sources to be covered include epidemic curve data, household-based observational data, and data arising from serosurveillance studies. We will also cover common computational algorithms for statistical inference and a few software packages that implement these algorithms. In addition, we will briefly introduce several advances in modeling frameworks to address challenges arising from the pandemic of COVID-19. Upon completion of this course, participants will recognize the various types of infectious disease data, common models designed to analyze these data, key parameters of epidemiological importance including intervention efficacies, and promising research directions in the field of infectious disease modeling.

Teaching Plan (Half-Day)

The course will be divided into three sessions each of 70min, with two 15-min breaks.

First session: History of infectious disease modeling; types of infectious disease data (case numbers, serology, household data including time of symptom onset) and the underlying hierarchy of information; Overview of transmission parameters of epidemiological importance such as the basic reproductive number, final attack rate, and secondary attack rate; Different measures of vaccine efficacies and effectiveness of vaccination programs.

Second session: Detail on classic models that are fitted to epidemic curve data, final size models with fixed and random infectious periods for close contact groups (e.g., households), discrete-time chain binomial models and continuous-time survival models for sequential data of symptom onsets or laboratory confirmations among close contact groups, statistical inference from serosurveillance data, and agent-based models.

Third session: Computational methods (EM and Monte Carlo EM algorithms, traditional MCMC, Approximate Bayesian Computing, Particle Filtering, and Hamiltonian Monte Carlo). We will introduce a few R packages (e.g. surveillance, transtat, serosolver) and show some data examples; recent advances in statistical transmission models to address challenges the a rose during the pandemic of COVID-19 (e.g., presymptomatic and asymptomatic infectiousness, under-testing, delayed reporting, etc.).

Prerequisites

Master-level statistical inference

Category

Methodology and applications

Short Course Presenter (contact in bold)

Presenter 1: Dr. Ira Longini *Email:*ilongini@ufl.edu
Affiliation: University of Florida
Biography:

Dr. Ira Longini is a professor of biostatistics in the College of Public Health and Health professions as well as Emerging Pathogens Institute at the University of Florida. He works on the mathematical modeling, stochastic processes and biostatistics applied to epidemiological infectious disease problems. He has specialized in the mathematical and statistical theory of epidemics--a process that involves constructing and analyzing mathematical models of disease transmission, disease progression and the analysis of infectious disease data based on these models. In addition, he works extensively in the design and analysis of vaccine and infectious disease prevention trials and observational studies.

*Presenter 2:***Dr. Yang Yang** *Email:* yangyang@ufl.edu
Affiliation: University of Florida
Biography:

Dr. Yang Yang is an associate professor of biostatistics in the College of Public Health and Health professions as well as Emerging Pathogens Institute at the University of Florida. His research focuses on statistical methods for disease transmission dynamics, efficacy evaluation, missing data and surveillance bias. He also works on ecological modeling and genetic association for clinical outcomes.

*Presenter 3:*Dr. Matt Hitchings *Email:* mhitchings@ufl.edu
Affiliation: University of Florida
Biography:

Dr. Matt Hitchings is an Assistant Professor in the Department of Biostatistics at the University of Florida. His primary focus is evaluating the effectiveness of interventions against infectious disease, through clinical trials, observational studies, and development and application of mathematical models. Recently he has been conducting observational studies of vaccine effectiveness using passive surveillance data in Brazil, and developing a framework for analysis of serological data for pathogens including SARS-CoV-2 and dengue virus.

ICSA Short Course Information

Course Title

Spatial analysis with Gaussian Markov random fields

Abstract

Gaussian Markov random fields have been applied with much success to account for discrete spatial variation in both lattice and areal unit data. Applications include astronomy, agriculture, computer vision, climate studies, epidemiology, image analysis, geology and other areas of environmental science. Lattice-based Gaussian Markov random fields are extremely adaptable to swift and uncomplicated statistical computations and provide ways to develop complex and hierarchical models through local specifications, and, for these reasons, have contributed to considerable success in the analysis of spatial data. This short course gives an introduction to spatial models based on Gaussian Markov random fields. The course covers statistical computation for spatial linear mixed models, particularly, residual maximum likelihood (REML) estimation and kriging or prediction. The course also presents statistical computation for general spatial mixed models using Markov Chain Monte Carlo (MCMC) sampling methods. Practicum sessions will introduce various R codes with applications from environmental sciences and geographical epidemiology.

The course will end with a summary of the topics and ideas covered and a list of further resources.

Teaching Plan (Half-Day)

Lecture 1: Introduction to spatial statistics, Gaussian Markov random fields, conditionals and intrinsic autoregressions.

Lecture 2: Spatial mixed models, REML, kriging, h-likelihood and MCMC computations.

Break

Lecture 3 and 4: Statistical calculations using R-codes. Applications from environmental sciences and geographical epidemiology.

Summary and further resources.

Recommended texts: NONE, but the following may be useful as secondary reading materials.

1. Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B (Methodological), 36(2), 192-225.

2. Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. Annals of the institute of statistical mathematics, 43(1), 1-20.

3. Besag, J., & Mondal, D. (2005). First-order intrinsic autoregressions and the de Wijs process. Biometrika, 92(4), 909-920.

4. Cressie, N. (2015). Statistics for spatial data. John Wiley & Sons.

5. Dutta, S., & Mondal, D. (2015). An h>likelihood method for spatial mixed linear models based on intrinsic auto>regressions. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77(3), 699-726.

6. Gelfand, A. E., Diggle, P., Guttorp, P., & Fuentes, M. (2010). Handbook of spatial statistics. CRC press.

7. Rue, H., & Held, L. (2005). Gaussian Markov random fields: theory and applications. Chapman and Hall/CRC.

Prerequisites

Students should have basic knowledge of statistics and inference, i.e., understanding of likelihood functions and linear regressions. A basic knowledge of probability and stochastic processes will be useful, but it is not a prerequisite.

Category

Theory and applications

Short Course Presenter (contact in bold)

Presenter: **Dr. Debashis Mondal** *Email:* mondal@wustl.edu

Affiliation: Washington University

Biography:

Debashis Mondal, PhD, is an associate professor in the Department of Mathematics and Statistics at Washington University in St Louis. Mondal's research interests include spatial statistics; computational science and machine learning; and applications in environmental sciences, ecology, including microbial ecology, and geographical epidemiology. Mondal won an NSF CAREER Award in 2013 and the International Indian Statistical Association's Young Researcher Award in 2015. He is also an elected member of the International Statistical Institute. Mondal earned his doctorate in statistics at the University of Washington, Seattle.

ICSA Short Course Information

Course Title

Bayesian Computational Tools for Clinical Data

Abstract

The Bayesian paradigm provides a structured and practical way of expressing complicated models through a sequence of simple conditional distributions making them useful for simple to complex data structures required to address multiple phases of clinical trials, particularly for those that involves different types of data irregularities (missing values, censored data, etc.). Over the recent years there have been tremendous efforts on developing Bayesian analytics for leveraging data from sources outside of prospectively designed study, referred to as external data such as various Real-World-Data (RWD) sources, historical clinical data, and data from multiple trials within a grand hierarchical structure. Thus, development of appropriate statistical models and related inference are warranted that are not only based on solid theoretical guarantees but also making sure that such complex models are estimable and interpretable in practical settings for modern clinical trials. Thus, one of the main goals of the proposed short course is to present the modern analytical tools that are easily accessible to practitioners by providing a glimpse of theoretical backgrounds supplemented by many practical examples derived from real case studies. This will be accomplished by illustrating numerous real-data examples (using software demos) ranging from two-arm trials to more complex hierarchical models that involves handling data irregularities commonly faced by practitioners.

Teaching Plan (Half-Day)

The first part of the short course will begin with a brief overview of Bayesian machine learning (BML) methods for randomized controlled trials (RCTs) using various study designs including sample size determination methods. In particular, it will showcase the use of Bayesian posterior predictive methods for properly handling missing and censored data, a feature that are not readily employed my routine ML methods. The second part of the course will involve more realistic and complex models that have recently emerged in the modern era used by pharmaceutical industries and regulatory agencies, and then showcase the use of modern BML methods through various real case studies. Throughout the tutorial practical applications and worked-out examples will be emphasized without getting into the theoretical underpinnings of the methods, but relevant literature will be provided for those wishing to learn more in-depth notions of BML tools. The concepts and methods discussed will be demonstrated using the popular software packages (R and SAS) developed by the presenters, but those are implementable by any other software capable of coding Markov Chain Monte Carlo (MCMC) methods. The two-parts of the course will consist of the following topics:

Part I - Introduction to Bayesian Methods for Clinical Trials

1. Basics of Bayesian Methods for RCTs (20min)
2. Predictive Distributions and Sample Size Determination (20min)
3. Computational Methods using Monte Carlo Methods (35min)
4. Primer on Bayesian Software (via R, Stan and SAS) (30min)

(15min break)

Part II – BML methods with real-data examples

1. Bayesian regression models using 'brms' R package (35min)
2. GLMs and Multi-level models PROC BGLIMM (40min)
3. Penalized regression models with data irregularities (30min)
4. Q&As and additional demos on demand (15min)

[If this proposal is selected, please assign this half-day short course in the afternoon session 1:30-5:30pm]

Prerequisites

Basic graduate level statistical inference, familiarity with SAS and R.

Category

methodology and technology training

Short Course Presenter (contact in bold)

*Presenter 1:***Prof. Sujit Ghosh** *Email:* sujit.ghosh@ncsu.edu

Affiliation: North Carolina State University

Biography:

Professor Sujit Kumar Ghosh has over 25 years of experience in conducting, applying, evaluating and documenting statistical analysis of biomedical and environmental data. Prof. Ghosh is actively involved in teaching, supervising and mentoring graduate students at the doctoral and master levels. He has supervised over 40 doctoral graduate students and published over 125 peer-reviewed journal articles in various areas of statistics with applications in biomedical and environmental sciences, econometrics and engineering. He has recently co-authored a book (with Dr. Reich) titled "Bayesian Statistical Methods," which is being used as a textbook at several universities. Prof. Ghosh has delivered over 180 invited lectures, seminars at national and international meetings. He has also delivered several short courses and served as short-term visiting professor at several institutions in various countries. Prof. Ghosh received the International Indian Statistical Association (IISA) Young Investigator Award in 2008; was elected a Fellow of the American Statistical Association (ASA) in 2009; was elected as the President of the NC Chapter of ASA in 2013 and also elected as the President of the IISA in 2017.

Presenter 2: Dr. Amy Shi *Email:* amyxyshi@hotmail.com
Affiliation: AstraZeneca Pharmaceutical
Biography:

Dr. Amy Shi is currently a Statistical Science Associate Director at AstraZeneca Pharmaceutical in the Late CVRM (Cardiovascular Renal Metabolism) group. Much of her work involves with taking part in clinical trials as a statistician and researching for innovative statistical methods. Before joining AstraZeneca, she was a Principal Research Statistician Developer in the Bayesian Modeling Group at SAS from 2010 to 2021. Her job responsibility was to enhance the Bayesian capabilities of SAS software, with a focus on generalized linear mixed models, multilevel hierarchical settings, variable selection, choice modeling, and machine learning. She developed a couple of SAS Bayesian procedures (PROC BCHOICE and PROC BGLIMM) and many functional packages. Dr. Shi has a MS in Statistics from the Michigan State University and a Ph.D. in Biostatistics from the University of North Carolina at Chapel Hill.