Bio³ Seminar Series

Jianguo Sun, Ph.D.
Professor at the Department of Statistics
University of Missouri

Title: Statistical Analysis of Interval-Censored Time-to-Event Data

Abstract: The analysis of failure time data plays an important and essential role in many studies, especially medical studies such as clinical trials and follow-up studies. One key feature of failure time data that separates the failure time data analysis from other fields, is censoring, which can occur in different forms. In this talk, we will discuss and review a general form, interval censoring, and the existing literature for the analysis of interval-censored data as well as some research topics.

FRIDAY, September 9, 2016
10:00am – 11:00am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Yongzhao Shao, Ph.D.
Professor of Population Health and Environmental Medicine, Deputy Director of New York University Cancer Institute Biostatistics Shared Resources

Title: Prognostic Accuracy for Semi-parametric Mixture Cure Models

Abstract: An unmet significant challenge in the treatment of many early-stage cancers is the lack of effective prognostic models to identify patients who are at high risk of disease progression from a large number of potentially cured patients. Semi-parametric mixture cure models can account for latent cure fractions in patient populations thus are more suitable prognostic models than standard survival models such as Cox Proportional Hazard models or Proportional Odds models that ignore the existence of latent cure fractions. Without the requirement of knowing who is surely cured, the semiparametric mixture cure models can be used to evaluate predictive utility of biomarkers on cure probability and on survival of uncured subjects. However, appropriate statistical metrics to evaluate prognostic efficiency in the presence of cured patients have been lacking. In this paper, we introduce concordance-based prognostic metrics for semi-parametric mixture cure models and develop consistent estimates. The asymptotic normality and confidence intervals of these estimates are also established. Finite sample applicability of the developed indices and estimates are investigated using numerical simulations and illustrated using a melanoma data set. This talk is based on joint work with Dr. Yilong Zhang at Merck.

FRIDAY, September 23, 2016
10:00am – 11:00am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Bio³ Seminar Series

Dennis Lin, Ph.D.
University Distinguished Professor of Statistics, Pennsylvania State University

Title: Dimensional Analysis and Its Applications in Statistics

Abstract: Dimensional Analysis (DA) is a fundamental method in the engineering and physical sciences for analytically reducing the number of experimental variables prior to the experimentation. The principle use of dimensional analysis is to reduce from a study of the dimensions of the variables on the form of any possible relationship between those variables. The method is of great generality. In this talk, an overview/introduction of DA will be first given. A basic guideline for applying DA will be proposed, using examples for illustration. Some initial ideas on using DA for Data Analysis and Data Collection will be discussed. Future research issues will be proposed.

FRIDAY, October 14, 2016
10:00am – 11:00am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Felix Elwert, Ph.D.
Associate Professor of Sociology
University of Wisconsin-Madison

Title: Graphical Causal Models

Abstract: This talk introduces the three central uses of directed acyclic graphs (DAGs) for causal inference in the observational biomedical and social sciences. First, DAGs provide clear notation for the researcher's theory of data generation, against which all causal inferences must be judged. Second, DAGs reveal to what extent the researcher's data-generating model can be tested. Third, researchers can inspect the DAG to determine whether a given causal question can be answered ("identified") from the data. After introducing basic building blocks, we will discuss a number of real examples to demonstrate how DAGs help solve thorny practical problems in causal inference.

FRIDAY, October 28th, 2016
10:00am – 11:00am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Keith Muller, Ph.D.
Associate Chair and Professor,
Institute for Child Health Policy
University of Florida

Title: Four Statistical Guidelines for Planning Reproducible Research

Abstract: Concerns about reproducibility in science are widespread. In response, the National Institutes of Health has changed review procedures and training requirements for applicants (https://www.nih.gov/research-training/rigor-reproducibility). The Director of NIH and his deputy outlined their plans in Collins and Tabak (2014). Key methodological concerns include poor study designs, incorrect statistical analyses, inappropriate sample size selection, and misleading reporting. Planners can avoid the concerns by following four statistical guidelines. 1) Explicitly control both Type I errors (false positives) and Type II errors (false negatives). 2) Align the scientific goals, study design, data analysis plan, and the sample size analysis. 3) Vary inputs to the sample size analysis to determine the sensitivity to the values assumed. 4) Account for statistical uncertainty in inputs to sample size computations. Extending the guidelines to sequences of studies requires careful allocation of exploratory and confirmatory analyses (leapfrog designs) and allows some forms of adaptive designs. We give examples in the talk for a variety of designs and hypotheses. Case studies include a randomized drug trial in kidney disease, an observational study of quality of care in Medicaid, and a neurotoxicology experiment in rats. Analytic and simulation results provide the foundation for the conclusions.

FRIDAY, November 11, 2016
10:00am – 11:00am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Mei-Ling Ting Lee, Ph.D.
Professor, Department of Epidemiology and Biostatistics
Director, Biostatistics and Risk Assessment Center
University of Maryland, College Park
School of Public Health
Editor-in-Chief, Lifetime Data Analysis

Title: Threshold Regression Models with Application in a Multiple Myeloma Clinical Trial

Abstract: This presentation reviews methods for comparative effectiveness research using Cox regression methods are well-known. It has, however, a strong proportional hazards assumption. In many medical contexts, a disease progresses until a failure event (such as death) is triggered when the health level first reaches a failure threshold. I'll present the Threshold Regression (TR) model for patient's latent health process that requires few assumptions and, hence, is quite general in its potential application. We use TR to analyze data from a randomized clinical trial of treatment for multiple myeloma. A comparison is made with a Cox proportional hazards regression analysis of the same data.

FRIDAY, January 13, 2017
9:30am – 10:30am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:15am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Goodarz Danaei, M.D. Sc.D.

Assistant Professor, Department of Global Health and Population, Department of Epidemiology
Harvard University, School of Public Health

Title: Observational data for comparative effectiveness research: An emulation of randomized trials of statins & primary prevention of coronary heart disease

Abstract: This presentation reviews methods for comparative effectiveness research using observational data. The basic idea is using an observational study to emulate a hypothetical randomised trial by comparing initiators versus non-initiators of treatment. After adjustment for measured baseline confounders, one can then conduct the observational analogue of an intention-to-treat analysis. We also explain two approaches to conduct the analogues of per-protocol and as-treated analyses after further adjusting for measured time varying confounding and selection bias using inverse-probability weighting. As an example, we implemented these methods to estimate the effect of statins for primary prevention of coronary heart disease (CHD) using data from electronic medical records in the UK. Despite strong confounding by indication, our approach detected a potential benefit of statin therapy. The analogue of the intention-to treat hazard ratio (HR) of CHD was 0.89 (0.73, 1.09) for statin initiators versus non-initiators. The HR of CHD was 0.84 (0.54, 1.30) in the per-protocol analysis and 0.79 (0.41, 1.41) in the as-treated analysis for 2 years of use versus no use. In contrast, a conventional comparison of current users versus never users of statin therapy resulted in a HR of 1.31 (1.04, 1.66). We provide a flexible and annotated SAS program to implement the proposed analyses.

FRIDAY, January 27, 2017
10:00am – 11:00am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Bio³ Seminar Series

Yinglei Lai, Ph.D.

Professor of Statistics,
Department of Statistics,
George Washington University

Title: Concordant integrative analysis of multiple two-sample genome-wide expression data sets

Abstract: The development of microarray and sequencing technologies enables biomedical researchers to collect and analyze large-scale molecular data. We will introduce our recent studies on the concordant integrative approach to the analysis of multiple related two-sample genome-wide expression data sets. A mixture model is developed and yields concordant integrative differential expression analysis as well as concordant integrative gene set enrichment analysis. As the number of data sets increases, it is necessary to reduce the number of parameters in the model. Motivated by the well-known generalized estimating equations (GEEs) for longitudinal data analysis, we focus on the concordant components and assume some special structures for the proportions of non-concordant components in the mixture model. The advantage and usefulness of this approach are illustrated on experimental data.

FRIDAY, February 10, 2017
9:30am – 10:30am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:15am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Peter Song, Ph.D.

Professor, Department of Biostatistics, University of Michigan

Title: Fusion Learning of Model Heterogeneity in Data Integration

Abstract: As data sets of related studies become more easily accessible, combining data sets of similar studies is often undertaken in practice to achieve a larger sample size and higher power. A major challenge arising from data integration pertains to data heterogeneity in terms of study population, study design, or study coordination. Ignoring such heterogeneity in data analysis may result in biased estimation and misleading inference. Traditional techniques of remedy to data heterogeneity include the use of interactions and random effects, which are inferior to achieving desirable statistical power or providing a meaningful interpretation, especially when a large number of smaller data sets are combined. In this paper, we propose a regularized fusion learning method that allows us to identify and merge inter-model homogeneous parameter clusters in regression analysis, without the use of hypothesis testing approach. Using the fused lasso, we establish a computationally efficient procedure to deal with large-scale integrated data. Incorporating the estimated parameter ordering in the fused lasso facilitates computing speed with no loss of statistical power. We conduct extensive simulation studies and provide an application example to demonstrate the performance of the new method with a comparison to the conventional methods. This is a joint work with Lu Tang.

FRIDAY, February 24, 2017
10:00 am – 11:00 am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Jenna Krall, Ph.D.

Assistant Professor, Global and Community Health,
College of Health and Human Services,
George Mason University

Title: Estimating Sources of Air Pollution and Their Impact on Human Health

Abstract: Exposure to particulate matter (PM) air pollution has been associated with increased mortality and morbidity. PM is a complex chemical mixture, and associations between PM and health vary by its chemical composition. Identifying which sources of PM, such as motor vehicles or wildfires, emit the most toxic pollution can lead to a better understanding of how PM impacts health. However, exposure to source-specific PM is not directly observed and must be estimated from PM chemical component data. Source apportionment models aim to estimate source-specific concentrations of PM and the chemical composition of PM emitted by each source. These models, while useful, have some limitations. Specifically, the models are not identifiable without additional information, the estimated source chemical compositions may not match known source compositions, and the models are difficult to apply in multicity studies. In this talk, I introduce source apportionment models and discuss current challenges and opportunities in their application. I estimate sources and their health effects in two studies: a study of commuters in Atlanta, GA and a multicity time series study of four U.S. cities.

FRIDAY, March 24, 2017
10:00am – 11:00am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center

Bio³ Seminar Series

Yuelin Li, Ph.D.

Associate Attending Behavioral Scientist,
Department of Epidemiology & Biostatistics,
Memorial Sloan Kettering Cancer Center

"Neurocognitive Impairment after Cancer, Aggression under Stress, and Supreme Court Justices' voting: What Do They Have in Common?"

Abstract: The Rasch Model (RM) is a classic IRT (Item Response Theory) model in psychometrics. RM is used to solve various applied problems including the measurement of a psychological construct, the scoring of patient-reported outcomes, and in understanding the politics in the high court. I will begin with the measurement of aggression as an example on how to fit a Rasch Model using Gibbs sampling. Next, the RM can be extended to quantify the politics of the Supreme Court. These first two examples will be brief. I will spend most of the time investigating a paradox—that many cancer survivors report memory deficits, yet their memory seems intact by standard neurocognitive tests. A Bayesian latent regression RM (Li, et al. 2016) helps to make sense of this apparent contradiction. I will provide a practical guide, on how to fit the Bayesian latent RM, and how to use the MCMC chains to derive empirical estimates that are harder to get with non-Bayesian methods. The overall goal is to show how psychometric methods have many useful applications across diverse disciplines. Hopefully, a brief introduction will stir discussions on how IRT models may be useful in your own research.

FRIDAY, April 28, 2017
10:00am – 11:00am

Warwick Evans Conference Room, Building D
Refreshments will be provided at 9:45am

Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center