

Bio³ Seminar Series



Hao Wang, Ph.D.

*Associate Professor of Oncology,
Division of Biostatistics and Bioinformatics,
Sidney Kimmel Comprehensive Cancer Center,
Johns Hopkins University*

Bias and Freezing Effect in Truncated Randomized Clinical Trials

Despite the wide use of the design with statistical stopping guidelines to stop a randomized clinical trial early for efficacy, there are unsettled debates of potential harmful consequences of such designs. These concerns include the possible over-estimation of treatment effects in early-stopped trials, and a newer argument of a “freezing effect” that will halt future RCTs on the same comparison since an early-stopped trial represents an effective declaration that randomization to the un-favored arm is unethical. We determine the degree of bias in designs that allow for early stopping, and assess the impact on estimation if indeed future experimentation is “frozen” by an early-stopped trial. We discuss methods to correct for the over-estimate in early-stopped trials. We demonstrate that superiority established in a RCT stopping early and designed with appropriate statistical stopping rules is likely a valid inference, even if the estimate may be slightly inflated.

FRIDAY, September 8, 2017

10:00am – 11:00am

Warwick Evans Conference Room, Building D

Refreshments will be provided at 9:45am

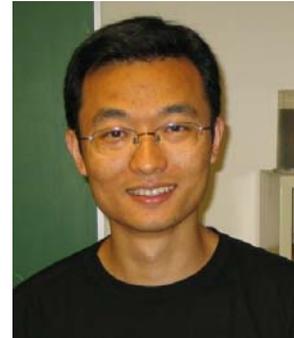


*Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.
Lombardi Comprehensive Cancer Center, Georgetown University Medical Center*

Bio³ Seminar Series

Jianhui Zhou, Ph.D.

*Associate Professor,
Department of Statistics,
University of Virginia*



A Single Index Model for Censored Quantile Regression

Quantile regression has been getting more attention recently in survival analysis due to its robustness and interpretability, and is considered as a powerful alternative to Cox proportional hazards model and accelerated failure time (AFT) model. Allowing a nonlinear relationship between survival time and risk factors, we study a single index model for censored quantile regression, and employ B-spline approximation for estimation. To avoid estimation bias caused by censoring, we consider the redistribution-of-mass to obtain a weighted quantile regression estimator. For high dimensional covariates, dimension reduction approach is adopted to alleviate the "curse of dimensionality". Furthermore, we penalize the developed estimator for variable selection. The proposed methods can be efficiently implemented using the existing weighted linear quantile regression algorithm. The asymptotic properties of the developed estimators are investigated, and their numerical performance is evaluated in simulation studies. We apply the proposed methods to dataset from a kidney transplant study.

FRIDAY, September 22, 2017
10:00am – 11:00am

Proctor Harvey Amphitheater, Med-Dent C105

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

Feng Cheng, Ph.D.

*Assistant Professor, Pharmaceutical Science,
University of South Florida- Tampa*

Identify Taxonomic Profiles of the Salivary Microbiome Associated with Type 1 Diabetes

Background: The oral cavity contains a diverse microbiome with over 700 bacterial species, many of which influence human health.

Objective: We hypothesize that features of the salivary microbiome will distinguish gingival health from disease and that these attributes will be more prevalent in those with Type 1 Diabetes (T1D).

1) Characterize the composition of the salivary microbiome from 16s sequencing. **2)** Identify features of the salivary microbiome that distinguish those with T1D from those without T1D.

Methods: Passive drool saliva samples and clinical data were obtained from 197 (97 with T1D and 100 without diabetes) adults attending the 12-year visit for the CACTI study. Salivary DNA was extracted and 16S amplicons were sequenced. 16S reads will be mapped and clustered into operational taxonomic units (OTUs). Multi testing analysis was used to identify associations between the taxonomic microbial profiles and T1D status.

Results: At the phylum level, the main constituents of the salivary microbiome in both T1D and non-T1D were *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*. However, we did find a significant increased abundance of Firmicutes in the saliva from T1D subjects (29%) compared to non-T1D subjects (25%, false-discovery rate (FDR)-adjusted $p=0.019$). At the genus level, the relative abundances of several genera were higher or lower in T1D compared to non-diabetics. The relative abundance of Prevotella was lower, and the relative abundances of Campylobacter and Streptococcus were higher in those with T1D compared to those without.

Conclusion: The composition of the salivary microbiome was largely made up with *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*. There is association between taxonomic Profiles of the Salivary Microbiome and Type 1 Diabetes. At the phylum level, T1D were enriched with Firmicutes compared to non-T1D. At the genus level, the relative abundances of several genera were higher or lower in T1D compared to non-diabetics.

FRIDAY, October 13, 2017

10:00am – 11:00am

New Research Building Auditorium

Refreshments will be provided at 9:45am

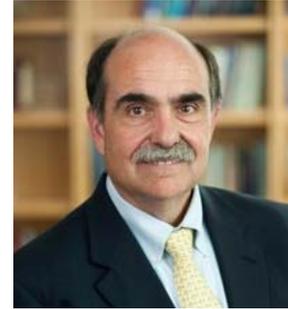


*Sponsored by the Department of Biostatistics, Bioinformatics, and Biomathematics
This seminar series meets the 2nd and 4th Friday of every month.
Lombardi Comprehensive Cancer Center, Georgetown University Medical Center*

Bio³ Seminar Series

Michael Stoto, Ph.D.

*Professor, Department of Health Systems
Administration and Population Health,
Georgetown University*



Meta-analysis for Drug Safety Assessment: Promises and Pitfalls

Meta-analysis has increasingly been used to identify adverse effects of drugs and vaccines, but the results have often been controversial. In one respect, metaanalysis is an especially appropriate tool in these settings. Efficacy studies are often too small to reliably assess risks that become important when a medication is in widespread use, so meta-analysis, which is a statistically efficient way to pool evidence from similar studies, seems like a natural approach. But, as the examples in this paper illustrate, different syntheses can come to qualitatively different conclusions, and the results of any one analysis are usually not as precise as they seem to be. There are three reasons for this: the adverse events of interest are rare, standard meta-analysis methods may not be appropriate for the clinical and methodological heterogeneity that is common in these studies, and adverse effects are not always completely or consistently reported. To address these problems, analysts should explore heterogeneity and use random effects or more complex statistical methods, and use multiple statistical models to see how dependent the results are to the choice of models.

FRIDAY, October 27, 2017

10:00am – 11:00am

New Research Building Auditorium

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



Eric Vance, Ph.D.

*Associate Professor of Applied Mathematics,
Director of Laboratory for Interdisciplinary
Statistical Analysis,
University of Colorado - Boulder*

Building a Global Network of Statisticians and Collaboration Laboratories

Statistics, analytics, and data science provide powerful methods, tools, and ways of thinking for solving problems and making decisions, but not everyone who could benefit from applying statistics and data science to their research has the knowledge or skills to apply it correctly. The Laboratory for Interdisciplinary Statistical Analysis (LISA) is a statistical collaboration laboratory recently created at the University of Colorado Boulder that generates, applies, and spreads new knowledge throughout the state, the nation, and the world. LISA's mission is to train statisticians to become interdisciplinary collaborators, provide research infrastructure to enable and accelerate high impact research, and engage with the community in outreach activities to improve statistical literacy. LISA has learned how to create statistical collaboration laboratories to train students to become effective statistical collaborators and to provide research infrastructure for the university. LISA is spreading this knowledge globally through the LISA 2020 Program to help scientists, government officials, businesses, and NGOs in developing countries discover local solutions to local problems through collaborations with statisticians from newly created statistical collaboration laboratories. The LISA 2020 goal is to build a global network of 20 statistical collaboration laboratories in developing countries by 2020. So far seven stat labs have been created in developing countries to train students to become effective interdisciplinary collaborators and enable researchers and government officials to solve problems and make better decisions.

FRIDAY, November 10, 2017

10:00am – 11:00am

Proctor Harvey Amphitheater, Med-Dent C105

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

Yixin Fang, Ph.D.

*Associate Professor,
Department of Mathematical Sciences,
New Jersey Institute of Technology*



“On Scalable Inference with Stochastic Gradient Descent”

Abstract: In many applications involving large dataset or online updating, stochastic gradient descent (SGD) provides a scalable way to compute parameter estimates and has gained increasing popularity due to its numerical convenience and memory efficiency. While the asymptotic properties of SGD-based estimators have been established decades ago, statistical inference such as interval estimation remains much unexplored. The traditional resampling method such as the bootstrap is not computationally feasible since it requires to repeatedly draw independent samples from the entire dataset. The plug-in method is not applicable when there are no explicit formulas for the covariance matrix of the estimator. In this paper, we propose a scalable inferential procedure for stochastic gradient descent, which, upon the arrival of each observation, updates the SGD estimate as well as a large number of randomly perturbed SGD estimates. The proposed method is easy to implement in practice. We establish its theoretical properties for a general class of models that includes generalized linear models and quantile regression models as special cases. The finite-sample performance and numerical utility is evaluated by simulation studies and two real data applications.

FRIDAY, January 12, 2018
10:00am – 11:00am

Building D, Warwick Evans Conference Room
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

Janet Sinsheimer, Ph.D.

Professor,
Departments of Human Genetics,
Biomathematics, and Biostatistics
University of California Los Angeles



“Linear Mixed Models and Maternal Offspring Gene Interactions”

Linear mixed effect models (LMMs) have a long history in genetics, going back at least as far as when R. A. Fisher proposed the polygenic model. They have become a mainstay in statistical modeling. Because these models can be computationally intense, LMMs were dropped in favor of simpler statistical tests in the whole genome era of genetics. However, quite recently LMMs surged in popularity for –omic studies and in particular for genome wide association studies. In my talk, I will review what makes these models so popular now in genomics, discuss my groups’ recent work with LMMs to detect maternal gene by offspring gene interactions, and then touch on some open questions that deserve consideration.

FRIDAY, January 26, 2018

10am – 11am

Warwick Evans 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



Kirk Wolter, Ph.D.

*Professor, Department of Statistics;
Executive Vice President: Survey Research,
NORC, University of Chicago*

“Sampling and Estimation Issues in the National Immunization Survey”

I will discuss the National Immunization Survey (NIS) conducted by the Centers for Disease Control and Prevention. NIS-Child monitors vaccination coverage for the population of children 19-35 months, NIS-Teen focuses on vaccination coverage of adolescents 13-17 years, and NIS-Flu monitors influenza vaccination coverage for the population of children 6 months – 17 years, all living in the U.S.A. This family of surveys is conducted annually in two phases: 1) a large survey of telephone numbers to identify households with age-eligible children, collect socio-demographic information about the child, and obtain contact information for the child’s immunization providers; and 2) a mail survey to collect immunization histories from identified providers for whom consent has been obtained to contact providers. Independent samples are selected quarterly in each of the 50 states and in specified urban areas. The first-phase sample is obtained by random digit dialing from a national sampling frame consisting of landline and cell-phone numbers. Because of insufficient validity of estimates derived from household-retained vaccination cards and parental recall, the NIS uses children with adequate provider-reported data to estimate vaccination coverage rates. I discuss the NIS dual-frame estimation procedure.

FRIDAY, February 9, 2018

10:00am – 11:00am

**Building D,
Warwick Evans Conference Room**

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

Alex Sverdlov, Ph.D.

*Director Statistical Scientist,
Early Development Biostatistics –
Translational Medicine,
Novartis Pharmaceuticals Corp.*



“Adaptive optimal designs for dose-finding studies with time-to-event outcomes”

Dose-response studies play an important role in clinical drug development. In this presentation, I will give an overview of some of my research work on optimal adaptive designs for dose-response studies with time-to-event outcomes. These designs utilize response adaptive allocation to most informative dose levels according to pre-defined statistical criteria. The designs can significantly improve estimation efficiency of the trial, which can potentially translate into reduction in study sample size. I will also emphasize some open topics in this field which can motivate some interesting research on optimal adaptive designs.

FRIDAY, February 23, 2018

10:00am – 11:00am

Building D, Warwick Evans Conference Room

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



Rajeshwari Sundaram, Ph.D.

*Senior Investigator, Biostatistics & Bioinformatics,
Division of Intramural Population Health Research,
National Institutes of Health (NIH)*

“Statistical Modeling of Progression of Spontaneous Labor in Women”

Defining labor progression in women has been a long-standing challenge for obstetricians. Cervical dilation, as integer-valued measurement, is a key indicator of the first stage of labor progression. Assessing the distribution of the time to per-unit increments of cervical dilation is of considerable interest in aiding obstetricians with better management of labor. Given that women are observed only intermittently for cervical dilation after they get admitted to hospital and that the observation frequency is very likely correlated to how fast/slow she dilates, one could view such data as panel count data with informative observation times and unknown time-zero. We propose semiparametric proportional rate models for the cervical dilation process and the observation process, with a multiplicative subject-specific frailty variable capturing the correlation between the two processes. Inference procedures for the gap times between consecutive events are proposed for both the scenarios with known and unknown time-zero using maximum likelihood approach and estimating equations. The methodology is assessed through simulation study and its large sample properties. A detailed analysis using the proposed method applied to the longitudinal cervical dilation data from the National Collaborative Perinatal Project from 1960s and the Consortium of Safe Labor of 2000s will be presented providing interesting comparisons across time. We will discuss other statistical challenges in studying labor progression including second stage of labor as well as neonatal and maternal morbidities.

FRIDAY, March 23, 2018

10:00am – 11:00am

**Building D,
Warwick Evans Conference Room**

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

Gengsheng Jeff Qin, Ph.D.

*Professor of Mathematics and Statistics;
Statistics Graduate Director, CDT Associate,
Georgia State University, Atlanta*



“Interval Estimation for the Correlation Coefficient”

The correlation coefficient (CC) is a standard measure of a possible linear association between two continuous random variables. The CC plays a significant role in many scientific disciplines. For a bivariate normal distribution, there are many types of confidence intervals for the CC, such as Z-transformation and maximum likelihood-based intervals. However, when the underlying bivariate distribution is unknown, the construction of confidence intervals for the CC is not well-developed. We discuss various interval estimation methods for the CC. We propose a generalized confidence interval for the CC when the underlying bivariate distribution is a normal distribution and two empirical likelihood-based intervals for the CC when the underlying bivariate distribution is unknown. We also conduct extensive simulation studies to compare the new intervals with existing intervals in terms of coverage probability and interval length. Finally, two real examples are used to demonstrate the application of the proposed methods.

FRIDAY, April 13, 2018

10:00am – 11:00am

Building D, Warwick Evans Conference Room

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



Ying Yuan, Ph.D.

*Professor, Department of Biostatistics,
Division of Science; University of Texas,
MD Anderson Cancer Ctr, Houston, TX*

BOIN: A Novel Platform for Designing Early Phase Clinical Trials

In this talk, I will introduce the Bayesian optimal interval (BOIN) design as a novel platform for designing various different types of early phase clinical trials, including single agent, combination, toxicity grades and late-onset toxicity, under a unified framework. The BOIN belongs a class of new designs, known as model-assisted designs, that use a model for efficient decision making like model-based designs, while their dose escalation and de-escalation rules can be tabulated before the onset of a trial as with algorithm-based designs. The BOIN design is easy to implement in a way similar to the 3+3 design, but is more flexible for choosing the target toxicity rate and cohort size and yields a substantially better performance that is comparable to that of more complex model-based designs, such as the continuous reassessment method. The BOIN design possesses intuitive Bayesian and frequentist interpretations with desirable finite-sample and large-sample properties, i.e., long-memory coherence and consistency. The software with graphical user interface will be demonstrated.

FRIDAY, April 27, 2018

10:00am – 11:00am

**Building D,
Warwick Evans Conference Room**

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*