Doctoral Dissertation Defense Announcement

“Statistical Methods to improve estimation efficiency for right-censored data under observational studies and clinical trials”

Xi Fang
Candidate for Doctor of Philosophy
Biostatistics
School of Graduate Studies
Medical College of Wisconsin

Committee in Charge:
Soyoung Kim, PhD (Mentor)
Kwang Woo Ahn, PhD
Brent R. Logan, PhD
Marten J. Michael, PhD
Wael Saber, MD, MS

Date: Monday, Feb 12, 2024
Time: 11:00 AM (CST)
**Defense Location:** A6520/A6628

**Zoom:** https://mcw-edu.zoom.us/j/99045766244?pwd=VG5lZXNjL1IeVlhiYTA0K2tSWEJ5dz09&from=addon

Meeting ID: 990 4576 6244
Passcode: 4v5uNDNb

**Graduate Studies:**

- Mathematical Stats II
- Statistical Model & Methods II
- Research Seminar
- Theory of Statistical Consult
- Applied Survival Analysis
- Statistical Consulting
- Intro to Bayesian Analysis
- Linear Models I
- Theory of Survival Analysis
- Intro Translational Bioinformatics
- Ethics & Integrity in Science
- Design & Analysis Clinical Trials
- Statistical Genetics
- Intro Statistical Machine Learning
- Advanced Statistical Computing
- Advanced Bayesian Analysis
- Advanced Statistics I
- Principals Database Systems
- Research Ethics Discussion Series
- Doctoral Dissertation
Dissertation

“Statistical Methods to improve estimation efficiency for right-censored data under observational studies and clinical trials”

The case-cohort study design provides a cost-effective study design for a large cohort study with competing risks outcomes. The subdistribution hazards model is widely used to estimate direct covariate effects on the cumulative incidence function for competing risks data. There are to commonly used subdistribution hazards model, proportional subdistribution hazards model and additive subdistribution hazards model. In biomedical studies, left truncation often occurs and brings extra challenges to the analysis. Existing inverse probability weighting methods for case-cohort studies with competing risks data not only have not addressed left truncation, but also are inefficient in regression parameter estimation for fully observed covariates. We propose an augmented inverse probability weighted estimating equation for left-truncated competing risks data to address these limitations of the current literature. We further propose a more efficient estimator when extra information from the other causes is available. The proposed estimators are consistent and asymptotically normally distributed. Simulation studies show that the proposed estimator is unbiased and leads to estimation efficiency gain in the regression parameter estimation. We analyze the Atherosclerosis Risk in Communities study data using the proposed methods.

The incorporation of historical data into a clinical trial analysis can improve the precision and efficiency of treatment evaluation if the historical data are exchangeable with clinical trial data. Evaluating the exchangeability of these two data sets is challenging, however, as an incorrect assessment of exchangeability yields invalid inference on the treatment effect that may produce bias and inflate the Type I error rate. To address this practical problem, we propose an adaptive fused group bridge penalty to evaluate the comparability of parameters between historical data and clinical trial data and make inferences on the treatment effect. The proposed penalty has oracle properties, including consistency for identifying the underlying model and the asymptotic normality of the estimators. Simulation studies show that the proposed method controls the Type I error rate better and has higher power than a competing method under both exchangeable and non-exchangeable settings. We apply the proposed method by reanalyzing a Phase III trial while also leveraging a corresponding historical data set.
Xi Fang

xfang@mcw.edu | 573-823-6409 | 8701 W Watertown Plank Rd, Milwaukee, WI 53226

**EDUCATION**

**PhD Candidate (Biostatistics) | GPA: [xx.xx]**  
**Medical College of Wisconsin**  
Jan 2020 - Feb 2024 (Expected)  
- **Dissertation topic:** Statistical Methods to Improve Estimation Efficiency for Right-censored Data under Observational Studies and Clinical Trials  
- **Advisors:** Drs. Soyoung Kim and Kwang Woo Ahn

**M.S (Biostatistics) | GPA: [xx.xx]**  
**Emory University**  
Aug 2015 - May 2017  
- **Thesis topic:** Improving Precision in Mediation Analysis Through Efficient Use of Case Data

**B.S (Biological Science) | GPA: [xx.xx]**  
**East China University of Science and Technology**  
Sept 2009 - May 2013

**SKILLS**

- **Programming Language:** R, SAS, Python, C++, Rcpp, SQL, JAGS, Stan, Java  
- **Software & tools:** Microsoft Office Suite (Word, Excel, Powerpoint), \LaTeX, Linux, Adobe Illustrator

**WORK EXPERIENCE**

**Experiential Internship**  
Abbvie INC., South San Francisco, CA  
05/2023 - 08/2023  
- Explored statistical methods and estimand strategies for intercurrent data following the ICH E9 (R1) guideline.  
- Conducted simulation studies to compare different adjustment methods for crossover treatments.  
- Drafted the user manual to provide tips for applying the frameworks and methods.

**Research Assistant**  
Medical College of Wisconsin, Milwaukee, WI  
01/2021 - 02/2024 (Expected)  
- **Statistical Methods to improve estimation efficiency for right-censored data under observational studies and clinical trials (MCW Cancer Center Pilot Grant)**  
  - Developing augmented inverse probability weighting methods to improve estimation efficiency for left-truncated competing risks data under the case-cohort design.  
  - Developing a penalized Cox model to incorporate historical data in the analysis of clinical trial data to potentially improve parameter estimation efficiency of the treatment effect.  
  - Applying the penalized Cox model to BMT CTN 1101 trial data and historical data.  
  - Writing the core algorithm using C++/Rcpp for a fast implementation and conducting extensive simulation studies using parallel computing.  
  - Developing an R package for the developed methods.  
  - Conducting literature reviews, presenting research outputs at college seminars, and writing two statistical manuscripts as the first author, where one of them was accepted by Biometrics.
Commensurate prior models with random effects for the survival outcome to incorporate historical data in the analysis of clinical trial data (FDA contract)
- Developing commensurate prior models, a power prior model, and a meta analysis method using a piecewise exponential model with random effects to analyze matched interval-censored historical data and clinical trial data.
- Performing extensive simulation studies using JAGS/Stan and parallel computing.
- Preparing and summarizing research findings for the report to FDA.
- One paper is under review for *Journal of Biopharmaceutical Statistics*. Currently drafting another paper as the first author.

Diffusion restriction comparison between Gleason 4 fused glands and Cribriform glands within patients using prostate pathology from MP-MRI
- Conducted statistical analysis using generalized linear mixed models, unsupervised clustering image segmentation, and Moran’s statistics for spatial data
- Performed simulation studies to compare the performance of different cluster-based bootstrap methods.
- Summarized the results for a peer-reviewed journal manuscript (Duenweg et al., 2022)

**Consulting Analyst**
Medical College of Wisconsin, Milwaukee, WI
01/2020 - 02/2024 (Expected)
- Performed statistical analysis for clinical trial data and experimental data including pressure surface, CIBMTR (Center for International Blood and Marrow Transplant Research), Warfarin studies.
- Conducted power analysis, Cox proportional hazards model, proportional cause-specific hazards model, and generalized linear mixed model using SAS for over 5 projects (Arshad et al., 2023).
- Communicated with PIs and collaborators and delivered written reports.

**Bioinformatic Analyst**
Eastern Virginia Medical School, Norfolk, VA
09/2017 - 12/2019
- Conducted statistical analysis for high-dimensional genetic data (Microarray and RNAseq data), including annotation, differential gene expression analysis, clustering analysis, and pathway analysis in contraceptive, HIV, and bacterial vaginosis studies among women.
- Developed R shiny webpages to integrate the Microarray analysis pipeline with a new annotation system.
- Performed power analysis and Pharmacokinetic (PK) analysis for anti-HIV activity drug development study.
- Summarized analysis results for the manuscript (Ouattara et al., 2022; Thurman et al., 2023; Zalenskaya et al., 2018) and presentations.

**Data Analyst**
CARE USA, Atlanta, GA
05/2016 - 05/2017
- Conducted statistical analysis for the women’s nutrition and wellness study in South Africa, and community-based children reading difficulty study, using generalized estimating equation, generalized linear mixed model, two-stage-least square regression, and mediation analysis.
- Developed a SAS macro for reproducibility.

**Publications**

with additive hazard subdistribution model under the case-cohort design. (In preparation)


AWARDS

1. JSM 2023 KISS Outstanding Student Paper Award.

2. 2023 LiDS Conference of the American Statistical Association Student Paper Award.

ABSTRACTS/PRESENTATIONS


2. Fang X, Kim S, Martens M, Logan BR, and Ahn KW. The Cox model with adaptive fused group bridge penalty to incorporate historical data into the analysis of clinical trials. 2023 Lifetime Data Science Conference, Raleigh, NC. May 2023. (Student Paper Award from 2023 LiDS Conference).