2017 Conference on Lifetime Data Science

Data Science, Precision Medicine and Risk Analysis with Lifetime Data

May 25 – May 27, 2017
University of Connecticut
Storrs, Connecticut

Special thanks to our sponsors

Takeda

BAYER

ASA

LIFE TIME

UCONN

LIDA 2017 Program

Contents

Cover **1**

Table of Contents **3**

Welcome and Overview **5**

Program Sketch **6**

Student Union Map **9**

Residence Hall Map **10**

Maps of Laurel Hall **11**

Residence Hall Map **14**

Welcoming Remarks **15**

Keynote Speakers **16**
- Dr. Niels Keiding . . . 16
- Dr. Lee-Jen Wei . . . 17

Detailed Program **18**
- Thursday 8:00-5:00 . . . 18
  - Short Courses . . . 18
- Friday 8:30–10:15 . . . 18
  - Opening (Keynote) Session . . . 18
- Friday 10:30–12:15 . . . 18
  1. Sample Size Considerations and Dynamic Prediction in Time-to-Event Studies . . . 18
  2. Special Session on Reverse Alignment in Survival Processes . . . 18
  3. Alternatives to the log-Rank Test for Clinical Trials . . . 18
  4. Complex Modeling of Survival and Longitudinal Data . . . 18
  5. Observational Survival Data . . . 19
  6. Asymptotics Methods on Censored and Truncated Data . . . 19
  7. Bayesian Survival Analysis . . . 19
  8. Causal Inference with Lifetime Data . . . 19
  9. Survival Analysis in Psychiatry and Neurological Disorders . . . 19
- Friday 1:45–3:30 . . . 19
  10. Joint Modeling with a View Towards Risk Predictions . . . 19
  11. Risk Prediction Models and Application . . . 20
  12. Risk Analysis in the Biomedical and Environmental Field . . . 20
  13. Recent Development in the Analysis of Complex Structured Survival Data . . . 20
  14. Outcome-Dependent Sampling in the Survival Analysis Context . . . 20
  15. Recent Developments in High-Dimensional Survival Analysis and Biased Sampling . . . 20
  16. Advances in Summarizing and Modeling Complex Survival Data . . . 21
  17. Semiparametric Statistical Methods for Complex Failure Time Data . . . 21
  18. Recent Developments in Statistical and Computational Methods for Biomedical Data . . . 21
- Friday 3:45–5:30 . . . 21
  19. From Functional to Neuroimaging Data . . . 21
  20. Recent Advances on Statistical Methods for Health Studies . . . 21
  21. Joint Modeling and Weighted Estimation for Survival Data Analysis . . . 22
  22. New Approaches for Analyzing Time to Event Data with Application in Cancer Studies . . . 22
  23. Quality of Life and Other Applications . . . 22
  24. Joint Modelling of Longitudinal Measurements and Event History Data . . . 22
  25. Time-to-Event Models for Human Health Risk Assessment . . . 22

26. Industry Perspectives on Lifetime Data Science . . . 23
27. the Win Ratio and Related Topics in Multiple Event Time Data . . . 23
Saturday 8:30–10:15 . . . 23
28. Recent Advances in Analyzing Multi-State and Family Data . . . 23
29. Measurement Error, Mediation Analysis, and Individualized Medicine . . . 23
30. Statistical Analysis of Recurrent, Competing Risks, and Current Status Data . . . 23
31. on Risk Prediction in the Presence of Competing Risks . . . 23
32. Challenges and New Methods of Complex Health Data . . . 24
33. New Statistical Methods for Complex Structures in Survival Data . . . 24
34. New Machine Learning Methods for Censored Survival and Competing Risks Data . . . 24
35. Advances in Statistical Modeling of Correlated Data . . . 24
36. Modeling Disease Natural History and Effects of Treatment: Applications to Prostate Cancer . . . 24
37. Student Paper Presentations . . . 25
Saturday 10:30–12:15 . . . 25
38. Tree-Based Methods for Survival Data . . . 25
39. Advances in Multi-State Models for Survival and Event History Analysis . . . 25
40. Advances in the Analysis of Clinical Trials with Lifetime Data using Restricted Mean Survival Times 25
41. New Methods for Censored Data . . . 25
42. Recent Advances in Survival Analysis for Cancer Research . . . 26
43. Risk Assessment and Prediction with Survival Data . . . 26
44. Recent Development of Time-to-Event Data Analysis Incorporating Disease Dynamics . . . 26
45. Analytical Challenges and New Advances in Assessing Time-to-Event Endpoints in Oncology Studies 26
46. Analysis of Complex Sampling Designs with Censored Data . . . 26
47. Survival Analysis for Family-Clustered Data . . . 27
Saturday 1:45–3:30 . . . 27
48. New Statistical Approaches on Medical Data . . . 27
49. Recent Advances in the Analysis of Competing Risks Data . . . 27
50. Bayesian Nonparametric Survival Analysis . . . 27
51. Recent Developments in Complex Survival and Longitudinal Data . . . 27
52. New Developments in Modeling Longitudinal, Recurrent Event and Survival Data . . . 28
53. Unconventional Usage of Classic Methods in Modern Medical Applications . . . 28
54. Recent Topics on Analysis of Interval Censored Data (with Covariate or Outcome Censored) . . . 28
55. New Development in Statistical Methods for Deriving and Validating Dynamic and Individualized Decision Rules . . . 28
Saturday 3:45–5:30 . . . 28
56. Advances in Semiparametric Regression Analysis of Panel Count Data . . . 28
57. a Recent Development on Competing and Semi-Competing Risks . . . 29
58. Recent Advances on Event History Analysis . . . 29
59. Recent Advances in Time-to-Event Analysis with High Dimensional, Heterogeneous, and/or Correlated Data . . . 29
60. Statistical Innovations for Data Science and Precision Medicine . . . 29
61. Risk Prediction in Survival Analysis . . . 29
62. Methods and Applications of Recurrent Event or Panel Count Data . . . 30
63. Statistical Challenges for Immunotherapy with Delayed Treatment Effects . . . 30

Abstracts of Invited Papers **31**

Abstracts of Posters **76**

LIDA 2017 Committees **79**
Program Committee . . . 79
Local Organizing Committee . . . 79
Student Paper and Poster Committee . . . 79
Student Committees . . . 79

LIDA 2017 Participants **80**

Welcome and Overview

It is our great pleasure to welcome you to the 2017 LIDA Conference held at the University of Connecticut, Storrs, May 25–27, 2017. The purpose of the conference is to promote and support the development and application of statistical methods for lifetime or time-to-event data in the very broad spectrum of data science, precision medicine and risk analysis. The area of lifetime data research includes methods, theory and applications of time-to-event data with censoring, truncation and competing risks; this research covers a wide array of topics including counting processes, multi-state models, multiple or clustered events, longitudinal biomarker histories and quality-of-life models. Importantly, the research also extends to interdisciplinary areas such as finance, economics, imaging, engineering, genomics and genetics.

The keynote speeches of the conference are delivered by two highly accomplished leaders in Biometry and Statistics: Professor Niels Keiding from the University of Copenhagen, Denmark, and Professor Lee-Jen Wei from Harvard Universioty, USA. Professor Keidings presentation is entitled "Survival Analysis Around a Cross-Section and Unobserved Heterogeneity." Professor Wei will give a lecture on "The Myth of Design and Analysis of Cancer Clinical Studies with PFS or OS as the Endpoint." Associated with the conference are two half-day and two full-day short courses to be held on Thursday, May 25. These courses are taught by five leading experts in their respective areas: Dr. Joanna Shih, "Modeling Clustered Failure Time Data with Application to Family/Genetic Studies"; Drs. Mitchell Gail and Ruth Pfeiffer, "Absolute Risk and Applications in Disease Prevention"; Dr. Xiao-Hua (Andrew) Zhou, "Statistical Methods in Diagnostic and Predictive Medicine"; and Dr. Joseph Ibrahim, "Applied Bayesian Survival Analysis."

This conference has received very enthusiastic support from researchers in the area, and 62 invited sessions have been created and organized; this substantially exceeds the original plan of approximately 40 invited sessions. The Local Organizing Committee, led by Drs. Ming-Hui Chen and Jun Yan from the University of Connecticut, has made significant effort to accommodate the conference with excellent facilities and equipment. We greatly appreciate the committees dedicated work on registration, website management, hotel and housing arrangements, program book editing, banquet planning and other conference related items. The Student Paper and Poster Committee is chaired by Dr. Zhezhen Jin (Columbia University), and includes committee members Guoqing Diao (George Mason University), Gang Li (UCLA), and Ian McKeague (Columbia University). We thank this committee for its excellent and efficient work in evaluating and judging student and post-doc papers and posters. The result of the student paper and poster competitions will be announced during the banquet on Friday (May 26).

Finally, one of us (JDK) would like to thank the other (MCW) for her work in putting together the scientific program for this conference. Mei-Cheng was the driving force in conceiving of and bringing the conference to fruition, and in recruiting the large and very distinguished Program Committee. Her enthusiasm and efforts have resulted in a meeting that promises to be a very significant contribution to this area of research and its applications.

The conference is sponsored by the Lifetime Data Analysis Interest Group (LIDA-IG), which operates under the rules of the Committee on Sections of the American Statistical Association (ASA). The Interest Group aims eventually to be approved as a Section of the ASA. We encourage all the conference participants to join the interest group and help promote research in the important area of lifetime data science. The registration form is included in the program book and can also be downloaded from the conference website. Please fill out the form and send it to: Mei-Cheng Wang, Chair (`mcwang@jhu.edu`) an email cc'd to Jonathan Siegel, Secretary (`jonathan.siegel@bayer.com`).

Finally, we welcome you to participate in and enjoy the conference!

Sincerely,

Jack Kalbfleisch	**Mei-Cheng Wang**
Conference Chair	Program Chair and LIDA-IG Chair

LIDA Conference Program

Thursday, May 25, 2017

Location: 3rd Floor, Student Union

Registration: 8:00am – 5:00pm

Dr. Xiao-Hua (Andrew) Zhou and Dr. Joseph Ibrahim will each give a full-day (6 hours) short course from 8:30–5:00.

Dr. Joanna Shih will give a 1/2 day course in the morning from 8:30am–12:00pm; Drs. Mitchell Gail and Ruth Pfeiffer will give a 1/2 day course in the afternoon from 1:30pm–5:00 pm.

Additionally, the following breaks are scheduled:

- Morning break 10:00am–10:30am
- Lunch 12:00pm–1:30pm
- Afternoon break 3:00pm–3:30pm

Dr. Joanna Shih, Biometric Research Program, National Cancer Institute
"Modeling Clustered Failure Time Data with Application to Family/Genetic Studies"
(1/2-day course: morning, SU 304C)

Dr. Mitchell Gail and **Dr. Ruth Pfeiffer**, Division of Cancer Epidemiology and Genetics, National Cancer Institute
"Absolute Risk and Applications in Disease Prevention"
(1/2-day course: afternoon, SU 304C)

Dr. Xiao-Hua (Andrew) Zhou, University of Washington
"Statistical Methods in Diagnostic and Predictive Medicine"
(Full-day course, SU 304A)

Dr. Joseph G. Ibrahim, University of North Carolina at Chapel Hill
"Applied Bayesian Survival Analysis"
(Full-day course, SU 304B)

Poster Session and Reception/Mixer

Time: 5:00pm – 8:30pm (Cash Bar available)

Location: Ballroom, Student Union

Friday, May 26, 2017

Locations:

Laurel Hall: registration, book exhibitions, all technical sessions, morning refreshment and AM and PM breaks
Student Union Market: Lunch
Chang's Garden: Conference Banquet

Registration: 8:00am – 5:00pm, LH 1st Floor

Opening Session: 8:30am – 10:15am, LH 102

- Chair: **Jack Kalbfleisch**, University of Michigan
- Keynote Speech: **Niels Keiding**, University of Copenhagen, Denmark
- Keynote Speech: **Lee-Jen Wei**, Harvard University

Morning Break: 10:15am-10:30am, LH 1st Floor Hall Way

Morning Concurrent Sessions: 10:30am-12:15pm, LH 1st – 3rd Floors

Lunch: 12:15pm-1:45pm

Afternoon Concurrent Sessions I: 1:45pm – 3:30pm, LH 1st – 3rd Floors

Afternoon Break: 3:30pm-3:45pm, LH 1st Floor Hall Way

Afternoon Concurrent Sessions II: 3:45pm-5:30pm, LH 1st – 3rd Floors

Banquet: 6:00pm-9:00pm

CHANG'S GARDEN

1244 Storrs Road
Storrs, CT 06268
860-487-1688

Saturday, May 27, 2017

Locations:

Laurel Hall: registration, book exhibitions, all technical sessions, morning refreshment and AM and PM breaks
Ballroom, Student Union: BBQ Lunch

Morning Concurrent Sessions I: 8:30am-10:15pm, LH 1st – 3rd Floors

Morning Break: 10:15am-10:30am, LH 1st Floor Hall Way

Morning Concurrent Sessions II: 10:30am-12:15pm, LH 1st – 3rd Floors

Lunch: 12:15pm-1:45pm

Afternoon Concurrent Sessions I: 1:45pm – 3:30pm, LH 1st – 3rd Floors

Afternoon Break: 3:30pm-3:45pm, LH 1st Floor Hall Way

Afternoon Concurrent Sessions II: 3:45pm-5:30pm, LH 1st – 3rd Floors

Room Facilities in Laurel Hall

Opened in 2011, Laurel Hall is centrally located along the "academic corridor" of campus. The building is air conditioned and contains a variety of bright and modern space. Classrooms are high tech ready; including computer, screen, projector, whiteboard(s) and document camera. Individual classroom configuration varies but typically includes seating at tables and has the capability to connect a personal laptop. However, pointers are not available in these rooms.

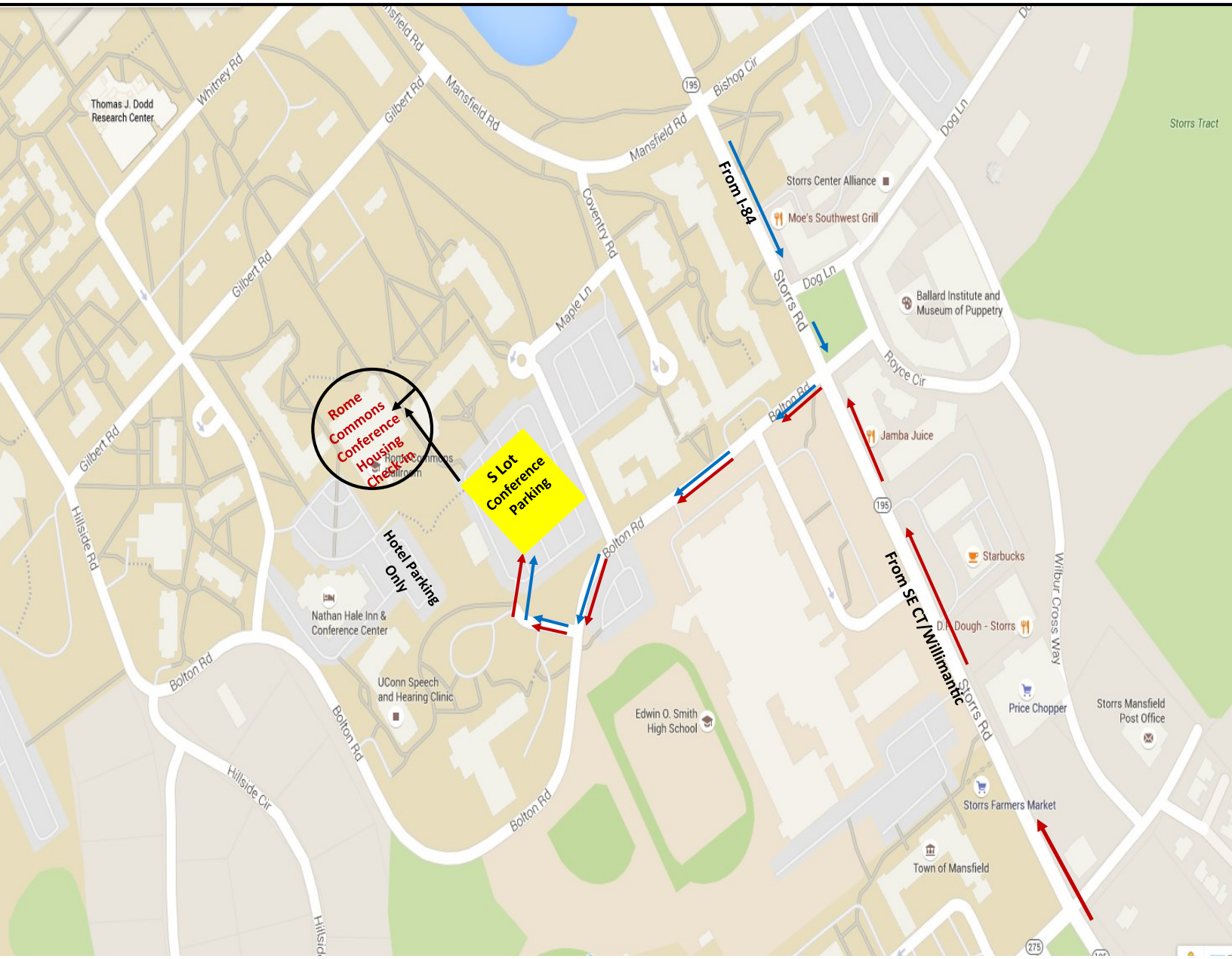
Student Union Map

LIDA Conference

Thursday, May 25, 2017

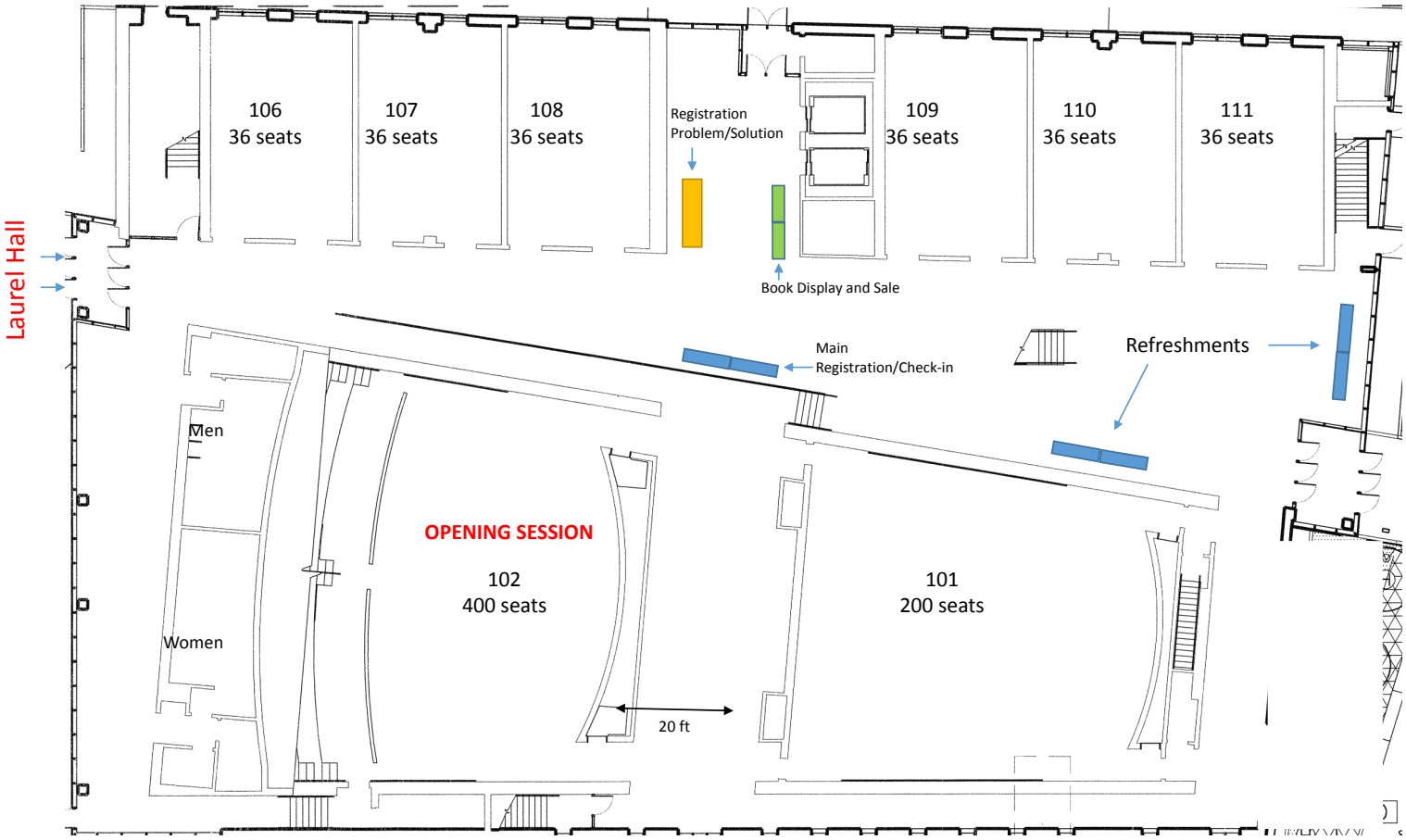
Student Union Building, Floor 3

Residence Hall Check-In

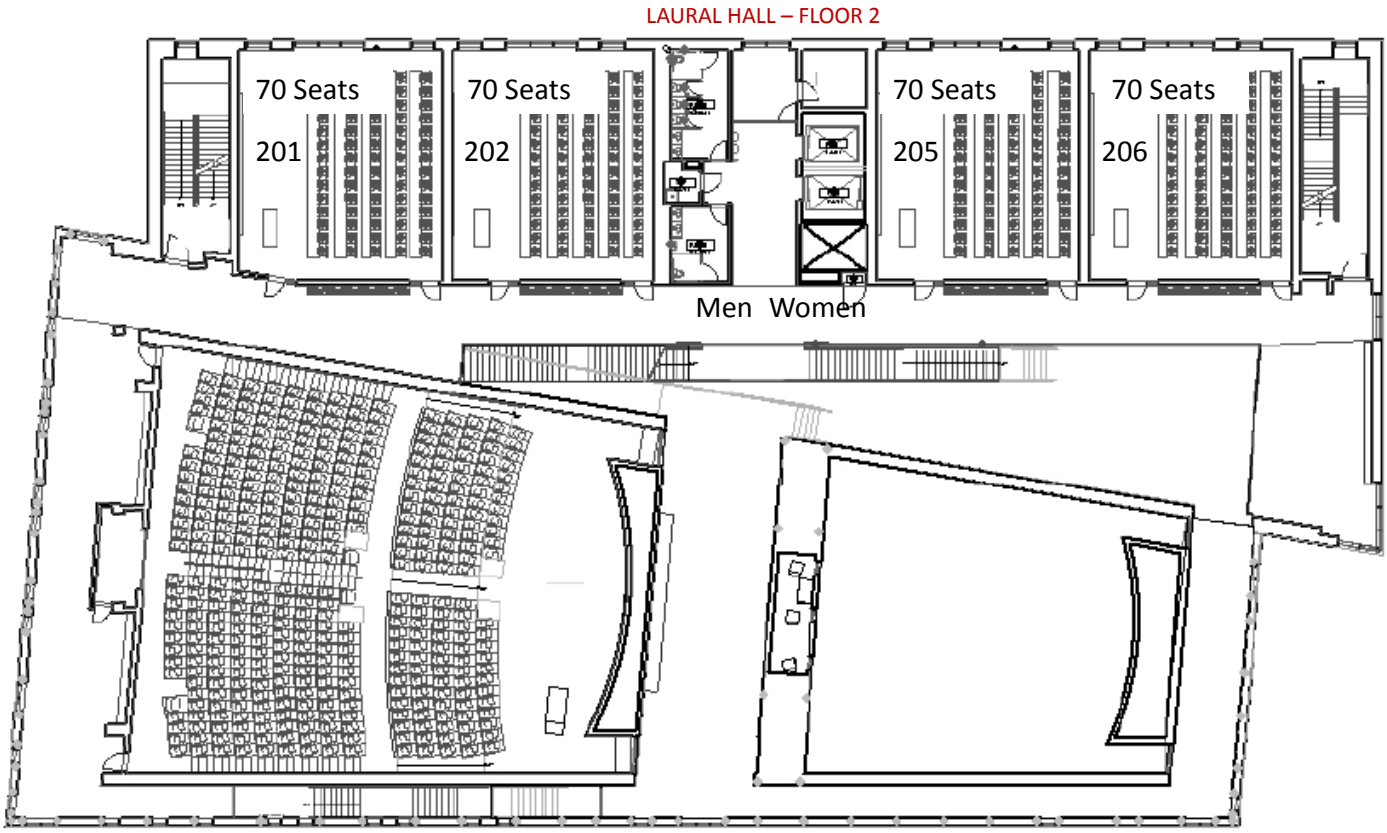


Maps of Laurel Hall

Laurel Hall Floor 1

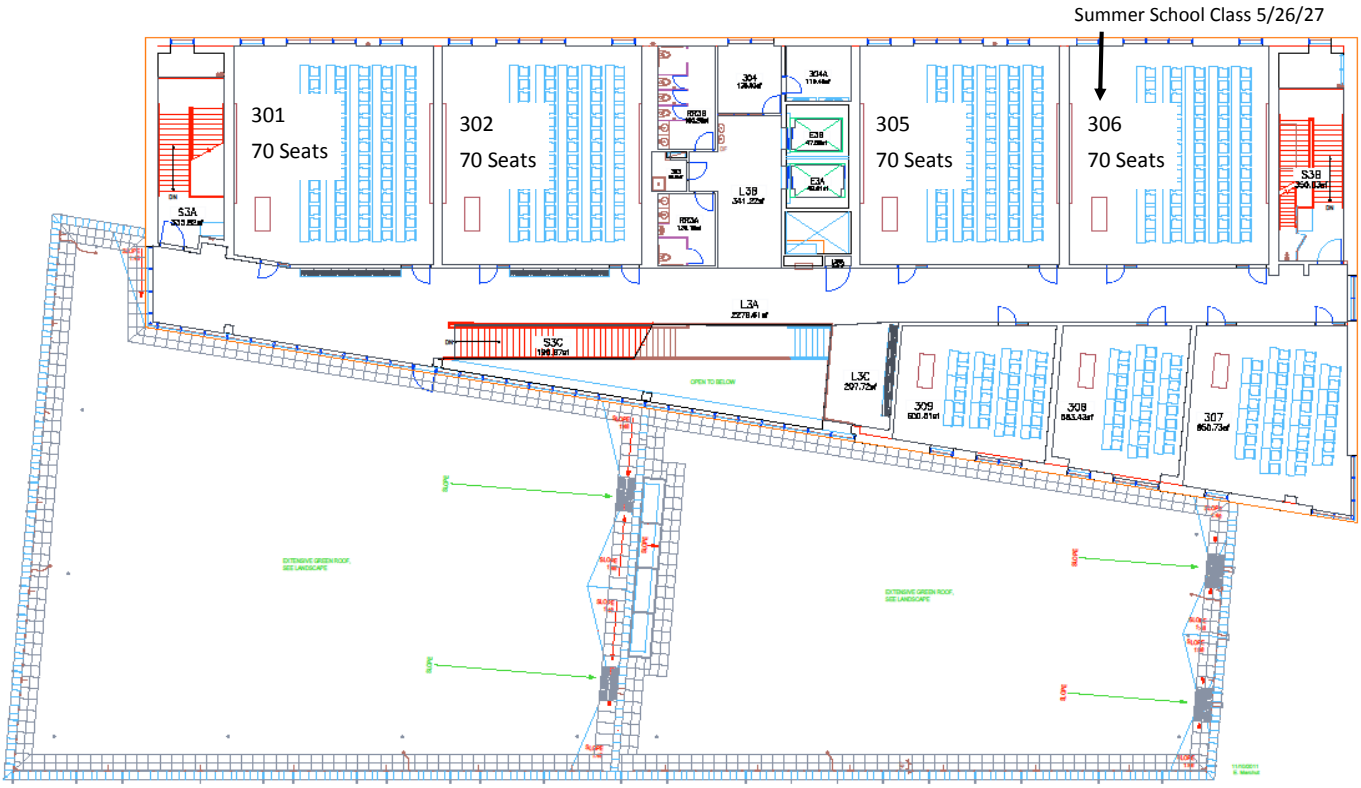


Laurel Hall Floor 2



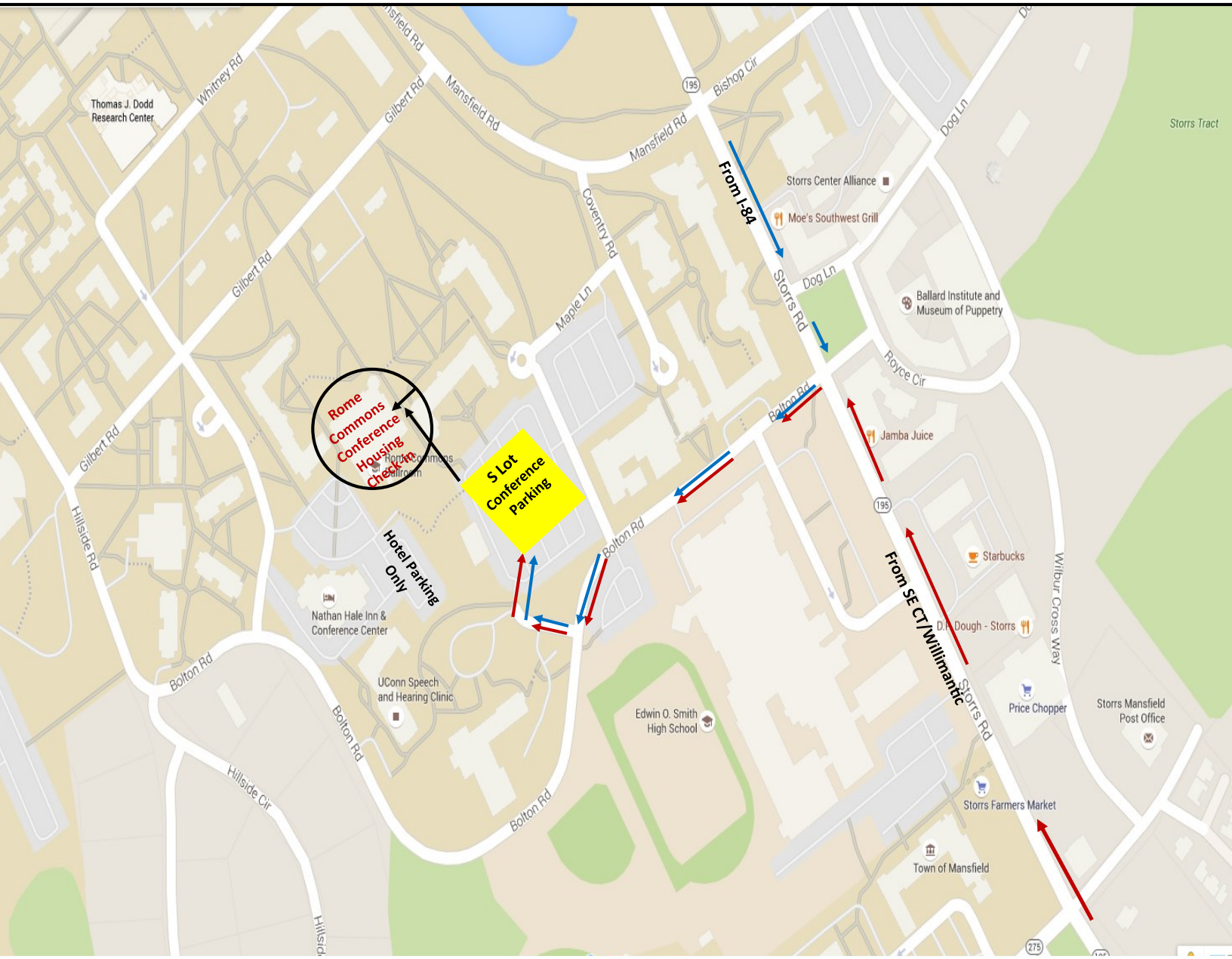
Laurel Hall Floor 3

LAUREL HALL - Floor 3



University of Connecticut West Classroom Building - Third Floor

Residence Hall Check-In



UCONN
UNIVERSITY OF CONNECTICUT
College of Liberal Arts & Sciences

University of Connecticut
Office of the Dean

May 10, 2017

Dear Scholars,

Welcome! We are pleased to host the 2017 Conference on Lifetime Data Science at the University of Connecticut. I understand that this year's conference will focus on data science, precision medicine, and risk analysis, all very important and timely issues in our world today.

The list of speakers and session presenters is quite impressive. I am certain that you will all benefit from engaging with your colleagues as you share research and discuss emerging issues.

Since its founding in 1939, the College of Liberal Arts and Sciences has been the academic heart of the University of Connecticut. We take seriously the foundations of a liberal education: We teach students to think creatively and analytically; to reason from evidence; to respect the views and experiences of all members of our diverse community; and to continue learning throughout their lives, wherever their professional and personal journeys take them. It is wonderful to hear that the LIDA upholds these same values.

Best wishes,

Davita Silfen Glasberg
Interim Dean of the College of Liberal Arts and Sciences
and Professor of Sociology

An Equal Opportunity Employer

215 Glenbrook Road, U-4158
Storrs, Connecticut 06269-4158

Telephone: (860) 486-1231
Facsimile: (860) 486-0296
Web: http://bsc.clas.uconn.edu

Keynote Speakers

Survival Analysis Around a Cross-Section and Unobserved Heterogeneity

Dr. Niels Keiding, University of Copenhagen, Denmark

Consider lifetimes originating at a series of calendar times $t_1, t_2, \ldots$. At a certain time t_0 a cross-sectional sample is taken, generating a sample of current durations (backward recurrence times) of survivors until t_0 and a prevalent cohort study consisting of survival times left-truncated at the current durations. A Lexis diagram is helpful in visualizing this situation. Survival analysis based on current durations and prevalent cohort studies is now well-established as long as all covariates are observed.

The general problems with unobserved covariates have been well understood for ordinary prospective follow-up studies, with the good help of hazard rate models incorporating frailties: as for ordinary regression models, the added noise generates attenuation in the regression parameter estimates. For current durations and prevalent cohort studies this attenuation remains, but in addition one needs to take account of the differential selection of the survivors from initiation t_i to cross-sectional sampling at t_0.

This talk intends to survey the recent development of these matters and the consequences for routine use of hazard rate models or accelerated failure time models in the many cases where unobserved heterogeneity may be an issue.

Dr. Keiding is now an Emeritus Professor of Biostatistics, University of Copenhagen. His research interests include methods for analysis of survival and event history data, particularly under non-standard observational plans, causal analysis and epidemiological methods, history of statistics and demography. Main applications of his research are reproductive epidemiology, particularly time-to pregnancy studies.

The Myth of Design and Analysis of Cancer Clinical Studies with PFS or OS as the Endpoint

Dr. Lee-Jen Wei, Harvard University

In a longitudinal clinical cancer study to compare a new treatment with a control, the primary end point is generally either the overall survival or progress-free survival time. The hazard ratio (HR) is routinely utilized to quantify a desirable treatment effect for sizing the study. The estimated total number of events needed to achieve a specific statistical power for the study can be obtained easily via a back-of-the-envelope calculation. However, since it is not clear how to interpret HR clinically, the specific HR value (e.g., 0.75) quantifying the desirable treatment effect is often justified via a certain degree of improvement from the treatment with respect to the median survival time (e.g., from 10 to 12 months). The median survival time is a clinically meaningful summary measure, however, it does not capture the long term survival profile well. Therefore, using the difference of two median survival times may not help us much to interpret a HR value at the design stage.

At the end of the study, the PFS/OS data are routinely analyzed via the HR estimation procedure and logrank test. This practice is more problematic at the analysis stage. The concerns and issues of using this summary measure have been discussed extensively in the literature. The validity of using HR depends on the proportional hazards assumption, that is, the hazard ratio for two groups is constant over the entire study period. This assumption is rarely valid in practice and the resulting HR estimate is difficult to interpret. In fact, in an interview article, being an extremely modest giant in our profession, Professor Cox stated that "of course, another issue is the physical or substantive basis for the proportional hazards model. I think thats one of its weaknesses..." To ease the difficulty of interpreting the estimated HR, the individual median survival time estimates for two groups are usually reported descriptively without formal comparisons in the study publications. However, since the median survival estimate is insensitivity to outliers and is unstable with respect to estimation precision, often the estimate for the difference of two medians results in an inconsistent conclusion regarding the treatment effect to that based on the HR estimate. It seems that the partnership between the HR and median survival is not working well for most conventional survival studies.

An alternative to the median survival is to use the restricted mean survival time (RMST) to summarize the survival profile. This measure has an intuitive, clinically meaningful interpretation. The procedure for estimating the difference of two RMSTs is always valid without any model assumption and is much stable than its median counterpart.

There is no single summary measure which can capture the entire survival profile of a group of patients. On the other hand, for the design and analysis of a study, a primary summary measure for the between-group-difference is needed. Moreover, the pre-specified analysis procedure for this measure should be robust, not heavily model-dependent, and will result in clinically interpretable conclusions about the treatment effect.

L.J. Weis research is in the area of developing statistical methods for the design and analysis of clinical trials. Dr. Wei has developed numerous methods for analyzing data with multiple outcome or repeated measurements obtained from study subjects. Currently, Wei and his colleagues are developing graphical and numerical methods for checking the adequacy of the Cox proportional hazards model, other semi-parametric survival models, parametric models, and random effects models for repeated measurements. Presently, Wei and his colleagues are working on various resampling methods for quantile regression, rank regression, and regression models for censored data. Dr. Wei is also a senior statistician at the Statistical and Data Analysis Center. He works closely with the medical investigators in Pediatrics AIDS clinical trials for evaluating new treatments for HIV patients.

Detailed Program

Thursday 8:00–5:00

Short Courses

- Dr. **Joanna Shih**, Biometric Research Program, National Cancer Institute
  "Modeling Clustered Failure Time Data with Application to Family/Genetic Studies"
  (1/2-day course)
- Dr. **Mitchell Gail** and Dr. **Ruth Pfeiffer**, Division of Cancer Epidemiology and Genetics, National Cancer Institute
  "Absolute Risk and Applications in Disease Prevention"
  (1/2-day course)
- Dr. **Xiao-Hua (Andrew) Zhou**, University of Washington
  "Statistical Methods in Diagnostic and Predictive Medicine"
  (Full-day course)
- Dr. **Joseph G. Ibrahim**, University of North Carolina at Chapel Hill
  "Applied Bayesian Survival Analysis"
  (Full-day course)

Friday 8:30–10:15

Opening (Keynote) Session

- *Chair:* **Jack Kalbfleisch**, University of Michigan
- *Keynote Speech:* **Niels Keiding**, University of Copenhagen, Denmark
- *Keynote Speech:* **Lee-Jen Wei**, Harvard University

Friday 10:30–12:15

1. Sample Size Considerations and Dynamic Prediction in Time-to-Event Studies

Organizer: **Adin-Cristian Andrei**, Department of Surgery, Northwestern University
Chair: **Mithat Gnen**, Memorial Sloan-Kettering Cancer Center

1. **Jeremy M.G. Taylor**, Department of Biostatistics, University of Michigan Ann Arbor
   "Dynamic Prediction of Event Times using Longitudinal Data"
2. **Guadalupe Gmez Melis**, Universitat Politcnica De Catalunya
   "Sample Size Considerations for Composite Endpoints"
3. **Adin-Cristian Andrei**, Department of Surgery, Northwestern University
   "Balancing Baseline Covariate Similarity Versus Sample Size Considerations when Matching Methods are Used in Observational Studies with Survival Endpoints"
4. **Joseph G. Ibrahim**, Department of Biostatistics, University of North Carolina Chapel Hill
   "Bayesian Sample Size Determination for Clinical Trials using Historical Data"

Laurel Hall Room 111

2. Special Session on Reverse Alignment in Survival Processes

Organizer and Chair: **Niels Keiding**, University of Copenhagen

1. **Peter McCullagh**, University of Chicago
   "Statistical Models for Survival Processes"
2. *Discussant:* **Gary Chan**, University of Washington, Seattle
3. *Discussant:* **Per Kragh Andersen**, University of Copenhagen

Laurel Hall Room 201

3. Alternatives to the log-Rank Test for Clinical Trials

Organizer and Chair: **Rick Chappell**, Departments of Statistics and of Biostatistics and Medical Informatics, University of Wisconsin

1. **Rick Chappell**, Departments of Statistics and of Biostatistics and Medical Informatics, University of Wisconsin
   "Drawbacks of the log-Rank Test and Alternatives to it"
2. **Lee McDaniel**, Louisiana State University School of Public Health
   "The Additive Hazards Model in Clinical Trials"
3. **Theodore Karrison**, Department of Public Health Sciences, University of Chicago
   "The Restricted Mean Life"
4. **Hajime Uno**, Departments of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and of Medicine, Harvard Medical School
   "Flexible Rank-Based Tests"

Laurel Hall Room 107

4. Complex Modeling of Survival and Longitudinal Data

Organizer: **Jane-Ling Wang**, University of California, Davis
Chair: **Gang Li**, UCLA

1. **Edsel De La Pena**, University of South Carolina
   "Dynamic Time-to-Event Models"
2. **Jianguo (Tony) Sun**, University of Missouri
   "Joint Analysis of Interval-Censored Failure Time Data and Panel Count Data"
3. **Jane-Ling Wang**, University of California, Davis
   "Functional Cox Model"
4. **Grace Yi**, University of Waterloo
   "Analysis of Survival Data with Covariate Measurement Error under the Additive Hazards Model"

Laurel Hall Room 205

5. Observational Survival Data

Organizer: **Nicholas P. Jewell**, University of California, Berkeley
Chair: **Jerry Lawless**, University of Waterloo, Canada

1. **Marco Carone**, University of Washington
   "Computerized Efficient Inference in Survival Problems"
2. **Ana Best**, National Cancer Institute
   "Nonparametric Survival Estimates from Mixed Incident and Prevalent Cohort Data"
3. **Nicholas P. Jewell**, University of California, Berkeley
   "Grouped Current Status Data"
4. **Jack Kalbfleisch**, University of Michigan
   "Profiling Medical Providers using Survival Data"

Laurel Hall Room 206

6. Asymptotics Methods on Censored and Truncated Data

Organizer and Chair: **Mounir Mesbah**, Universit Pierre et Marie Curie, P6

1. **Abdelkader Tatachak**, Universit Houari Boumedienne, Algiers, Algeria
   "Nonparametric Regression Estimation for Associated Data under Censoring or Truncation: Strong Consistencies and Rates"
2. **Jimmy Efird**, University of Newcastle, Australia
   "Risk Stratification using a Weibull Proportional Hazards Model: Asymptotic Efficiency of a Counting Process Approach"
3. *Discussant:* **Mounir Mesbah**, Universit Pierre et Marie Curie, P6

Laurel Hall Room 106

7. Bayesian Survival Analysis

Organizer and Chair: **Lynn Kuo**, University of Connecticut

1. **Victor Hugo Lachos Davila**, University of Connecticut
   "A Multivariate Student-t Regression Model with Measurement Errors for Censored Data"
2. **Kyu Ha Lee**, The Forsyth Institute
   "Accelerated Failure Time Models for Semi-Competing Risks Data in the Presence of Complex Censoring"
3. **Sungduk Kim**, National Institutes of Health
   "A Joint Model Approach for Longitudinal Data with No Time Zero and Time-to-Event with a Competing Risk"
4. **Danjie Zhang**, Gilead Sciences
   "Bayesian Model Assessment in Joint Modeling of Longitudinal and Survival Data with Applications to Cancer Clinical Trials"

Laurel Hall Room 202

8. Causal Inference with Lifetime Data

Organizer and Chair: **Richard Cook**, Department of Statistics and Actuarial Science, University of Waterloo

1. **Douglas Schaubel**, University of Michigan
   "Estimating the Effect of a Time-Dependent State Change on Correlated Recurrent and Terminal Events"
2. **Kjetil Roysland**, University of Oslo
   "Causal Local Independence Models"
3. **David Stephens**, McGill University
   "Survival Dynamic Treatment Regimes with a Cured Fraction"
4. **Aksel Jensen**, University of Copenhagen
   "A Marginal Structural Model for Recurrent Events in the Presence of Time-Dependent Confounding: Non-Specific Effects of Vaccines on Child Hospitalisations"

Laurel Hall Room 302

9. Survival Analysis in Psychiatry and Neurological Disorders

Organizer and Chair: **Sharon Xiangwen Xie**, Department of Biostatistics and Epidemiology, University of Pennsylvania

1. **Rebecca Betensky**, Departments of Biostatistics, Harvard University
   "Methods for Analyzing Time-to-Event Data with Time-Varying Biomarkers Measured Only at Study Entry: Application to Imaging Markers in Alzheimer's Disease"
2. **Mengjie Zheng** and **Sujuan Gao**, Department of Biostatistics, Indiana University School of Medicine
   "Joint Modeling of Longitudinal and Survival Outcomes using Generalized Estimating Equations"
3. **Jeffrey D. Long**, Department of Psychiatry, Carver College of Medicine and Department of Biostatistics, College of Public Health, University of Iowa
   "Random Survival Forests for Exploratory Analysis in Neurological Disorders"
4. **Ralitza Gueorguieva**, Department of Biostatistics, Yale University
   "Joint Modeling of Symptom Severity and Competing Risk Dropout in Psychiatry"

Laurel Hall Room 301

Friday 1:45–3:30

10. Joint Modeling with a View Towards Risk Predictions

Organizer: **Rajeshwari Sundaram**, Biostatistics Branch, NICHD, NIH
Chair: **Mei-Cheng Wang**, Johns Hopkins University

1. **Paul S. Albert**, Senior Investigator and Chief, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute
   "Predicting Poor Pregnancy Outcomes from Multivariate Ultrasound Fetal Growth Data"

2. **Danping Liu**, Investigator, Biostatistics and Bioinformatics Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development
"A Joint Modelling Approach for Informative Cluster Size and Gap Time in Longitudinal Data with Application to a Repeated Pregnancy Study"
3. **Sheng Luo**, University of Texas-Houston
"Dynamic Prediction of Alzheimers Disease Progression with Longitudinal Functional Joint Model"
4. **Debajyoti Sinha**, Florida State University
"Bayesian Estimation of Recurrent Events and Dependent Terminal Event"

Laurel Hall Room 111

11. Risk Prediction Models and Application

Organizer: **Nilanjan Chatterjee**, Bloomberg School of Public Health, Johns Hopkins University
Chair: **Guoqing Diao**, George Mason University

1. **Jinbo Chen**, Perelman School of Medicine, University of Pennsylvania
"Absolute Risk Prediction Through Integration of Data from Multiple Sources"
2. **Summer Han**, Stanford University
"Simulating Risk Factors for Lung Cancer to Optimize Lung Screening Guidelines"
3. **Rajeshwari Sundaram**, Nichd
"Prediction of Infertility Based on Behavior and Biology: a Joint Modeling Approach"
4. **Parichoy Pal Choudhury**, Bloomberg School of Public Health, Johns Hopkins University
"Model Evaluation using Missing Data Approach in Two-Phase Studies"

Laurel Hall Room 201

12. Risk Analysis in the Biomedical and Environmental Field

Organizer and Chair: **Catherine Huber-Carol**, University Paris Descartes

1. **Olivier Bouaziz**, University Paris Descartes
"A Change-Point Model for Detecting Heterogeneity in Ordered Survival Responses"
2. **Sneh Gulati**, Florida International University
"Methods to Estimate Probable Maximum Loss for the Florida Public Hurricane Loss Model"
3. **Pascale Tubert-Bitter**, Unit Inserm Mixte, Universit De Versailles Saint-Quentin-en-Yvelines
"Parametric Maximum Likelihood Estimation of Time-to-Onset Distribution from Adverse Drug Reaction Spontaneous Reporting Databases"
4. **Min-Ge Xie**, Rutgers University
"iFusion - Fusion Learning from Individual to Clique with an Application to Recurrent Events"

Laurel Hall Room 106

13. Recent Development in the Analysis of Complex Structured Survival Data

Organizer: **Wenqing He**, University of Western Ontario
Chair: **Yichuan Zhao**, Georgia State University

1. **Jerry Lawless**, University of Waterloo
"Some Aspects of Life History Analysis from Observational Data Bases"
2. **Zhezhen Jin**, Department of Biostatistics, Columbia University
"Survival Time-Related Cut-Point with Censored Data"
3. **Sundar Subramanian**, New Jersey Institute of Technology
"Function-Based Hypothesis Testing via Plug-in Empirical Likelihood in Censored Location-Scale Families"
4. **Wenqing He**, University of Western Ontario
"Analysis of Multivariate Survival Data under Semiparametric Copula Models with/without Measurement Error"

Laurel Hall Room 202

14. Outcome-Dependent Sampling in the Survival Analysis Context

Organizer and Chair: **Sebastien Haneuse**, Harvard T.H. Chan School of Public Health

1. **James Dai**, Fred Hutchinson Cancer Research Center
"Augmented Case-Only Designs for Randomized Clinical Trials with Failure Time Endpoints"
2. **Michelle Zhou**, Mississippi State University
"Assessing Incremental Value of Biomarkers with Multi-Phase Nested Case-Control Studies"
3. **Ina Jazic**, Harvard T.H. Chan School of Public Health
"Analysis of Semi-Competing Risks Data from a Nested Case-Control Study"
4. *Discussant:* **Enrique Schisterman**, Eunice Kennedy Shriver National Institute of Child Health and Human Development

Laurel Hall Room 205

15. Recent Developments in High-Dimensional Survival Analysis and Biased Sampling

Organizer: **Gang Li**, UCLA
Chair: **Yu Cheng**, University of Pittsburgh

1. **Ian W. McKeague**, Department of Biostatistics, Columbia University
"Tests for Stochastic Ordering under Biased Sampling"
2. **Donglin Zeng**, Department of Biostatistics, University of North Carolina at Chapel Hill
"Predicting Survival Event using Sparsely Measured High Dimensional Biomarkers"
3. **Yi Li**, Department of Biostatistics University of Michigan
"Conditional Screening for High Dimensional Survival Outcome Data"

4. **Gang Li**, Department of Biostatistics, UCLA
"Broken Adaptive Ridge Regression for High Dimensional Survival Data"

Laurel Hall Room 206

16. Advances in Summarizing and Modeling Complex Survival Data

Organizer: **Amita Manatunga**, Emory University, Atlanta, Georgia
Chair: **Eugene Huang**, Emory University, Atlanta, Georgia

1. **Chenxi Li**, Department of Epidemiology and Biostatistics Michigan State University
"A Semiparametric Multi-State Model for Correlated Interval-Censored Life-History Data in Caries Research"
2. **Limin Peng**, Emory University, Atlanta, Georgia
"A New Flexible Dependence Measure for Semi-Competing Risks Data"
3. **Shuling Liu**, Center for Outcomes Research and Evaluation, Yale University
"A Modeling Approach for Multivariate Survival Data with Random Length"
4. **David Oakes**, University of Rochester Medical Center
"Matched Pairs Survival Data a New Look at an Old Problem"

Laurel Hall Room 301

17. Semiparametric Statistical Methods for Complex Failure Time Data

Organizer: **Yanqing Sun**, The University of North Carolina at Charlotte
Chair: **Jianguo (Tony) Sun**, University of Missouri

1. **Yichuan Zhao**, Georgia State University
"Empirical Likelihood Inference for the Odds Ratio of Two Survival Functions under Right Censoring"
2. **Jiajia Zhang**, University of South Carolina
"Computationally Efficient Estimation for the Generalized Odds Rate Mixture Cure Model with Interval Censored Data"
3. **Tony Sit**, Department of Statistics, The Chinese University of Hong Kong
"Semiparametric Survival Models under General Biased Sampling Scheme"
4. **Donna Spiegelman**, Harvard School of Public Health
"Causal Estimation of Direct and Indirect Effects in Studies with Clustering"

Laurel Hall Room 302

18. Recent Developments in Statistical and Computational Methods for Biomedical Data

Organizer and Chair: **Jong H. Jeong**, University of Pittsburgh

1. **Chung-Chou H. Chang**, Department of Medicine and Department of Biostatistics, University of Pittsburgh
"Development of Model-Based Surrogate Endpoint for Sepsis Studies"
2. **Pang Du**, Department of Statistics, Virginia Tech
"Promotion Time Cure Rate Model with Nonparametric Form of Covariate Effects"
3. **Bin Nan**, Department of Biostatistics, University of Michigan
"Semiparametric z-Estimation for Bundled Parameters and Case-Cohort Design"
4. **Ying Wei**, Department of Biostatistics, Columbia University
"Handling Missing Data in Electronic Health Records"

Laurel Hall Room 107

Friday 3:45–5:30

19. From Functional to Neuroimaging Data

Organizer and Chair: **Jane-Ling Wang**, University of California, Davis

1. **Hans-Georg Mueller**, University of California, Davis
"Dynamic Modeling of Longitudinal Snippets"
2. **Damla Senturk**, University of California, Los Angeles
"A Multi-Dimensional Functional Principal Components Analysis of Eeg Data"
3. **Yue Wang**, University of North Carolina, Chapel Hill
"Functional Linear Models via Partial Least Square"
4. **Yehua Li**, Iowa State University
"Joint Modeling of Longitudinal Drug using Pattern and Time to First Relapse in Cocaine Dependence Treatment Data "

Laurel Hall Room 111

20. Recent Advances on Statistical Methods for Health Studies

Organizer and Chair: **DoHwan Park**, University of Maryland, Baltimore County

1. **Liang Zhu**, University of Texas, Houston
"A Semiparametric Likelihood-Based Method for Regression Analysis of Mixed Panel-Count Data"
2. **Haiying Wang**, University of New Hampshire, Durham
"Focused and Model Average Estimation for Regression Analysis of Panel Count Data"
3. **Dong-Yun Kim**, National Institute of Health
"Sequential Patients Recruitment Monitoring (SPRM)"

Laurel Hall Room 106

21. Joint Modeling and Weighted Estimation for Survival Data Analysis

Organizer and Chair: **Haocheng Li**, Departments of Oncology and Community Health Sciences, University of Calgary

1. **Hua Shen**, University of Calgary
   "Analysis of the Interval-Censored Life History Data with Missing Information"
2. **Ying Yan**, University of Calgary
   "Optimally Weighted Estimation in Case-Cohort Studies"
3. **Dongdong Li**, Simon Fraser University
   "Modelling and Analysis of Event Times with Observations Subject to Informative Censoring"
4. **Liqun Diao**, University of Waterloo
   "Copula-Based Models for Recurrent Exacerbations"

Laurel Hall Room 107

22. New Approaches for Analyzing Time to Event Data with Application in Cancer Studies

Organizer and Chair: **Xiaonan Xue**, Albert Einstein College of Medicine

1. **Charles B. Hall**, Albert Einstein College of Medicine
   "Piecewise Exponential Survival Models with Change Points for Modeling Changes in Relative Risk over Follow Up"
2. **Yongzhao Shao**, New York University School of Medicine
   "Measures for Prognostic Accuracy in Semi-Parametric Mixture Cure Models"
3. **Xiaonan Xue**, Albert Einstein College of Medicine
   "A Censored Quantile Regression Approach for the Analysis of Time to Event Data"
4. **Arthur Berg**, Penn State University
   "Reduced Bias in Nonparametric Censored Density and Hazard Estimation"

Laurel Hall Room 201

23. Quality of Life and Other Applications

Organizer: **Mei-Ling Ting Lee**, University of Maryland
Chair: **Xin He**, University of Maryland

1. **Catherine Huber**, Luniversit Paris Descartes
   "Survival and Quality of Life"
2. **Yen-Tsung Huang**, Academia Sinica, Taiwan
   "Mendelian Randomization for Survival Outcomes using Semi-parametric Transformation Models"
3. **Victoria Wang**, Dana Farber Cancer Institute
   "Meta-Stepp: Subpopulation Treatment Effect Pattern Plot for Individual Patient Data Meta-Analysis"
4. **Weiliang Qiu**, Harvard Medical School
   "Association of Pre-Diagnosis Bmi Measurements to Prostate Cancer Mortality"

Laurel Hall Room 202

24. Joint Modelling of Longitudinal Measurements and Event History Data

Organizer: **Wei Liu**, Department of Mathematics and Statistics, York University
Chair: **X. Joan Hu**, Department of Statistics and Actuarial Science, Simon Fraser University

1. **Wei Liu**, Department of Mathematics and Statistics, York University
   "Two-Step and Likelihood Methods for HIV Viral Dynamic Models with Covariate Measurement Errors and Missing Data"
2. **Paul Y. Peng**, Department of Public Health Sciences, Queen's University
   "Joint Modeling of Longitudinal Proportional Measurements and Survival Time and Its Application to a Breast Cancer Clinical Trial"
3. **Guohua Yan**, Department of Mathematics and Statistics, University of New Brunswick
   "A Flexible Approach for Multivariate Mixed-Effects Models in the Presence of Non-Ignorable Missingness and Measurement Error"
4. **Hongbin Zhang**, Department of Epidemiology and Biostatistics, City University of New York
   "Joint Inference of Nlme and Glmm Models with Nonignorable Drop-Outs"

Laurel Hall Room 205

25. Time-to-Event Models for Human Health Risk Assessment

Organizer and Chair: **Michael Pennell**, The Ohio State University

1. **Joel Schwartz**, T.H. Chan School of Public Health, Harvard University
   "A Marginal Structural Additive Quantile Survival Analysis to Estimate Years of Life Lost Attributable to Risk Factors"
2. **Polyna Khudyakov**, T.H. Chan School of Public Health, Harvard University
   "Survival Analysis with Measurement Error in a Cumulative Exposure Variable: Radon Progeny in Relation to Lung Cancer Mortality"
3. **Dustin Long**, University of Alabama-Birmingham,
   "Erroneous Hormetic Effect Identification in Cox Models Due to Polynomial Splines"

Laurel Hall Room 206

26. Industry Perspectives on Lifetime Data Science

Organizer and Chair: **Jonathan Siegel**, Bayer US

1. **Satrajit Roychoudhury**, Director of Statistics, Novartis
   "Audit Strategy for Blinded Independent Central Review of Progression in Cancer Clinical Trials"
2. **Aparna Anderson**, Statistics Collaborative
   "Co-Primary Endpoints: Scientific and Regulatory Points to Consider"
3. **Wayne Nelson**, Statistical Consultant
   "Get more Information from Recurrent Events Data: Product Repairs, Disease Recurrences, and Other Applications"
4. **Christian Kappeler**, Bayer Ag
   "Implementing Overall Survival Crossover Adjustment Methods: Clinical Evidence to Support the Common Treatment Effect Assumption"

Laurel Hall Room 301

27. the Win Ratio and Related Topics in Multiple Event Time Data

Organizer and Chair: **David Oakes**, University of Rochester

1. **David A. Schoenfeld**, Department of Biostatistics, Harvard T.H. Chan School of Public Health
   "Deconstructing the Win Ratio"
2. **Lu Mao**, University of Wisconsin-Madison
   "Regression and Causal Models for the Composite Endpoint"
3. **Xiaoding Luo**, Sanofi Inc
   "Weighted Win Loss Approach for Analyzing Prioritized Outcomes"
4. **Changyong Feng**, University of Rochester
   "Using Wei Lin and Weissfelds Approach in Estimating a Win Ratio"

Laurel Hall Room 302

Saturday 8:30–10:15

28. Recent Advances in Analyzing Multi-State and Family Data

Organizer and Chair: **Ying Yan**, Department of Mathematics and Statistics, University of Calgary

1. **Baojiang Chen**, University of Nebraska Medical Center
   "Using the Accelerated Failure Time Model to Analyze Current Status Data with Misclassified Covariates"
2. **Yun-Hee Choi**, Western University
   "Joint Nested Frailty Models for Screening Visits and Survival Data Arising from Lynch Syndrome Families"
3. **Karen Kopciuk**, Alberta Health Services and University of Calgary
   "Risk Estimation in Family Data via Multi-State Models"
4. **Leilei Zeng**, University of Waterloo
   "Modelling Multistate Data under Prevalent Cohort Sampling"

Laurel Hall Room 106

29. Measurement Error, Mediation Analysis, and Individualized Medicine

Organizer and Chair: **Ying Qing Chen**, Fred Hutchinson Cancer Research Center

1. **Yijian Huang**, Emory University
   "Cox Regression with Covariate Measurement Error"
2. **Cheng Zheng**, University of Wisconsin, Milwaukee, WI
   "Mediation Analysis on Time-to-Event Outcome Data with Unmeasured Confounding and Measurement Error"
3. **Yingqi Zhao**, Fred Hutchinson Cancer Research Center
   "Estimating Individualized Treatment with Censored Data"
4. *Discussant:* **Ross Prentice**, Fred Hutchinson Cancer Research Center

Laurel Hall Room 111

30. Statistical Analysis of Recurrent, Competing Risks, and Current Status Data

Organizer and Chair: **Edsel A. Pena**, University of South Carolina

1. **Gary Chan**, University of Washington
   "Modeling and Estimating the Terminal Behavior of Recurrent Marker Processes Before Failure Events"
2. **Michael Pennell**, Ohio State University
   "Bayesian Threshold Regression for Multivariate Current Status Data with Informative Censoring"
3. **Alexander McLain**, University of South Carolina
   "Analysis of Current Duration Data: Modeling Strategies and Competing Risks"
4. **Piaomu Liu**, Bentley University
   "Model Diagnostics of a Class of Joint Dynamic Models of Recurrent Competing Risks and a Terminal Event"

Laurel Hall Room 201

31. On Risk Prediction in the Presence of Competing Risks

Organizer and Chair: **Joanna H. Shih**, National Cancer Institute

1. **Guoqing Diao**, Department of Statistics, the Volgenau School of Engineering, George Mason University
   "Time-Varying Coefficient Risk Prediction Models for Competing Risks Data"
2. **Thomas H. Scheike**, Department of Biostatistics, University of Copenhagen
   "Estimation of Dependence Parameters for Competing Risks Data"
3. **Malka Gorfine**, School of Mathematical Sciences, Tel Aviv University
   "Calibrated Predictions for Multivariate Competing Risks Models"

4. **Per Kragh Andersen**, Department of Biostatistics, University of Copenhagen
"Evaluation of Models for Predicting the Cumulative Incidence"

Laurel Hall Room 202

32. Challenges and New Methods of Complex Health Data

Organizer and Chair: **Grace Y. Yi**, University of Waterloo

1. **Yanqing Sun**, The University of North Carolina at Charlotte
"Analysis of Two-Phase Sampling Data with Semiparametric Additive Hazards Models"
2. **Tibor Schuster**, McGill University
"Challenges in Simulating High-Dimensional Health Data for Inference Method Evaluation"
3. **Michal Abrahamowicz**, McGill University
"Flexible Modeling of Survival Curves Conditional on Time-Dependent and/or Non-Linear Effects of Prognostic Factors"
4. **Joan Hu**, Simon Fraser University
"Risk Classification and Prediction with Event History Data"

Laurel Hall Room 205

33. New Statistical Methods for Complex Structures in Survival Data

Organizer: **Donglin Zeng**, University of North Carolina at Chapel Hill
Chair: **Jianwen Cai**, University of North Carolina at Chapel Hill

1. **Rui Song**, North Carolina State University
"On Estimation of Optimal Treatment Regimens for Survival Data"
2. **Qingxia Chen**, Vanderbilt University
"Quantifying the Average of the Time-Varying Hazard Ratio via a Class of Transformations"
3. **Qiang Sun**, Yale University
"Counting Process Based Dimension Reduction for Censored Outcome"
4. **Fang-Shu Ou**, Mayo Clinic
"Quantile Regression Models for Interval-Censored Failure Time Data"

Laurel Hall Room 206

34. New Machine Learning Methods for Censored Survival and Competing Risks Data

Organizer and Chair: **Robert Strawderman**, Department of Biostatistics & Computational Biology, University of Rochester

1. **Jon Steingrimsson**, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health.
"Censoring Unbiased Regression Trees and Forests"
2. **Youngjoo Cho**, Department of Biostatistics & Computational Biology, University of Rochester
"Regression Trees under Competing Risks"
3. **Ruoqing Zhu**, Department of Statistics, University of Illinois at Urbana-Champaign
"Consistency of Survival Tree Models with Martingale-Based Splitting Rules"
4. *Discussant:* **Annette Molinaro**, Department of Neurological Surgery, University of California at San Francisco

Laurel Hall Room 301

35. Advances in Statistical Modeling of Correlated Data

Organizers: **Zhengqing Ouyang**, The Jackson Laboratory for Genomic Medicine, and **Yuping Zhang**, University of Connecticut
Chair: **Zhengqing Ouyang**, The Jackson Laboratory for Genomic Medicine

1. **Antai Wang**, New Jersey Institute of Technology
"A New Estimator of Baseline Hazard Function in Bivariate Frailty Models"
2. **Zuoheng (Anita) Wang**, Yale University
"Joint Statistical Modeling of Multiple Phenotypes in Related Samples"
3. **Ji Meng Loh**, New Jersey Institute of Technology
"A Single-Index Model for Inhomogeneous Spatial Point Patterns"
4. **Yuping Zhang**, University of Connecticut
"A Statistical Framework for Data Integration through Graphical Models with Application to Cancer Genomics"

Laurel Hall Room 302

36. Modeling Disease Natural History and Effects of Treatment: Applications to Prostate Cancer

Organizer and Chair: **James Dignam**, Department of Public Health Sciences, University of Chicago

1. **Alexander Tsodikov**, Department of Biostatistics, University of Michigan
"Modeling Cancer Natural History and Mortality"
2. **Yolanda Hagar**, Department of Applied Mathematics, University of Colorado
"Flexible Modeling of the Hazard Rate to Capture Short and Long-Term Treatment Effects"
3. **Meredith Regan**, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute
"Identification and Validation of Surrogate Time-to-Event Endpoints for Use in Localized Prostate Cancer Randomized Trials using Individual Patient Data Meta-Analysis"
4. **Loic Ferrer**, Inserm, Universit De Bordeaux, France
"Joint Modelling of Longitudinal and Multi-State Processes: Application to Clinical Progressions in Prostate Cancer"

Laurel Hall Room 107

37. Student Paper Presentations

Organizer and Chair: **Mei-Cheng Wang**, Johns Hopkins University

1. **Ling-Wan Chen**, University of Pittsburgh
   "Cumulative Incidence Regression for Dynamic Treatment Regimens"
2. **Fei Gao**, University of North Carolina
   "Semiparametric Regression Analysis of Interval-Censored Data with Informative Dropout"
3. **Haitao Huang**, Georgia State University
   "Empirical Likelihood for the Bivariate Survival Function under Univariate Censoring"
4. **Nathalie C. Moon**, University of Waterloo
   "Tracing Studies in Cohorts with Loss-to-Follow-Up: Selection Models for Optimal Efficiency"
5. **Jin Wang**, University of North Carolina
   "Single Index Models in Proportional Hazard Regression for Precision Medicine"
6. **Wenjie Wang**, University of Connecticut
   "Extended Cox Model by Ecm Algorithm for Uncertain Survival Records Due to Imperfect Data Integration"

Laurel Hall Room 108

Saturday 10:30–12:15

38. Tree-Based Methods for Survival Data

Organizer: **Denis Larocque**, HEC Montral
Chair: **Ruoqing Zhu**, University of Illinois at Urbana-Champaign

1. **Adele Marshall**, Queen's University Belfast
   "A Tree-Based Method for the Coxian Phase-Type Distribution"
2. **Philippe Brot**, University Paris-Saclay
   "Bagging Improper Survival Trees for Prediction and Variable Selection"
3. **Jeffrey S. Simonoff**, New York University
   "Conditional Inference Survival Trees for Nonstandard Data"
4. **Rodney Sparapani**, Medical College of Wisconsin
   "Nonparametric Recurrent Events Analysis with Bart and an Application to the Hospital Admissions of Patients with Diabetes"
5. **David Vock**, University of Minnesota
   "Adapting Trees and Other Machine Learning Techniques to Censored Time-to-Event Health Record Data"
6. **Denis Larocque**, HEC Montral
   "L1 Splitting Rules in Survival Forests"

Laurel Hall Room 108

39. Advances in Multi-State Models for Survival and Event History Analysis

Organizer: **Lihui Zhao**, Northwestern University
Chair: **X. Joan Hu**, Simon Fraser University

1. **Candemir Cigsar**, Memorial University of Newfoundland
   "Assessment of Dynamic Models for Multi-State Models with Recurrent Events: a Copula Approach"
2. **Elizabeth Juarez-Colunga**, University of Colorado Denver
   "Analysis of Recurrent Pulmonary Exacerbations in Cystic Fibrosis Children: Early Pseudomonas Infection Control (Epic) Observational Study"
3. **Lihui Zhao**, Northwestern University
   "A Semi-Markov Model for Event History Data with Time-Dependent Covariates"
4. **Peihua Qiu**, University of Florida
   "Effective Comparison of Two or more Hazard Rate Functions"

Laurel Hall Room 106

40. Advances in the Analysis of Clinical Trials with Lifetime Data using Restricted Mean Survival Times

Organizer and Chair: **Ludovic Trinquart**, Department of Biostatistics, Boston University School of Public Health

1. **Ivan Diaz**, Department of Biostatistics and Epidemiology, Weill Cornell Medical College, Cornell University
   "Improved Precision in the Analysis of Randomized Trials with Survival Outcomes, without Assuming Proportional Hazards"
2. **Kevin H. Eng**, Department of Biostatistics and Bioinformatics, Roswell Park Cancer Institute
   "Continuous Biomarker Strategy Characterization by Restricted Mean Survival Curve"
3. **Lu Tian**, Department of Health Research and Policy, Stanford University School of Medicine
   "Optimal Stratification in Outcome Prediction using Baseline Information"
4. **Sarah Conner**, Department of Biostatistics, Boston University School of Public Health
   "Adjusted Restricted Mean Survival Times in Observational Studies"

Laurel Hall Room 111

41. New Methods for Censored Data

Organizer and Chair: **Bin Nan**, University of Michigan at Ann Arbor

1. **Jong-Hyeon Jeong**, University of Pittsburgh
   "Semiparametric Regression on Quantile Life Lost: Application to Phase Iii Breast Cancer Data"
2. **Ying Qing Chen**, Fred Hutchinson Cancer Research Center
   "Modeling the Trend of Recurrent Event Data with Weak Comparability"
3. **Douglas Schaubel**, University of Michigan at Ann Arbor
   "Modeling Restricted Mean Survival Time under General Censoring Mechanisms"
4. **Ross Prentice**, Fred Hutchinson Cancer Research Center
   "Self-Consistent Nonparametric Estimation of the Multivariate Survivor Function"

Laurel Hall Room 201

42. Recent Advances in Survival Analysis for Cancer Research

Organizer: **Andy Ni**, Memorial Sloan Kettering Cancer Center
Chair: **Jianwen Cai**, University of North Carolina at Chapel Hill

1. **Qi Long**, Department of Biostatistics, University of Pennsylvania
   "Addressing Issues Associated with Evaluating Prediction Models for Survival Endpoints Based on the Concordance Statistic"
2. **Glenn Heller**, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center
   "The Concordance Probability Estimate to Assess Model Performance with Survival Data"
3. **Noorie Hyun**, Division of Cancer Epidemiology and Genetics, National Cancer Institute
   "Absolute Cause-Specific Risk Calculation from Interval-Censored Electronic Health Records"
4. **Yixin Wang**, Fred Hutchinson Cancer Research Center
   "Proportional Mean Residual Life Model for both Incident and Prevalent Data"

Laurel Hall Room 202

43. Risk Assessment and Prediction with Survival Data

Organizer: **Jing Ning**, Department of Biostatistics, The University of Texas MD Anderson Cancer Center
Chair: **Chunyan Cai**, Department of Internal Medicine, The University of Texas Health Science Center at Houston

1. **Xuelin Huang**, Department of Biostatistics, the University of Texas MD Anderson Cancer Center
   "Using Longitudinal Biomarker Data to Dynamically Predict Time to Disease Progression"
2. **Yingye Zheng**, Biostatistics and Biomathematics Program, Fred Hutchinson Cancer Research Center
   "Efficient Evaluation of the Incremental Values of Novel Biomarkers: Design and Analysis"
3. **Liang Li**, Department of Biostatistics, the University of Texas MD Anderson Cancer Center
   "On the Existence and Data Generation of Landmark Cox Model"
4. **Ying Ding**, Biostatistics Department, University of Pittsburgh
   "Progression Risk Estimation with Copula Model in Age-Related Macular Degeneration Patients"

Laurel Hall Room 205

44. Recent Development of Time-to-Event Data Analysis Incorporating Disease Dynamics

Organizer and Chair: **Chen Hu**, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine

1. **Zhi He**, University of Michigan, Ann Arbor
   "Gateaux Differential-Based Boosting for Fitting Large-Scale Survival Data with Non-Proportional Hazards"
2. **Bin Zhang**, Cincinnati Children's Hospital Medical Center
   "Bayesian Approach for Clustered Interval-Censored Data with Time-Varying Covariate Effects"
3. **Sijin Wen**, West Virginia University Health Science Center
   "A Bayesian Multivariate Joint Frailty Model for Disease Recurrences and Survival"
4. **Li C. Cheung**, George Washington University
   "Concordance Indices for Composite Survival Outcomes"

Laurel Hall Room 107

45. Analytical Challenges and New Advances in Assessing Time-to-Event Endpoints in Oncology Studies

Organizer and Chair: **Chiung-Yu Huang**, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University

1. **Chen Hu**, Johns Hopkins University
   "A New Method for Summarizing Treatment Effect on Multiple Endpoints in Hematopoietic Stem Cell Transplantation Studies"
2. **Wei-Ting Hwang**, University of Pennsylvania
   "Predictive Value of Continuous Markers for Censored Survival Data: a Likelihood Ratio Approach and Extension to Competing Risk Framework"
3. **Hongyuan Cao**, University of Missouri
   "On the Proportional Hazards Model with Last Observation Carried Forward Covariates"
4. **Hongwei Zhao**, Texas A&M University
   "Improved Survival Analysis in Two Personalized Breast Cancer Studies"

Laurel Hall Room 206

46. Analysis of Complex Sampling Designs with Censored Data

Organizer and Chair: **Jon Steingrimsson**, Department of Biostatistics Johns Hopkins Bloomberg School of Public Health

1. **Youyi Fong**, Vaccine and Infectious Disease Division and Public Health Sciences Division Fred Hutchinson Cancer Research Center
   "Calibration Weighted Estimation of Semiparametric Transformation Models for Two-Phase Sampling"
2. **Olli Saarela**, Division of Biostatistics University of Toronto
   "A New Weighted Partial Likelihood Method for Estimating Marginal Structural Hazard Models"

3. **Takumi Saegusa**, Department of Mathematics, University of Maryland
   "Survival Analysis under Multiple-Frame Sampling"
4. **Andy Ni**, Memorial Sloan Kettering Cancer Center
   "Tuning Parameter Selection in Cox Proportional Hazards Model with Diverging Number of Parameters"

Laurel Hall Room 301

47. Survival Analysis for Family-Clustered Data

Organizer and Chair: **Niels Keiding**, University of Copenhagen

1. **Frank Eriksson**, University of Copenhagen
   "Analysis of Cluster Truncated Register Data"
2. **Luise Cederkvist**, University of Copenhagen & Danish Cancer Society Research Center
   "Modelling the Cumulative Incidence Function of Multivariate Competing Risks Data While Allowing for Within-Family Dependence of Risk and Timing"
3. **Jeanine Houwing-Duistermaat**, Department of Statistics, University of Leeds
   "Modelling Mortality in Long Lived Families"
4. *Discussant:* **Malka Gorfine**, Tel Aviv University, Israel

Laurel Hall Room 302

Saturday 1:45–3:30

48. New Statistical Approaches on Medical Data

Organizer and Chair: **Mounir Mesbah**, Universit Pierre et Marie Curie,P6

1. **Jean-Franois Dupuy**, Insa, Rennes
   "Estimation of Extreme Quantiles of a Conditional Survival Distribution with Right-Censoring"
2. **Amelie Anota**, University Hospital of Besanon
   "Time to Health Related Quality of Life Score Deterioration as a Tool of Longitudinal Analyses in Oncology: Is it the Optimal Statistical Modeling?"
3. **Agathe Guilloux**, Universit Pierre et Marie Curie, Paris 6
   "C-Mix: a High Dimensional Mixture Model for Censored Durations, with Applications to Biomedical Data"
4. *Discussant:* **Mounir Mesbah**, Universit Pierre et Marie Curie, P6

Laurel Hall Room 111

49. Recent Advances in the Analysis of Competing Risks Data

Organizer: **Ronghui (Lily) Xu**, University of California, San Diego
Chair: **Leilei Zeng**, University of Waterloo, Canada

1. **Arthur Allignol**, University of Ulm, Germany
   "Estimation of the Cumulative Incidence Function in the Presence of Dependent Left-Truncation"
2. **Qing Yang**, Duke University
   "Sample Size Calculation for Joint Testing of Competing Risk Survival Data"
3. **Jelena Bradic**, University of California, San Diego
   "Inference for Competing Risks in High-Dimensional Settings"

Laurel Hall Room 201

50. Bayesian Nonparametric Survival Analysis

Organizer: **Lynn Kuo**, University of Connecticut
Chair: **Victor Hugo Lachos Davila**, Campinas State University

1. **Peter Mueller**, University of Texas at Austin
   "A Bayesian Nonparametric Approach for Semi-Competing Risk"
2. **Athanasios Kottas**, University of California, Santa Cruz
   "Nonparametric Bayesian Modeling for Mean Residual Life Regression"
3. **William Cipolli**, Colgate College
   "Accelerated Failure Time Models via Smoothed, Approximate Polya Trees"
4. **Lynn Kuo**, University of Connecticut
   "Model Selection for Bayesian Nonparametric Survival Models using Bregman Divergence Measure"

Laurel Hall Room 202

51. Recent Developments in Complex Survival and Longitudinal Data

Organizer and Chair: **Yifei Sun**, Johns Hopkins University

1. **Yize Zhao**, Division of Biostatistics and Epidemiology, Department of Healthcare Policy and Research, Weill Cornell Medical College, Cornell University
   "Hierarchical Feature Selection Incorporating Known and Novel Biological Information: Identifying Genomic Features Related to Prostate Cancer Recurrence"
2. **Ming-Yueh Huang**, Department of Biostatistics, University of Washington
   "Sufficient Dimension Reduction for Censored Survival Data"
3. **Shao-Hsuan Wang**, Division of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University
   "Optimal Sufficient Dimension Reduction Score with Censored Survival Data"
4. **Yong Chen**, Department of Biostatistics and Epidemiology, University of Pennsylvania
   "Analysis of Longitudinal Data under Biased Sampling"

Laurel Hall Room 106

52. New Developments in Modeling Longitudinal, Recurrent Event and Survival Data

Organizer: **Hong Zhu**, Division of Biostatistics, Department of Clinical Sciences, The University of Texas Southwestern Medical Center
Chair: **Jing Ning**, University of Texas MD Anderson Cancer Center

1. **Chunyan Cai**, Department of Internal Medicine, the University of Texas Health Science Center at Houston
   "Time-Varying Dependence Measure of Bivariate Recurrent Event Processes"
2. **Sehee Kim**, Department of Biostatistics, University of Michigan, Ann Arbor
   "Joint Partially Linear Model for Longitudinal Data with Informative Drop-Outs"
3. **Chi Hyun Lee**, Department of Biostatistics, the University of Texas MD Anderson Cancer Center
   "Semiparametric Regression Model for Bivariate Alternating Recurrent Event Data"
4. **Shanshan Li**, Department of Biostatistics, Indiana University School of Public Health
   "Joint Modeling of Recurrent Event Processes and Intermittently Observed Time-Varying Binary Covariate Processes"

Laurel Hall Room 205

53. Unconventional Usage of Classic Methods in Modern Medical Applications

Organizer: **Yanyuan Ma**, Penn State University
Chair: **Sibo Zhao**, New York University

1. **Tanya Garcia**, Texas A&M University
   "Cox Regression with Exclusion Frequency-Based Weights: Application to a Neuro-Imaging Study of Huntington's Disease"
2. **Fei Jiang**, University of Hongkong
   "A Second Order Semiparametric Method for Survival Analysis, with Application to an Aids Clinical Trial Study"
3. **Ying Liu**, Medical College of Wisconsin
   "Estimating Personalized Diagnostic Rules Depending on Individualized Characteristics"
4. **Shijun Zhu**, University of Maryland School of Nursing
   "Joint Modeling of Multivariate Longitudinal Data and Recurrent Events and It's Application"

Laurel Hall Room 206

54. Recent Topics on Analysis of Interval Censored Data (with Covariate or Outcome Censored)

Organizer and Chair: **Xiangrong Kong**, Johns Hopkins University

1. **Paul Bernhardt**, Villanova University
   "A Fast Em Algorithm for Fitting Joint Models of a Binary Response and Multiple Longitudinal Covariates Subject to Detection Limits"
2. **Jing Qian**, Department of Biostatistics and Epidemiology, University of Massachusetts - Amherst
   "Regression Analysis with Randomly Censored Covariates"
3. **Abdus Sattar**, Case Western Reserve University
   "A Parametric Survival Model when a Covariate is Subject to Left-Censoring"
4. **Qiang Zhao**, Texas State University
   "Estimation of Survival Function for Interval-Censored Std Data with Auxiliary Diaries"

Laurel Hall Room 301

55. New Development in Statistical Methods for Deriving and Validating Dynamic and Individualized Decision Rules

Organizer and Chair: **Yingye Zheng**, Fred Hutchinson Cancer Research Center

1. **Dandan Liu**, Department of Biostatistics, Vanderbilt University Medical Center
   "Target Population Risk Prediction with Time-to-Event Outcome"
2. **Li Hsu**, Department of Biostatistics, Public Health Sciences Division, Fred Hutchinson Cancer Research Center
   "Recommendation of when to Start Intervention: From Binary Decision to Time-to-Treatment Decision"
3. **Marlena Maziarz**, Biostatistics Branch, National Cancer Institute
   "Evaluating Longitudinal Markers under Two-Phase Study Designs"
4. **Yunro Chung**, Department of Biostatistics, Public Health Sciences Division, Fred Hutchinson Cancer Research Center
   "Estimation of Disease Progression Rate using Longitudinal Surrogate Outcomes in Non-Randomized Validation Subsample"

Laurel Hall Room 302

Saturday 3:45–5:30

56. Advances in Semiparametric Regression Analysis of Panel Count Data

Organizer and Chair: **Guoqing Diao**, Department of Statistics, George Mason University

1. **Ao Yuan**, Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University
   "Maximum Likelihood Estimation and Em Algorithms with Panel Count Data"
2. **Xingqiu Zhao**, Department of Applied Mathematics, the Hong Kong Polytechnic University
   "Semiparametric Regression Analysis of Multivariate Longitudinal Data with Informative Observation Times"
3. **Bin Yao**, FMD K&L Inc
   "Joint Modeling of Panel Count Data and Interval-Censored Data with Application to Sexually Transmitted Infections"

4. **Xin He**, Department of Epidemiology and Biostatistics, School of Public Health, University of Maryland
"Semiparametric Partially Linear Varying Coefficient Models with Panel Count Data"

Laurel Hall Room 111

57. a Recent Development on Competing and Semi-Competing Risks

Organizer and Chair: **Qingxia Chen**, Department of Biostatistics, Vanderbilt University Medical Center

1. **Chung-Chou H. Chang**, Univeristy of Pittsburgh
"Regression Model for Data with Competing Risks under Random Signs Censoring"
2. **Ruosha Li**, The University of Texas, Health Science Center at Houston
"Flexible Association Modelling and Prediction with Semi-Competing Risks Data"
3. **Yu Cheng**, Departments of Statistics and Psychiatry, University of Pittsburgh
"Novel Diagnostic Accuracy Analysis for Competing Risks Outcomes with Roc Surface"
4. **Ming-Hui Chen**, Department of Statistics, University of Connecticut
"A Bayesian Cure Rate Frailty Model for Survival Data in Presence of Semi-Competing and Competing Risks"

Laurel Hall Room 201

58. Recent Advances on Event History Analysis

Organizer and Chair: **Ling Ma**, Department of Mathematical Sciences, Clemson University

1. **Zhigang Zhang**, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center
"A Sensitivity Study of Latent Frailty Terms in Modeling Dependence Structures in Failure Time Data"
2. **Yifei Sun**, Department of Biostatistics, Johns Hopkins University
"Missing Information Principle: a Unified Approach for General Left-Truncated and/or Right-Censored Survival Data Problems"
3. **Ling Ma**, Department of Mathematical Sciences, Clemson University
"Joint Modeling of Longitudinal Functional Features and Discrete Time-to-Event"
4. **Qingning Zhou**, Department of Biostatistics, University of North Carolina at Chapel Hill
"Case-Cohort Studies with Interval-Censored Failure Time Data"

Laurel Hall Room 202

59. Recent Advances in Time-to-Event Analysis with High Dimensional, Heterogeneous, and/or Correlated Data

Organizer and Chair: **Elizabeth Schifano**, Department of Statistics, University of Connecticut

1. **Dipak K. Dey**, Department of Statistics, University of Connecticut
"A Flexible Cure Rate Model for Spatially Correlated Survival Data Based on Generalized Extreme Value Distribution and Gaussian Process Priors"
2. **Shuangge Steven Ma**, Department of Biostatistics, Yale University
"Robust Analysis of Cancer Prognosis Data with Gene-Environment Interactions"
3. **Sebastien Haneuse**, Department of Biostatistics, Harvard T.H. Chan School of Public Health
"Hierarchical Models for Semi-Competing Risks Data with Application to Quality of End-of-Life Care for Pancreatic Cancer"
4. **Jian Kang**, Department of Biostatistics, University of Michigan
"Conditional Screening for Ultrahigh Dimensional Covariates with Survival Outcomes"

Laurel Hall Room 205

60. Statistical Innovations for Data Science and Precision Medicine

Organizer and Chair: **Yuping Zhang**, University of Connecticut

1. **Xiao-Li Meng**, Harvard University
"Personalized Treatment: Sounds Heavenly, but where on Earth Did They Find the Right Guinea Pig for Me?"
2. **Jun Xie**, Purdue University
"New Statistical Methods of Large-Scale Inference with Applications on Genomics Data for Precision Medicine"
3. **Haiqun Lin**, Yale University
"Mediation Analysis with Latent Class Mediators"
4. **Zhengqing Ouyang**, The Jackson Laboratory for Genomic Medicine
"Statistical Modeling of Genome-wide Chromatin Interaction Data to Elucidate Spatial Organizations of Genomes"

Laurel Hall Room 206

61. Risk Prediction in Survival Analysis

Organizer and Chair: **Qian (Michelle) Zhou**, Mississippi State University

1. **Joel Dubin**, University of Waterloo, Canada.
"Similarity-Based Time-to-Event Prediction"
2. **Yan Yuan**, School of Public Health, University of Alberta, Canada.
"A Threshold-Free Prospective Prediction Accuracy Measure for Censored Time to Event Data"

3. **Jessica Minnier**, Ohsu-Psu School of Public Health, Oregon Health & Science University
"Automated Feature Selection for Prediction with Electronic Medical Records Data"

Laurel Hall Room 301

62. Methods and Applications of Recurrent Event or Panel Count Data

Organizer and Chair: **Jun Yan**, Department of Statistics, University of Connecticut

1. **Steven Chiou**, Harvard University
"Semiparametric Regression Scale-Change Model for Panel Count Data with Information Observation Times"
2. **Gongjun Xu**, School of Statistics, University of Minnesota
"Joint Scale-Change Models for Recurrent Events and Failure Time"
3. **Ran Duan**, Diabetes Business Unit, Eli Lilly and Company
"Estimate Variable Importance for Recurrent Event Outcomes with an Application to Identify Hypoglycemia Risk Factors"
4. *Discussant:* **Haim Bar**, Department of Statistics, University of Connecticut

Laurel Hall Room 302

63. Statistical Challenges for Immunotherapy with Delayed Treatment Effects

Organizer and Chair: **Alan Chiang**, Eli Lilly and Company

1. **Tianle Hu**, Eli Lilly and Company
"What Do We Learn from Non-Proportional Hazard in PD-(L)1 Trials?"
2. **Wen Zhou**, Novartis
"Statistical Challenges in Quantitation of Clinical Benefits of Immunotherapeutic Drugs"
3. **Zhenzhen Xu**, FDA
"Designing Therapeutic Cancer Vaccine Trials with Random Delayed Treatment Effect"
4. **Alan Chiang**, Eli Lilly and Company
"Practical Considerations for Delayed Treatment Effects in Clinical Trials"

Laurel Hall Room 106

Abstracts of Invited Papers

Friday 10:30–12:15

1. Sample Size Considerations and Dynamic Prediction in Time-to-Event Studies

- **Bin Nan**, University of Michigan
  "Semiparametric z-Estimation for Bundled Parameters and Case-Cohort Design"
  Bin Nan

  Many semiparametric models can be parameterized by two types of parameters a Euclidean parameter of interest and an infinite-dimensional nuisance parameter, and the two parameters are bundled together, i.e., the nuisance parameter is an unknown function that contains the parameter of interest as part of its argument. In this talk, I will present a general semiparametric Z-estimation theory for a class of problems where the estimating function for the Euclidean parameter of interest is constructed by replacing the infinite-dimensional nuisance parameter with a reasonable estimator in some random map. This is motivated by the increasingly used outcome-dependent sampling designs for censored survival data – the case-cohort studies, for which the commonly used counting process stochastic integrals approach lacks theoretical justification for outcome-dependent weighted methods due to non-predictability.

3. Alternatives to the log-Rank Test for Clinical Trials

- **Rick Chappell**, University of Wisconsin Madison
  "Drawbacks of the log-Rank Test and Alternatives to It."
  Rick Chappell

  The hazard ratio (HR) and the associated log-rank test are often used to describe differences in survival time endpoints for clinical trials. When the underlying hazards are proportional, the HR is a simple quantity to estimate and test. However, Cox himself pointed out that it usually lacks an underlying scientific interpretation in medical contexts; furthermore, its estimate can be biased even in the presence of randomization and proportionality. This talk will elaborate on these criticisms and present some simple alternative measures.

- **Hajime Uno**, Dana Faber Cancer Institute
  "On Testing Based on Restricted Mean Survival Time for Time-to-Event Outcomes"
  Miki Horiguchi, Masahiro Takeuchi, Hajime Uno

  Panel count data occur in studies where the study subjects are observed only periodically or at discrete examination times. In contrast to most of the existing approaches that assume independence between the recurrent event process and the examination time process, we address the issue of informative examination time process by considering a scale-change model for the underlying recurrent event process and allow the two processes to be correlated through a shared frailty. Our model can be viewed as an accelerated failure time type model for recurrent events, and the regression parameters have a simple marginal interpretation of modifying the time scale of the event process. A novel estimation procedure for the regression parameters and the baseline rate function is proposed, which, in contrast to existing methods, is robust in the sense that it does not require the strong Poisson-type assumption for the underlying recurrent event process, or impose a parametric assumption on the distribution of the unobserved frailty. Large-sample properties of the estimators are established, and their variances are estimated by a model-based smoothed bootstrap procedure. Numerical studies demonstrated that the proposed point estimator and variance estimator perform well with practical sample sizes. The methods are applied to data from a skin cancer chemoprevention trial.

4. Complex Modeling of Survival and Longitudinal Data

- **Grace Y. Yi**, University of Waterloo
  "Analysis of Survival Data with Covariate Measurement Error under the Additive Hazards Model"
  Ying Yan and Grace Y. Yi

  Covariate measurement error has attracted extensive interest in survival analysis. Since Prentice (1982), a large number of inference methods have been developed to handle error-contaminated data, and most methods are addressed to proportional hazards models. In contrast to proportional hazards models, additive hazards models offer a flexible alternative to delineate survival data. However, there is relatively less research on measurement error effects under such models, although some authors investigated this problem. In this talk, I will discuss several methods to correct for measurement error effects under additive hazards models. These methods will be justified both theoretically and empirically.

- **Jianguo Sun**, University of Missouri
  "Joint Analysis of Interval-Censored Failure Time Data and Panel Count Data"
  Da Xu and Jianguo Sun

  Interval-censored failure time data and panel count data are two types of incomplete data that commonly occur in event history studies and many methods have been developed for their analysis separately (Sun, 2006; Sun and Zhao, 2013). Sometimes one may be interested in or need to conduct their joint analysis such as in the clinical trials with composite endpoints, for which it does not seem to exist an established approach in the literature. In this paper, a sieve maximum likelihood approach is developed for the joint analysis and in the proposed method, Bernstein polynomials are used to approximate unknown functions. The asymptotic properties of the resulting estimators are established and in particular, the proposed estimators of regression parameters are shown to be semiparametrically efficient. In addition, an extensive simulation study was conducted and the proposed method is applied to a set of real data arising from a skin cancer study.

- **Jane-Ling Wang**, University of California, Davis
  "Functional Cox Model"
  Simeng Qu, Jane-Ling Wang and Xiao Wang

  Functional covariates are common in many medical, biodemographic, and neuroimaging studies. The aim of this paper is to study functional Cox models with right-censored data in the presence of both functional and scalar covariates. We study the asymptotic properties of the maximum partial likelihood estimator and establish the asymptotic normality and efficiency of the estimator of the finite-dimensional estimator. Under the framework of reproducing kernel Hilbert space, the estimator of the coefficient function for a functional covariate achieves the minimax optimal rate of convergence under a weighted L_2-risk. This optimal rate is determined jointly by the censoring scheme, the reproducing kernel and the covariance kernel of the functional covariates. Implementation of the estimation approach and the selection of the smoothing parameter are discussed in detail. The finite sample performance is illustrated by simulated examples and a real application.

- **Jianguo Sun**, University of Missouri
  "Joint Analysis of Interval-Censored Failure Time Data and Panel Count Data"
  Da Xu and Jianguo Sun

  Interval-censored failure time data and panel count data are two types of incomplete data that commonly occur in event history studies and many methods have been developed for their analysis separately (Sun, 2006; Sun and Zhao, 2013). Sometimes one may be interested in or need to conduct their joint analysis such as in the clinical trials with composite endpoints, for which it does not seem to exist an established approach in the literature. In this paper, a sieve maximum likelihood approach is developed for the joint analysis and in the proposed method, Bernstein polynomials are used to approximate unknown functions. The asymptotic properties of the resulting estimators are established and in particular, the proposed estimators of regression parameters are shown to be semiparametrically

efficient. In addition, an extensive simulation study was conducted and the proposed method is applied to a set of real data arising from a skin cancer study.

5. Observational Survival Data

- **Lida Gharibvand**, Loma Linda University
  "The Association Between Ambient Fine Particulate Matter and Adenocarcinoma Subtype of Lung Cancer. Results from a Cohort Study. "
  Lida Gharibvand; David Shavlik; W. Lawrence Beeson; Raymond Knutsen; Synnove F. Knutsen

  Background: Adenocarcinoma (AC) is the most common lung cancer among non-smokers, but few studies have assessed the effect of PM2.5 on AC among never smokers. The purpose of this study was to assess the association between ambient PM2.5 and incident lung AC in the Adventist Health and Smog Study-2 (AHSMOG-2), a cohort of 80,044 non-smokers (81Methods: Incident lung AC was identified through linkage with U.S. state cancer registries. Ambient PM2.5 and ozone (O3) levels at subjects' residence were estimated for the years 2000 and 2001, immediately prior to study start. Results: A total of 164 incident lung AC occurred during follow-up. Each 10 g/m3 increment in PM2.5 was associated with an increase in the hazard rate of lung AC [HR=1.31 (95Conclusions: Increased risk of AC was observed for each 10-g/m3 increment in ambient PM2.5 concentrations. The risk was higher among those without prevalent NMSC and those who spent more than 1 hr/day outdoors.

- **Nicholas p Jewell**, University of California, Berkeley
  "Grouped Current Status Data"
  Nicholas P. Jewell, Lucia C. Petito

  Grouped testing, first introduced by Dorfman (1943), has been used as a method to reduce costs when estimating the prevalence of a binary characteristic based on a screening test of k groups that include n independent individuals in total. If the unknown prevalence is low, and the screening test suffers from misclassification, it is also possible to obtain more precise prevalence estimates than those obtained from testing all n samples separately (Tu et al., 1994). In some applications, the individual binary response corresponds to whether an underlying time to event variable T is less than an observed screening time C, a data structure known as current status data. Given sufficient variation in the observed C s, it is possible to estimate the distribution function, F, of T nonparametrically, at least at some points in its support, using the pool-adjacent-violators algorithm (Ayer et al., 1955). Here, we consider similar nonparametric estimation of F based on group tested current status data for k groups where the group tests positive" if and only if any individual's unobserved T is less than its corresponding observed C. We investigate the performance of the group-based estimator as compared to the individual test nonparametric maximum likelihood estimator, and show that the former can be more precise in the presence of misclassification for low values of F(t). Potential applications include testing for the presence of various diseases from pooled samples where interest focuses on the age at incidence distribution rather than overall prevalence. We apply this estimator to the age-at-incidence curve for hepatitis C infection in a sample of U.S. women who gave birth to a child in 2014, where group assignment is done at random and based on maternal age.

6. Asymptotics Methods on Censored and Truncated Data

- **Jimmy T. Efird, Phd, Msc**, University of Newcastle, Australia
  "Risk Stratification using a Weibull Proportional-Hazards Model: Asymptotic Efficiency of a Counting Process Approach"
  Jimmy T.Efird, PhD, MSc

  Although commonly used to assess disease risk, adjusted survival estimates generated from a Cox proportional-hazards model tend to deviate from their true values with increasing time from exposure. Using a counting process martingale approach, Altshuler (Nelson-Aalen) estimates have been shown

to have asymptotically robust properties when censoring is nominal. In this talk, we explore an alternative counting process method that aims to increase the asymptotic efficiency of Altshulers survival estimates. This technique involves using a generalized (2-parameter) Weibull model to identify censor-to-failure transition jump points that can then be incorporated into the Altshuler estimate. Given a set of covariates, we show how to estimate the Hessian and Fisher information function corresponding to the Weibull model. We further provide an example demonstrating how the maximum-likelihood parameter estimates from this model can be used in a localized iterative fashion to identify asymptotically efficient transition jump points. Key Words: Risk exposure, event-time models, generalized Weibull distribution, censoring

7. Bayesian Survival Analysis

- **Dooti Roy**, University of Connecticut
  "Analysis of Multivariate Survival Data Based on Vine Copulas"
  Dooti Roy, Vivekananda Roy, Dipak Dey

  Our paper introduces a novel copula based methodology to analyze right censored multivariate survival data. In practice, implementation of existing methodologies for analyzing multivariate survival data often leads to challenges with respect to evaluation of the likelihood and other computational issues. Using a vine copula structure, we propose a computationally tractable Bayesian modeling approach for the analysis of multivariate survival data. We apply our method on two multivariate data sets.

- **Sung Duk Kim**, National Cancer Institute
  "A Joint Model Approach for Longitudinal Data with No Time Zero and Time-to-Event with a Competing Risk"
  Sungduk Kim, Olive D. Buhule, and Paul S. Albert

  Station is a digitalized measure of how low the fetus head is positioned in the pelvis of a pregnant woman. It is measured from -3 to +4, where a value of -3 implies a fetus is still very high in the pelvis and not close to delivery, while +4 implies the fetus is below the pelvis and is due for delivery. The fetus is delivered vaginally (spontaneous or vacuum) or through a C-section. It is of interest to predict the timing and delivery type using individualized longitudinal assessments of station. Importantly, women enter the hospital at different station measurements, resulting in no clear time zero for use as a reference point for valid statistical inferences and predictions. We develop a shared random parameter model that links together model components for the longitudinal station process (with no time-zero) and both the time and type of delivery. Specifically, each model component includes random effects that are shared between these three components, inducing realistic dependence between these three data elements. The goal in constructing this model is to develop an adaptive predictor to predict both the timing and type of delivery based on repeated station values. Markov chain Monte Carlo sampling is used to carry out Bayesian posterior computation. The approach is illustrated using a longitudinal cohort of digitized station measurements and the timing and type of delivery in an international cohort.

- ,
  "",

- **Kyu Ha Lee**, The Forsyth Institute
  "Accelerated Failure Time Models for Semi-Competing Risks Data in the Presence of Complex Censoring"
  Kyu Ha Lee, Virginie Rondeau, Sebastien Haneuse

  Statistical analyses that investigate risk factors for Alzheimer's disease (AD) are often subject to a number of challenges. Some of these challenges arise due to practical considerations regarding data collection such that the observation of AD events is subject to complex censoring including left-truncation and either interval or right-censoring. Additional challenges arise due to the fact that study participants

under investigation are often subject to competing forces, most notably death, that may not be independent of AD. Towards resolving the latter, researchers may choose to embed the study of AD within the "semi-competing risks" framework for which the recent statistical literature has seen a number of advances including for the so-called illness-death model. To the best of our knowledge, however, the semi-competing risks literature has not fully considered analyses in contexts with complex censoring, as in studies of AD. This is particularly the case when interest lies with the accelerated failure time (AFT) model, an alternative to the traditional multiplicative Cox model that places emphasis away from the hazard function. In this work we outline a new Bayesian framework for estimation/inference of an AFT illness-death model for semi-competing risks data subject to complex censoring. An efficient computational algorithm that gives researchers the flexibility to adopt either a fully parametric or a semi-parametric model specification is developed and implemented. The proposed methods are motivated by and illustrated with an analysis of data from the Adult Changes in Thought study, an on-going community-based prospective study of incident AD in western Washington State.

8. Causal Inference with Lifetime Data

- **Kjetil rysland**, University of Oslo
  "Causal Local Independence Models"
  Kjetil Rysland, Vanessa Didelez, Pl Christie Ryalen, Mats Julius Stensrud

  Survival analysis has become one of the fundamental fields of biostatistics. Such analyses are almost always subject to censoring. This necessitates special statistical techniques and forces statisticians to think more in terms of stochastic processes. The theory of stochastic integrals and martingales have therefore been important for the development of such techniques. Causal inference has lately had a huge impact on how statistical analyses based on non-experimental data are done. The idea is to use data from a non-experimental scenario that could be subject to several spurious effects and then fit a model that would govern the frequencies we would have seen in a related hypothetical scenario where the spurious effects are eliminated. This opens up for using health registries to answer new and more ambitious questions. However, there has not been so much focus on causal inference based time-to-event data or survival analysis. The now well established theory of causal Bayesian networks is for instance not suitable for handling such processes. Motivated by causal inference event-history data from the health registries, we have introduced causal local independence models. We show that they offer a generalization of causal Bayesian networks that also enables us to carry out causal inference based on non-experimental data when there is continuous-time processes involved. The main purpose of this work in collaboration with Vanessa Didelez, is to provide new tools for determining identifiability of causal effects of event history data that is subject to censoring. It builds on previous work on local independence graphs and delta-separation by Vanessa Didelez and previous work on causal inference for counting processes by Kjetil Rysland. We provide a new result that gives quite general graphical criteria for when causal validity of a local independence model is preserved in sub-models. If the observable variables, or processes, form a causally valid sub-model, then we can identify most relevant causal effects by re-weighting the actual observations. This is used to prove that the continuous time marginal structural models for event history analysis, based on martingale dynamics, are valid in a much more general context than what has been known previously.

- **Aksel Karl Georg Jensen**, University of Copenhagen
  "A Marginal Structural Model for Recurrent Events in the Presence of Time-Dependent Confounding: Non-Specific Effects of Vaccines on Child Hospitalisations"
  Aksel KG Jensen, Henrik Ravn, Signe Srup, Per K Andersen

  Using a Danish register-based study on childhood vaccination and hospitalisation as motivation, a marginal structural model for recurrent events is studied. The model addresses a number of challenges which may be seen more generally in large register-based cohort studies. One is to adjust for a time-dependent confounder when studying the effect of a time-varying exposure on a recurrent event based on an analysis in continuous time. Another is to report results via a measure which is easy to interpret and communicate even though quite elaborate treatment regimes are considered. Lastly,

the implementation of continuously updated weights implies a substantial computationally demanding workload.

9. Survival Analysis in Psychiatry and Neurological Disorders

- **Ralitza Gueorguieva**, Yale University
  "Joint Modeling of Symptom Severity and Competing Risk Dropout in Psychiatry"
  Ralitza Gueorguieva, Robert Rosenheck, Haiqun Lin

  Dropout in longitudinal studies occurs for different reasons: inefficacy (likely informative), side effects (likely missing at random), outside circumstances (likely missing completely at random). Using information on dropout improves inferences and provides better understanding of the association between dropout and outcome processes. We use a fully parametric approach for joint analysis of repeatedly measured outcome and competing risk dropout: a linear mixed model component for the repeated measures outcome, cause-specific dropout processes that allow testing of different hazard alternatives for each dropout reason, association between the longitudinal series and the competing risks via shared random effects. In this approach both time-independent and time-dependent covariates are allowed, the model is well suited for interval-censored data, parameter estimates for some of the special-case models have intuitive interpretation and SAS PROC NLMIXED can be used for estimation. We illustrate the approach on data from the CATIE clinical trial in schizophrenia. Compared to the original analyses of the CATIE data, our approach improves power for detecting treatment differences, simultaneously tests the effects of treatments on the repeated measures and survival outcomes and allows quantitative assessment of the strength of the relationship between dropout due to various reasons and the efficacy outcome. We show via a simulation study that our method reduces bias in estimation of treatment effects when dropout is informative.

- **Mengjie Zheng** , Indiana University School of Medicine
  "Phd Student "
  Mengjie Zheng, Sujuan Gao

  The joint modeling framework for longitudinal and time-to-event data has been introduced to study the association between repeatedly measured exposures and the risk of an event. The motivating idea is to model the survival outcome using the repeated longitudinal measures while accounting for special features of the longitudinal data. Existing estimation methods include the two-stage approach, Bayesian and maximum likelihood estimation methods (MLEs). The two-stage method is computationally straightforward but introduces biases, while the Bayesian and MLEs rely on the joint likelihood of longitudinal and survival processes and can be computationally intensive. In this paper, we propose a joint generalized estimating equation framework using an inverse intensity weighting approach to correct the biases from a two-stage method. The proposed method is capable of handling longitudinal outcomes from the exponential family of distributions and is computationally efficient to carry out. The performance of the proposed method is assessed through simulation studies. The proposed method is applied to data from a longitudinal cohort to determine the association of longitudinal low-density lipoprotein (LDL), high-density lipoprotein (HDL) measures and the risk of coronary artery disease (CAD).

- **Jeffrey D. Long**, University of Iowa
  "Random Survival Forests for Exploratory Analysis in Neurological Disorders"
  Jeffrey D. Long

  Random survival forest (RSF) is an ensemble tree method for the analysis of right-censored survival data. The advantages of RSF are that it does not require restrictive modeling assumption (e.g., proportional hazards), it can automatically assess nonlinear and complex interaction effects among the predictors, it has a built-in model validation capability, and it is resistant to over-fitting. The algorithm grows a survival tree on a bootstrap sample of the data, and at each node, splitting is determined by maximizing the log-rank statistic based on a random sample of predictors. Splitting stops when the terminal node has a pre-specified proportion of events. A forest of trees is grown (e.g., 2000), error

and importance measures from all trees are aggregated, and an ensemble cumulative hazard estimate is computed. The worth of predictors can be determined from the discrepancy in the prediction error of a variable with and without randomly assigning daughter nodes (greater discrepancy = greater worth). Worth can also be determined by how deep in the trees the nodes tend to be for which a predictor is important in splitting (shallower depth = greater worth). An application of RSF to Huntingtons disease (HD) is presented. A set of 32 predictors measured at baseline of a large observational study were used to predict time to HD diagnosis. Subgroups of predictors were compared using cross-validation prediction accuracy evaluated with the time-dependent Brier score. Results show that a subgroup of clinical, genetic, and exposure variables provide good prediction accuracy and constitute an improvement over the traditional set of predictors.

- **Rebecca Betensky**, Harvard T.h. Chan School of Public Health
  "Time-to-Event Data with Time-Varying Biomarkers Measured Only at Study Entry, with Applications to Alzheimer's Disease"
  Catherine Lee, Rebecca A. Betensky

  Relating time-varying biomarkers of Alzheimer's disease (AD) to time-to-event using a Cox model is complicated by the fact that AD biomarkers are sparsely collected, typically only at study entry; this is problematic since Cox regression with time-varying covariates requires observation of the covariate process at all failure times. The analysis might be simplified by using study entry as the time origin and treating the time-varying covariate measured at study entry as a fixed baseline covariate. In this paper, we first derive conditions under which using an incorrect time origin of study entry results in consistent estimation of regression parameters when the time-varying covariate is continuous and fully observed. We then derive conditions under which treating the time-varying covariate as fixed at study entry results in consistent estimation. We provide methods for estimating the regression parameter when a functional form can be assumed for the time-varying biomarker, which is measured only at study entry. We demonstrate our analytical results in a simulation study and apply our methods to data from the Rush Religious Orders Study and Memory and Aging Project, and data from the Alzheimer's Disease Neuroimaging Initiative.

Friday 1:45–3:30

10. Joint Modeling with a View Towards Risk Predictions

- **Sheng Luo**, The University of Texas Health Science Center at Houston
  "Dynamic Prediction of Alzheimers Disease Progression with Longitudinal Functional Joint Model"
  Kan Li, Sheng Luo

  The current joint models involve only scalar variables. Functional exposures are commonly measured longitudinally. We propose a functional joint model (FJM) that consists of a longitudinal regression model with longitudinal functional exposure (high dimensional MRI) and a survival model for event time. We also develop methods for model-based personalized dynamic predictions of future outcome trajectories and risks of target events. Our proposed model is motivated and applied to the Alzheimer's Disease Neuroimaging Initiative (ADNI), a motivating clinical study to determine the relationships among the clinical, cognitive, imaging, genetic and biochemical biomarkers characteristics of the entire spectrum of Alzheimers disease (AD).

- **Danping Liu**, National Institutes of Health
  "A Joint Modelling Approach for Informative Cluster Size and Gap Time in Longitudinal Data with Application to a Repeated Pregnancy Study"
  Joe Bible, Danping Liu, Paul S. Albert

  Transition models are useful in estimating the probability of recurrent events in longitudinal studies. However, direct application of a transition model may suffer from two complications, informative cluster size and informative gap time between observations. For example, Consecutive Pregnancy Study

(CPS)is a retrospective cohort study aiming at understanding the recurrence patterns and predictors of adverse pregnancy outcomes, such as preterm birth. The number of pregnancies observed and the gap time may be both indicative of a womens underlying fertility, and hence correlated with the pregnancy outcomes. We propose a shared random effect structure for jointly modelling the transition model with the informative observation process. The gap time is modelled by a parametric distribution with right censoring; the cluster size is characterized by a continuation ratio model. We also investigated the estimation and interpretation of two transition probabilities: one adjusted for gap time and the other marginalized over gap time. Through extensive simulation studies and analyses of the CPS data, we show that naive approaches ignoring the cluster size model, the gap time model, or both, could lead to seriously biased inference.

11. Risk Prediction Models and Application

- **Rajeshwari Sundaram**, National Institutes of Health
  "Prediction of Infertility Based on Biology and Behavior using a Joint Modeling Approach"
  Rajeshwari Sundaram, Sedigheh Mirzaei, Germaine Buck Louis

  Human fecundity, defined as the biologic capacity of men and women for reproduction irrespective of pregnancy intentions, is of considerable public health interest, given growing evidence supporting its worldwide decline. As clinical and public health groups work toward the implementation of preconception clinical guidance for couples at risk for pregnancy, it is imperative to understand determinants of fecundity and fertility. To accomplish this goal, statistical models that incorporate biological features of both members of the couple are needed. This is challenging since complex multivariate longitudinal modeling techniques need to be developed to adequately reflect the behavior and biology. In this talk, I will discuss methods that allow us to assess some easily available longitudinal processes assessing behavior and menstrual cycle characteristics and show how they can be used to predict time-to-pregnancy dynamically and consequently infertility. All the approaches will be illustrated on data arising through various prospective pregnancy study designs.

12. Risk Analysis in the Biomedical and Environmental Field

- **Olivier Bouaziz**, University Paris Descartes, Laboratory Map5
  "A Change-Point Model for Detecting Heterogeneity in Ordered Survival Responses"
  Olivier Bouaziz, Grgory Nuel

  In survival analysis it is quite common that heterogeneity between patients results in various survival response distributions. This heterogeneity can be controlled through known covariates (such as date of birth, age at diagnosis, gender, treatment, co-exposure, BMI, etc.) using regression-type models such as the Cox proportional hazard model and by performing stratified analyses or by incorporating a random effect in a frailty model. Other types of heterogeneous dataset arise when the incidence rate changes over the calendar time in a cohort study and specific models like age-period-cohort have been extensively studied to take into account this kind of heterogeneity. While theses models have proved to be most useful, it is however likely that unaccounted latent heterogeneity remains in the survival signal. This might be due for example to an unknown interaction between a treatment and some exposure, or to some unaccounted heterogeneity of the disease itself (for example an unknown cancer sub-type). For instance, age at diagnosis might be associated with a higher chance to receive a new treatment or BMI might be associated with a specific exposure. In the present work, we suggest a new approach considering survival heterogeneity as a breakpoint model in an ordered sequence of survival responses. The survival responses might be ordered according to any numerical covariate (ties are possible) like age at diagnosis, BMI, etc. The basic idea being that heterogeneity will be detected as soon as it is associated with the chosen covariate. In such a model, we aim at two objectives: first we want to estimate the hazard rates and the proportional factors in each homogenous region through a Cox model. Secondly, we want to accurately provide the number and location of the breakpoints. Recently a constrained Hidden Markov Model (HMM) method was suggested in the context of breakpoint

analysis (see Luong et al, 2013). This method allows to perform a full change-point analysis in a segment-based model (one parameter by segment) providing linear EM estimates of the parameter and a full specification of the posterior distribution of change points. In this talk we adapt this method to the context of survival analysis, where the estimation is performed through the EM algorithm to provide update of the hazard rate estimates and the posterior distribution at each iteration step. The method will be illustrated on the dataset on diabetic patients from the Steno Memorial hospital in Copenhagen (dataset from Andersen et al., 1993), where the event times are ordered with respect to the calendar time of disease onset. On this dataset, the years of disease onset of the patients range from 1933 to 1972. A two breakpoint model is found from our method and survival functions and hazard ratios are estimated on each three segment. Our results clearly indicate a general medical improvement over time for Danish diabetic patients.

- **Min-Ge Xie**, Rutgers University-New Brunswick
  "Individualized Fusion Learning (i-Fusion) for Individualized Inference"
  Minge Xie*, Jieli Shen and Regina Liu

  Learnings from different data sources can often be fused together to yield more powerful findings than those from individual sources alone. This talk presents a new fusion learning approach, named i-Fusion, for drawing efficient individual inference by fusing leanings from relevant data sources. This i-Fusion is robust for handling heterogeneity arising from diverse sources in big data, and is ideally suited for goal-directed applications such as precision medicine. Specifically, i-Fusion summarizes individual inference information in confidence distributions (CDs), then adaptively forms a clique of individuals that bear relevance to the target individual, and finally combines the CDs from those relevant individuals and draws inference for the target individual from the combined CD. In essence, i-Fusion borrows strength from relevant individuals to improve efficiency while retaining inference validity. Computationally, i-Fusion is parallel in nature and scales up well in comparison with existing competitors. Examples in simulations and real applications in financial forecasting are presented.

13. Recent Development in the Analysis of Complex Structured Survival Data

- **Wenqing He**, University of Western Ontario
  "Analysis of Multivariate Survival Data under Semiparametric Copula Models with/without Measurement Error"
  Wenqing He, Grace Y. Yi and Naisyin Wang

  There has been extensive research on univariate survival data with covariate measurement error. However, there is limited attention on the impact of covariate measurement error on analysis of multivariate survival data. In this paper, we discuss semiparametric linear transformation marginal models for multivariate survival data, with covariate measurement error incorporated. The local linear kernel method is employed to approximate the marginal linear transformation model, and a two stage maximum likelihood method and a working three stage estimation are developed. The impact of misspecification of the joint copula model on estimation of covariate effects is investigated, and a simulation-based method is explored to correct for measurement error effect on parameter estimation. Extensive simulation studies are conducted to assess the performance of the proposed methods for a variety of scenarios, and the Busselton Health study is analyzed with the proposed methods.

- **Sundar Subramanian**, New Jersey Institute of Technology
  "Function-Based Hypothesis Testing via Plug-in Empirical Likelihood in Censored Location-Scale Families"
  Sundar Subramanian

  The empirical characteristic function as a tool to address function-based hypothesis testing for location-scale models works well for uncensored data. For censored location-scale models, when two independent samples are each subjected to independent right censoring, a formal test of adequacy is lacking. A plug-in empirical likelihood combined with a large-sample analysis leads to a test which, however, is not asymptotically distribution free. Hence for practical situations bootstrap is necessary for performing the

test. A multiplier bootstrap is employed to compute requisite critical values from the null distribution. Although minimum distance estimators of the location and scale are employed for the plug-in, the empirical likelihood method allows any consistent estimators. Some numerical results are reported.

14. Outcome-Dependent Sampling in the Survival Analysis Context

- **Qian (Michelle) Zhou**, Mississippi State University
  "Assessing Incremental Value of Biomarkers with Multi-Phase Nested Case-Control Studies"
  Qian M. Zhou, Yingye Zheng, Lori B. Chibnik, Elizabeth W. Karlson, and Tianxi Cai

  In the Nurses Health Study, to examine the effects of several biomarkers and genetic markers on the risk of rheumatoid arthritis (RA), a three-phase nested case-control (NCC) design was conducted, in which two sequential NCC subcohorts were formed with one nested within the other, and one set of new markers measured on each of the subcohorts. One objective of the study is to evaluate clinical values of novel biomarkers in improving upon existing risk models because of potential cost associated with assaying biomarkers. In this paper, we develop robust statistical procedures for constructing risk prediction models for RA and estimating the incremental value of new markers based on three-phase NCC studies. Our method also takes into account possible time-varying effects of biomarkers in risk modeling, which allows us to more robustly assess the biomarker utility and address the question of whether a marker is better suited for short-term or long-term risk prediction.

15. Recent Developments in High-Dimensional Survival Analysis and Biased Sampling

- **Yi Li**, U of Michigan
  "Integrated Powered Density (Ipod): Screening Ultrahigh-Dimensional Covariates with Survival Outcomes"
  Yi Li

  Modern biomedical studies have yielded abundant survival data with high-throughput predictors. Variable screening is a crucial first step in analyzing of such data, for the purpose of identifying predictive biomarkers, understanding biological mechanisms and making parsimonious predictions. To nonparametrically quantify the relevance of each candidate variable to the survival outcome, we propose integrated powered density (IPOD), which compares the differences in the covariate-stratified distribution functions. This proposed new class of statistics, with a flexible weighting scheme, is general and includes the Kolmogorov statistic as a special case. Moreover, the method does not rely on rigid regression model assumptions and can be easily implemented. We show that our method possesses sure screening properties, and confirm the utility of the proposal with extensive simulation studies. We apply the method to analyze a multiple myeloma study on detecting gene signatures for cancer patients' survival.

- **Ian Mckeague**, Columbia University in the City of New York
  "Nonparametric Ordering of Survival Functions and Non-Inferiority Testing"
  Hsin-wen Chang and Ian W. McKeague

  New tests for the nonparametric ordering of multiple survival functions are developed with the possibility of right censorship taken into account. The motivation comes from non-inferiority trials with multiple treatments. The proposed tests are based on nonparametric likelihood ratio statistics, which are known to provide more powerful tests than Wald-type procedures, but in this setting have only been studied for pairs of survival functions or in the absence of censoring. We introduce a novel type of pool adjacent violator algorithm that leads to a complete solution of the problem. The limit distributions can be expressed as weighted sums of squares involving projections of certain Gaussian processes onto the given ordered alternative. A simulation study shows that the new procedures have superior power to a competing Wald-type Cox model approach, even under proportional hazards. We illustrate the proposed methods using data from a three-arm non-inferiority trial.

- **Donglin Zeng**, University of North Carolina
  "Time-Varying Coefficient Hazards Model with Sparsely Measured High Dimensional Biomarkers"
  Yuanjia Wang, Xiang Li, Quefeng Li, Donglin Zeng

  Clinical studies with time-to-event outcomes often collect measurements of a large number of dynamic covariates over time (e.g., clinical assessments or neuroimaging biomarkers) to build time-sensitive prognostic model. An emerging challenge is that due to resource intensive or invasive (e.g., lumbar puncture) data collection process, biomarkers may be measured infrequently and thus not available at every observed event time point. We propose a kernel-smoothing based approach to borrow information across subjects to remedy infrequent and unbalanced biomarker measurements under a time-varying hazards model. A penalized pseudo-likelihood function is proposed for estimation, and an efficient augmented penalization minimization algorithm related to the alternating direction method of multipliers (ADMM) is adopted for computation. Under some regularity conditions, we show that even in the presence of ultra-high dimensionality, the proposed method selects important biomarkers with high probability. Through extensive simulation studies, we demonstrate superior performance in terms of estimation and selection performance compared to alternative methods. Finally, we apply the proposed method to analyze a recently completed real world study to model time to disease conversion using longitudinal, whole brain structural magnetic resonance imaging (MRI) biomarkers.

16. Advances in Summarizing and Modeling Complex Survival Data

- **Chenxi Li**, Michigan State University
  "A Semiparametric Multi-State Model for Correlated Interval Censored Life-History Data in Caries Research"
  Daewoo Pak, Chenxi Li, David Todem

  We propose a semiparametric multi-state Markov frailty model for interval-censored life-history data. The proposed model is motivated by an attempt to describe the life course of dental caries at the tooth level, taking into account the multiplicity of the disease states and the intra-oral clustering of observations made only at periodic time points. In particular, the model is intended to investigate the intra-oral symmetries for caries formation at the quadrant level, and whether any of the symmetries vary with gender. We assume a proportional hazards form for the transition intensities conditional on a subject-level frailty in the model and approximate the log baseline intensities by linear splines. The estimation of the model is conducted using a penalized likelihood where the smoothing parameters are estimated as reciprocal variance components under a mixed-model representation. We develop a Bayesian method for predicting tooth-level caries transition probabilities, which can be used for tailoring tooth-level treatment plans. Intensive simulation studies indicate that the model-fitting and prediction methods perform reasonably well in samples with realistic sizes. The practical utility of the methods is illustrated using data generated from a unique longitudinal study on oral health among children from low-income families residing in the city of Detroit, Michigan.

- **David Oakes**, University of Rochester
  "Matched Pair Survival Data - a New Look at an Old Problem "
  David Oakes

  Kartsonaki and Cox (Biometrika, 103, pp. 219-224, 2016) recently gave a succinct account including some new approaches to the analysis of matched pair survival data in the absence of censoring. We discuss possible modifications of their methods to matched pair survival data subject to censoring

- **Shuling Liu**, Yale University
  "Statistician Ii"
  Shuling Liu, Amita Manatunga, Limin Peng, Michele Marcus

  In many biomedical studies that involve correlated data, an outcome is often repeatedly measured for each individual subject along with the number of these measurements, which is also treated as an observed outcome. This type of data has been referred as multivariate random length data by

Barnhart and Sampson (1995). A common approach to handling such type of data is to jointly model the multiple measurements and the random length. In previous literature, a key assumption is the multivariate normality for the multiple measurements. Motivated by a reproductive study, we propose a new copula-based joint model which relaxes the normality assumption. Specifically, we adopt the ClaytonOakes model for multiple measurements with flexible marginal distributions specified as semi-parametric transformation models. The random length is modeled via a generalized linear model. We develop an approximate EM algorithm to derive parameter estimators and standard errors of the estimators are obtained through bootstrapping procedures and the finite-sample performance of the proposed method is investigated using simulation studies. We apply our method to the Mount Sinai Study of Women Office Workers (MSSWOW), where women were prospectively followed for 1 year for studying fertility.

17. Semiparametric Statistical Methods for Complex Failure Time Data

- **Yichuan Zhao**, Georgia State University
  " Empirical Likelihood Inference for the Odds Ratio of Two Survival Functions under Right Censoring"
  Meng Zhao, Yichuan Zhao, Ian McKeague

  This paper develops a new method for estimating survival odds ratios under right censoring using the censored empirical likelihood. Results of a simulation study show that under small sample sizes the proposed method outperforms the corresponding Wald-type CI and the log-transformed Wald-type CI.

18. Recent Developments in Statistical and Computational Methods for Biomedical Data

- **Chung-Chou H. Chang**, University of Pittsburgh-Pittsburgh Campus
  "Development of Model-Based Surrogate Endpoint for Sepsis Studies"
  Chung-Chou H. Chang, Zhongying Xu, Victor Talisa, Hernando Gomez

  Sepsis is a high risk and life-threatening syndrome caused by the bodys overwhelming inflammatory response to infection. Traditionally, sepsis has been considered as one syndrome with clinical presentations varying only by severity. However, recent data has challenged this paradigm, and has suggested that sepsis probably encompasses multiple phenotypes. In this study, we use several unsupervised clustering methods to explore diverse patterns of sepsis in a cohort of 3,267 patients admitted to the ICU with suspected sepsis. To assess the clinical relevance of the resulting clusters we analyzed the association of each cluster with survival outcomes including ICU, hospital, 90-day, and 1-year mortality.

Friday 3:45–5:30

19. from Functional to Neuroimaging Data

- **Hans-Georg mller**, University of California, Davis
  "Dynamic Modeling of Longitudinal Snippets"
  Matthew Dawson, Hans-Georg Mller

  Longitudinal data are often plagued with sparsity of time points where measurements are available. The functional data analysis perspective provides an effective and flexible approach when measurements are sparse but their times are randomly distributed over an interval. A different kind of sparsity occurs when one has longitudinal snippets, which are very short stretches of longitudinal measurements. For each subject the stretch of available data is much shorter than the time frame of interest. An added challenge is introduced if a meaningful time proxy such as time since disease onset is not available. We approach this problem through conditional quantile trajectories for monotonic processes that arise as solutions of a dynamic system and discuss an application to shrinking brain volumes in Alzheimers patients. This talk is based on joint work with Matt Dawson, UC Davis.

20. Recent Advances on Statistical Methods for Health Studies

- **Shijun Zhu**, University of Maryland Baltimore
  "Joint Modeling of Multivariate Longitudinal Data and Recurrent Events and It's Application"
  Shijun Zhu, Erika Friedmann, DoHwan Park

  In the longitudinal study the outcomes are often measured at the routine clinical visits along with the episode of endpoints events. An example of such study is the ongoing longitudinal study of urea cycle disorders (UCD) which was initiated in 2006 among 14 sites in the US, Europe and Canada, collecting data on the outcomes at the routine clinical visits along with the episode of endpoints events (Batshaw et al, 2014, Lee et al, 2015). The glutamine and ammonia levels were recorded in each clinical visit and the acute hyperammonemic episodes in UCDs recorded in the patients hospitalizations. Analysis of the longitudinal study where two or more outcomes correlated with recurrent events often involves treating the longitudinal outcomes as covariates in the time-to-event models. In contrast, joint modeling of recurrent events with longitudinal data assessed changes of these outcomes over time and time-to-recurrent events in the same model. It could provide greater insight of the relationships of various dimensions of the measurements. Patient-level modeling of the recurrent events jointly with the trajectory data would also inform the event antecedents therefore allowing characterization of factors associated with event recurring. Joint modeling with random effects by Zeng and Lin (2007) has been extended to propose a joint modeling of multivariate longitudinal data with recurring time. Multivariate normal random variables will be included to account for the heterogeneity among the repeated measures and among the recurrent events within each participant. Nonparametric maximum likelihood estimation (NPMLE) and expectation-maximization (EM) algorithm are used to obtain the estimates and their variances along with the recursive for the jumps of the cumulative intensity. The simulation studies showed that the proposed method performs fairly well and the NPMLEs are shown to be consistent, asymptotically normal, and efficient. The method is then applied to analyze the urea cycle disorder patients' hyperammonemia episodes. Reference: 1. Batshaw M.L., Tuchman M., Summar M., et al., 2014, A longitudinal study of urea cycle disorders, Molecular Genetics and Metabolism, 113, 127-130 2. Lee B., Diaz G.A., Rhead W., et al., 2015, Blood ammonia and glutamine as predictors of hyperammonemic crises in patients with urea cycle disorder, Genetic in Medicine, 17, 561-568 3. Zeng D. and Lin D.Y., 2007, Maximum likelihood estimation in semiparametric regression models with censored data, J. R. Stat. Assoc. Ser. B, 69, 507-564

21. Joint Modeling and Weighted Estimation for Survival Data Analysis

- **Ying Yan**, University of Calgary
  "Optimally Weighted Estimation in Case-Cohort Studies"
  Ying Yan

  The case-cohort study design is used to reduce financial cost of assembling and measuring expensive covariates in large cohort studies. Various weighted estimators have been proposed, where the known weights are commonly estimated by available information. These existing weighted estimators usually do not effectively make use of the information available in the whole cohort. In this talk, we propose an optimally weighted estimator. The known weights are not required to be estimated. We show that the proposed estimator is optimal among the existing weighted estimators. The advantage of the proposed estimator is demonstrated in numerical studies.

- **Liqun Diao**, University of Waterloo
  "Copula-Based Models for Recurrent Exacerbations"
  Liqun Diao, Richard Cook and Leilei Zeng

  In some chronic disease settings, processes are not progressive but rather exhibit a continual risk of periodic episodic conditions. Examples include asthma, infectious disease, and psychotic disorders. For each of these examples, the recurrent events have non-ignorable durations associated with them and are better characterized as recurrent exacerbations. We assume that the dependence structure

of successive gap times is induced by a copula-based model. We propose inverse probability weighted composite likelihood to correct biases due to induced dependent censoring, which builds a bridge to the standard practice in the context of copula modeling when complete data are available.

22. New Approaches for Analyzing Time to Event Data with Application in Cancer Studies

- **Masood Anwar**, Comsats Institute of Information Technology, Islamabad, Pakistan
  "The Half-Logistic Lomax Distribution for Lifetime Modeling"
  Masood Anwar, Javeria Zahoor

  In this paper, we introduce a new two-parameter lifetime distribution called the half-logistic Lomax (HLL) distribution. The proposed distribution is obtained by compounding half-logistic and Lomax distributions. We obtain some mathematical properties of the proposed distribution such as the survival and hazard rate function, quantile function, mode, median, moments and moment generating functions, mean deviations from mean and median, mean residual life function, order statistics and entropies. The estimation of parameters is performed by maximum likelihood and provide formulas for the elements of the Fisher information matrix. A simulation study is provided to access the performance of maximum likelihood estimators (MLEs). The flexibility and potentiality of the proposed model is illustrated by means of a real data set.

- **Arthur Berg**, Penn State Hershey
  "Reduced Bias Nonparametric Censored Density and Hazard Estimation"
  Arthur Berg, Dimitris Politis, Kagba Suaray, Hui Zeng

  Kernel-based nonparametric hazard rate estimation is considered with a special class of infinite-order kernels that achieves favorable bias and mean square error properties. A fully automatic and adaptive implementation of a density and hazard rate estimator is proposed for randomly right censored data. Careful selection of the bandwidth in the proposed estimators yields estimates that are more efficient in terms of overall mean squared error performance, and in some cases achieves a nearly parametric convergence rate. Additionally, rapidly converging bandwidth estimates are presented for use in second-order kernels to supplement such kernel-based methods in hazard rate estimation. Simulations illustrate the improved accuracy of the proposed estimator against other nonparametric estimators of the density and hazard function. A real data application is also presented on survival data from 13,166 breast carcinoma patients.

23. Quality of Life and Other Applications

- **Catherine Huber-Carol**, University of Paris Descartes (Paris, France)
  "Survival and Quality of Life"
  Catherine Huber-Carol

  The aim of this talk is to show how one can deal with the censoring and truncation inconveniences, in both case of only one terminal event and also when several different outcomes may happen. But, as the aim of survival analysis is not only to try to increase the length of the pure survival time (to death) but also the survival time "free of disease", this leads to estimate the number of "years free of disease" lost due to environmental or behavioural factors for the sake of prevention. For that purpose, the FHT model (First Hitting Time model) is very useful.

- **Weiliang Qiu**, Harvard Medical School
  "Association of Pre-Diagnosis Bmi Measurements to Prostate Cancer Mortality"
  Meng Yang, Weiliang Qiu*, Haiyan Zhang, Brandon Guo, Changzhen Yuan, Jorge E Chavarro, Jing Ma

  Prostate cancer is the most common cancer, besides skin cancer, in men in the United States. It is the second leading cause of death from cancer in men. Prostate cancer often has no early symptoms and usually grows very slowly. While some types of prostate cancer grow slowly and may need minimal

or no treatment, other types are aggressive and can spread quickly. Routine screening with either a DRE or PSA is not supported by the evidence as there is no mortality benefit from screening. In some people, screening may lead to over-treatment, hence unnecessary disruption and possibly harmful consequences. Hence, it is imperative to find effective risk factors of mortality before prostate cancer diagnosis. It has been reported that pre-diagnostic BMI is positively associated to prostate cancer mortality. However, it is unclear yet if the change of pre-diagnostic BMI is associated to prostate cancer mortality or not. This study aims to evaluate the association of longitudinal pre-diagnostic BMI measurements to prostate cancer mortality by using the Physicians Health Study (PHS), in which BMI for 3581 subjects, who were diagnosed with prostate cancer during the follow-up, were measured at years 1982, 1990, 1991, 1992, 1993, 1994, 1995. In this talk, we will report the results of our data analysis.

24. Joint Modelling of Longitudinal Measurements and Event History Data

- **Guohua Yan**, University of New Brunswick, Canada
  "A Flexible Approach for Multivariate Mixed-Effects Models in the Presence of Non-Ignorable Missingness and Measurement Error"
  Juxin Liu, Wei Liu, Lang Wu and Guohua Yan

  We propose a flexible model approach for the distribution of random effects when both response variables and covariates have non-ignorable missing values in a longitudinal study. A Bayesian approach is developed with a choice of nonparametric prior for the distribution of random effects. We apply the proposed method to a real data example from a national long-term survey by Statistics Canada. We also design simulation studies to further check the performance of the proposed approach. The result of simulation studies indicates that the proposed approach outperforms the conventional approach with normality assumption when the heterogeneity in random effects distribution is salient.

25. Time-to-Event Models for Human Health Risk Assessment

- **Polyna Khudyakov**, Harvard T.h. Chan School of Public Heal
  "Survival Analysis with Measurement Error in a Cumulative Exposure Variable: Radon Progeny in Relation to Lung Cancer Mortality"
  Polyna Khudyakov, Jonathan Samet, Charles Wiggins, Molin Wang, Xiaomei Liao, Angela Meisner, and Donna Spiegelman

  Exposure variables in occupational and environmental epidemiology are usually measured with error, typically flattening estimated exposure-response relationships. In this work, we extend the risk set regression calibration (RRC) method for Cox models of cumulative exposure variables to obtain consistent point and interval estimates of relative risks corrected for exposure measurement error. We show that the RRC methodology originally developed for use with an external validation study can be generalized to internal validation study designs as well. We then analyze the New Mexico uranium miners study with follow-up extended from 1957 to 2012. The exposure data were collected using several different methods of ascertainment, some of which had a substantial amount of error. We compared results from the standard analysis of the effect of cumulative radon progeny exposure on lung cancer mortality to measurement error corrected results in a subset of 2,337 miners observed during the time period when exposure was measured either by work area samples or personal exposure estimates provided by the mining companies. The correlation between these two methods of measurements was 0.33. After adjusting for bias due to exposure measurement error, the multivariate-adjusted hazard ratio for lung cancer mortality in relation to cumulative radon exposure (100 WLM) was estimated to be 4.69 (95

26. Industry Perspectives on Lifetime Data Science

- **Wayne B. Nelson**, Consultant
  "Get more Information from Recurrent Events Data: Product Repairs, Disease Recurrences, and Other Applications"
  Wayne B. Nelson

  Recurrent events data arise in biomedical, engineering, sociological, business, and many other applications. Despite over 25 years of literature articles and books on analysis of such data, some current medical studies of recurrent diseases still use only each patient's first recurrence time and ignore subsequent ones, thus losing information. This article provides an introduction to modern analysis of recurrent events data and gives new results on how much added information and accuracy are gained by the use of all recurrences.

- **Aparna Anderson**, Statistics Collaborative, Inc.
  "Co-Primary Endpoints: Scientific and Regulatory Points to Consider"
  Aparna Anderson

  A universal challenge in drug development is optimizing speed to market while at the same time taking a measured approach to understand a drugs toxicity profile and produce robust efficacy results. In oncology development, where phase 1 studies are conducted in patients rather than healthy volunteers, stand-alone phase 2 studies are being abandoned in favor of phase 1 dose expansion cohorts to establish proof of concept. And when development progresses to phase 3 registrational studies, it is not uncommon to see trial designs with co-primary endpoints such as progression-free survival, an early (surrogate) endpoint to support accelerated approval, alongside an overall survival endpoint to enable conversion from accelerated to full approval or, importantly, to offer a second shot on goal with statistical adjustment for type 1 error if the early endpoint is not positive. Such choices in the overarching development strategy can meaningfully shrink the time to regulatory approval, but they can also have unintended consequences. This presentation will describe the scientific and regulatory implications of co-primary endpoints using oncology examples, though the considerations can be generalized to other therapeutic areas.

- **Satrajit Roychoudhury**, Novartis Pharmaceutical Company
  "Audit Strategy for Blinded Independent Central Review of Progression in Cancer Clinical Trials "
  Satrajit Roychoudhury

  Progression-free survival (PFS) is a commonly used primary endpoint in oncology phase 3 trials. In most occasions, regulatory agencies have generally required a complete-case blinded independent central review (BICR) of PFS to assess and reduce potential bias in the local site evaluation (LE). However, recent publications showed a high correlation between LE and BICR assessments of the PFS treatment effect in many disease settings, which questions whether complete-case BICR is necessary. One potential alternative is a sample based BICR as an audit tool to detect evaluation bias in the LE. In this talk, a BICR audit strategy is proposed as an alternative to a full BICR to provide assurance of the treatment effect. The proposed model uses a Bayesian approach to quantify the assurance of treatment effect, given the data of BICR for a selective sample and the data of LE from the complete sample. If the probability crosses the pre-specified threshold value, then no complete BICR is needed. Otherwise, go for complete BICR. Performance and implementation of this method are evaluated using two real data applications.

- **Christian Kappeler**, Bayer Ag
  "Implementing Overall Survival Crossover Adjustment Methods: Clinical Evidence to Support the Common Treatment Effect Assumption"
  Christian Kappeler, Andrea Wagner, Joachim Kalmus, Marcia S Brose, Martin Schlumberger, George D Demetri

  Background: In controlled trials that allow patients to crossover from placebo (PBO) to active treatment (ACT) after progression, OS in PBO patients is confounded by ACT. Statistical methods to

adjust OS for crossover rely on an assumption of common treatment effect (i.e. PBO to ACT crossover after progression has the same impact as ACT started at randomization). This exploratory analysis of two phase 3 trials (PBO-controlled and allowing crossover) with regorafenib in GIST and sorafenib in thyroid cancer provides evidence to support this assumption. Methods: Target lesions were assessed at baseline and at regular intervals post baseline. Progression-free survival (PFS) was assessed during double-blind (DB) treatment and secondary PFS (SPFS) in PBO patients during open-label (OL) treatment. Changes in target lesion diameter over time were estimated by a parametric model. Early tumor growth rate (TGR), defined as percent change from baseline/month in the sum of target lesion diameters, was calculated as the model curve slope at earliest time on the TGR curve. Results: SPFS during OL in PBO patients is closer to DB PFS in ACT patients than to DB PFS in PBO patients (Table). Early TGR during OL in PBO patients is closer to DB early TGR in ACT patients than to DB early TGR in PBO patients. Conclusions: Currently available statistical methods to adjust for the impact of treatment crossover are supported by clinical results from two independent randomized, controlled phase 3 studies. These results provide validity to crossover correction analyses. Clinical trial registration: NCT01271712; NCT00984282

27. the Win Ratio and Related Topics in Multiple Event Time Data

- **David Schoenfeld**, Harvard University
  "Deconstructing the Win Ratio"
  David A. Schoenfeld, Dianne M. Finkelstein

  The win ratio has been suggested as an estimate of treatment benefit for clinical trials with a combined endpoint, especially when the statistical test is the Generalized Gehen Wilcoxon Test that we introduced in 1999 is used. This test considers all pairs of patients and ranks them first on survival and secondly on a secondary criteria. The win ratio is the ratio of the proportion of time the active treatment wins divided by the proportion of times the control treatment wins. An issue with this estimate is that as the follow up decreases the proportion of comparisons due to the secondary endpoint increases so that the win ratio is a function of the follow up time. We introduce graphical techniques for plotting the win ratio as a function of follow up time and decomposing it into the proportion of the ratio due to survival and the proportion due to the other endpoints. These graphs are illustrated for examples in oncology, cardiology and neurology, each of which has a different type of secondary endpoint.

- **Xiaodong Luo**, Sanofi Us
  "Weighted Win Loss Approach for Analyzing Prioritized Outcomes"
  Xiaodong Luo, Junshan Qiu, Steven Bai and Hong Tian

  To analyze prioritized outcomes, Buyse (2010) and Pocock et al. (2012) proposed the win loss approach. This approach uses a layered comparison procedure to account for the order of priorities of different outcomes therefore has attracted a lot of interests in methodological development and real-life application. In this talk, I will first study the relationship between the win loss approach and the traditional survival analysis on the time to the first event. We then propose the weighted win loss statistics to improve the efficiency of the un-weighted methods. Contribution index is used to supplement the win loss approach for a better interpretation of the results. Simulation studies and real data analysis demonstrated the characteristics of the proposed statistics.

- **Changyong Feng**, University of Rochester
  "Using Wei Lin and Weissfeld's Approach in Estimating a Win Ratio"
  Changyong Feng, David Oakes, Xiang Lu

  The win ratio is a simple nonparametric method to study the treatment effect on prioritized multiple outcomes. For simple time-to-event data, the calculation of the win-ratio depends on the distribution of censoring time. The win-ratio is free of censoring times iff the survival functions in two groups satisfy the proportional hazards model (PHM) and then the hazard ratio is exactly the win-ratio. For two dimensional survival data (denoted as fatal and non-fatal events), if the joint survival functions are

in the bivariate Lehmann family, the common parameter also has the interpretation of the win ratio. We can estimate the win-ratio either using all fatal event data or the non-fatal event data (given that the fatal events have not occurred. Neither of them is efficient). We use the WLW method to obtain an optimal linear combination of these two estimators. Simulation study shows that the combined estimator has smaller variance than both estimators. We study the effects on the estimation when the Lehmann family assumption is not satisfied.

Saturday 8:30–10:15

28. Recent Advances in Analyzing Multi-State and Family Data

- **Leilei Zeng**, University of Waterloo
  "Design and Analysis of Clinical Trials with Composite Endpoints: Consideration of Intermittent Observation Scheme"
  Lan Wen, Ker-Ai Lee, Richard Cook

  Clinical trials has been routinely designed and analyzed in recent years on the basis of the composite endpoint of event-free survival, where the event of interest may represent a complication, relapse, or progression. This is viewed as enabling a more timely and cost-effective approach to assessing the clinical benefit novel interventions. Given that the non-terminal event such as progression is typically only assessed periodically, the composite endpoint is thus subject to a dual censoring scheme involving interval censoring for progression and right censoring for death. We highlight statistical issues associated with the conventional approach of using right endpoint imputation in the progression free survival analysis, including the biases in the estimation of survival and loss of power in detecting treatment effect. We also considers design of cancer trials directed at the composite event of progression-free survival. In particular we derive sample size criteria based on an illness-death model that considers cancer progression and death jointly while accounting for the fact that progression is assessed only intermittently.

- **Baojiang Chen**, Universtiy of Texas Health Science Center at Houston
  "Using the Accelerated Failure Time Model to Analyze Current Status Data with Misclassified Covariates"
  Baojiang Chen, Jing Qin and Ao Yuan

  Current status data arises commonly in applications when there is only one feasible observation time to check if the failure time has occurred or not by the observation time, but the exact failure time remains unknown. To accommodate the covariate effect on failure time, the accelerated failure time (AFT) model has been widely used to analyze current status data, where the failure time is assumed to follow some parametric distribution. The limitation of this approach is that it is sensitive to the distribution assumption of the failure time. Covariate misclassification is very common in current status data, and in general, a naive method employing the misclassified variable can lead to invalid inferences. Little study has been concentrated on the covariate misclassification in current status data. In this paper, we consider a semiparametric AFT model to analyze current status data with misclassified covariates and eliminate the bias caused by the misclassification of covariates. This model is also robust to the misspecification of the failure time compared to the parametric AFT model, as we assume an unknown distribution of the failure time in the proposed model. Furthermore, incorporating the covariate effect on the failure time increases the flexibility of the model. We adapted the EM algorithm for the estimation, which guarantee the convergence of the estimate. Both theory and empirical studies show the consistency of the estimator.

- **Karen Kopciuk**, University of Calgary
  "Risk Estimation in Family Data via Multi-State Models"
  Karen Kopciuk, Laurent Briollais, Yun-Hee Choi, Narges Nazeri Rad, Jerry Lawless

  Estimating risk in families who harbour a strong genetic mutation predisposing them to several types of cancer presents many statistical challenges: (a) families are identified and selected directly from

population disease registries or high risk cancer clinics or from two-stage sampling designs, (b) missing genetic information is common or may be unknown for putative genes, (c) residual familial correlation exists when additional risk genes or environmental factors are shared, and (d) multiple cancers are frequent. Multi-state models need to take into account these features for age-at-onset outcomes, as well as direct interventions to the disease process. Disease risk models for family data we have developed will be described as will our new R package, FamEvent.

29. Measurement Error, Mediation Analysis, and Individualized Medicine

- **Yijian Huang**, Emory University
  "Cox Regression with Dependent Error in Covariates"
  Yijian Huang, C. Y. Wang

  Many survival studies have error-contaminated covariates due to the lack of a gold standard of measurement. Furthermore, the error distribution can depend on the true covariates but the structure may be difficult to characterize; heteroscedasticity is a common manifestation. We suggest a novel dependent measurement error model with minimal assumptions on the dependence structure, and propose a new functional modeling method for Cox regression when an instrumental variable is available. This proposal accommodates much more general error contamination than existing approaches including nonparametric correction methods of Huang and Wang (2000, 2006). The estimated regression coefficients are consistent and asymptotically normal, and a consistent variance estimate is provided for inference. Simulations demonstrate that the procedure performs well even under substantial error contamination. Illustration with a clinical study is provided.

- **Cheng Zheng**, University of Wisconsin-Milwaukee
  "Mediation Analysis on Time-to-Event Outcome Data with Unmeasured Confounding and Measurement Error"
  Cheng Zheng

  Mediation analysis is an important topic as it helps researchers to understand why an intervention works. Most previous mediation analyses define effects in the mean scale and require a binary or continuous outcome. Recently, possible ways to dene direct and indirect effects for causal mediation analysis with survival outcome were proposed. However, these methods mainly rely on the assumption of sequential ignorability, which implies no unmeasured confounding. To handle the potential confounding between the mediator and the outcome, in this article, we proposed a structural additive hazard model for mediation analysis with failure time outcome and derived estimators for controlled direct effects and controlled mediator effects. Our methods allow time-varying effects. Simulations showed that our proposed estimator is consistent in the presence of unmeasured confounding while the traditional additive hazard regression ignoring unmeasured confounding produces biased results. We also extend our method to a structural additive sub-distribution hazard model for competing risk data and using regression calibration to handle measurement error in the exposure/mediator.

30. Statistical Analysis of Recurrent, Competing Risks, and Current Status Data

- **Michael Pennell**, The Ohio State University
  "Bayesian Threshold Regression for Multivariate Current Status Data with Informative Censoring"
  Tao Xiao, Michael Pennell

  In animal carcinogencity studies, tumors are only observable at time of natural death or terminal sacrifice. Thus, time-to-tumor is subject to current status observation. When the observation time is independent of the event time, current status data can be analyzed using standard methods for interval censored data. However, time-to-death is likely related to time-to-tumor making this assumption unreasonable. To further complicate matters, multiple types of tumors are observed and analysis models should account for the relationships between these tumors to avoid potential biases. In this presentation, we will present a Bayesian approach which models the times to each type of tumor using

latent Wiener processes which fail when they hit a threshold for the first time. We incorporate shared random effects into the drifts of the latent processes to account for relationships in the tumor risks. Informative observation time is accounted for by modeling time-to-death using a latent Wiener process whose time scale is affected by the occurrence of each tumor.

- **Kwun Chuen Gary Chan**, University of Washington
  "Modeling and Estimating the Terminal Behavior of Recurrent Marker Processes Before Failure Events"
  Kwun Chuen Gary Chan and Mei-Cheng Wang

  Recurrent event processes with marker measurements are mostly studied with forward time models starting from an initial event. Interestingly, the processes could exhibit important terminal behavior during a time period before occurrence of the failure event. A natural and direct way to study recurrent events prior to a failure event is to align the processes using the failure event as the time origin and to examine the terminal behavior by a backward time model. We studied regression models for backward recurrent marker processes by counting time backward from the failure event. A three-level semiparametric regression model is proposed for jointly modeling the time to a failure event, the backward recurrent event process, and the marker observed at the time of each backward recurrent event. By jointly modeling the three components, estimating equations can be constructed for marked counting processes to estimate the target parameters in the three-level regression models. The proposed models and methods are illustrated by a community-based AIDS clinical trial to examine opportunistic infections among HIV infected individuals in the last six months of life.

31. on Risk Prediction in the Presence of Competing Risks

- **Sai Dharmarajan**, University of Michigan
  "A Semiparametric Mixture Component Model with Random Effects for the Analysis of Clustered Competing Risks Data"
  Sai Dharmarajan, Douglas Schaubel

  We propose a semiparametric random effects model for data in the clustered competing risks setting. Specifically, we propose direct modeling of cluster and covariate effects on the cumulative incidence functions of each risk through semiparametric additive regression models containing cluster-specific random effects. A unique feature of our approach is that we model the dependency of failure times both within and across causes across individuals within a cluster by allowing for the correlation of cluster-specific random effects across causes. By decomposing the cause-specific cumulative incidence functions using a mixture model representation, we are able to estimate model parameters associated with all competing risks under consideration while ensuring that the additivity constraint for cumulative incidence functions is satisfied. We develop estimating equations for parameters and test our estimation procedure via simulations. We apply our method to multicenter competing risks data from the Scientific Registry of Transplant Recipients.

- **Malka Gorfine**, Tel Aviv University, Israel
  "Calibrated Predictions for Multivariate Competing Risks Models"
  Malka Gorfine, Li Hsu, David M. Zucker, Giovanni Parmigiani

  Prediction models for time-to-event data play a prominent role in assessing the individual risk of a disease, such as cancer. Accurate disease prediction models provide an efficient tool for identifying individuals at high risk, and provide the groundwork for estimating the population burden and cost of disease and for developing patient care guidelines. We focus on risk prediction of a disease in which family history is an important risk factor that reflects inherited genetic susceptibility, shared environment, and common behavior patterns. In this work family history is accommodated using frailty models, with the main novel feature being allowing for competing risks, such as other diseases or mortality. We show through a simulation study that naively treating competing risks as independent right censoring events results in non-calibrated predictions, with the expected number of events overestimated. Discrimination performance is not affected by ignoring competing risks. Our proposed

prediction methodologies correctly account for competing events, are very well calibrated, and easy to implement.

- **Guoqing Diao**, George Mason University
  "Time-Varying Coefficient Risk Prediction Models for Competing Risks Data"
  Guoqing Diao

  We propose a new stratified Cox model which allows the underlying cumulative hazard function to vary at different values of one or more discrete and/or continuous predictors and investigate its connections with the existing varying-coefficient models. We develop likelihood-based estimation and inference procedures with right-censored data and competing risks. Such procedures are smoothing-free and statistically efficient. Furthermore, the proposed estimator of the cause-specific cumulative hazard function is guaranteed to be non-increasing. We establish the asymptotic properties of the proposed inference procedures. Simulation studies demonstrate that the proposed methods perform well in practical settings. An application to a real study is provided.

32. Challenges and New Methods of Complex Health Data

- **Tibor Schuster**, Mcgill University, Montreal
  "Challenges in Simulating High-Dimensional Health Data for Inference Method Evaluation"
  Tibor Schuster

  The development of advanced methods for statistical inference commonly involves large-scale empirical simulation studies comparing the performance of the competing approaches. However, particularly in high-dimensional data settings, as commonly present in complex administrative health databases, simulation of realistic data scenarios is not straightforward. I will provide a review of recent challenges (and some solutions) related to the set-up of simulation studies in the context of effectiveness research using large complex health data. The potential benefits and limitations of topical plasmode simulation studies and dimension-reducing approaches will be discussed.

33. New Statistical Methods for Complex Structures in Survival Data

- **Fang-Shu Ou**, Mayo Clinic
  "Quantile Regression Models for Interval-Censored Failure Time Data"
  Fang-Shu Ou, Jianwen Cai, Donglin Zeng, David Couper

  Case II interval-censored data arises when the event of interest can not be monitored continuously but only observed at discrete times. This situation occurs commonly in longitudinal data and oncology research where the events were observed when participants had a clinic visit. Censored quantile regression offers a semiparametric alternative to the popular Cox proportional hazards model for describing the relationship between the response and covariates. In this presentation, we propose a quantile regression model to analyze case II interval-censored data. Our model assumes that the conditional quantile of failure time is a linear function of the covariates and the failure time and observation times are independent conditioned on the covariates. An M-estimator is developed for parameter estimation which is computed using a concave-convex procedure and its confidence intervals are constructed using a subsampling method. The estimators are consistent and have a cube-root convergence rate. The proposed method performed well in simulation studies. Finally, we apply the proposed method to analyze data from the Atherosclerosis Risk in Communities (ARIC) Study.

- **Qingxia Chen**, Vanderbilt University
  "Quantifying the Average of the Time-Varying Hazard Ratio via a Class of Transformations"
  Qingxia Chen

  The hazard ratio derived from the Cox model is a commonly used summary statistic to quantify a treatment effect with a time-to-event outcome. The proportional hazards assumption of the Cox model, however, is frequently violated in practice and many alternative models have been proposed

in the statistical literature. Unfortunately, the regression coefficients obtained from different models are often not directly comparable. To overcome this problem, we propose a family of weighted hazard ratio measures that are based on the marginal survival curves or marginal hazard functions, and can be estimated using readily available output from various modeling approaches. The proposed transformation family includes the transformations considered by Schemper et al. (Statist Med 28:24732489, 2009) as special cases. In addition, we propose a novel estimate of the weighted hazard ratio based on the maximum departure from the null hypothesis within the transformation family, and develop a KolmogorovSmirnov type of test statistic based on this estimate. Simulation studies show that when the hazard functions of two groups either converge or diverge, this new estimate yields a more powerful test than tests based on the individual transformations recommended in Schemper et al. (Statist Med 28:24732489, 2009), with a similar magnitude of power loss when the hazards cross. The proposed estimates and test statistics are applied to a colorectal cancer clinical trial.

34. New Machine Learning Methods for Censored Survival and Competing Risks Data

- **Jon a Steingrimsson**, Johns Hopkins University
  "Censoring Unbiased Regression Trees and Forests"
  Jon Steingrimsson, Rob Strawderman, Liqun Diao, Annette Molinaro

  Survival trees use recursive partitioning to separate patients into distinct risk groups when some observations are right-censored. Survival forests average multiple survival trees creating more flexible prediction models. In the absence of censoring, the algorithms rely heavily on the choice of loss function used in the decision making process. Motivated by semiparametric efficiency theory, we replace the loss function used in the absence of censoring by doubly robust loss functions. We derive properties of these loss functions and discuss practical issues related to the implementation of the algorithms. If time permits, the performance of the resulting survival trees and forests will be evaluated through simulation studies and analyzing data on death from myocardial infraction.

- **Youngjoo Cho**, University of Rochester
  "Regression Trees for Cumulative Incidence Functions"
  Youngjoo Cho, Annette M. Molinaro, Chen Hu, and Robert L. Strawderman

  The use of cumulative incidence functions for characterizing the risk of one type of event in the presence of others has become increasingly popular over the past decade. The problems of modeling, estimation and inference have been treated using parametric, nonparametric and semi-parametric methods. Efforts to develop suitable extensions of machine learning methods, such as regression trees and related ensemble methods, have begun only recently. In this paper, we develop a new approach to building regression trees for estimating cumulative incidence curves in a competing risks setting. Following Steingrimsson et al. (2016), the proposed methods employ augmented estimators of the Brier score risk as the primary basis for building and pruning trees. The proposed methods are easily implemented using the rpart procedure, available as part of the R statistical software package. Simulation studies demonstrate the utility of our approach in the competing risks setting. Data from the Radiation Therapy Oncology Group (trial 9410) is used to illustrate these new methods.

35. Advances in Statistical Modeling of Correlated Data

- **Zuoheng Wang**, Yale University
  "Joint Statistical Modeling of Multiple Phenotypes in Related Samples"
  Zuoheng Wang, Zhong Wang

  Genetic association studies have routinely been conducted to search for variants associated with diseases and quantitative phenotypes. Clinical and epidemiological studies typically collect data on a set of correlated phenotypes that may share common environmental and/or genetic factors. Such phenotypes contain more information than univariate phenotypes. Thus joint modeling of multiple phenotypes can potentially have increased power to detect association and increased precision of parameter estimation

than univariate analysis. In this study, we develop novel statistical methods for multivariate association mapping in samples that contain arbitrarily related individuals. The proposed methods are based on retrospective analysis that is less dependent on model assumptions on phenotypes, thus they are robust to trait model misspecification. The new methods can potentially accommodate the external biological information by integrating graphical models and multivariate analysis, and are computationally affordable. We evaluate the proposed methods by simulation studies and real data application.

- **Yuping Zhang**, University of Connecticut
  "A Statistical Framework for Data Integration Through Graphical Models with Application to Cancer Genomics"
  Yuping Zhang, Zhengqing Ouyang, Hongyu Zhao

  Recent advances in high-throughput biotechnologies have generated various types of genetic, genomic, epigenetic, transcriptomic and proteomic data across different biological conditions. It is likely that integrating data from diverse experiments may lead to a more unified and global view of biological systems and complex diseases. We present a coherent statistical framework for integrating various types of data from distinct but related biological conditions through graphical models. Specifically, our statistical framework is designed for modeling multiple networks with shared regulatory mechanisms from heterogeneous high-dimensional datasets. The performance of our approach is illustrated through simulations and its applications to cancer genomics.

- **Ji Meng Loh**, New Jersey Institute of Technology
  "A Single-Index Model for Inhomogeneous Spatial Point Patterns"
  Ji Meng Loh, Yixin Fang

  I will introduce a single index model for the intensity of an inhomogeneous spatial point process, relating the intensity function to an unknown function of a linear combination of measurements of a p-dimensional spatial covariate process. Such a model extends and generalizes a commonly used model where is known. Under some regularity assumptions, consistency and asymptotic normality of estimates of can be shown. I will present results of some simulation studies showing the effectiveness of the procedure. Finally, results from applying the procedure to a dataset of fast food restaurant locations in New York City will be shown.

- **Ji Meng Loh**, New Jersey Institute of Technology
  "A Single-Index Model for Inhomogeneous Spatial Point Patterns"
  Ji Meng Loh, Yixin Fang

  I will introduce a single index model for the intensity of an inhomogeneous spatial point process, relating the intensity function to an unknown function of a linear combination of measurements of a p-dimensional spatial covariate process. Such a model extends and generalizes a commonly used model where is known. Under some regularity assumptions, consistency and asymptotic normality of estimates of can be shown. I will present results of some simulation studies showing the effectiveness of the procedure. Finally, results from applying the procedure to a dataset of fast food restaurant locations in New York City will be shown.

36. Modeling Disease Natural History and Effects of Treatment: Applications to Prostate Cancer

- **Yolanda Hagar**, University of Colorado Boulder
  "Modeling of Prostate Cancer Recurrence"
  Yolanda Hagar, James Dignam, Vanja Dukic

  Prostate cancer patients can live for many years post diagnosis, with long records of survivor follow-up. This pattern can lead to treatment effects that vary over time, making inference difficult. We examine RTOG clinical trials data using the multi-resolution hazard model, which is a Bayesian, semi-parametric survival model for estimation of the hazard rate and the effects of covariates on survival time. The model has been extended to incorporate non-proportional hazards as well as periods of time

with sparsely observed failures. We compare the results of the MRH model to those from other popular packages in R.

- **Loc Ferrer**, Inserm U1219, University of Bordeaux, France
  "Joint Modelling of Longitudinal and Multi-State Processes: Application to Clinical Progressions in Prostate Cancer"
  Loc Ferrer, Ccile Proust-Lima

  Joint modelling of longitudinal and survival data is increasingly used in clinical trials on cancer. In prostate cancer, these models permit to account for the link between longitudinal measures of prostate-specific antigen (PSA) and time of clinical recurrence when studying the risk of relapse. In practice, multiple types of relapse may occur successively. Distinguishing these transitions between health states would allow to evaluate, for example, how PSA trajectory and classical covariates impact the risk of dying after a distant recurrence post radiotherapy, or to predict the risk of one specific type of clinical recurrence post-radiotherapy, from the PSA history. In this context, we present a joint model for a longitudinal process and a multi-state process, which is divided into two sub-models: a linear mixed sub-model for longitudinal data and a multi-state sub-model with proportional hazards for transition times, both linked by a function of shared random effects. Parameters of this joint multi-state model are estimated within the maximum likelihood framework. It is implemented under R, by combining and extending mstate and JM packages. The estimation program is validated by simulations and applied on pooled data from two cohorts of men with localized prostate cancer. Thanks to the classical covariates available at baseline and the repeated PSA measurements, we are able to assess the biomarkers trajectory, define the risks of transitions between health states and quantify the impact of the PSA dynamics on each transition intensity. Another aspect of this work seeks to check the Markov assumption. Indeed, in joint multi-state modelling it could be assumed that some individuals are more likely to experience a succession of events than others, even after adjusting for covariates. In prostate cancer for example, for two patients with the same measured characteristics, one of them could experience several relapses of the disease, and the other none. This can be accounted for by including a frailty term (i.e. an individual random effect) in the multi-state process. To test the necessity of such random effect in a our joint multi-state model, we propose a score test which only requires the estimation of the model under the null hypothesis, that is excluding the frailty term. The type-I and type-II errors of the score test are evaluated in simulations.

- **Meredith Regan**, Dana-Farber Cancer Institute
  "Identification and Validation of Surrogate Time-to-Event Endpoints for Use in Localized Prostate Cancer Randomized Trials using Individual Patient Data Meta-Analysis"
  Meredith Regan, Wanling Xie, Christopher Sweeney; ICECaP Working Group

  Adjuvant therapy for localized prostate cancer (CaP) reduces recurrence and death from CaP. The Intermediate Clinical Endpoints in CaP (ICECaP) Working Group is conducting an individual patient data meta-analysis of potential surrogate endpoints for localized CaP trials. Surrogates for overall survival (OS) could expedite the evaluation of new adjuvant therapies. We hypothesized that disease-free survival (DFS) and metastasis-free survival (MFS) are valid surrogate endpoints for OS. We systematically identified 102 completed or ongoing eligible randomized trials comparing treatments in localized CaP and collected individual patient data from 28 trials with 28,905 patients. We evaluated the surrogacy of two ICEs (MFS: time from randomization to first evidence of distant metastasis or death from any cause, or censored at date of last follow-up; and DFS: time from randomization to first evidence of loco-regional or distant recurrence or death from any cause) with OS (time from randomization to death from any cause). We used a 2-stage meta-analytic validation model by determining whether the ICE and OS were correlated (patient-level surrogacy) and whether the ICE and OS treatment effects were correlated (trial-level surrogacy). We estimated a surrogate threshold effect from the weighted linear regression of OS treatment effects on the surrogate treatment effects and investigated the implications for future study design.

37. Student Paper Presentations

- **Ling-Wan Chen**, University of Pittsburgh-Pittsburgh Campus
  "Cumulative Incidence Regression for Dynamic Treatment Regimens"
  Ling-Wan Chen, Idil Yavuz, Yu Cheng and Abdus S. Wahed

  Recently dynamic treatment regimens (DTRs) have drawn considerable attention, as an effective tool for personalizing medicine. Sequential Multiple Assignment Randomized Trials (SMARTs) are often used to gather data for making inference on DTRs. In this paper, we focus on regression analysis of DTRs from a two-stage SMART for competing-risk censored outcomes based on cumulative incidence functions (CIFs). Even though there are extensive works on the regression problem for DTRs, no research has been done on modeling the CIF for SMART trials. We extend existing CIF regression models to handle covariate effects for DTRs. Asymptotic properties are established for our proposed estimators. The models can be implemented using existing software by an augmented-data approximation. We show the improvement provided by our proposed methods by simulation, and illustrate its practical utility through an analysis of a SMART neuroblastoma study, where disease progression is subject to competing-risk censoring by death.

- **Fei Gao**, University of North Carolina at Chapel Hill
  "Semiparametric Regression Analysis of Interval-Censored Data with Informative Dropout"
  Fei Gao, Donglin Zeng, D. Y. Lin

  Interval-censored data arise when the event of interest can only be ascer- tained through periodic examinations. In medical studies, subjects may not complete the examination schedule for reasons related to the event of interest. In this paper, we develop a semiparametric approach to adjust for such informative dropout in regression analysis of interval-censored data. Specifically, we propose a broad class of joint models, under which the event time of interest follows a transformation model with a random effect and the dropout time follows a different transformation model but with the same random effect. We consider nonparametric maximum likelihood estimation and develop an EM algorithm that involves simple and stable calculations. We prove that the resulting estimators of the regression parameters are consistent, asymptotic normal, and asymptotically efficient with a covariance matrix that can be consistently estimated through profile likelihood. In addi- tion, we show how to consistently estimate the survival function when dropout represents voluntary withdrawal and the cumulative incidence function when dropout is an unavoid- able terminal event. Furthermore, we assess the performance of the proposed numerical and inferential procedures through extensive simulation studies. Finally, we provide an application to data on the incidence of diabetes from a major epidemiological cohort study.

Saturday 10:30–12:15

38. Tree-Based Methods for Survival Data

- **Rodney Sparapani**, Medical College of Wisconsin
  "Nonparametric Recurrent Events Analysis with Bart and an Application to the Hospital Admissions of Patients with Diabetes"
  Rodney Sparapani, Lisa Rein, Sergey Tarima, Tourette Jackson, John Meurer

  Much of survival analysis is concerned with absorbing events, i.e., events for which only a single event can be experienced. Here we focus on recurrent events: events which subjects are capable of experiencing multiple times. Recurrent events have been studied by many, however, most rely on the linear proportional intensity assumption. We propose a new method for recurrent events based on Bayesian Additive Regression Trees (BART) which does not employ restrictive assumptions such as linearity or proportionality. We explore this new method via a motivating example of hospital admissions for diabetes patients and simulated data sets.

- **Jeffrey S. Simonoff**, New York University
  "Conditional Inference Survival Trees for Nonstandard Data"
  Wei Fu, Jeffrey S. Simonoff

  Tree methods (recursive partitioning) are a popular class of nonparametric methods for analyzing data. One extension of the basic tree methodology is the survival tree, which applies recursive partitioning to censored survival data. This has mainly been designed for right-censored data. We discuss application of the conditional inference survival tree method to two important but less standard data types, left-truncated and right-censored (LTRC) data and interval-censored data. Further, we show that LTRC trees can be used to analyze survival data with time-varying covariates, essentially building a time-varying covariates survival tree. Implementation of the methods is easy, and simulations and real data analysis results show that the proposed methods work well from both a predictive point of view and in uncovering tree structure in the underlying survival process.

- **Philippe brot**, Inserm and Faculty of Medicine, University Paris-Saclay
  "Bagging Improper Survival Trees for Prediction and Variable Selection"
  Phiippe Brot, Cyprien Mbogning

  For clinical studies with high-dimensional datasets, tree-based ensemble methods offer a powerful solution for variable selection and prediction taking into account the complex interrelationships between explanatory variables. One of the key components of the tree-building process is the splitting criterion. For survival data, the classical splitting criterion is the Logrank statistic. However, the presence of a fraction of non-susceptible patients in the studied population is advocating for considering a criterion tailored to this peculiar situation. In this context, we propose a bagging improper survival tree procedure for variable selection and prediction where the survival tree-building process relies on a splitting criterion that explicitly focuses on time-to-event survival distribution among susceptible patients. Different criteria for evaluating the importance of the explanatory variables and the prediction performance are also proposed. This procedure will be presented and illustrated on a genomic dataset with gene expression measurements from early breast cancer patients.

- **Denis Larocque**, Hec Montreal
  "L1 Splitting Rules in Survival Forests"
  Hoora Moradian, Denis Larocque, Franois Bellavance

  The log-rank test is used as the split function in many commonly used survival trees and forests algorithms. However, the log-rank test may have a significant loss of power in some circumstances, especially when the hazard functions or when the survival functions cross each other in the two compared groups. We investigate the use of the integrated absolute difference between the two children nodes survival functions as the splitting rule. Simulations studies and applications to real data sets show that forests built with this rule produce very good results in general, and that they are often better compared to forests built with the log-rank splitting rule.

- **David M. Vock**, University of Minnesota
  "Inverse Probability of Censoring Weighted Bagging: a General Approach for Adapting Machine Learning Methods to Right-Censored Data"
  David M. Vock, Ales Kotalik, Julian Wolfson, Patrick J. O'Connor

  Risk prediction plays an important role in prioritizing limited healthcare resources, guiding treatment decisions, and motivating patients to remain adherent to those recommended treatments. Modern machine learning techniques provide a wide variety of ways to classify or predict outcomes, especially suited to the setting of large and complex datasets, such as those derived from administrative health records databases. These techniques tend to be more flexible and computationally efficient than traditional regression methods, but they typically do not account for the fact that event status may be censored, leading to biased estimators of individual risk. We propose a universal and intuitive method, inverse probability of censoring weighted bagging, that can be used in conjunction with any ML technique to estimate risk in the presence of censored data. We demonstrate our method by considering the problem of predicting 5-year cardiovascular risk using electronic health data (EHD) from a large Midwestern

health system, where over 50% of subjects have fewer than 5 years of follow-up. We compare the performance of our method against several previously proposed ad hoc methods of handling censoring in machine learning problems. Our general purpose approach, unlike other ad hoc approaches previously proposed, leads to well-calibrated risk predictions while maintaining discrimination performance.

39. Advances in Multi-State Models for Survival and Event History Analysis

- **Candemir Cigsar**, Memorial University of Newfoundland
  "Assessment of Dynamic Multi-State Models with Recurrent Events: a Copula Approach"
  Candemir Cigsar

  There has been a recent interest in the analysis of event history data through dynamic models. These models are intensity based stochastic process models, and can be used to examine the effects of past event occurrences on the present or future evolution of a process. The adequacy of dynamic models is an important issue in event history settings. Robust inference procedures based on marginal features of event history models such as rate or mean functions are not applicable. Furthermore, the models based on the strong independence assumption of the inter-event (sojourn) times between successive events may result in substantial bias in parameter estimates. To address these issues, we apply a semiparametric copula model, in which the marginal distributions of inter-event times are nonparametrically estimated. We therefore use nonparametric estimates of marginal distributions of the second and subsequent inter-event times to assess the plausibility of a dynamic model. In addition, we discuss how to assess the validity of the assumed copula model. We illustrate our methods to analyze a data set from a study involving recurrent asthma attacks in children.

- **Lihui Zhao**, Northwestern University
  "Restricted Mean Survival Time in Event History Analysis"
  Lihui Zhao, Lu Tian, Lee-Jen Wei

  In a longitudinal study with the time to a specific event as the primary end point, the restricted mean survival time (RMST) is an easily interpretable, clinically meaningful summary of the survival function in the presence of censoring. The RMST is the mean survival time of all subjects in the study population followed up to a time point t and can be estimated consistently by the area under the Kaplan-Meier curve over [0, t]. In this research, we discuss the extension of RMST in the more general setting of event history analysis, which includes classical competing risks and semi-competing risks. The methods are illustrated with the data from two clinical studies.

- **Peihua Qiu**, University of Florida
  "Professor"
  Peihua Qiu

  Effective Comparison of Two or More Hazard Rate Functions Peihua Qiu Department of Biostatistics University of Florida In clinical trials and other medical studies, we often need to compare two or more hazard rate functions concerning treatment effects. In the literature, most existing methods assume that the related hazard rate curves are parallel to each other or they do not cross. However, in practice it is common that the hazard rate curves cross each other, representing different treatment effects in different time periods. In this talk, we will present our recent research on comparison of two or more crossing hazard rate curves. Both theoretical arguments and numerical results show that they work well in practice.

- **Elizabeth Juarez-Colunga**, University of Colorado Denver
  "Analysis of Recurrent Pulmonary Exacerbations in Cystic Fibrosis Children: Early Pseudomonas Infection Control (Epic) Observational Study"
  Elizabeth Juarez-Colunga, Brandie Wagner, Edith Zemanick

  Cystic Fibrosis (CF), a hereditary disease that is marked by irreversible loss of lung function, affects approximately 30,000 individuals in the United States leading to expenditures of approximately 1.4

billion dollars a year. Pulmonary exacerbations (PE), episodes of acute worsening of respiratory symptoms, in CF are a marker of loss of lung function. Understanding the recurrence of PE is of vital importance in early CF. The ongoing EPIC observational has collected data from young CF children for several years on PE and time-varying factors potentially associated with occurrence of PE. To address the intermittent aspect in recording of time-varying factors, we propose a joint model using shared frailties. We present results of such analysis in the EPIC observational study.

40. Advances in the Analysis of Clinical Trials with Lifetime Data using Restricted Mean Survival Times

- **Lu Tian**, Stanford University
  "Exact Inference on the Restricted Mean Survival Time"
  Lu Tian and LJ Wei

  In a randomized clinical trial with the time to event as the primary endpoint, one often evaluates the treatment effect by comparing the survival distributions from two groups. This can be achieved by for example estimating the hazard ratio under the popular proportional hazards (PH) model. However, when the hazard rate is very low, e.g., in safety studies, there may be too few observed events to warrantee the valid asymptotic inferences based on the PH model. The exact inference including hypothesis testing and constructing 95

- **Ivn daz**, Weill Cornell Medicine
  "Improved Precision in the Analysis of Randomized Trials with Survival Outcomes, without Assuming Proportional Hazards"
  Ivn Daz, Elizabeth Colantuoni, and Michael Rosenblum

  We present a new estimator of the restricted mean survival time in randomized trials where there is right censoring that may depend on treatment and baseline variables. The proposed estimator leverages prognostic baseline variables to obtain equal or better asymptotic precision compared to traditional estimators. Under regularity conditions and random censoring within strata of treatment and baseline variables, the proposed estimator has the following features: (i) it is interpretable under violations of the proportional hazards assumption; (ii) it is consistent and at least as precise as the Kaplan-Meier estimator under independent censoring; (iii) it remains consistent under violations of independent censoring (unlike the Kaplan-Meier estimator) when either the censoring or survival distributions are estimated consistently; and (iv) it achieves the nonparametric efficiency bound when both of these distributions are consistently estimated. We illustrate the performance of our method using simulations based on resampling data from a completed, phase 3 randomized clinical trial of a new surgical treatment for stroke; the proposed estimator achieves a 12

- **Sarah Conner**, Boston University School of Public Health
  "Adjusted Restricted Mean Survival Times in Observational Studies "
  Sarah Conner, Lisa Sullivan, Sandro Galea, Ludovic Trinquart

  Observational studies primarily report the hazard ratio (HR), commonly estimated with a multivariable Cox proportional hazards (PH) model, as the preferred effect size for time-to-event outcomes. A model-free alternative measure is the restricted mean survival time (RMST) to a fixed time horizon. An effect size can then be measured as the difference in RMST between groups. Adjustment for covariates is necessary to address potential confounding bias in the comparison of groups of patients. We present a method for estimating adjusted differences in RMST from observational studies, based on integrating Inverse Probability Weighted (IPW) Kaplan-Meier estimators in the groups of interest. The associated variance takes the weighting scheme into account. We assessed the statistical performance (in terms of bias, relative bias, mean squared error, and coverage probability) of the approach through a simulation study under PH and non-PH settings. We compared it to the pseudo-observation regression-based approach by Andersen and colleagues. We found that the proposed approach was superior to the pseudo-observation approach in terms of bias, relative bias, and coverage, although the pseudo-observation approach performed better in terms of mean squared error for imbalanced data. Finally, we

illustrated these methods on the Framingham Coronary Heart Disease (CHD) 10-year Risk Score. We estimated the difference in adjusted RMST for CHD between higher and lower cholesterol groups and the corresponding HR from a multivariable Cox model. In men and women, we found that the adjusted HR for higher vs. lower total cholesterol ranged from 1.3 to 2.1. However, we found a difference of -0.10 to 3.66 months in time without CHD, on average, between lower and higher total cholesterol over 10 years. In conclusion, the proposed approach is congruent with adjusted Kaplan-Meier curves and both would merit widespread reporting in observational studies. We provide R code to implement the approach.

41. New Methods for Censored Data

- **Ying Qing Chen**, Fred Hutchinson Cancer Research Center
  "Modeling the Trend of Recurrent Event Data with Weak Comparability"
  Ying Qing Chen

  The HPTN 052 Study is a high-profile randomized clinical trial to establish the prominent efficacy of Treatment-as-Prevention (TasP) in HIV prevention among HIV sero-discordant couples. Such an efficacy nevertheless relies on the HIV-infected patients adherence to their prescribed antiretroviral therapies. We hence aim to develop a new statistical method to understand the trend of drug adherence measured by pill counts in the HPTN 052 Study. The new method focuses on modeling the durations alternating between high and low adherences during the follow-up. In the talk, we will present a new concept of weak comparability in modeling recurrent gap times, and discuss the analysis results and their practical implication.

- **Jong-Hyeon Jeong**, University of Pittsburgh-Pittsburgh Campus
  "Statistical Inference on Life Lost"
  Jong-Hyeon Jeong

  Time-to-event data may be analyzed based on (1) cumulative information up to a specific time point through the hazard function or the survival function, or on (2) residual information beyond that time point through the mean or quantile residual life function. In this talk, a new summary measure, the quantile life lost, is introduced, which has several advantages over the existing summary measures for censored time-to-event data. The distribution of life lost is characterized by the reversed hazard function. Nonparametric inference on the quantile life lost for one-sample and two-sample cases are presented, together with some extension to a regression setting. Asymptotic distributions of the estimators and test statistics are derived and some simulation results are presented to demonstrate their finite sample behaviors. The proposed method is illustrated with a real data set from a breast cancer study.

- **Ying Qing Chen**, Fred Hutchinson Cancer Research Center
  " Modeling the Trend of Recurrent Event Data with Weak Comparability"
  Peng Liu, Yijian Huang, Kwun Chuen Gary Chan, Ying Qing Chen

  Recurrent event data are frequently observed in medical research, where each subject experiences more than one event. One particular aspect of recurrent event is the presence or absence of time trend. Trend refers to systematic variation among the occurrence rates of times between events, it can be used as a measure of disease progression. Wang and Chen (Biometrics, 2000) proposed a strong comparability concept to study the trend in recurrent event data. In this paper we propose weak comparability under the same assumption as Wang and Chen (2000). Our proposed concept can produce more comparable pairs and thus result in a more efficient estimate. Monte Carlo simulation as well as real data analyse are performed to validate the effectiveness of the new method.

- **Ross Prentice**, Fred Hutchinson Cancer Research Center
  "Self-Consistent Nonparametric Estimation of the Multivariate Survivor Function"
  Ross Prentice, Shanshan Zhao

A brief review will be given of nonparametric estimators of the bivariate survivor function. This will be followed by a description of a new estimator based on Dabrowska's estimator in conjunction with self-sufficiency to include doubly-censored data, and corresponding simulation results. Potential regression generalizations focusing on Cox model marginal hazard rates and a multiplicative cross ratio regression model will also be briefly mentioned.

42. Recent Advances in Survival Analysis for Cancer Research

- **Qi Long**, University of Pennsylvania
  "Addressing Issues Associated with Evaluating Prediction Models for Survival Endpoints Based on the Concordance Statistic"
  Ming Wang, Qi Long

  Prediction models for disease risk and prognosis play an important role in biomedical research, and evaluating their predictive accuracy in the presence of censored data is of substantial interest. The standard concordance (c) statistic has been extended to provide a summary measure of predictive accuracy for survival models. Motivated by a prostate cancer study, we address several issues associated with evaluating survival prediction models based on c-statistic with a focus on estimators using the technique of inverse probability of censoring weighting (IPCW). Compared to the existing work, we provide complete results on the asymptotic properties of the IPCW estimators under the assumption of coarsening at random (CAR), and propose a sensitivity analysis under the mechanism of noncoarsening at random (NCAR). In addition, we extend the IPCW approach as well as the sensitivity analysis to high-dimensional settings. The predictive accuracy of prediction models for cancer recurrence after prostatectomy is assessed by applying the proposed approaches. We find that the estimated predictive accuracy for the models in consideration is sensitive to NCAR assumption, and thus identify the best predictive model. Finally, we further evaluate the performance of the proposed methods in both settings of low-dimensional and high-dimensional data under CAR and NCAR through simulations.

- **Noorie Hyun**, National Cancer Institute
  "Risk Prediction Models for Left- Or Interval-Censored Data from Electronic Health Records"
  Noorie Hyun, Li C. Cheung, Qing Pan, Mark Schiffman, Hormuzd A. Katki

  Electronic health records are a large and cost-effective data source for developing risk-prediction models. However, for screen-detected diseases, standard risk models (such as Kaplan-Meier or Cox models) do not account for key issues encountered with electronic health record data: left-censoring of pre-existing (prevalent) disease, interval-censoring of incident disease, and ambiguity of whether disease is prevalent or incident when definitive disease ascertainment is not conducted at baseline. Furthermore, researchers might conduct novel screening tests only on a complex two-phase subsample. We propose a family of weighted mixture models that account for left/interval-censoring and complex sampling via inverse-probability weighting in order to estimate current and future absolute risk: we propose a weakly-parametric model for general use and a semi-parametric model for checking goodness of fit of the weakly-parametric model. We demonstrate asymptotic properties analytically and by simulation. We used electronic health records to assemble a cohort of 33,295 human papillomavirus (HPV) positive women undergoing cervical cancer screening at Kaiser Permanente Northern California (KPNC) that underlie current screening guidelines. The next guidelines would focus on HPV typing tests, but reporting 14 HPV types is too complex for clinical use. National Cancer Institute along with KPNC conducted an HPV typing test on a complex subsample of 9,258 women in the cohort. We used our model to estimate the risk due to each type and grouped the 14 types (the 3-year risk ranges 21.9-1.5) into 4 risk-bands to simplify reporting to clinicians and guidelines. These risk-bands could be adopted by future HPV typing tests and future screening guidelines.

43. Risk Assessment and Prediction with Survival Data

- **Ying Ding**, University of Pittsburgh-Pittsburgh Campus
  "Progression Risk Prediction with Copula Model for Age-Related Macular Degeneration (Amd) Patients"
  Ying Ding, Yi Liu, Tao Sun, Wei Chen

  Age-related Macular Degeneration (AMD) is a polygenic and progressive neurodegenerative disease, which is a leading cause of blindness in developed countries. Some patients with AMD maintain good vision for a long time with little disease progression over time, while others quickly advance to vision-threatening late AMD. The progressions of two eyes within the same subject are often correlated. In this research, we first develop a computationally efficient copula-based score test, of which the dependence between bivariate progression times is explicitly modeled, to identify susceptible demographical/environmental and genetic risk factors associated with AMD progression. Then, using a large randomized trial data, Age-related Eye Disease Study (AREDS), we establish copula-based prediction models to predict the joint progression-free probability of two eyes within a subject. Finally, we evaluate and validate the prediction models using another independent large randomized trial -AREDS2.

- **Liang Li**, Md Anderson Cancer Center
  "Dynamic Prediction with Landmark Survival Models"
  Liang Li

  Dynamic prediction is the personalized, real-time prediction of the risk of clinical events using longitudinal data. For any subject in the population under study, the prediction can be made at any time during the follow-up, adaptive to the time-varying at-risk population, personalized etiological history, and predictor-outcome association. We will review two popular approaches for dynamic prediction, joint modeling of longitudinal and survival data and landmark survival models, and will present some recent methodological development of the latter. First, we present a semi-parametric kernel based analytical framework for the landmark Cox model, including innovations to model formulation, estimation and prediction accuracy assessment. Second, we show that there exists a joint distribution of the longitudinal and survival data that satisfies the modeling assumptions of the landmark Cox model. This result suggests that the landmark Cox model should not be viewed only as a working model or a prediction algorithm; it has probability basis and rigorous asymptotic analysis and simulation work can be conducted. We illustrate the proposed methodology by data from the African American Study of Kidney Disease and Hypertension.

44. Recent Development of Time-to-Event Data Analysis Incorporating Disease Dynamics

- **Zhi (Kevin) He**, University of Michigan
  "Gateaux Differential-Based Boosting for Fitting Large-Scale Survival Data with Non-Proportional Hazards"
  Kevin He, Yanming Li, Ji Zhu and Yi Li

  Non-proportional hazards model is a flexible and powerful tool for modeling the dynamic changes of covariate effects in survival analysis. However,such models are often difficult to model. The computational burden increases quickly as the sample size or the number of predictors grows. which prohibits the application of existing statistical methods to large-scale data. In view of this gap, we propose a new Gateaux differential-based boosting procedure for simultaneously selecting and automatically determining the potential non-proportional effects. Specifically, our procedure allows that in each boosting learning step only the best-fitting base-learner (and therefore the most informative covariate) is added to the predictor, and consequently encourages sparsity. In addition, our method controls smoothness, which is crucial for improving the predictive performance. The performance of the proposed method is examined by simulations and by applications to analyze national kidney transplant data.

- **Li Cheung**, National Cancer Institute
  "Concordance Indices for Composite Survival Outcomes"
  Qing Pan, Noorie Hyun, Hormuzd Katki

  Composite outcomes are increasingly common in clinical trials to increase power, reduce cost, or to fully account for multifaceted diseases. Harrells Concordance (C) index, widely used to evaluate predictions from regression models for univariate survival outcomes, cannot be used for composite outcomes. We propose two extensions to the C index for composite survival outcomes that account for frequency of outcome occurrence and outcome severity or importance. A weighted C index is proposed for disease processes with multiple primary endpoints, and a most severe comparable C index is proposed for disease processes with a rare primary outcome and a more common correlated secondary outcome. Asymptotic properties are derived based on theorems for U-statistics. Our simulation studies show that these extensions are efficient at identifying true prognostic variables. We illustrate the value of these new concordance indices by applying them to two large diabetes cohorts.

45. Analytical Challenges and New Advances in Assessing Time-to-Event Endpoints in Oncology Studies

- **Wei-Ting Hwang**, University of Pennsylvania
  "Predictive Value of Continuous Markers for Censored Survival Data: a Likelihood Ratio Approach and Extension to Competing Risk Framework"
  Wei-Ting Hwang, Andrew Smith

  The likelihood ratio function (LR) provides an easily interpretable measure for the value of risk prediction markers, and has been explored for both binary and continuous markers of a binary outcome (e.g., diseased or not) (Gu and Pepe, 2011). In this work, we extend the concept of LR to analyze survival outcomes that are subject to censoring. Estimation for the time-dependent LR (TD-LR) will be done using Kaplan-Meier estimation and a univariate Cox proportional hazards (PH) model to quantify the update of the risk prediction due to the knowledge of the marker value. A scale invariant" approach based on marker quantiles is discussed to allow comparison of predictive values between markers with different scales. Relationships to time-dependent receiver-operator characteristic (ROC) curves (Heagerty, Lumley, and Pepe, 2000), area under the curve (AUC), and optimal cut-off values for dichotomizing a continuous marker are considered. We also focus the interpretation TD-LR using a graphical presentation to ease communication with clinical investigators. Extension to competing risk setting will be also discussed. The proposed methods will be applied to data from a bladder cancer clinical trial to determine whether the neutrophil-to-lymphocyte ratio is a valuable biomarker for predicting overall survival following surgery or combined chemotherapy and surgery and from a breast cancer clinical trial with the endpoint of breast cancer recurrence that competing with the event of death without recurrence.

- **Hongwei Zhao**, Texas A&M University
  "Professor"
  Gang Han, Hongwei Zhao

  The development of personalized cancer therapy leads to clinical trials with smaller sample sizes compared with trials involving complete disease entities. The use of exponential survival modeling has the potential of gaining 30

- **Hongyuan Cao**, University of Missouri
  "Assistant Professor"
  Hongyuan Cao, Li Chen, Jason P. Fine

  Long term follow-up with longitudinal data is common in many medical investigations. In such studies, some longitudinal covariates can be omitted for various reasons. In cross sectional studies, coefficient estimation of a covariate is unbiased if the covariate is orthogonal to the omitted covariate. This is not true in longitudinal data analysis, where omission of time dependent covariate can lead to

biased coefficient estimate even if the corresponding covariate is orthogonal to the omitted longitudinal covariate. In this article, we propose a new unbiased estimation method to accommodate omitted longitudinal covariate. In addition, if the omitted longitudinal covariate is asynchronous with the longitudinal response, we propose a two stage approach for valid statistical inference. Asymptotic properties of the proposed parameter estimates are established. Extensive simulation studies provide numerical support for the theoretical

findings. We illustrate the performance of our method on a dataset from an HIV study.

46. Analysis of Complex Sampling Designs with Censored Data

- **Andy Ni**, Memorial Sloan Kettering Cancer Center
  "Tuning Parameter Selection in Cox Proportional Hazards Model with a Diverging Number of Parameters"
  Andy Ni, Jianwen Cai

  Regularized variable selection is a powerful tool for identifying the true regression model from a large number of candidates by applying penalties to the objective functions. The penalty functions typically involve a tuning parameter that control the complexity of the selected model. The ability of the regularized variable selection methods to identify the true model critically depends on the correct choice of the tuning parameter. In this study we develop a consistent tuning parameter selection method for regularized Cox's proportional hazards model with a diverging number of parameters. The tuning parameter is selected by minimizing the generalized information criterion. We prove that, for any penalty function that possesses the oracle property, the proposed tuning parameter selection method identifies the true model with probability approaching one as sample size goes to infinity. Its finite sample performance is evaluated by simulations. Its practical use is demonstrated in the Cancer Genome Atlas (TCGA) breast cancer data.

- **Olli Saarela**, University of Toronto
  "A Weighted Partial Likelihood Method for Estimating Flexible Parametric Marginal Structural Hazard and Intensity Models"
  Olli Saarela, Zhihui Liu

  Parameters in marginal structural Cox models can be estimated through weighting risksets in a Cox partial likelihood. Similar easily applicable and computationally convenient method for estimating flexible parametric marginal structural hazard models appears to be lacking. Herein we propose a new weighted partial likelihood function for this purpose, based on case-base sampling of person-moments. This results in weighted logistic/multinomial regression with an offset term, with time effects modeled through regression splines. The proposed method can accommodate continuous-time weights, and generalizes to estimating intensity functions for recurrent event outcomes. We show that the resulting estimating equation is unbiased, motivate the asymptotic distribution of the resulting estimator, and study the properties of the estimator through simulations.

47. Survival Analysis for Family-Clustered Data

- **Luise Cederkvist**, University of Copenhagen & Danish Cancer Society Research Center
  "Modelling the Cumulative Incidence Function of Multivariate Competing Risks Data While Allowing for Within-Family Dependence of Risk and Timing"
  Luise Cederkvist, Klaus K. Holst, Klaus K. Andersen and Thomas H. Scheike

  Data arising from family studies on chronic diseases are multivariate competing risks data where the competing failure cause death may prevent disease occurrence. In the competing risks setting, one has the choice between analysing the multivariate data on the hazard scale focusing on the cause-specific hazard or on the probability scale focusing on the cause-specific cumulative incidence. Both approaches are equally relevant and may complement each other. However, in analysis of family studies, there is

often a strong interest in describing age at disease onset and how it is affected by within-family dependence. The distribution of age at disease onset is directly described by the cause-specific cumulative incidence. We propose to model the cause-specific cumulative incidence function of multivariate competing risks data using a random effects model that allows for within-family dependence with regard to both risk and timing. That is, the model allows the risk level (in terms of absolute risk) and the failure time distribution to vary between families. Under the proposed model, the cumulative incidence functions of all failure causes are modelled. Consequently, left-truncation as well as right-censoring are easily dealt with and specification of censoring distributions avoided. The proposed model is assessed using simulation studies and applied in analysis of Danish register-based family data on breast cancer.

- **Frank Eriksson**, University of Copenhagen
  "Semiparametric Inference of Clustered Left-Truncated Data"
  Frank Eriksson, Torben Martinussen, Thomas Scheike

  The Nordic twin registers offer the possibility to study the genetic and environmental underpinnings of diseases. Delayed entry occurs as only twin pairs with both twins alive after a given date are included. For example, in the Finnish twin cohort a twin pair is included only if both twins were alive in 1974 when the registration started. It is well known how to handle delayed entry when considering univariate survival times. However, when analyzing clustered survival data, there are limited results on semiparametric inference. Existing methods mostly focus on parametric modelling. In this talk I will discuss different types of left-truncation and suggest estimators for semiparametric hazard models under specific truncation schemes. I will consider both conditional and marginal hazard models. The large sample properties of the estimators are established. Small sample properties are investigated via simulation studies and the suggested estimators are used in a study of prostate cancer based on the Finnish, Danish and Norwegian twin cohorts.

- **Jeanine Houwing-Duistermaat**, Dept of Statistics, University of Leeds, Uk
  "Modelling Mortality in Long Lived Families"
  Jeanine J Houwing-Duistermaat, Mar Rodriguez Girondo

  Analyses of age at onset outcomes in families is challenging because of the presence of correlation between family members due to genetic and shared environmental effects and the ascertainment of the families in the study. Two commonly used family designs are the proband family design and the (multiple) case family design. In the first design probands are selected and family members of these probands are recruited. In the second design the selection is based on the distribution of an outcome in the family. A third complication is survival data subject to delayed entry. The state of the art method is not applicable to the general situation of clusters for which not all members are observed. Our work is motivated by ongoing recruitment of historical of case and control families from the Netherlands (HSN) and the Leiden Longevity study (LLS) comprising 420 sibships with at least two nonagenarian siblings who are alive and willing to participate. Since the siblings need to be alive the survival time in this second study is subject to left truncation. The goals of both studies are to assess clustering of life span within families, to estimate genetic and environmental contributions to life span and to detect markers for healthy ageing. We will discuss how to correct for ascertainment of the families. To obtain unbiased parameter estimates for LLS, we propose a weighted shared frailty model. Here, the weights represent the probability for the family to be in the sample. They depend on the original family size of eligible family members. For both studies it appears that the study design yielded a gain in life span of the family members compared to the general population. Analyses of life span of the siblings of the LLS showed a relatively small estimate of the variance of the frailty in the 90+ subjects.

Saturday 1:45–3:30

49. Recent Advances in the Analysis of Competing Risks Data

- **Qing Yang**, Duke University
  "Sample Size Determination for Jointly Testing a Cause-Specific Hazard and the any-Cause Hazard in

the Presence of Competing Risks"
Qing Yang, Wing K. Fung, Gang Li

This research considers sample size determination for jointly testing a cause-specific hazard and the any-cause hazard for competing risks data. The cause-specific hazard and the any-cause hazard jointly characterize important study endpoints such as the disease-specific survival and overall survival, which are commonly used as co-primary endpoints in clinical trials. Specifically, we derive sample size calculation methods for two-group comparisons based on an asymptotic chi-square joint test and a maximum joint test of the aforementioned quantities, taking into account of censoring due to lost to follow-up as well as staggered entry and administrative censoring. Our simulations demonstrate that the proposed methods can produce substantial sample size savings over the classical Bonferroni adjustment method and generally have satisfactory finite sample performance. We illustrate the application of the proposed methods using the 4-D (Die Deutsche Diabetes Dialyse Studie) clinical trial.

50. Bayesian Nonparametric Survival Analysis

- **William Cipolli**, Colgate College
  "Accelerated Failure Time Models via Smoothed, Approximate Polya Trees"
  William Cipolli

  A smoothed, approximate Polya tree is developed and implemented in the accelerated failure time (AFT) model, including a heteroscedastic generalization. Early Bayesian nonparametric approaches used Dirichlet process mixtures (Kuo and Mallick, 1997; Kottas and Gelfand, 2001) and Polya trees (Walker and Mallick, 1999; Hanson and Johnson, 2002). We introduce a discrete approximation to a mixture of Polya trees prior that enjoys surprisingly simple and efficient conjugate updating. The discrete approximation is then smoothed with Gaussian kernels to provide a differentiable density for use with continuous data; fast, conjugate updating is retained, thus obviating poor MCMC mixing often encountered with "regular" mixtures of Polya tree models. The smoothed approximation is illustrated in the context of density estimation, as well is in a survival context on recent data involving breast cancer survival in Louisiana. Finally, the AFT model is generalized to a baseline distribution that changes smoothly with continuous covariates, i.e. a "continuously stratified" heterscedastic AFT model.

- **Peter Mueller**, Ut Austin
  "A Bayesian Nonparametric Approach for Semi-Competing Risks"
  Peter Mueller, Yanxun Xu, Mike Daniels, Daniel Scharfstein

  We develop a Bayesian nonparametric (BNP) model to assess the treatment effect in semi-competing risks, where a nonterminal event may be censored by a terminal event, but not vice versa. Semi-competing risks are common in brain cancer trials with death being censored by cerebellar progression. We propose a flexible BNP approach to model the joint distribution of progression and death events, thereby effectively inferring the marginal distributions of progression time and death time, characterizing within-subject dependence structure, predicting the progression and death times given a patient's covariate, and quantifying uncertainties of all estimates. More importantly, we define a causal effect of treatment, which can be estimated from the data and has a nice causal interpretation. We perform extensive simulation studies to evaluate the proposed BNP model. The simulations show that the proposed model can accurately estimate the treatment effect in semi-competing risks setup. We also implement the proposed BNP model on data from a brain cancer Phase II trial.

51. Recent Developments in Complex Survival and Longitudinal Data

- **Yize Zhao**, Weill Cornell Medicine, Cornell University
  "Hierarchical Feature Selection Incorporating Known and Novel Biological Information: Identifying Genomic Features Related to Prostate Cancer Recurrence"
  Yize Zhao, Matthias Chung, Brent A. Johnson, Carlos S. Moreno, Qi Long

Our work is motivated by a prostate Q1 cancer study aimed at identifying mRNA and miRNA biomarkers that are predictive of cancer recurrence after prostatectomy. It has been shown in the literature that incorporating known biological information on pathway memberships and interactions among biomarkers improves feature selection of high-dimensional biomarkers in relation to disease risk. Biological information is often represented by graphs or networks, in which biomarkers are represented by nodes and interactions among them are represented by edges; however, biological information is often not fully known. For example, the role of microRNAs (miRNAs) in regulating gene expression is not fully understood and the miRNA regulatory network is not fully established, in which case new strategies are needed for feature selection. To this end, we treat unknown biological information asmissing data (i.e., missing edges in graphs), different from commonly encountered missing data problems where variable values are missing. We propose a new concept of imputing unknown biological information based on observed data and define the imputed information as the novel biological information. In addition,we propose a hierarchical group penalty to encourage sparsity and feature selection at both the pathway level and the within-pathway level, which, combined with the imputation step, allows for incorporation of known and novel biological information. While it is applicable to general regression settings, we develop and investigate the proposed approach in the context of semiparametric accelerated failure time models motivated by our data example. Data application and simulation studies show that incorporation of novel biological information improves performance in risk prediction and feature selection and the proposed penalty outperforms the extensions of several existing penalties.

52. New Developments in Modeling Longitudinal, Recurrent Event and Survival Data

- **Chi Hyun Lee**, The University of Texas Md Anderson Cancer Center
  "Semiparametric Regression Method for Bivariate Alternating Recurrent Event Data"
  Chi Hyun Lee, Xianghua Luo, Chiung-Yu Huang, Gongjun Xu

  Alternating recurrent event data arise frequently in clinical and epidemiologic studies, where two types of events such as hospital admission and discharge occur alternately over time. The two alternating states defined by these recurrent events, in the previous example the care period and the break period, could each carry important distinct information about a patient's underlying health condition and/or the quality of care. In this paper, we propose a semiparametric method for evaluating covariate effects on the two alternating states jointly. The proposed methodology accounts for the dependence among the alternating states as well as the heterogeneity across patients via a frailty with unspecified distribution. Moreover, the estimation procedure, which is based on smooth estimating equations, not only properly addresses challenges such as induced dependent censoring and intercept sampling bias commonly confronted in serial event gap time data, but also is more computationally tractable than the existing rank-based methods. The proposed methods are evaluated by simulation studies and illustrated by analyzing psychiatric contacts from the South Verona Psychiatric Case Register.

- **Shanshan Li**, Indiana University R.m. Fairbanks School of Public Health
  "Joint Modeling of Recurrent Event Processes and Intermittently Observed Time-Varying Binary Covariate Processes"
  Shanshan Li

  When conducting recurrent event data analysis, it is common to assume that the covariate processes are observed throughout the follow-up period. In most applications, however, the values of time-varying covariates are only observed periodically rather than continuously. A popular ad-hoc approach is to carry forward the last observed covariate value until it is measured again. This simple approach, however, usually leads to biased estimation. To tackle this problem, we propose to model the covariate effect on the risk of the recurrent events through jointly modeling the recurrent event process and the longitudinal measures. Despite its popularity, estimation of the joint model with binary longitudinal measurements remains a challenge, because the standard linear mixed effects model approach is not appropriate for binary measures. In this talk, we postulate a Markov model for the binary covariate process and a random-effect proportional intensity model for the recurrent event process. We use a Markov chain Monte Carlo algorithm to estimate all the unknown parameters. The performance of

the proposed estimator is evaluated via simulations. The methodology is applied to an observational study designed to evaluate the effect of Group A streptococcus on pharyngitis among school children in India.

- **Chunyan Cai**, The University of Texas Health Science Center at Houston
  "Time-Varying Dependence Measure of Bivariate Recurrent Event Processes"
  Chunyan Cai, Liang Zhu, Jing Ning

  Bivariate or multivariate recurrent event processes are often encountered in longitudinal studies in which more than one type of event is of interest. There has been much research on regression analysis for such data, but little has been done to measure the dependence between recurrent event processes. We use a time-dependent measure, termed the rate ratio, to assess the local dependence between two types of recurrent event processes. The commonly used shared random effect models assume that the dependence between different types of recurrent events is a constant over time and may not be adequate to model the observed multivariate recurrent events data. We propose a class of flexible shared random effect models to allow for the time-varying dependence. We develop an expectationmaximization (EM) algorithm for the model fitting.

- **Shanshan Li**, Indiana University R.m. Fairbanks School of Public Health
  "Joint Modeling of Recurrent Event Processes and Intermittently Observed Time-Varying Binary Covariate Processes"
  Shanshan Li

  When conducting recurrent event data analysis, it is common to assume that the covariate processes are observed throughout the follow-up period. In most applications, however, the values of time-varying covariates are only observed periodically rather than continuously. A popular ad-hoc approach is to carry forward the last observed covariate value until it is measured again. This simple approach, however, usually leads to biased estimation. To tackle this problem, we propose to model the covariate effect on the risk of the recurrent events through jointly modeling the recurrent event process and the longitudinal measures. Despite its popularity, estimation of the joint model with binary longitudinal measurements remains a challenge, because the standard linear mixed effects model approach is not appropriate for binary measures. In this talk, we postulate a Markov model for the binary covariate process and a random-effect proportional intensity model for the recurrent event process. We use a Markov chain Monte Carlo algorithm to estimate all the unknown parameters. The performance of the proposed estimator is evaluated via simulations. The methodology is applied to an observational study designed to evaluate the effect of Group A streptococcus on pharyngitis among school children in India.

- **Sehee Kim**, University of Michigan
  "Joint Partially Linear Model for Longitudinal Data with Informative Drop-Outs"
  Sehee Kim, Donglin Zeng, Jeremy M.G. Taylor

  In biomedical research, a steep rise or decline in longitudinal biomarkers may indicate latent disease progression, which may subsequently cause patients to drop out of the study. Ignoring the informative drop-out can cause bias in estimation of the longitudinal model. In such cases, a full parametric specification may be insufficient to capture the complicated pattern of the longitudinal biomarkers. For these types of longitudinal data with the issue of informative drop-outs, we develop a joint partially linear model, with an aim to find the trajectory of the longitudinal biomarker. Specifically, an arbitrary function of time along with linear fixed and random covariate effects is proposed in the model for the biomarker, while a flexible semiparametric transformation model is used to describe the drop-out mechanism. Advantages of this semiparametric joint modeling approach are the following: 1) it provides an easier interpretation, compared to standard nonparametric regression models, and 2) it is a natural way to control for common (observable and unobservable) prognostic factors that may affect both the longitudinal trajectory and the drop-out process. We describe a sieve maximum likelihood estimation procedure using the EM algorithm, where the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are considered to select the number of knots. We show that the proposed estimators achieve desirable asymptotic properties through empirical process theory. The

proposed methods are evaluated by simulation studies and applied to prostate cancer data.

54. Recent Topics on Analysis of Interval Censored Data (with Covariate or Outcome Censored)

- **Jing Qian**, University of Massachusetts - Amherst
  "Regression Analysis with Randomly Censored Covariates"
  Jing Qian

  Censored covariates arise frequently in biomarker assessement in clinical studies and in family history studies of disease. While there is a large literature on regression models when the outcome variable is subject to censoring, there is a more limited literature on the treatment of censored covariates, especially for randomly censored ones. We develop threshold regression approaches for linear regression models with a covariate subject to random censoring. Compared with existing methods, the proposed ones are simple but effective as they avoid complicated modeling in dealing with censored covariate values. In addition to estimating the regression coefficient of the censored covariate, the threshold regression methods can be used to test whether the effect of the censored covariate is significant. We discuss the choice of optimal threshold which yields the most powerful test. We conduct simulation studies and apply the proposed methods to an Alzheimers disease study. Comparison with alternative approaches that we developed recently for censored covariates, including multiple imputation method and reverse survival regression, will be discussed as well.

- **Qiang Zhao**, Texas State University
  "Estimation of Survival Function for Interval-Censored Std Data with Auxiliary Diaries"
  Qiang Zhao

  Interval-censored data arises naturally in follow-up clinical trials or epidemiological studies. Recently, patients in some studies on sexually transmitted diseases (STDs) were asked to record their coital activities in diaries. This study focuses on the estimation of the survival function of time to an STD infection based on interval-censored data with auxiliary diaries. We propose a self-consistency procedure that incorporates the coital diary data in the estimation. Comparisons to Turnbull's self-consistency algorithm and an existing imputation method are made in a simulation study. The proposed method is then applied to a set of STD data for illustration.

55. New Development in Statistical Methods for Deriving and Validating Dynamic and Individualized Decision Rules

- **Jin Wang**, University of North Carolina, Chapel Hill
  "Single Index Models in Proportional Hazard Regression for Precision Medicine"
  Jin Wang, Donglin Zeng, Danyu Lin

  We propose a flexible single-index model to allow complex treatment-covariate interactions and to derive a simple linear treatment rule for personalized treatment decision. Our model extends the proportional hazards models by using a monotone regression function. The inference procedures are based on sieve estimation that includes single-index parameters in the basis expansion and partiali likelihood for the baseline hazard function. We obtain the theoretical results by deriving the necessary rates of convergence for the nonparamatric estimator of the arbitrary regression function. We provide simultaneous inference on single-index parameters and other regression parameters. Simulation studies are conducted to assess the finite sample performance. An application to a multiple-type cancer study is presented to illustrate our methods.

- **Yunro Chung**, Fred Hutchinson Cancer Research Center
  "Estimation of Disease Progression Rate using Longitudinal Surrogate Outcomes in Non-Randomized Validation Subsample"
  Yunro Chung, Tianxi Cai, Yingye Zheng

For a current status data where an event of interest is observed before or after a random monitoring time without knowing the exact event time, an outcome is often missing when not all subjects are monitored. Particularly when a small subset of subjects is non-randomly selected to be monitored, a standard complete case analysis that uses the observed outcome only may be inefficient and biased. Here, we develop an estimated likelihood method for estimating survival function in such setting by additionally using an error-prone (longitudinal) surrogate outcome that can be easily assessed for all subjects. Under the missing at random assumption, we use the inverse probability weighting kernel estimator to adjust biases from the non-random selection and imperfect surrogate outcome. We apply our method to estimate disease progression rate for the prostate cancer active surveillance study.

- **Marlena Maziarz**, National Cancer Institute, Nih
  "Evaluating Longitudinal Markers under Two-Phase Study Designs"
  Marlena Maziarz, Yingye Zheng

  Little attention has been given to the design of efficient studies to evaluate longitudinal biomarkers. Measuring longitudinal markers on an entire cohort is cost prohibitive and, especially for rare outcomes such as cancer, may be infeasible. Thus, methods for evaluation of longitudinal biomarkers using efficient and cost-effective study designs are needed. Case-cohort (CCH) and nested case-control (NCC) studies allow investigators to evaluate biomarkers rigorously and at reduced cost, with only a small loss in precision. In this paper we develop estimators of several measures to evaluate the accuracy and discrimination of predicted risk under CCH and NCC study designs. We use double inverse probability weighting (DIPW) to account for censoring and sampling bias in estimation and inference procedures. To facilitate inference using two-phase longitudinal data, we develop valid resampling-based variance estimation procedures under CCH and NCC. We evaluate the performance of our estimators under CCH and NCC using simulation studies and illustrate them on a nested case-control study within the Hepatitis C Antiviral Long-Term Treatment against Cirrhosis (HALT-C) clinical trial. Our estimators and inference procedures perform well under CCH and NCC, provided that the sample size at the time of prediction (effective sample size) is reasonable. These methods are widely applicable, efficient and cost-effective, and can be easily adapted to other study designs used to evaluate prediction rules in a longitudinal setting.

Saturday 3:45–5:30

56. Advances in Semiparametric Regression Analysis of Panel Count Data

- **Ao Yuan**, Georgetown University
  "Analysis of Panel Count Data using Semiparametric Regression Model"
  Jing Qin, Ao Yuan, Guoqing Diao

  A semiparametric regression model is proposed and studied for panel count data, in which the time distribution function is nonparametric only assumed to be monotone increasing, and the regression part is specified parametric. The sempiarametric maximum likelihood estimate are used to estimate both the regression and the time distribution function, via the EM-algorithm. The nonparametric part is estimated in closed form in each of the EM iteration. Asymptotic results are investigated, and simulation studies are conducted to evaluate the performance of the proposed method, the results are promising. The method will be used to analyze a real data when available.

- **Xingqiu Zhao**, The Hong Kong Polytechnic University
  "Hypothesis Testing for Panel Count Data"
  Xingqiu Zhao and Ying Zhang

  In semiparametric and nonparametric statistical inference, the weak convergence theory on the asymptotic distribution of estimators have been widely used to establish the asymptotic normality of the estimators when the estimators are $n^{1/2}$-consistent. In many applications, nonparametric estimators are not able to achieve this rate. We propose a general theorem on the asymptotic normality of nonparametric M-estimators which can be used no matter whether the rate of convergence of an estimator

is $n^{-1/2}$ or slower. We apply the proposed theory to study the asymptotic distribution of sieve estimators of functionals of a mean function from a counting process, and develop nonparametric tests for the problem of treatment comparison with panel count data. The test statistics are constructed by using statistically and computationally efficient spline likelihood estimators instead of nonparametric likelihood estimators. The new tests have a more general and simpler structure and thus are easy to implement. Simulation studies show that the proposed tests perform well even for small sample sizes. Moreover, we find that a new test is always powerful for all the situations considered and thus robust.

57. a Recent Development on Competing and Semi-Competing Risks

- **Ruosha Li**, University of Texas Health Science Center at Houston
  "Flexible Association Modelling and Prediction with Semi-Competing Risks Data"
  Ruosha Li

  Semi-competing risks data involve a non-terminal event time, such as time to disease progression, and a terminal event time such as time to death. Existing methods for handling semi-competing risks data often assume that the underlying association between the two event times follows a pre-specified copula with unknown association parameters, which often correspond to the strength of association. In this article we propose a flexible association model that does not require pre-specifying a copula. Therefore our methods facilitate a convenient and robust evaluation of the underlying association pattern, as well as the association strength. Furthermore the proposed association model leads to a robust estimator for the conditional survival probability of the terminal event given the non-terminal event. The methods were also extended to handle left-truncation. Both the association and survival estimators were shown to feature desirable asymptotic properties and satisfactory numerical performance. Our methods were successfully applied to a diabetes data set to study the association between time to diabetic nephropathy and time to death, and to predict the mortality rate given the onset time of nephropathy.

- **Yu Cheng**, University of Pittsburgh
  "Novel Diagnostic Accuracy Analysis for Competing Risks Outcomes with Roc Surface"
  Yu Cheng

  Many medical conditions are marked by a sequence of events or statuses that are associated with continuous changes in some biomarkers. However, few works have been done to assess the overall accuracy of a biomarker on separating various competing events. Existing methods usually focus on a single cause and compare it with the event-free controls each time. In our study, the concept of ROC surface and the volume under the ROC surface (VUS) has been extended to competing risks outcomes, given its capability to handle ordinal multi-category outcomes. We propose two methods to estimate the VUS. The first method is based on the correct classification probabilities (CCPs) for the subjects who have experienced different cause-specific events given a pair of threshold points from the distribution of a diagnostic marker. The second method is to measure concordance between the marker and competing outcomes. Since the samples are often subject to independent censoring, inverse probability of censoring weigh is introduced to handle censored outcomes. Asymptotic results are derived using counting process techniques. The practical performances of the proposed estimators have been evaluated through numerical studies.

- **Gongjun Xu**, University of Michigan
  "Joint Scale-Change Models for Recurrent Events and Failure Time"
  Gongjun Xu, Sy Han Chiou, Chiung-Yu Huang, Mei-Cheng Wang, Jun Yan

  Recurrent event data arise frequently in various fields such as biomedical sciences, public health, engineering, and social sciences. In many instances, the observation of the recurrent event process can be stopped by the occurrence of a correlated failure event, such as treatment failure and death. In this talk, we propose a joint scale-change model for the recurrent event process and the failure time, where a shared frailty variable is used to model the association between the two types of outcomes. In contrast to the popular Cox-type joint modeling approaches, the regression parameters in the proposed joint scale-change model have marginal interpretations. The proposed approach is robust in the sense that

no parametric assumption is imposed on the distribution of the unobserved frailty and that we do not need the strong Poisson-type assumption for the recurrent event process. We establish consistency and asymptotic normality of the proposed semiparametric estimators under suitable regularity conditions. To estimate the corresponding variances of the estimators, we develop a computationally efficient resampling-based procedure. Simulation studies and an analysis of hospitalization data from the Danish Psychiatric Central Register illustrate the performance of the proposed method.

- **Sy Han Chiou**, Harvard University
  "Postdoctoral Researcher"
  Sy Han Chiou, Gongjun Xu, Jun Yan, Chiung-Yu Huang

  Panel count data occur in studies where the study subjects are observed only periodically or at discrete examination times. In contrast to most of the existing approaches that assume independence between the recurrent event process and the examination time process, we address the issue of informative examination time process by considering a scale-change model for the underlying recurrent event process and allow the two processes to be correlated through a shared frailty. Our model can be viewed as an accelerated failure time type model for recurrent events, and the regression parameters have a simple marginal interpretation of modifying the time scale of the event process. A novel estimation procedure for the regression parameters and the baseline rate function is proposed, which, in contrast to existing methods, is robust in the sense that it does not require the strong Poisson-type assumption for the underlying recurrent event process, or impose a parametric assumption on the distribution of the unobserved frailty. Large-sample properties of the estimators are established, and their variances are estimated by a model-based smoothed bootstrap procedure. Numerical studies demonstrated that the proposed point estimator and variance estimator perform well with practical sample sizes. The methods are applied to data from a skin cancer chemoprevention trial.

- **Ran Duan**, Eli Lilly and Company
  "Estimate Variable Importance for Recurrent Event Outcome with an Application to Hypoglycemia Events"
  Ran Duan; Haoda Fu

  Recurrent event data are an important data type for medical research. In particular, many safety endpoints are recurrent outcomes, such as hypoglycemic events. For such a situation, it is important to identify the factors causing these events and rank these factors by their importance. Traditional model selection methods are not able to provide variable importance in this context. Methods that are able to evaluate the variable importance, such as gradient boosting and random forest algorithms, cannot directly be applied to recurrent events data. In this paper, we propose a two-step method that enables us to evaluate the variable importance for recurrent events data. We evaluated the performance of our proposed method by simulations and applied it to a data set from a diabetes study.

- **Chung-Chou H. Chang**, University of Pittsburgh-Pittsburgh Campus
  "Regression Models for Data with Competing Risks under Random Signs Censoring"
  Chung-Chou H. Chang, Jonathan G. Yabes, Tianxiu Wang

  Since 2002, the Pediatric End-Stage Liver Disease (PELD) score has been used to estimate the 90-day mortality rate for pediatric patients who need a liver transplant and are on the United Network for Organ Sharing (UNOS) transplant waiting list. However, it has been shown that the PELD score underestimates the mortality rate, physicians routinely overadjust the score by adding about 5 points to it. The original PELD model was based on a standard Cox regression. To account for identifiability issue of competing risks, we propose a new model to estimate death without transplantation. The proposed model is a Cox proportional hazards regression under the random signs censoring (RSC) principle, which assumes that the main event failure time is independent of the indicator that the main event precedes the competing event. Unlike identifying assumptions that are typically imposed in practice, RSC is verifiable via stochastic ordering in the observed data. We showed that the model is not only easy to implement but also covariate effects have desirable asymptotic properties. We evaluated the estimator's finite sample performance through simulations.

58. Recent Advances on Event History Analysis

- **Qingning Zhou**, University of North Carolina at Chapel Hill
  "Case-Cohort Studies with Interval-Censored Failure Time Data"
  Qingning Zhou, Haibo Zhou, Jianwen Cai

  The case-cohort design has been widely used as a means of cost reduction in assembling or measuring expensive covariates in large cohort studies. The existing literature on the case-cohort design is mainly focused on right-censored data. In practice, however, the failure time is often subject to interval-censoring; it is known only to fall within some random time interval. In this paper, we consider the case-cohort study design for interval-censored failure time and develop a sieve semiparametric likelihood approach for analyzing data from this design under the proportional hazards model. We construct the likelihood function using inverse probability weighting and build the sieves with Bernstein polynomials. The consistency and asymptotic normality of the resulting regression parameter estimator are established and a weighted bootstrap procedure is considered for variance estimation. Simulations show that the proposed method works well for practical situations, and an application to data from the Atherosclerosis Risk in Communities (ARIC) study is provided.

59. Recent Advances in Time-to-Event Analysis with High Dimensional, Heterogeneous, and/or Correlated Data

- **Shuangge Ma**, Yale University
  "Robust Analysis of Cancer Prognosis Data with Gene-Environment Interactions"
  Shuangge Ma

  For the prognosis of cancer, it has been increasingly established that the interactions between genetic (G) and environmental (E) factors have an important impact beyond the main G and E effects. Most of the existing studies adopt standard likelihood-based estimation, which is nonrobust. In practice, data contamination and model mis-specification may happen, which demand robust estimation. Such robust analysis gets complicated with the high data dimensionality. In our studies, we have developed multiple robust approaches for the analysis of cancer prognosis data with the presence of G-E interactions. Methodological, theoretical, and numerical studies have been extensively conducted for multiple cancer types.

- **Sebastien Haneuse**, Harvard T.h. Chan School of Public Health
  "Hierarchical Models for Semi-Competing Risks Data with Application to Quality of End-of-Life Care for Pancreatic Cancer"
  Sebastien Haneuse, Kyu Ha Lee, Francesca Dominici, Deborah Schrag

  Readmission following discharge from an initial hospitalization is a key marker of quality of health care in the United States. For the most part, readmission has been studied among patients with 'acute' health conditions, such as pneumonia and heart failure, with analyses based on a logistic-Normal generalized linear mixed model. Naive application of this model to the study of readmission among patients with "advanced" health conditions such as pancreatic cancer, however, is problematic because it ignores death as a competing risk. A more appropriate analysis is to imbed such a study within the semi-competing risks framework. To our knowledge, however, no comprehensive statistical methods have been developed for cluster-correlated semi-competing risks data. To resolve this we propose a novel hierarchical modeling framework for the analysis of cluster-correlated semi-competing risks data that permits parametric or non-parametric specifications for a range of components giving analysts substantial flexibility as they consider their own analyses. Estimation and inference is performed within the Bayesian paradigm since it facilitates the straightforward characterization of (posterior) uncertainty for all model parameters, including hospital-specific random effects. Model comparison and choice is performed via the deviance information criterion and the log-pseudo marginal likelihood statistic, both of which are based on a partially marginalized likelihood. An efficient computational scheme, based on the Metropolis-Hastings-Green algorithm, is developed and had been implemented in

the SemiCompRisks R package. A comprehensive simulation study shows that the proposed framework performs very well in a range of data scenarios, and outperforms competitor analysis strategies. The proposed framework is motivated by and illustrated with an on-going study of the risk of readmission among Medicare beneficiaries diagnosed with pancreatic cancer. Using data on n=5,298 patients at J=112 hospitals in the six New England states between 2000-2009, key scientific questions we consider include the role of patient-level risk factors on the risk of readmission and the extent of variation in risk across hospitals not explained by differences in patient case-mix.

- **Dipak K. Dey**, University of Connecticut
  "Flexible Link Functions in Nonparametric Binary Regression with Gaussian Process Priors"
  Dipak K. Dey, Dan Li, and Xia Wang

  n many scientific fields, it is a common practice to collect a sequence of 0-1 binary responses from a subject across time, space, or a collection of covariates. Researchers are interested in finding out how the expected binary outcome is related to covariates, and aim at better prediction in the future 0-1 outcomes. Gaussian processes have been widely used to model nonlinear systems; in particular to model the latent structure in a binary regression model allowing nonlinear functional relationship between covariates and the expectation of binary outcomes. A critical issue in modeling binary response data is the appropriate choice of link functions. Commonly adopted link functions such as probit or logit links have fixed skewness and lack the flexibility to allow the data to determine the degree of the skewness. To address this limitation, we propose a flexible binary regression model which combines a generalized extreme value link function with a Gaussian process prior on the latent structure. Bayesian computation is employed in model estimation. Posterior consistency of the resulting posterior distribution is demonstrated. The flexibility and gains of the proposed model are illustrated through detailed simulation studies and two real data examples. Empirical results show that the proposed model outperforms a set of alternative models, which only have either a Gaussian process prior on the latent regression function or a Dirichlet prior on the link function.

60. Statistical Innovations for Data Science and Precision Medicine

- **Zhengqing Ouyang**, The Jackson Laboratory for Genomic Medicine
  "Statistical Modeling of Genome-Wide Chromatin Interaction Data to Elucidate Spatial Organizations of Genomes"
  Chenchen Zou, Yuping Zhang

  Recent high-throughput genomic technologies (such as Hi-C) are revolutionizing biomedical research by allowing genome-wide characterization of chromatin interactions and structures. While massive data sets have been generated by Hi-C and related technologies, few statistical approaches exist for effectively modeling the spatial organizations of genomes. Here, we introduce a model-based approach for analyzing chromatin interaction data. Through simulation and real data applications, we demonstrate accurate and robust reconstruction of 3D chromatin structures at high resolution and genome-wide level.

61. Risk Prediction in Survival Analysis

- **Jimmy T. Efird**, Chairperson, Medical Statistics, University of Newcastle, Australia
  "Risk Statification using a Weibull Proportional-Hazards Model: Asymptotic Efficiency of a Counting Process Approach"
  Jimmy T. Efird

  Although commonly used to assess disease risk, adjusted survival estimates generated from a Cox proportional-hazards model tend to deviate from their true values with increasing time from exposure. Using a counting process martingale approach, Altshuler (Nelson-Aalen) estimates have been shown to have asymptotically robust properties when censoring is nominal. In this talk, we explore an alternative counting process method that aims to increase the asymptotic efficiency of Altshulers survival estimates.

This technique involves using a generalized (2-parameter) Weibull model to identify censor-to- failure transition jump points that can then be incorporated into the Altshuler estimate. Given a set of covariates, we show how to estimate the Hessian and Fisher information function corresponding to the Weibull model. We further provide an example demonstrating how the maximum- likelihood parameter estimates from this model can be used in a localized iterative fashion to identify asymptotically efficient transition jump points.

- **Jimmy T. Efird**, University of Newcastle
  "Risk Statification using a Weibull Proportional-Hazards Model "
  Jimmy T. Efird, Natasha Weaver, Daniel Barker

  Abstract: Although commonly used to assess disease risk, adjusted survival estimates generated from a Cox proportional-hazards model often deviate significantly from their true values with increasing time from exposure. In this talk, we present a family of survival models based on the generalized Weibull distribution for performing risk stratification of time to event data. Given a set of covariates, we derive the corresponding density, cumulative distribution, hazard, and survival function for the Weibull proportional-hazards model. Additionally, we show how to estimate the Hessian and Fisher information function corresponding to model beta coefficients and how to test the null hypothesis that beta=0. With new methods appearing in the literature for efficiently imputing censored time to event data, the parametric Weibull proportion-hazard model presents a practical alternative approach for analyzing such data.

- **Yan Yuan**, University of Alberta
  "Measuring the Prediction Performance for Risk Scores in the Era of Clinical Preventive Care"
  Yan Yuan. Qian Michelle Zhou, Bingying Li, Henrui Cai, Eric Chow, Greg Armstrong

  Prediction performance of a risk scoring system needs to be carefully assessed before its adoption in clinical practice. Clinical preventive care often uses risk scores to *screen* asymptomatic population. The primary clinical interest is to predict the risk of having an event by a pre-specified *future* time t_0. Prospective accuracy measures such as positive predictive values have been recommended for evaluating the predictive performance. However, for commonly used continuous or ordinal risk score systems, these measures require a subjective cut-off threshold value that dichotomizes the risk scores. The need for a cut-off value created barriers for practitioners and researchers. In this talk, we discuss a threshold-free summary index of positive predictive values that accommodates time-dependent event status. We develop a nonparametric estimator that accommodates competing risk events and provide an inference procedure for comparing two risk scores for censored time to event data. We conduct a simulation study to examine the finite-sample performance of the proposed estimation and inference procedures. Lastly, we illustrate the use of this measure on a real data example, comparing two risk score systems for predicting heart failure in childhood cancer survivors.

63. Statistical Challenges for Immunotherapy with Delayed Treatment Effects

- **Zhenzhen Xu**, Fda
  "Designing Therapeutic Cancer Vaccine Trials with Delayed Treatment Effect"
  Zhenzhen Xu, Boguang Zhen, Yongsoek Park and Bin Zhu

  Arming the immune system against cancer has emerged as a powerful tool in oncology during recent years. Instead of poisoning a tumor or destroying it with radiation, therapeutic cancer vaccine, a type of cancer immunotherapy, unleashes the immune system to combat cancer. This indirect mechanism-of-action of vaccines poses the possibility of a delayed onset of clinical effect, which results in a delayed separation of survival curves between the experimental and control groups in therapeutic cancer vaccine trials with time-to-event endpoints. This violates the proportional hazard assumption. As a result, the conventional study design based on the regular log-rank test ignoring the delayed effect would lead to a loss of power. In this paper, we propose two innovative approaches for sample size and power calculation using the piecewise weighted log-rank test to properly and efficiently incorporate the delayed effect into the study design. Both theoretical derivations and empirical studies demonstrate

that the proposed methods, accounting for the delayed effect, can reduce sample size dramatically while achieving the target power relative to standard practice.

Abstracts of Posters

- "Cumulative Incidence Regression for Dynamic Treatment Regimens"
  Ling-Wan Chen, Idil Yavuz, Yu Cheng Abdus S. Wahed, *University of Pittsburgh*

  Recently dynamic treatment regimens (DTRs) have drawn considerable attention, as an effective tool for personalizing medicine. Sequential Multiple Assignment Randomized Trials (SMARTs) are often used to gather data for making inference on DTRs. In this paper, we focus on regression analysis of DTRs from a two-stage SMART for competing-risk censored outcomes based on cumulative incidence functions (CIFs). Even though there are extensive works on the regression problem for DTRs, no research has been done on modeling the CIF for SMART trials. We extend existing CIF regression models to handle covariate effects for DTRs. Asymptotic properties are established for our proposed estimators. The models can be implemented using existing software by an augmented-data approximation. We show the improvement provided by our proposed methods by simulation, and illustrate its practical utility through an analysis of a SMART neuroblastoma study, where disease progression is subject to competing-risk censoring by death.

- "Statistical Modeling and SIR Modeling in New York State Counties of Syndrome X based on Text Based Indicators from Social Media and Food Environment"
  Samantha DAlonzo

  The rising incidence of obesity in America is causing health issues for the obesogenic population, including hypertension, diabetes, and heart disease, and financial issues for the nation, is resulting in escalation of health care costs. In this study, multivariate logistic regression models were constructed to determine the marginal impact of Twitter data on the propensity for obesity in 36 New York State counties. These counties were divided into five quintiles based on 2013 obesity data. Two regression models, including and excluding social media data, were estimated for each quintile. Data from the United States Department of Agriculture Economic Resource Services (USDA ERS) on food environment were used as control variables when applicable. Geo-enabled tweets from the Tweepy API were divided into four categories using custom keywords. The results from the third quintile, i.e. counties with obesity rates between 25.96% - 27.38%, indicated that social media may be a valid way to track Syndrome X. This regression model, which included four control variables selected using a purposeful approach method, and the Tweepy data, yielded a p-value of .09. The p-values of the nine other logistic regression models ranged from .00 to .51. Two SIR models, including and excluding Twitter data, were estimated for Richmond County to depict future obesity trends. The beta and gamma values for both were calculated using the logistic regressions. More robust models are being estimated to capitalize on these successful regression models, which showed that social media data may be a viable method for tracking (and ultimately preventing) communicable diseases.

- "Semiparametric Regression Analysis of Interval-Censored Data With Informative Dropout"
  Fei Gao, Donglin Zeng, and D. Y. Lin, *University of North Carolina at Chapel Hill*

  Interval-censored data arise when the event of interest can only be ascertained through periodic examinations. In medical studies, subjects may not complete the examination schedule for reasons related to the event of interest. In this paper, we develop a semiparametric approach to adjust for such informative dropout in regression analysis of interval-censored data. Specifically, we propose a broad class of joint models, under which the event time of interest follows a transformation model with a random effect and the dropout time follows a different transformation model but with the same random effect. We consider nonparametric maximum likelihood estimation and develop an EM algorithm that involves simple and stable calculations. We prove that the resulting estimators of the regression parameters are consistent, asymptotic normal, and asymptotically efficient with a covariance matrix that can be consistently estimated through profile likelihood. In addition, we show how to consistently estimate the survival function when dropout represents voluntary withdrawal and the cumulative incidence function when dropout is an unavoidable terminal event. Furthermore, we assess the performance of the proposed numerical and inferential procedures through extensive simulation studies. Finally, we provide an application to data on the incidence of diabetes from a major epidemiological cohort study.

- "Censoring Unbiased Regression Trees and Ensembles"
  Jon Arni Steingrimsson, Liquin Diao, and Robert L. Strawderman, *Johns Hopkins University, University of Waterloo*

  This paper proposes a novel approach to building regression trees and ensemble learning in survival analysis. By first extending the theory of censoring unbiased transformations, we construct observed data estimators of full data loss functions in cases where responses can be right-censored. This theory is used to construct two specific classes of methods for building regression trees and regression ensembles that respectively make use of Buckley-James and doubly robust estimating equations for a given full data risk function. For the particular case of squared error loss, we further show how to implement these algorithms using existing software (e.g., CART, random forests) by making use of a related form of response imputation. Comparisons of these methods to existing ensemble procedures for predicting survival probabilities are provided in both simulated settings and through applications to four datasets. It is shown that these new methods either improve upon, or remain competitive with, existing implementations of random survival forests, conditional inference forests, and recursively imputed survival trees.

- "Missing Information Principle: A Unified Approach for General Left-Truncated and/or Right-Censored Survival Data Problems"
  Yifei Sun, Jing Qin, and Chiung-Yu Huang, *Johns Hopkins University*

  It is well known that truncated survival data are subject to sampling bias, where the sampling weight depends on the underlying truncation time distribution. Recently, there has been a rising interest in developing methods to better exploit the information about the truncation time, thus the sampling weight function, to obtain more efficient estimation. In this paper, we propose to treat truncation and censoring as missing data mechanism and apply the missing information principle to develop a unified framework for analyzing left-truncated and right-censored data with unspecified or known truncation time distributions. Our framework is structured in a way that is easy to understand and enjoys a great flexibility for handling different types of models. Moreover, a new test for checking the independence between the underlying truncation time and survival time is derived along the same line. The proposed hypothesis testing procedure utilizes all observed data and hence can yield a much higher power than the conditional Kendalls tau test that only involves comparable pairs of observations under truncation. Numerical simulation studies with practical sample sizes are conducted to compare the performance of the proposed method with its competitors. The proposed methodologies are applied to a dementia study and a nursing house study for illustration.

- "Case-cohort studies with interval-censored failure time data"
  Qingning Zhou, Haibo Zhou, and Jianwen Cai, *University of North Carolina at Chapel Hill*

  The case-cohort design has been widely used as a means of cost reduction in assembling or measuring expensive covariates in large cohort studies. The existing literature on the case-cohort design is mainly focused on right-censored data. In practice, however, the failure time is often subject to interval-censoring; it is known only to fall within some random time interval. In this paper, we consider the case-cohort study design for interval-censored failure time and develop a sieve semiparametric likelihood approach for analyzing data from this design under the proportional hazards model. We construct the likelihood function using inverse probability weighting and build the sieves with Bernstein polynomials. The consistency and asymptotic normality of the resulting regression parameter estimator are established and a weighted bootstrap procedure is considered for variance estimation. Simulations show that the proposed method works well for practical situations, and an application to data from the Atherosclerosis Risk in Communities (ARIC) study is provided.

- "Empirical likelihood for the bivariate survival function under univariate censoring"
  Haitao Huang and Yichuan Zhao, *Georgia State University*

  The empirical likelihood method is developed for constructing confidence intervals for a bivariate survival function in the presence of univariate censoring. It is shown that the empirical log likelihood ratio has an asymptotic standard chi-square distribution with one degree of freedom. Simulation study

shows that the proposed method outperforms the conventional normal approximation method in finite samples. The Diabetic Retinopathy Data are analyzed for illustration of the proposed procedure.

- "Tracing studies in cohorts with loss-to-follow-up: selection models for optimal efficiency"
  Nathalie C Moon, Leilei Zeng, and Richard J Cook

- "Single Index Models in Proportional Hazard Regression for Precision Medicine"
  Jin Wang, Donglin Zeng, and D.Y. Lin, University of North Carolina at Chapel Hill

  We propose a flexible single-index model to allow complex treatment-covariate interactions and to derive a simple linear treatment rule for personalized treatment decision. Our model extends the proportional hazards models by using a monotone regression function. The inference procedures are based on sieve estimation that includes single-index parameters in the basis expansion and partial likelihood for the baseline hazard function. We obtain the theoretical results by deriving the necessary rates of convergence for the nonparametric estimator of the arbitrary regression function. We provide simultaneous inference on single-index parameters and other regression parameters. Simulation studies are conducted to assess the finite sample performance. An application to a multiple-type cancer study is presented to illustrate our methods.

- "Extended Cox Model by ECM Algorithm for Uncertain Survival Records Due to Imperfect Data Integration"
  Wenjie Wang, Kun Chen, and Jun Yan, University of Connecticut

  In the era of big data, there has been an increasing need to use data integrated from disparate sources to conduct statistical analysis. The potential benefits from data integration, however, may be compromised by the induced data uncertainty due to incomplete/imperfect linkage, causing potential bias in statistical inference. Thus, it is pivotal to take into account the uncertainty associated with data integration. Motivated by a suicide prevention study, we consider a survival analysis setup to handle uncertainty event records arising from data integration. Specifically, a survival dataset constructed from hospital discharge fails to capture the events of interest for all the subjects, and the missing events may be recovered from a complete death record database that contains all the event records of a much larger population. Nonetheless, the original dataset can only be linked to the database by matching basic characteristics of subjects. As such, a censored subject from the original dataset could find multiple possible event times in the second database, which may or may not contain the true event time. We propose an extended Cox regression approach, in which such uncertainty and mismeasurement of survival data are modeled probabilistically. The estimation procedure is derived in the spirit of expectation conditional maximization (ECM) algorithm and profile likelihood function. It reduces to the regular Cox model when there is not uncertainty in the data. The estimation and data fitting performance is evaluated through simulation studies. The proposed method outperforms the naive approaches under slight and severe censoring when the data matching leads to more true outcomes than noise. We show that the extended Cox model is attractive in practice by applying it to the 20052012 suicide attempt data from the State of Connecticut, which suggests interesting and insightful results. Key words: EM algorithm; Expectation conditional maximization; Incomplete data; Profile likelihood; Suicide prevention.

LIDA 2017 Committees

Program Committee

- *Chair*
  - **Mei-Cheng Wang**, Johns Hopkins University
- *Committee Members*
  - **Jianwen Cai**, University of North Carolina at Chapel Hill
  - **Tianxi Cai**, Harvard University
  - **Nilanjan Chatterjee**, Johns Hopkins University
  - **Chrys Caroni**, National Technical University of Athens, Greece
  - **Richard J Cook**, University of Waterloo, Canada
  - **Jason P. Fine**, University of North Carolina, Chapel Hill
  - **Mitchell Gail**, NCI
  - **Lupe Gomez**, Technical University Catalunya, Spain
  - **David Harrington**, Harvard University
  - **Joan Hu**, Simon Fraser University, Canada
  - **Philip Hougaard**, Denmark
  - **Catherine Huber-Carol**, Paris Ren Descartes University, France
  - **Nicholas P. Jewell**, UC Berkeley
  - **Zhezhen Jin**, Columbia University
  - **Niels Keiding**, University of Copenhagen, Denmark
  - **Lynn Kuo**, University of Connecticut
  - **Mei-Ling Ting Lee**, University of Maryland, College Park
  - **Amita Manatunga**, Emory University.
  - **Mounir Mesbah**, University of Pierre et Marie Curie, France
  - **David Oakes**, University of Rochester
  - **Edsel Pena**, University of South Carolina
  - **Ross Prentice**, Fred Hutchinson Cancer Research Center
  - **Douglas Schaubel**, University of Michigan, Ann Arbor
  - **Jonathan Siegel**, Bayer
  - **Robert Strawderman**, University of Rochester
  - **Jianguo Sun**, University of Missouri
  - **Jane-Ling Wang**, UC Davis
  - **Grace Yi**, University of Waterloo, Canada

Local Organizing Committee

- *Co-Chairs*
  - **Ming-Hui Chen**, University of Connecticut
  - **Jun Yan**, University of Connecticut
- *Committee Members*
  - **Dipak K. Dey**, University of Connecticut
  - **Ray Liu**, Takeda Pharmaceuticals International Co.
  - **Shuangge (Steven) Ma**, Yale University
  - **Weiliang Qiu**, Harvard University
  - **Elizabeth D. Schifano**, University of Connecticut
  - *Web:* **Henry Linder**, University of Connecticut

Student Paper and Poster Committee

- *Chair*
  - **Zhezhen Jin**, Columbia University
- *Committee Members*
  - Guoqing Diao, George Mason University
  - Gang Li, UCLA
  - Ian McKeague, Columbia University

Student Committees

- *Coordinator*
  - Hao Li
- *IT*
  - *Head:* Disheng Mao
  - Pedro Ramos
  - Kai Peng
  - Wei Shi
  - Yaohua Zhang
  - Siddhesh Kulkarni
  - Abhishek Bishoyi
  - Daoyuan Shi
  - Liwei Ye
  - Yifan Cao
  - Yizhou Xie
- *Posters*
  - *Head:* Jing Wu
- *Registration*
  - *Head:* Yeongjin Gwon
  - Sudeep Bapat
  - Kangyan Liu
  - Chaoyu Lin
  - Shuyi Liang
  - Yulia Sidi
  - Ziqi Yang
  - Xuan Ren
  - Jun Hu
  - Hakema Hussein
  - Chongliang Luo
  - Fan Zhang
- *Photo*
  - *Aritra Halder*
  - Wenxin Wang
  - Zhe Sun

LIDA 2017 Participants

George Agogo
Arthur Allignol
Per Kragh Andersen
Aparna Anderson
Adin Andrei
Amlie Anota
Masood Anwar
Sudeep Bapat
Alice Benson
Arthur Berg
Paul Bernhardt
Ana Best
Rebecca Betensky
Md Al Masum Bhuiyan
Abhishek Bishoyi
Olivier Bouaziz
Philippe Broet
Chunyan Cai
Yifan Cao
Hongyuan Cao
Marco Carone
Luise Cederkvist
Kwun Chuen Gary Chan
Chung-Chou Chang
Richard Chappell
Qingxia Chen
Yong Chen
Ling-Wan Chen
Baojiang Chen
Ming-Hui Chen
Kun Chen
Ying Chen
Jinbo Chen
Yu Cheng
Li Cheung
Alan Chiang
Steven Chiou
Youngjoo Cho
Yunhee Choi
Parichoy Pal Choudhury
Yunro Chung
Candemir Cigsar
William Cipolli
Rebecca Clark
Sarah Conner
Richard Cook
James Dai
Victor Hugo Lachos Davila
Charmaine Dean
Dipak Dey
Sai Dharmarajan
Guoqing Diao
Liqun Diao
Ivan Diaz
James Dignam
Ying Ding
Pang Du
Ran Duan
Joel Dubin
Jean-franois Dupuy
Jimmy Efird
Kevin Eng
Frank Eriksson
Changyong Feng
Loc Ferrer
Dianne Finkelstein
Youyi Fong
Mitchell Gail
Fei Gao
Tanya Garcia
Lida Gharibvand
Malka Gorfine
Ralitza Gueorguieva
Yeongjin Gwon
Yolanda Hagar
Charles Hall
Summer Han
Ling Han
Sebastien Haneuse
Daniel Haverstock
Zhi (Kevin) He
Wenqing He
Xin He
Glenn Heller
Jiayi Hou
Jeanine Houiwing-Duistermaat
Li Hsu
Jun Hu
Chen Hu
Tianle Hu
Joan Hu
Ming-Yueh Huang
Xuelin Huang
Chiung-Yu Huang
Yen-Tsung Huang
Yijian Huang
Jie Huang
Catherine Huber
Catherine Huber-Carol
Nidal Huniti
Hakema Hussein
Wei-Ting Hwang
Noorie Hyun
Joseph Ibrahim
David James
Ina Jazic
Aksel Jensen
Jong Jeong
Nicholas Jewell
Fei Jiang
Zhezhen Jin
Elizabeth Juarez-Colunga
Jack Kalbfleisch
Christian Kappeler
Theodore Karrison
Eric Kawaguchi
Polyna Khudyakov
Sung Duk Kim
Dong-Yun Kim
Sehee Kim
Uwe Koehn
Xiangrong Kong
Karen Kopciuk
Athanasios Kottas
Siddhesh Kulkarni
Lynn Kuo
Chia-Ling Kuo
Denis Larocque
Jerry Lawless
Youjin Lee
Chi Hyun Lee
Kyu Ha Lee
Hye-Seung Lee
Yehua Li
Liang Li
Chenxi Li
Gang Li
Shanshan Li
Dongdong Li
Hao Li
Yi Li
Yehua Li
Shuyi Liang
Chaoyu Lin
Henry Linder
Piaomu Liu
Dandan Liu
Danping Liu
Kangyan Liu
Piaomu Liu
Ruiyiin Liu
Shuling Liu
Wei Liu
Ying Liu
Shuling Liu
Kangyan Liu
Ji Meng Loh
Dustin Long
Qi Long
Jeffrey Long
Xiang Lu
Xiaodong Luo
Sheng Luo
Chongliang Luo
Shuangge Ma
Ling Ma
Amita Manatunga
Lu Mao
Disheng Mao
Adele Marshall
Marlena Maziarz
Peter McCullagh
Lee McDaniel
Ian McKeague
Alexander McLain
Guadalupe Gomez Melis
Mounir Mesbah
Jessica Minnier
Annette Molinaro
Nathalie Moon
Peter Mueller
Hans-Georg Mller
Terrence Murphy
Bin Nan
Wayne Nelson
Ai Ni
Jing Ning
David Oakes
Fang-Shu Ou
Zhengqing Ouyang
Katherine Panageas
Edsel Pena
Limin Peng
Yingwei Peng
Kai Peng
Michael Pennell
Ruth Pfeiffer
Valerie Poynor
Ross Prentice
Jing Qian
Weiliang Qiu
Peihua Qiu
Pedro Ramos
Meredith Regan
Xuan Ren
Dooti Roy
Satrajit Roychoudhury
Kjetil Rysland
Pl Christie Ryalen
Olli Saarela
Takumi Saegusa
Abdus Sattar
Douglas Schaubel
Thomas Scheike
Thomas Scheike
Elizabeth Schifano
David Schlueter
Tibor Schuster
Damla Senturk
Yongzhao Shao
Hua Shen
Daoyuan Shi
Wei Shi
Yulia Sidi
Jonathan Siegel
Jeffrey Simonoff
Debajyoti Sinha
Tony Sit
Rodney Sparapani
Donna Spiegelman
Jon Steingrimsson
David Stephens
Robert Strawderman
Sundar Subramanian
Yanqing Sun
Yifei Sun
(Tony) Jianguo Sun
Zhe Sun
Rajeshwari Sundaram
Jeremy Taylor
Lu Tian
Ludovic Trinquart
Terrance Tritz
Alex Tsodikov
Pascale Tubert-Bitter
Hajime Uno
Bella Vakulenko-Lagun
Morten Valberg
David Vock
Mei-Cheng Wang
Cuiling Wang
Haiying Wang
Jane-Ling Wang
Jin Wang
Wenxin Wang
Xin Wang
Yue Wang
Zuoheng Wang
Shao-Hsuan Wang
Ying Wei
Sijin Wen
Jing Wu
Rong Wu
Yizhou Xie
Sharon Xie
Jun Xie
Minge Xie
Zhenzhen Xu
Gongjun Xu
Yingying Xu
Xiaonan Xue
Ying Yan
Guohua Yan
Jun Yan
Qing Yang
Ziqi Yang
Yuchen Yang
Lei Yang
Bin Yao
Liwei Ye
Grace Yi
Yildiz Yilmaz-Cigsar
Yan Yuan
Ao Yuan
Wenlin Yuan
Donglin Zeng
Leilei Zeng
Tianyu Zhan
Jiajia Zhang
Yiding Zhang
Yaohua Zhang
Panpan Zhang
Danjie Zhang
Hongbin Zhang
Ying-Ying Zhang
Jiwei Zhang
Bin Zhang
Fan Zhang
Hongbin Zhang
Jiajia Zhang
Jiwei Zhang
Panpan Zhang
Yaohua Zhang

Yiding Zhang
Ying-Ying Zhang
Yuping Zhang
Yichuan Zhao
Yize Zhao
Hongwei Zhao

Sibo Zhao
Yingqi Zhao
Lihui Zhao
Xingqiu Zhao
Lili Zhao
Qiang Zhao

Cheng Zheng
Mengjie Zheng
Jie Zhou
Qingning Zhou
Qian (Michelle) Zhou
Wen Zhou

Liang Zhu
Shijun Zhu
Ruoqing Zhu
Zhaoyin Zhu