COURSE SYLLABUS

**BST/STA 222 Survival Analysis, Fall Quarter 2020 (4 units)**

**Time:** Tuesday/Thursday, 5:10 pm – 7:00pm

**Location:**  Online

**Instructor:**

Lihong Qi, Ph.D.

Division of Biostatistics, Department of Public Health Sciences

Email: [lhqi@ucdavis.edu](mailto:lhqi@ucdavis.edu)

**Office hours:** by appointment.

**Course TA:**

Xiner Zhou

Email: [xezhou@ucdavis.edu](mailto:xezhou@ucdavis.edu)

**Office hours:** 7:00pm – 8:00pm, Thursdays

**Lab:** 6:10-7:00 pm, Thursdays

For both Lab and OH, please use the zoom link with passcode: survival  
<https://ucdstats.zoom.us/j/9681353237?pwd=ZzhLRHlhNG92bHFUaHkwM0Fib1pqQT09>  
Meeting ID: 968 135 3237

**COURSE INFORMATION**

1. **Course Description**:

This course deals with statistical methods analyzing data-to-event data and both theoretical and applied materials will be covered. In particular, this is **not** a course only on how to run a SAS program and interpret its output; you also need to know underlying statistical procedures, including their modeling assumptions and possibly their derivations. we will cover non-parametric, semi-parametric, and parametric models and two-sample test techniques. In addition, we will also demonstrate mathematical and graphical methods for evaluating goodness of fit and introduce the concept of dependent censoring/competing risk. Both SAS and R will be used.

**Organization:**

Tues: Lectures; Thurs: Lecture from 5:10 – 6:00pm; Lab from 6:10pm-7:00pm

1. **Course Prerequisites:**

Prerequisites: best if having taken STA 231A, B (Mathematical Statistics I, II, or STA200 ABC for MS students), BST223: Generalized Linear Models, or equivalent. Students are supposed to know concepts of distributions, statistical estimations and hypothesis testing; linear regression, generalized linear regression, maximum likelihood theory, and to have some experiences in using SAS and R.

# **Course Goals and Objectives**

This course deals with methods analyzing survival times or time-to-event data, which may be censored and/or truncated. The main topics are: 1) estimating a survival curve; 2) comparing two (or more) survival curves; and 3) regression analysis. Specifically,

* Recognize the characteristics of survival data, e.g. censoring and truncation.
* Determine the proper method to be used in analyzing time-to-event data (e.g., parametric, semi-parametric or non-parametric method).
* Understand the assumptions and the theory for the method chosen to analyze the data.
* Apply mathematical and graphical methods to check goodness of fit.
* Perform survival analysis using a computer statistical software package.
* Interpret computer outputs.
* Assess the quality of survival analysis conducted in published research papers.

# **Methods of Instruction and Work Expectations**

* Online lectures are the main method of instruction. Students are expected to attend online class and participate in discussions, read the textbook, finish assignments, 2 projects, and midterm and final summary reports.
* **Zoom attendance is mandatory**.  Requests for exceptions must be made to the instructor at least 3 days prior to the expected absence, or as soon as possible in the case of an emergency.
* Late assignment is not accepted unless with **in advance** permission from the instructor (see below) or with other legitimate reasons (such as illness).
* Your UCD e-mail address is regarded as your official e-mail address. Updates, class information, assignments, and all other correspondence will be sent to your official UCD e-mail address.

1. **Course Text and Readings:**

* **Required:** J.P. Klein and M.L. Moeschberger, **Survival Analysis,** 2nd edition. Springer, 2003. (**KM**)
* Optional:
* P.O Allison, **Survival Analysis using the SAS System**. SAS Institute Inc. 2010
* **Applied Survival Analysis Using R**, Dirk F. Moore, Springer.

# [John D. Kalbfleisch](https://www.google.com/search?hl=en&q=John+D.+Kalbfleisch&stick=H4sIAAAAAAAAAONgVuLRT9c3zEguKKnIMEh_xGjOLfDyxz1hKb1Ja05eY9Tg4grOyC93zSvJLKkUkuJig7IEpPi4UDTyLGIV9srPyFNw0VPwTsxJSstJzSxOzgAArdI6aGAAAAA) and [Ross L. Prentice](https://www.google.com/search?hl=en&q=Ross+L.+Prentice&stick=H4sIAAAAAAAAAONgVuLVT9c3NEzLMjFPrigyecRowS3w8sc9YSn9SWtOXmPU5OIKzsgvd80rySypFJLmYoOyBKX4uVB18ixiFQjKLy5W8NFTCChKBSpLTgUA8CnLq18AAAA). **The Statistical Analysis of Failure Time Data,**2nd Edition. Wiley-Interscience, 2002.

* David G. Kleinbaum and Mitchel Klein. **Survival Analysis: A Self-Learning Text**, Third Edition. Springer, 2011.
* **Software**: SAS and R

1. **Evaluation and Grading**

Course evaluation will be based on course participation, homework assignments, project 1 and project 2, midterm and final summary reports, with weights 15%, 25%, 25% and 25%, 5% and 5% respectively.

* **Course Participation**: Online discussions and webinars. If a student **actively** participates all online discussions and webinars, the student will get 15% of the final grade for participations (5% for attending but not actively participate discussions. Scales will be used to evaluate the level of participation.)
* **Homework assignments:** involve some theoretical problems and running SAS or R programs to analyze data. Please note that it is required to include your SAS or R programs, **only** relevant parts of output, major steps of hand calculations, and necessary interpretations/conclusions.
* Students are allowed to discuss homework problems, however, each one is expected to program and write up **independently. Copying other's work, including computer programs, is cheating or plagiarism, which will lead to an automatic "F" and possible reporting to the University office.**
* Project 1 (25%) will include the following tasks:
* Identify a dataset
* Specify the hypothesis
* Define event and time and censoring/truncation mechanism
* Propose appropriate survival analysis and then perform the analysis to analyze the data
* Writing a final report using the *Publication Manual of the American Psychological Association* format (<https://apastyle.apa.org/>)
* Include software codes in Appendix.
* In the method section, **describe** in detail the primary analysis methods used, including **assumptions, and theoretical derivation**. In the result section, include **interpretations** of your results.
  + Students **CANNOT** collaborate on project 1
  + **Students need to submit a milestone report during the 6th week**
* Project 2 (25%) will include the following tasks:
* Identify a medical paper published in journals such as *JAMA, JAMA Oncol, Cancer, Clinical Trials, Cancer Epidemiol Biomarkers Prev, Journal of the Academy of Nutrition and Dietetics, American Journal of Epidemiology, American Journal of Public Health*.
* Understand event and time and censoring/truncation mechanism in the paper and the primary hypothesis of the paper
* Simulate the data using the values presented in the paper
* Analyze the simulated data using various methods you have learned and explain why you choose these methods.
* Derive formulas for the primary analysis method including assumptions.
* Re-generate major results reported in the paper, including tables and figures.
* Write a report using Statistics in Medicine format (<https://onlinelibrary.wiley.com/page/journal/10970258/homepage/la_tex_class_file.htm>) for Project 2
* Include software codes in Appendix.
* Students **CANNOT** collaborate on project 2
* Project presentation: choose one project for a 10 minutes presentation. **The final reports of both projects will be due on 12/12/2020. More details about the projects will be given in class.**
* Midterm summary report (5%): review and summarize useful things you have learned so far and how you plan to apply them. Due by November 2, 2020
* Final summary report (5%): review and summarize useful things you have learned in this course and how you have applied them and/or plan to apply them in your future research or career. Due by December 12, 2020

1. **Course Communication:**

Course communication will be through Canvas. All course announcements will be posted on the course site so check out the course site constantly!

**ACADEMIC INTEGRITY AND STUDENT BEHAVIOR**

This Code of Academic Conduct exists to support high standards of behavior and to ensure fair evaluation of student learning. Students who violate the Code of Academic Conduct are subject to disciplinary sanctions that include censure, probation, deferred separation, suspension or dismissal from the University of California. Unless specifically authorized by the instructor in writing, misconduct includes, but is not limited to the following:

➤ **Academic misconduct on exams or other coursework**

* Copying or attempting to copy from another student, allowing another student to copy, or collaborating on an exam
* Displaying or using any unauthorized material such as notes, cheat-sheets, or electronic devices
* Looking at another student’s exam
* Not following an instructor's directions
* Talking, texting or communicating during an exam
* Altering assignments or exams for re-grading purposes
* Bringing pre-written answers to an exam
* Having another person take an exam for you, or taking an exam for another student
* Theft of academic work
* Unexcused exit and re-entry during an exam period

➤  **Plagiarism**

* Taking credit for any work created by another person; work includes, but is not limited to books, articles, experimental methodology or results, compositions, images, lectures, computer programs, internet postings
* Copying any work belonging to another person without indicating that the information is copied and properly citing the source of the work
* Using another person’s presentation of ideas without putting such work in your own words or form and/or failing to provide proper citation
* Creating false citations that do not correspond to the information you have used
* Representing your previous work as if it is new work

➤  **Unauthorized collaboration**

* Working together on evaluated coursework without permission of the instructor
* Working with another student beyond the limits set by the instructor
* Providing or obtaining unauthorized assistance on graded coursework

➤  **Misuse of an instructor’s course materials or the materials of others**

* Posting, purchasing, obtaining, sharing, or copying any course materials of an instructor without the explicit written permission of that instructor
* Unauthorized use of another student's work

➤  **Lying or fraud**

* Giving false excuses to obtain exceptions for deadlines, to postpone an exam, or for other reasons
* Forging signatures or submitting documents containing false information
* Making false statements regarding attendance at class sessions, requests for late drops, incomplete grades, or other reasons

➤ **Intimidation or disruption**

* Pressuring an instructor or teaching assistant to regrade work, change a final grade, or obtain an exception such as changing the date of an exam, extending a deadline, or granting an incomplete grade
* Refusing to leave an office when directed to do so
* Physically or verbally intimidating or threatening an instructor, teaching assistant or staff person, including invading personal space, or engaging in any form of harassment
* Repeatedly contacting or following an instructor, teaching assistant or staff person when directed not to do so
* Misusing a classroom electronic forum by posting material unrelated to the course
* Interfering with an instructor’s or teaching assistant’s ability to teach a class, or interfering with other students’ participation in a class

➤ **Misuse of University properties or resources**

* Purchasing or selling (or attempting to) one’s registration in a class

**DISCLAIMER**

This syllabus is a guide for the course and may be changed at the discretion of the course director. If changes are made, you will be notified with ample time to complete additional tasks and will have no resulting penalties.

**Course Calendar / Schedule**

|  |  |  |
| --- | --- | --- |
| **Date** | **Topic/Lessons** | **Reading and Assignment**  *Assignment is due by Monday midnight of the following week* |
| Sep 8th~13th  (Labor Day) | * Overview of course * Introduction of survival data, including examples * Review of logistic and Poisson regression for analyzing survival data * Introduction of NCDB database | KM Ch. 1; DM Ch. 1  HW #1 |
| Sep 14th~20th | * Survival functions: survival, hazard and cumulative hazard functions * Types of censoring and truncation * Likelihood function of survival data | KM Ch. 2~3; DM Ch. 2 |
| Sep 21st~27th (Nonparametric) | * Estimation of survival, hazard and cumulative hazard functions * Confidence interval for survival function | KM Ch. 4~6; DM Ch. 3  HW #2 |
| Sep 28th~Oct 4th | Hypothesis testing:   * Log-rank and Wilcoxon tests * Trend and stratified tests * Power and sample size calculations | KM Ch. 7; DM Ch. 4 & 11  HW #3 |
| Oct 5th~11th  (Semi-parametric) | Proportional hazards (PH) model   * Model characteristics and specification * Model estimation * Hypothesis testing | KM Ch. 8.1~8.6; DM Ch. 5  HW #4 |
| Oct 12th~18th | Model building and selection   * Variable selection * Model selection * Model evaluation | KM Ch. 8.7; DM Ch. 6  HW #5 |
| **Oct 19th~25th** | * **Midterm (closed book)** | **KM Ch. 1~8; DM Ch. 1~6** |
| Oct 26th~Nov 1st | Model diagnostics   * Assess goodness of fit using residuals * Evaluate PH assumption | KM Ch. 11; DM Ch. 7 |
| Nov 2nd~8th | Extension of PH model   * Time dependent covariates * Stratified PH model   Additive hazard regression | KM Ch. 9; DM Ch. 8  HW #6 |
| Nov 9th~15th | Analysis of competing risk data   * Cause-specific hazard function * Cumulative incidence function * Sub-distribution hazard regression | KM Ch. 2.7, 4.7; DM Ch. 9  HW #7 |
| Nov 16th~22nd  (Parametric model) | * Parametric survival models * Recurrent event data * Frailty model (clustered survival data) | KM Ch. 12~13; DM Ch. 10 |
| Nov 23rd~29th  (Thanksgiving) | * Random forest for survival data * Model validation | ***Note: No Webinar!*** |
| Nov 30th~Dec 6th | Final project   * 10~15 min online presentation for scientific researchers * 15~20 page scientific final report | **Online presentation on Thursday (12/3) 9:00~10:30AM** |