

# Data Studio

1:30–3:00pm, Wednesday, 20 May 2020

Videoconference: <https://stanford.zoom.us/j/91103071808>

**Investigator:** Lihua Lei Postdoctoral Fellow, Statistics  
**Investigator:** Emmanuel Candès Professor, Statistics  
**Investigator:** Chiara Sabatti Professor, Biomedical Data Science

**Title:** Survival Analysis of COVID-19 Patients

## Summary:

The COVID-19 pandemic has placed extraordinary demands on health systems. These unprecedented demands create the urgent need for medical resource allocation. Our goal is to build a transparent data-driven system that will assist decision making for front-line workers. The first step of our project is to build a patient-level risk assessment tool based on survival analysis. We integrated over 15 publicly available data sources and performed extensive data cleaning, de-duplication, and filtering to build a database for more than 10 countries and regions in the world. For each subgroup, we performed elementary survival analyses and presented our results on a website (<https://zren.shinyapps.io/covid19/>). The main challenge we are facing is to handle the missing values (e.g. age and gender) in those publicly available datasets that are unlikely to be missing at random. We therefore use the population statistics and apply inverse probability weighting techniques to adjust for the non-random missing values under some assumptions about the missingness mechanism. We further propose a method without those assumptions based on partial identification, although it is conservative and computationally intensive.

## Questions:

1. It appears that the missing values are ubiquitous in retrospective cohort studies. What is the standard way to address this issue?
2. Are there better ways to utilize population statistics that rely on fewer assumptions to adjust for non-random missing values?
3. Recovery and death seem to be competing risks.
  - (a) Is it appropriate to perform competing risk analysis in this setting?
  - (b) If we analyze each of them separately, is our analysis biased?
4. We will likely get access to some EHR data soon. This will provide more covariates but meanwhile renders the parametric survival models more implausible. What are reliable and interpretable techniques for survival data with potentially complicated covariates?

## **Zoom Meeting Information**

Topic: Workshop: Data Studio

Time: May 20, 2020 13:30 Pacific Time (US and Canada)

Meeting ID: 911 0307 1808

Join from PC, Mac, Linux, iOS or Android:

<https://stanford.zoom.us/j/91103071808>

Or iPhone one-tap (US Toll):

+18333021536,,91103071808# or

+16507249799,,91103071808#

Or Telephone:

Dial: +1 650 724 9799 (US, Canada, Caribbean Toll) or

+1 833 302 1536 (US, Canada, Caribbean Toll Free)

Meeting ID: 911 0307 1808

International numbers available:

<https://stanford.zoom.us/j/adEC2ZdjuK>

SIP: 91103071808@zoomcrc.com

**For more information about Data Studio:**

<https://med.stanford.edu/dbds/programs/data-studio.html>