Jonathan Luu

Curriculum Vitae

408-889-3972
jluu@g.harvard.edu
github.com/luuj
jonathanluu.com

Education

Harvard University 2019-Present
Graduate School of Arts and Sciences
Doctor of Philosophy Candidate in Biostatistics

University of Southern California 2017-2019
Keck School of Medicine
Master of Science in Biostatistics

University of Southern California 2013-2017
Viterbi School of Engineering
Bachelor of Science in Computer Engineering and Computer Science

Research Experience

Expanding the two-part model for clustered semi-continuous data truncated by death (2022)
Principal investigator: Sebastien Haneuse

- Context: Is there a better method for comparing cost and healthcare utilization data in nursing homes?
- Developed new methodology to analyze clustered semi-continuous data that incorporates the semi-competing risk of death
- Made a Bayesian semi-parametric framework for random effects in a logistic-log-normal model
- Created joint metrics that incorporate the two-part nature of the data
- Applied new model and metrics to Medicare dataset consisting of multistate, multi-year, longitudinal data of 20 million nursing home residents
- URL: https://github.com/luuj/Semi-continuous-Bayesian-Modeling

Duration of viral shedding and culture positivity with post-vaccination breakthrough delta variant infections (2021)
Principal investigator: Mark J. Siedner

- Context: Isolation and distancing practices are fundamental elements of COVID-19 epidemic control. Should we extend the recommended 5 days of isolation after a positive test?
- Collected longitudinal viral load, viral culture samples, and CT values on MGH employees who tested positive for SARS-CoV-2
- Analyzed differences between delta and non-delta variants and vaccine types (Pfizer, Moderna, J&J)
- Ran survival analyses on negative viral culture, CT values >30, and undetectable viral load. Kaplan-Meier and trajectory spaghetti plots were made to summarize the data
- Quadratic and cubic splines were used in a simple linear regression to create a predictive line for delta and non-delta plots. Hazard ratios were calculated using cox-proportional hazards models
- URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8855795/

Estimating the Treatment Effect in Randomized Trials with Correlated Time-to-event Outcomes (2020)
Principal investigator: Rui Wang

- Context: For unclustered randomized clinical trials with correlated individuals, are naïve analyses just as effective?
- Compared the performance of various analysis strategies, including naive analysis ignoring correlation, robust sandwich variance, and randomization-based inference, for survival outcomes through simulation
- URL: https://github.com/luuj/Clustered-RCT-simulations

LOFT-HF Sample Size Re-estimation (2020)
Principal investigator: Rui Wang

- Context: Can we re-calculate sample sizes for power estimation using blinded aggregate data?
- Re-estimated nuisance parameters that had an impact on the power calculation of a trial.
- Ran a comprehensive simulation study, designed to mimic the LOFT trial, to determine the impact of increasing the sample size and/or follow-up time on study power and overall type I error.
- Simulated additional scenarios by changing the year of follow-up after all patients were accrued, as well as increasing the number of participants after 1 year of follow-up.
- URL: https://github.com/luuj/Recurring-endpoints---SS-reestimation

A Phase I / II Study of E7389 Halichondrin B analog (NSC # 707389) in Metastatic Urothelial Tract Cancer and Renal Insufficiency (2019)
Principal investigator: Susan Groshen

- Context: How effective is this new cancer drug in treating bladder cancer?
- Analyzed phase II trial data to generate descriptive baseline and demographic statistics
- Condensed adverse events into toxicity tables for DSMC report
- Created response tables and Kaplan-Meir plots for progression-free survival and overall survival
- Ran multivariate cox regression for progression-free survival
- URL: https://github.com/luuj/Urothelial-carcinoma-study

A Simulation Evaluation of the Effectiveness and Usability of the 3+3 Rules-based Design for Phase I Clinical Trials (2019)
Principal investigator: Susan Groshen

- Context: How efficient is the 3+3 algorithm for phase I clinical trials?
- Created a simulation program using the 3+3 algorithm to evaluate the design's properties in various phase I clinical trial scenarios
- Calculated descriptive statistics and generated plots
- Ran linear, Poisson, and logistic regression on six endpoints for prediction modeling and hypothesis testing
- Checked simulation validity with goodness of fit test
- URL: https://github.com/luuj/3-3-Simulation

Teaching Experience

Teaching Assistant - Applied Survival Analysis (BST223) 2021-2022
Professor: Sebastien Haneuse

- Taught weekly virtual labs and attended lectures
- Put together labs and homework assignments for students to complete
- Put together rubric and solutions for homework assignments
- Graded homework assignments, midterm, and final exam
- Held weekly office hours and met during weekly TA meetings

Teaching Assistant – Intro to Data Science (BST260) 2021
Professor: Heather Mattie

- Taught weekly labs (both in-person and virtual) and attended lectures
- Helped students get setup with GitHub
- Wrote R scripts to automate setup procedures
- Helped students get setup and familiar with R and RStudio
- Graded homework assignments, midterm, and final exam
- Held weekly office hours and met during weekly TA meetings

Teaching Assistant - Survival Methods in Clinical Research (BST224) 2022
Professor: Long Ngo

- Helped students get familiar with R and RStudio
- Graded weekly quizzes and project
- Held weekly office hours and met during weekly TA meetings

Biostatistics Consulting Center – Harvard T.H. Chan School of Public Health 2021-2022
Supervisor: Marcello Pagano

- Free consulting service for students and post-docs from HSPH and HMS
- Offered guidance on study design, analysis planning, statistical programming, etc.
- Assisted with research projects, grant submissions, and dissertations
- Participated in bi-monthly meetings where we presented and discussed client submissions

StatStart - Harvard T.H. Chan School of Public Health 2021-2022
Supervisor: Marcello Pagano

- Summer program for high school students interested in data science and computing
- Taught programming in R and basic statistics in the form of lectures and lab
- Helped develop computational thinking and problem-solving skills
- Guided students in a final project and presentation

Work Experience

Research Assistant – Enguídanos Lab 2018-2019
Supervisor: Susan Enguídanos

- Context: Ran clinical trial to compare hospital vs. home-based palliative care
- Managed excel files containing patient data sent from Blue Shield
- Created conditional logic surveys and scripts with REDCap
- Recorded and monitored new inpatient referrals with REDCap
- Wrote scripts to summarize demographic information, ineligibility criteria, and patient concerns for presentation to funding agencies
- URL: https://github.com/luuj/Palliative-care-clinical-trial

CIO's Assistant – USC Credit Union 2015-2016
Supervisor: David Schauer-West

- Context: Student IT worker
- Managed employee accounts with Active Directory/Microsoft Exchange
- Kept banking applications updated with Configuration Manager
- Completed help-desk tickets using Kayako and VNC Viewer

Computer Science Projects

Personal Website

- Wrote up HTML/CSS code to build a personal portfolio
- GitHub URL: https://github.com/luuj/luuj.github.io
- URL: www.jonathanluu.com

OSRS Plugins 2021

- Created plugins that can be used in a video game called RuneScape
- URL: https://github.com/luuj/BlueLite-Inferno-Plugin

Polar Deep Search Engine 2016

Principal investigator: Chris Mattmann

- Crawled the deep-web using Apache Nutch to collect polar-related data
- Indexed collected data with Apache Solr to setup database for queries
- Created data visualizations using Banana, Facetview, and D3.js APIs
- Developed USC-branded website with Wicket and Twitter Bootstrap
- URL: http://www-scf.usc.edu/~sanchitl/ufo.usc.edu-gh-pages/html/index.html

Destructo-Block 2015

- Developed an animated Android puzzle game
- Constructed GUI using Android Studio
- Implemented working leaderboard and notification services
- URL: https://github.com/luuj/Destructo-Block

Pokemon Battle Simulator 2015

- Utilized Java Swing to create animated battling simulator
- Applied multi-threading and networking for multiplayer battle & live chat

- Generated player stats with MySQL database and networking protocol
- URL: https://github.com/luuj/Battle-Simulator

Web Parser 2015
- Implemented Google's web parsing algorithm to crawl the internet
- URL: https://github.com/luuj/Web-Parser

Digital Neuron 2014
- Assembled digital neuron that fired signal upon receiving input combination
- Built input memory and combinational logic using MOS VLSI circuit design
- Used Cadence to create schematics/layouts out of PMOS/NMOS transistors
- Ran Spectre simulations to test for optimal clock speed and temperature
- URL: https://github.com/luuj/Arduino-Projects

Skills

Programming Languages
- Java
- C++
- Python
- HTML/CSS

Statistical Programming Languages
- R
- SAS
- Stata

Other
- LaTeX
- Git + GitHub
- Cadence

Relevant Coursework

Harvard University
- Probability I (BST230)
- Methods I (BST232)
- Statistical Inference I (BST231)
- Intro Data Structures & Algorithms (BST234)
- Adaptive Clinical Trials (BST254 II)
- Cancer Genome Data Science (BST283)
- Intro Social & Bio Networks (BST267)
- Responsible Conduct of Research (HPM548)
- Topics in Clinical Trials (BST238)
- Stat & Quant Methods for Pharmaceutical Regulatory Science (BST217)
- Pharmacoepidemiology (EPI221)
- Adv Regression and Statistical Learning (BST235)
- Stat Computing & Learning (STAT221)
- Analysis of Multivariate & Longitudinal Data (BST245)

- Bayesian Methodology (BST249)
- Health Survey Samples (BST239)
- Consultation (BST312)
- Applied Survival Analysis (BST223)
- Intro to Data Science (BST260)
- Analysis of Failure Time Data (BST244)

University of Southern California

- Principles of Biostatistics (PM510)
- Principles of Epidemiology (PM512)
- Data Analysis (PM511A)
- Statistical Methods for Epidemiology (PM518A)
- Design of Clinical Studies (PM523)
- Statistical Programming in R (PM560)
- Data Analysis (PM511B)
- Introduction to the Theory of Probability (PM522A)
- Introduction to the Theory of Inference (PM522B)
- Advanced Statistical Computing (PM520)
- Experimental Designs (PM513)
- Introduction to Programming (CSCI103)
- Calculus II (MATH126)
- Discrete Methods in Computer Science (CSCI170)
- Data Structures and Object-Oriented Design (CSCI104)
- Principles of Software Development (CSCI201)
- Calculus III (MATH226)
- Linear Algebra and Linear Differential Equations (MATH225)
- Introduction to Algorithms and Theory of Computing (CSCI270)
- Design and Construction of Large Software Systems (CSCI401)
- Parallel and Distributed Computation (EE451)
- Introduction to Computer Networks (EE450)
- Introduction to Probability and Statistics (EE364)
- Professional C++ (ITP435)
- Programming Graphical user Interfaces (ITP368)
- Computer Systems Organization (EE457)
- Introduction to Operating Systems (CSCI350)