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 (2023)
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
- GitHub: https://github.com/luuj/Semi-continuous-Bayesian-Modeling

HaSET program: Analyzing stunting of newborns in Ethiopia (2022)
Principal investigators: Frederick Goddard, Grace Chan

- Context: Is there a significantly higher rate of stunting in newborns in Ethiopia?
- Collected longitudinal height and weight data of over 20,000 newborns, starting from birth up to 24 months, in Ethiopia
- Analyzed data with significant measurement error to approximate prevalence, incidence, and reversal of stunting in Ethiopia
- Modeled data using a generalized linear mixed model with piecewise splines to remove likely incorrect observations
- Paper: https://www.medrxiv.org/content/10.1101/2023.05.20.23290246v1

Addressing incomplete and missing electronic health records data in implementation science (2022)
Principal investigator: Cristina Huebner Torres

- Context: How can we analyze health center screening data to improve screening procedures at Caring Health Center?

- Collected screening data with over 250,000 observations with the goal of answering questions about social determinants of health
- Performed exploratory analysis focusing on race, ethnicity and language
- Looked at a variety of missingness definitions based on which screening questions were answered
- Paper: In progress

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
- Paper: 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
- GitHub: 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.
- GitHub: https://github.com/luuj/Recurring-endpoints---SS-reestimation
- Paper: https://www.thelancet.com/journals/landia/article/PIIS2213-8587(19)30346-8/fulltext

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
- GitHub: https://github.com/luuj/Urothelial-carcinoma-study
- Paper: https://pubmed.ncbi.nlm.nih.gov/22198425/

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
- GitHub: https://github.com/luuj/3-3-Simulation

Work Experience

Intern – Bristol Myers Squibb 2023
Exploring correlation between surrogate endpoints and overall survival in cancer trials
Principal Investigator(s): Guotao Chu, Charles Xiaochen Zhu

- Context: Quantify the relationship between surrogate endpoints (progression-free survival and objective response rate) and overall survival
- Compared several correlation coefficients (Pearson's, Spearman's Rho, Kendall's Tau, Harrell's C-index) to check Prentice criteria of a valid surrogate
- Applied weighted OLS model to trial level data and Bayesian normal-induced copula model to subject level data
- Analyzed cancer data stratified by indication and treatment type (chemo vs. immunotherapy vs. combination of both)

Research Assistant – Boston College 2022-2023
Vaccine hesitancy among Latinx adults - a cluster-randomized crossover trial (2023)
Principal Investigator(s): Kirsten Davison, Sebastien Haneuse

- Context: Can we improve COVID-19 vaccine uptake for Latinx adults using motivational interviewing and behavioral health services?
- Used electronic prompting to notify clinicians when to perform intervention, while providing easy hand-off to nurses with vaccination access at the point of care
- Designed as a multiple-period cluster-randomized crossover trial within four programs at EBNHC
- Examined theory-based elements of vaccine hesitancy on the causal pathway between the intervention and vaccine uptake
- Created the data infrastructure and GitHub codebase for the beginning stages of the trial

Research Assistant – Enguídanos Lab 2018-2019
Expanding access to home-based palliative care: a randomized controlled trial protocol
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
- GitHub: https://github.com/luuj/Palliative-care-clinical-trial
- Paper: https://pubmed.ncbi.nlm.nih.gov/31486727/

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

Teaching Experience

Teaching Assistant - Applied Survival Analysis (BST223) 2021-2023

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) 2021-2022

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 2021-2023

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

Computer Science Projects

Personal Website 2023

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

OSRS Plugins 2021

- Created plugins that can be used in a video game called RuneScape
- GitHub: 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
- Website: 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
- GitHub: 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
- GitHub: https://github.com/luuj/Battle-Simulator

Web Parser 2015

- Implemented Google's web parsing algorithm to crawl the internet
- GitHub: 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
- GitHub: https://github.com/luuj/Arduino-Projects

Skills

Programming (from most proficient to least): C++, Java, R, Python, SAS, HTML/CSS, Stata, C, C#, Ruby, Julia, Stan

Software: Microsoft Office, Adobe Suite, AutoHotkey, Terminal, Linux, Bootcamp, Git/GitHub, LaTex, Cadence, IT experience

Typing WPM: 175

Poster Presentations

Pfizer Pharmaceutical Careers & Postdoctoral Opportunities Educational Event	2022
International Conference on Health Policy Statistics	2023
ENAR	2023
JSM	2023