Jonathan Luu

Curriculum Vitae

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

Education

Harvard University 2019-2024
Graduate School of Arts and Sciences
Doctor of Philosophy 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

Comparing methods for multiple period cluster randomized crossover trials 2024
Principal investigators: Sebastien Haneuse, Kirsten Davison

- Context: Are individual-level methods valid for analyzing MPCRCTs with few clusters
- Compared performance of marginal models via GEE and mixed effects models with relevant small-sample adjustments through simulation
- Used simulation results to support an analysis plan for an ongoing clinical trial looking to improve vaccine uptake for Latinx adults through motivational interviewing

Measuring performance for costs and healthcare utilization in nursing homes 2024
Principal investigators: Sebastien Haneuse, Sharon-Lise Normand

- Context: How can we better compare and profile nursing homes by expanding upon existing metrics?
- Proposed six novel metrics to compare hospitalization costs while acknowledging the effect of mortality in the frail environment of nursing homes
- Highlighted differences in performance between proposed and existing metrics through simulation
- Applied metrics on CMS national dataset consisting of longitudinal data of nursing home residents

Hierarchical models for longitudinal clustered semi-continuous data subject to mortality 2023
Principal investigators: Sebastien Haneuse, Nina Joyce

- Context: How can we create a model to analyze complex dependent data and apply it towards a national nursing home dataset?
- Created a novel model in a discrete-time framework that jointly acknowledges clustering, a non-standard outcome distribution, and the semi-competing risk of mortality
- Estimation was run using a flexible Bayesian framework with a combination of random-scan Gibbs sampling and Metropolis-Hastings
- Applied model on CMS national dataset consisting of longitudinal data of nursing home residents

Prevalence, incidence, and reversal pattern of childhood stunting from birth to age 2 years 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

Determinants of social risk screening and response equity in community health centers 2022
Principal investigator: Cristina Huebner Torres

- Context: How can we analyze health center screening data to improve screening procedures at a community 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 while considering a variety of missingness definitions based on which screening questions were answered
- Paper: https://www.ajpmonline.org/action/showPdf?pii=S0749-3797%2823%2900094-6

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.
- Paper: https://www.thelancet.com/journals/landia/article/PIIS2213-8587(19)30346-8/fulltext

A phase I / II study of E7389 Halichondrin B analog in metastatic urothelial tract cancer 2019
Principal investigator: Susan Groshen

- Context: How effective is Halichondrin B analog for treating bladder cancer?
- Analyzed phase II trial data to generate descriptive statistics and summarize adverse events to report to the DSMC
- Created Kaplan-Meir plots and ran Cox regression for progression-free survival and overall survival
- 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 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
- Paper: https://www.proquest.com/openview/88a141f4bfd493b69231606fc30a02cb/1?pq-origsite=gscholar&cbl=18750&diss=y

Work Experience

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

- Context: How can we quantify the relationship between surrogate endpoints (e.g. 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 2022-2024
Vaccine hesitancy among Latinx adults
Principal Investigators: Kirsten Davison, Sebastien Haneuse

- Context: Can we improve COVID-19 vaccine uptake for Latinx adults using motivational interviewing and behavioral health services?
- Helped design electronic prompting to notify clinicians through EPIC when to perform intervention
- Ran power calculations and calculated operating characteristics through simulation for this multiple-period cluster-randomized crossover trial conducted at four health clinics in Boston
- 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: A university-based clinical trial comparing hospital vs. home-based palliative care
- Managed excel files containing patient data sent from Blue Shield insurance
- Recorded new inpatient referrals using conditional logic surveys through REDCap
- Wrote scripts to summarize demographic information, ineligibility criteria, and patient concerns for presentation to funding agencies
- Paper: https://pubmed.ncbi.nlm.nih.gov/31486727/

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

- Context: Student IT worker
- Helped manage employee accounts through Active Directory
- Completed help-desk tickets

Teaching Experience

Teaching Assistant - Applied Survival Analysis (BST223) 2021-2023
Supervisor: Sebastien Haneuse

- Taught weekly labs and attended lectures
- Put together labs, homework assignments, grading rubrics, and assignment solutions
- Graded homework assignments, midterm, and final exam
- Held weekly office hours and debriefed during weekly TA meetings

Teaching Assistant – Intro to Data Science (BST260) 2021-2022
Supervisor: Heather Mattie

- Taught weekly labs and attended lectures
- Graded homework assignments, midterm, and final exam
- Held weekly office hours and debriefed during weekly TA meetings
- Helped students get setup with R, RStudio, and GitHub
- Wrote R scripts to automate setup procedures for grading purposes

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

- Helped students get familiar with R and RStudio
- Graded weekly quizzes and monthly projects
- Held weekly office hours and debriefed 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 Harvard School of Public Health and Harvard Medical School
- Offered guidance on study design, analysis planning, and statistical programming
- Assisted with research projects, grant submissions, and dissertations

StatStart - Harvard T.H. Chan School of Public Health 2021-2023
Supervisor: Marcello Pagano

- StatStart is a summer program for high school students interested in data science and computing
- Taught statistical programming in R through lecture and applied labs
- Helped develop computational thinking and problem-solving skills by guiding students through a final project and presentation

Computer Science Projects

Personal Website 2023

- Built a personal portfolio to highlight recent works
- 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 Old School RuneScape
- GitHub: https://github.com/luuj/BlueLite-Inferno-Plugin

Polar Deep Search Engine 2016

Principal investigator: Chris Mattmann

- Crawled the deep-web using a combination of Apache Nutch, Solr, Banana, and D3.js APIs to collect and index polar-related data
- 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 with working leaderboard and notification services
- GitHub: https://github.com/luuj/Destructo-Block

Pokemon Battle Simulator 2015
- Utilized Java Swing and MySQL to create an animated Pokemon-inspired battling simulator
- Used multi-threading and networking for multiplayer battle, live chat, and generating player stats
- 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 in Cadence and Spectre that fired signal upon receiving input combination using MOS VLSI circuit design
- GitHub: https://github.com/luuj/Arduino-Projects

Poster Presentations

ASA Joint Statistical Meeting (JSM) 2023
Eastern North American Region Spring Meeting (ENAR) 2023
International Conference on Health Policy Statistics 2023
Pfizer Pharmaceutical Careers & Postdoctoral Opportunities Educational Event 2022