

# Nima Hejazi

✉ [nhejazi@berkeley.edu](mailto:nhejazi@berkeley.edu)  
🌐 [nimahejazi.org](http://nimahejazi.org)  
in [nshejazi](#)  
🐦 [nshejazi](#)  
🌐 [nhejazi](#)

## Curriculum Vitæ

---

updated: July 2018

### Current Positions

- *Doctoral Student*, Interdepartmental Group in Biostatistics, University of California, Berkeley  
Advisors: Mark van der Laan and Alan Hubbard

### Research Interests

- Methodology nonparametric statistics, causal inference, machine learning, semiparametric theory, targeted minimum loss-based estimation, data-adaptive inference, survival analysis, statistical genomics, reproducible research, statistical software development, data science.
- Applications precision medicine, computational biology and bioinformatics, epidemiology, randomized clinical trials, vaccine efficacy trials, healthcare policy, public policy.

### Education

- 2016–present **Doctor of Philosophy (Ph.D.) in Biostatistics**, *University of California, Berkeley*.  
Designated emphasis in Computational and Genomic Biology (Center for Computational Biology)  
*Committee*: [Mark van der Laan](#) (co-chair), [Alan Hubbard](#) (co-chair), [Nicholas Jewell](#), [Fernando Pérez](#)  
*Dissertation*: Topics in Nonparametric Causal Inference and Statistical Data Science
- 2016–2017 **Master of Arts (M.A.) in Biostatistics**, *University of California, Berkeley*.  
*Committee*: Mark van der Laan (co-chair), Alan Hubbard (co-chair), Martyn Smith  
*Thesis*: Generalized Application of Empirical Bayes Statistics to Asymptotically Linear Parameters
- 2011–2015 **Bachelor of Arts (B.A.)**, *University of California, Berkeley*.  
Molecular & Cell Biology, Psychology, Public Health (triple major)

#### Accreditation, Open Education, and Training

- 10/2017 **Instructor Training Program Certification**, *Software Carpentry*.
- 04–11/2016 **Genomic Data Science Specialization**, *Johns Hopkins University via Coursera*.
- 02–12/2015 **Data Science Specialization**, *Johns Hopkins School of Public Health via Coursera*.

### Research Experience

- 01/2017–present **Graduate Student Researcher**, *University of California, Berkeley, Berkeley, CA*.  
Advisor: Prof. Mark J. van der Laan, Division of Biostatistics & Department of Statistics
- Development of theory and methods for nonparametric statistical and causal inference, including stochastic treatment regimes, machine learning, survival analysis, and data-adaptive inference.
  - Development and founding member of a software ecosystem for targeted statistical learning and inference in collaboration with an initiative of the Bill and Melinda Gates Foundation.

- 08/2015–  
present **Graduate Student Researcher**, *University of California, Berkeley*, Berkeley, CA.  
Advisor: Prof. Alan E. Hubbard, Division of Biostatistics
- Development of methods for robust statistical inference with applications in high-dimensional biology, including data-adaptive hypothesis testing and moderated statistics for targeted statistical learning.
  - Statistical analysis of high-dimensional genomics and epigenetics data, as well as bioinformatical support for experimental work in molecular biology and environmental epidemiology.
- 03/2014–  
05/2015 **Undergraduate Student Researcher**, *University of California, Berkeley*, Berkeley, CA.  
Supervised by Prof. Mark T. D’Esposito, Helen Wills Neuroscience Institute
- Study of the properties of resting-state voxel connectivity in the visual system using fMRI data.
- 05/2012–  
02/2015 **Undergraduate Student Researcher**, *UCSF Medical Center*, San Francisco, CA.  
Supervised by Dr. Paul S. Larson, Department of Neurological Surgery
- Analysis of outcomes from deep brain stimulation surgery using interventional MR imaging.
- 08/2012–  
05/2014 **Undergraduate Research Assistant**, *University of California, Berkeley*, Berkeley, CA.  
Supervised by Prof. Robert T. Knight, Helen Wills Neuroscience Institute
- Study of disruptions in lexical selection from neurological disease based on human EEG recordings.
- 08/2011–  
05/2013 **Undergraduate Research Assistant**, *University of California, Berkeley*, Berkeley, CA.  
Supervised by Prof. Leslea J. Hlusko, Department of Integrative Biology
- Data collection for studies in quantitative genetics concerning the evolution of hominid dentition.

## Industry Experience

- 08/2018–  
present **Biostatistical Analyst (Intern)**, *Kaiser Permanente Division of Research*, Oakland, CA.
- Statistical methods for ensemble machine learning, causal inference, and data visualization applied to large-scale observational healthcare studies examining diabetes onset and related disorders.
- 01/2016–  
03/2016 **Data Scientist (Intern)**, *Cogitativo, Inc.*, Berkeley, CA.
- Applied modern statistical methods, data analytic tools, and data product design to problems in the healthcare services space, producing client-side tools for interactive data mining and visualization.
- 03/2016–  
present **Private Consultant**.
- Consultation on problems in statistics, data analysis, and software development for data science.

## Awards, Honors, Scholarships

- June 2018 Travel Scholarship, BioC 2018: Where Software and Biology Connect (annual conference), Bioconductor: Open Source Software for Bioinformatics
- March 2018 The Wellness Scholarship in Honor of Chin Long Chiang, School of Public Health, University of California, Berkeley
- 2017–2018 NIH BD2K Training Grant (T32LM012417-02), Biomedical Big Data Training Program, University of California, Berkeley
- July 2017 Travel Scholarship, Summer Institute in Statistics for Big Data, Department of Biostatistics, University of Washington
- May 2017 Honorable mention (rank 3 – 6), Thomas R. Ten Have Award for junior researchers, for the poster “Nonparametric variable importance for continuous exposures with applications in high-dimensional biology”, Atlantic Causal Inference Conference
- June 2015 Travel Scholarship, Summer Institute in Biostatistics, Department of Biostatistics & Medical Informatics, University of Wisconsin — Madison

## Publications

### Statistical Applications

- [1] **N. Hejazi**, C. Schiffman, and X. Zhou, "Review of 'statistical analysis of numerical preclinical radiobiological data'," *ScienceOpen Research*, 2017. doi: 10.14293/s2199-1006.1.sor-stat.afhtwc.v1.rxidzs. [Online]. Available: <https://dx.doi.org/10.14293%2Fs2199-1006.1.sor-stat.afhtwc.v1.rxidzs>

### Refereed Statistical Software

- [1] **N.S. Hejazi**, W. Cai, and A.E. Hubbard, "biotml: Targeted Learning for biomarker discovery," *The Journal of Open Source Software*, vol. 2, no. 15, 2017. doi: 10.21105/joss.00295. [Online]. Available: <https://dx.doi.org/10.21105/joss.00295>
- [2] J.R. Coyle and **N.S. Hejazi**, "origami: A generalized framework for cross-validation in R," *The Journal of Open Source Software*, vol. 3, no. 21, 2018. doi: 10.21105/joss.00512. [Online]. Available: <https://dx.doi.org/10.21105/joss.00512>
- [3] W. Cai, A. Hubbard, and **N.S. Hejazi**, "adaptest: Data-adaptive statistics for high-dimensional multiple hypothesis testing in R," *The Journal of Open Source Software*, submitted.

### Currently in Preparation

- [1] **N.S. Hejazi**, R.V. Phillips, A.E. Hubbard, and M.J. van der Laan, "methyvim: Targeted, robust, and model-free analysis of differential methylation in R."
- [2] **N.S. Hejazi**, S. Kherad-Pajouh, M.J. van der Laan, and A.E. Hubbard, "Variance stabilization of targeted estimators of causal parameters in high-dimensional settings." [Online]. Available: <https://arxiv.org/abs/1710.05451>
- [3] **N.S. Hejazi**, K. Benac, and N.P. Jewell, "Immortal time bias analyses: Time-varying survival regression models and delayed entry approaches."
- [4] **N.S. Hejazi**, M.J. van der Laan, H.E. Janes, P.B. Gilbert, and D.C. Benkeser, "Efficient estimation and robust inference on treatment effects of stochastic interventions under two-phase sampling designs."
- [5] **N.S. Hejazi** and M.J. van der Laan, "Optimal nonparametric estimation of constrained functional parameters."
- [6] —, "Targeted maximum likelihood estimation with constraint-specific paths."
- [7] **N.S. Hejazi**, J.R. Coyle, and M.J. van der Laan, "On the scalability of the Highly Adaptive Lasso estimator."
- [8] **N.S. Hejazi**, A.E. Hubbard, and M.J. van der Laan, "Adaptive and targeted identification of differentially methylated positions and regions with ensemble machine learning."
- [9] D.C. Benkeser, A. Chambaz, and **N.S. Hejazi**, "A guided tour of Targeted Learning."
- [10] W. Cai, **N.S. Hejazi**, and A.E. Hubbard, "Data-adaptive statistics for multiple hypothesis testing in high-dimensional settings." [Online]. Available: <https://arxiv.org/abs/1704.07008>

### Dissertations and Theses

- [1] **N.S. Hejazi**, "Generalized application of empirical Bayes statistics to asymptotically linear parameters," Master's thesis, University of California, Berkeley, 2017.

---

## Software

### *Open Source Software and Projects*

- [1] **N.S. Hejazi** and A.E. Hubbard, “biotmle: Targeted Learning for biomarker discovery with moderated statistics,” <https://bioconductor.org/packages/biotmle>, 2017. [Online]. Available: <https://dx.doi.org/10.18129/B9.bioc.biotmle>
- [2] **N.S. Hejazi** and M.J. van der Laan, “methyvim: Targeted data-adaptive estimation and inference for differential methylation analysis,” <https://bioconductor.org/packages/methyvim>, 2017. [Online]. Available: <https://dx.doi.org/10.18129/B9.bioc.methyvim>
- [3] D.C. Benkeser and **N.S. Hejazi**, “survtmle: Compute targeted minimum loss-based estimates in right-censored survival settings,” <https://CRAN.R-project.org/package=survtmle>, 2017. [Online]. Available: <https://dx.doi.org/10.5281/zenodo.835868>
- [4] J.R. Coyle and **N.S. Hejazi**, “origami: Generalized framework for cross-validation,” <https://CRAN.R-project.org/package=origami>, 2017. [Online]. Available: <https://dx.doi.org/10.5281/zenodo.1155901>
- [5] J.R. Coyle, **N.S. Hejazi**, I. Malenica, and O. Sofrygin, “sl3: Modern Super Learning with pipelines,” <https://github.com/tlverse/sl3>, 2017.
- [6] W. Cai, A.E. Hubbard, and **N.S. Hejazi**, “adaptest: Data-adaptive statistics for high-dimensional multiple testing,” <https://bioconductor.org/packages/adaptest>, 2018. [Online]. Available: <https://dx.doi.org/10.18129/B9.bioc.adaptest>

---

## Presentations

### *Talks — Invited and Contributed*

- [1] **N.S. Hejazi**, “Fair inference through semiparametric-efficient estimation over constraint-specific paths,” August 2018, Joint Statistical Meetings (contributed). [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/present/2018\\_jsm\\_fairtmle.pdf](https://statistics.berkeley.edu/~nhejazi/present/2018_jsm_fairtmle.pdf)
- [2] **N.S. Hejazi**, A.E. Hubbard, and M.J. van der Laan, “Data-adaptive estimation and inference for differential methylation analysis,” July 2018, BioC 2018: Where Software and Biology Connect, Bioconductor (contributed). [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/present/2018\\_bioc\\_methyvim.html](https://statistics.berkeley.edu/~nhejazi/present/2018_bioc_methyvim.html)
- [3] **N. Hejazi**, “Robust nonparametric inference for stochastic interventions under multi-stage sampling,” April 2018, Biostatistics Seminar Series, Division of Biostatistics, UC Berkeley (contributed). [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/present/2018\\_berkeley\\_txshift.pdf](https://statistics.berkeley.edu/~nhejazi/present/2018_berkeley_txshift.pdf)
- [4] **N. Hejazi** and K. Benac, “Efficient estimation of survival prognosis under immortal time bias,” March 2018, Berkeley Statistics Annual Research Symposium, Department of Statistics, UC Berkeley (contributed). [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/present/2018\\_itb\\_bstars.pdf](https://statistics.berkeley.edu/~nhejazi/present/2018_itb_bstars.pdf)
- [5] **N. Hejazi**, “Data-adaptive estimation and inference for differential methylation analysis,” November 2017, Annual Computational and Genomic Biology Retreat, Center for Computational Biology, UC Berkeley (contributed). [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/present/2017\\_berkeley\\_methyvim.pdf](https://statistics.berkeley.edu/~nhejazi/present/2017_berkeley_methyvim.pdf)

- [6] —, “Finite-sample inference and moderated statistics for asymptotically linear parameters,” March 2017, Biostatistics Seminar Series, Division of Biostatistics, UC Berkeley (contributed). [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/present/2017\\_berkeley\\_biotmle.pdf](https://statistics.berkeley.edu/~nhejazi/present/2017_berkeley_biotmle.pdf)
- [7] A.E. Hubbard, **N. Hejazi**, W. Cai, and A. Decker, “Targeted learning for high-dimensional variable importance,” July 2016, Statistical Causal Inference and Applications to Genetics, Centre de Recherches Mathématiques (invited). [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/present/2016\\_crm\\_limmatmle.pdf](https://statistics.berkeley.edu/~nhejazi/present/2016_crm_limmatmle.pdf)

#### *Posters — Invited and Contributed*

- [1] **N.S. Hejazi**, M.J. van der Laan, and D.C. Benkeser, “Robust nonparametric inference for stochastic interventions under multi-stage sampling,” May 2018, Atlantic Causal Inference Conference. [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/posters/2018\\_acic.pdf](https://statistics.berkeley.edu/~nhejazi/posters/2018_acic.pdf)
- [2] **N. Hejazi**, K. Benac, and N.P. Jewell, “Efficient estimation of survival prognosis under immortal time bias,” March 2018, Berkeley Statistics Annual Research Symposium. [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/posters/2018\\_bstars.pdf](https://statistics.berkeley.edu/~nhejazi/posters/2018_bstars.pdf)
- [3] **N.S. Hejazi**, I. Malenica, A.K. Waschka, A.E. Hubbard, and M.J. van der Laan, “Nonparametric variable importance for continuous exposures with applications in high-dimensional biology,” May 2017, Atlantic Causal Inference Conference. [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/posters/2017\\_acic.pdf](https://statistics.berkeley.edu/~nhejazi/posters/2017_acic.pdf)
- [4] **N. Hejazi** and A.E. Hubbard, “Super Learner for optimal prediction,” July 2016, Statistical Causal Inference and Applications to Genetics, Centre de Recherches Mathématiques. [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/posters/2016\\_crm-comp.pdf](https://statistics.berkeley.edu/~nhejazi/posters/2016_crm-comp.pdf)

#### *Other Talks and Presentations*

- [1] **N. Hejazi**, “Targeted biomarker discovery,” April 2018, Career Exploration Committee, Statistics Undergraduate Student Association, UC Berkeley. [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/present/2018\\_susa\\_berkeley/2018\\_susa\\_berkeley.html](https://statistics.berkeley.edu/~nhejazi/present/2018_susa_berkeley/2018_susa_berkeley.html)
- [2] —, “Targeted biomarker discovery,” March 2017, Graduate Student Admit Day, Group in Biostatistics, UC Berkeley. [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/present/2017\\_admitday\\_berkeley/2017\\_admitday\\_berkeley.html](https://statistics.berkeley.edu/~nhejazi/present/2017_admitday_berkeley/2017_admitday_berkeley.html)
- [3] **N. Hejazi** and E. Muzzall, “Ensemble machine learning with Super Learner and h2o in R,” December 2016, The Hacker Within, Berkeley Institute for Data Science, UC Berkeley. [Online]. Available: [https://statistics.berkeley.edu/~nhejazi/present/2016\\_talk\\_h2oSL\\_THW.pdf](https://statistics.berkeley.edu/~nhejazi/present/2016_talk_h2oSL_THW.pdf)

## Professional Activities

### *Ad-Hoc Reviewer*

- 2018–present IEEE/ACM Transactions on Computational Biology and Bioinformatics
- 2018–present The Journal of Open Source Software

### *Education and Outreach*

- 2018–present Lesson Maintainer, “Version Control with Git” (workshop materials), Software Carpentry

### *University and Departmental Committees*

Feb. 2018 Member, Admissions Committee for Graduate Studies (M.A./Ph.D.), Interdepartmental Group in Biostatistics, University of California, Berkeley; Berkeley, CA, USA

### Affiliations

2017–present International Biometrics Society (Western North American Region)

2016–present Berkeley Institute for Data Science

2016–present American Statistical Association

---

## Teaching Experience

### Teaching Assistantship

Spring 2018 *Targeted Learning in Biomedical Big Data* (PBHLTH 290), with Prof. Mark J. van der Laan, Division of Biostatistics, University of California, Berkeley

- o Developed all laboratory materials and homework assignments, provided weekly laboratory instruction, and supervised course projects, as the sole graduate student instructor for the pilot offering of this course.
  - Course website: <https://vanderlaan-group.github.io/tlbbd-sp2018/>
  - Laboratory materials: <https://github.com/tlbbd-spring2018>

### Short Course and Workshop Instructorship

2017–present *Software Carpentry* — workshop instruction aiming to promote and foster skills, best practices, and workflows for software development and scientific computing in research settings.

- o Jul. 2018; “Scientific Computing with Bash, Git, and Python”; workshop instructor (co-instructor: K. Marwaha); Berkeley Institute for Data Science; website: <https://bids.github.io/2018-07-16-bids/>
- o Jan. 2018; “Scientific Computing with Bash, Git, and R”; workshop helper with K. Ottoboni & T. Hart; Berkeley Institute for Data Science; website: <http://www.kellieottoboni.com/2018-01-11-bids/>

2017–present *Data Carpentry* — workshop instruction centering on core skills and best practices for the application of data analytic principles and tools for scientific computing in research.

- o May 2018; “Software Tools for Genomic Data Analysis”; workshop instructor (co-instructor: A. Orr); Lawrence Berkeley National Laboratory; website: <https://code.nimahejazi.org/2018-05-03-LBNL/>
- o Aug. 2017; “Software Tools for Genomic Data Analysis”; workshop helper with K. Hertweck & E. Becker; Lawrence Berkeley National Laboratory; website: <https://k8hertweck.github.io/2017-08-07-LBNL/>

Aug. 2018 Module Instructor, *R Bootcamp*, with Dr. Christopher J. Paciorek, Department of Statistics, University of California, Berkeley

- o Presented modules on modern data manipulation (via the “tidyverse”) and data visualization, contributed new instructional materials, and assisted in one-on-one instruction of participants.
  - Short course description: <https://statistics.berkeley.edu/computing/r-bootcamp>
  - Workshop materials: <https://github.com/berkeley-scf/r-bootcamp-2018>

Aug. 2017 Module Instructor, *R Bootcamp*, with Dr. Christopher J. Paciorek, Department of Statistics, University of California, Berkeley

- o Presented a module on programming concepts in R (iteration, flow control, writing functions), contributed new materials on modern functional programming, and assisted in one-on-one instruction of participants.
  - Short course description: <https://statistics.berkeley.edu/computing/r-bootcamp>
  - Workshop materials: <https://github.com/berkeley-scf/r-bootcamp-2017>

### Miscellaneous Teaching

2017–present Private Tutor in Statistics (introductory, computing, and intermediate-level theory courses)

2015–2016 Tutor in Mathematics & Statistics, Athletic Study Center, University of California, Berkeley

2013–2015 Tutor in General Chemistry, Student Learning Center, University of California, Berkeley

## ■ Computing Skills

Systems: macOS (OS X), Linux (Ubuntu, Fedora, Red Hat), UNIX

Languages: R, PYTHON, JULIA, C++, shell scripting, L<sup>A</sup>T<sub>E</sub>X, SQL, MATLAB

Apps/Other: Amazon Web Services (EC2), Google Compute Engine, Git, Microsoft Office