

# Nima Hejazi

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

## Curriculum Vitae

---

updated: November 2017

### Current Positions

*Doctoral Student*, Division of Biostatistics, University of California, Berkeley

Advisor: Mark J. van der Laan

*Biomedical Big Data Pre-doctoral Fellow*, UC Berkeley Biomedical Big Data Training Program

### Research Interests

**Methodology** nonparametric statistics, causal inference, machine learning, robust data-adaptive inference, targeted maximum likelihood estimation, stochastic treatment regimes, statistical genomics, survival analysis, vaccine sieve analysis, reproducible research, statistical/scientific computing.

**Applications** precision medicine, computational biology and bioinformatics, clinical trials, epidemiology, healthcare policy, health economics, behavioral sciences and public policy.

### Education

2016–present **Doctor of Philosophy (Ph.D.) in Biostatistics**, *University of California, Berkeley*.

Designated emphasis (graduate minor) in Computational and Genomic Biology

2016–2017 **Master of Arts (M.A.) in Biostatistics**, *University of California, Berkeley*.

*Committee*: Mark J. van der Laan (co-chair), Alan E. Hubbard (co-chair), Martyn T. 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

08/2018 **Biomedical Big Data Training Certification**, *University of California, Berkeley*.

10/2017 **Instructor Training Program Certification**, *Software and Data Carpentries*.

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 statistics and causal inference, including high-dimensional machine learning, survival analysis, and data-adaptive inference with applications in vaccine efficacy studies and computational biology, among other areas.
- Provided statistical and data analytic support for research efforts in precision medicine and public health in collaboration with initiatives of the Bill and Melinda Gates Foundation.
- Development of a software ecosystem for statistical inference in the framework of Targeted Learning.

- 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 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 **Research Assistant (undergraduate)**, *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 **Research Assistant (undergraduate)**, *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 **Research Assistant (undergraduate)**, *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 **Research Assistant (undergraduate)**, *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

- 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 science, and software development for data analysis.

## Fellowships and Scholarships

- 2017-2018 Biomedical Big Data Pre-doctoral Fellowship, Biomedical Big Data Training Program (NIH BD2K), University of California, Berkeley
- 2017 Travel and Registration Scholarships, Summer Institute in Statistics for Big Data, Department of Biostatistics, University of Washington
- 2015 Travel and Registration Scholarships, Summer Institute in Biostatistics, Department of Biostatistics & Medical Informatics, University of Wisconsin — Madison

## Awards and Honors

- 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

## Conferences and Workshops

### Contributed or Invited

- May 2017 Atlantic Causal Inference Conference; Chapel Hill, NC, USA
- July 2016 Statistical Causal Inference and Applications to Genetics, Centre de Recherches Mathématiques, Université de Montréal; Montréal, QC, Canada

### Attended or Visited

- July 2017 Summer Institute for Statistics in Big Data, Department of Biostatistics, School of Public Health, University of Washington; Seattle, WA, USA
- July 2016 Computational Genomics Summer Institute, Institute for Pure and Applied Mathematics, University of California, Los Angeles; Los Angeles, CA, USA
- June 2016 BioC: Where Software and Biology Connect; Stanford, CA, USA
- June 2015 Summer Institute in Biostatistics, Department of Biostatistics & Medical Informatics, University of Wisconsin — Madison; Madison, WI, USA

---

## Professional Affiliations

- 2017–present Western North American Region of the International Biometrics Society (member)
- 2016–present Berkeley Institute for Data Science (member)
- 2016–present American Statistical Association (member)

---

## Teaching Experience

### Teaching Assistantship

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

### Volunteer Instructorship

- 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.
- 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.

### Miscellaneous Teaching

- 2017–present Private Tutor in Statistics (introductory, computing, and intermediate-level theory courses)
  - Aug. 2017 Module Instructor, R Bootcamp, Department of Statistics, University of California, Berkeley
  - Aug. 2017 Teaching Assistant, Genomics Workshop, Lawrence Berkeley Nat'l Laboratory, Data Carpentry
- 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 (Mac OS X), Linux (Ubuntu, Fedora, Red Hat), Chrome OS, UNIX
- Languages: R, PYTHON, C++, shell scripting, L<sup>A</sup>T<sub>E</sub>X, JULIA, MATLAB, SQL
- Apps/Other: Git, Amazon EC2, Microsoft Office

---

## 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 Software Projects

- [1] **N.S. Hejazi**, W. Cai, and A.E. Hubbard, "biotmle: 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. submitted, 2017.

### Manuscripts in Preparation

- [1] **N.S. Hejazi**, R.V. Phillips, A.E. Hubbard, and M.J. van der Laan, "methyvim: Targeted and model-free analysis of differential methylation in R," 2017.
- [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**, A.E. Hubbard, and M.J. van der Laan, "Data-adaptive and nonparametric analysis of differential methylation with variable importance measures."
- [4] 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>
- [5] **N.S. Hejazi** and N.P. Jewell, "Immortal time bias analyses: Time-varying survival regression models and delayed entry approaches."
- [6] W. Cai, A. Hubbard, and **N.S. Hejazi**, "adaptest: Data-adaptive statistics for high-dimensional multiple hypothesis testing in R."
- [7] D.C. Benkeser, A. Chambaz, and **N.S. Hejazi**, "A guided tour of Targeted Learning."

### 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 Packages 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.835603>
- [5] J.R. Coyle, **N.S. Hejazi**, I. Malenica, and O. Sofrygin, “sl3: Scalable and efficient Super Learning with machine learning pipelines,” <https://github.com/jeremycoyle/sl3>, 2017.
- [6] W. Cai, A.E. Hubbard, and **N.S. Hejazi**, “adaptest: Data-adaptive statistics for high-dimensional testing,” <https://github.com/wilsoncai1992/adaptest>, 2017.

## Presentations

### *Talks — Invited and Contributed*

- [1] **N. Hejazi**. Data-adaptive estimation and inference for differential methylation analysis. Annual Computational and Genomic Biology Retreat, Center for Computational Biology, UC Berkeley, November 2017.
- [2] **N. Hejazi**. Empirical Bayes moderation of asymptotically linear parameters. Biostatistics Seminar Series, Division of Biostatistics, UC Berkeley, March 2017.
- [3] **N. Hejazi**. Targeted biomarker discovery. Graduate student admit day, Division of Biostatistics, UC Berkeley, March 2017.
- [4] **N. Hejazi** and E. Muzzall. Ensemble machine learning with Super Learner and h2o in R. The Hacker Within, Berkeley Institute for Data Science, UC Berkeley, December 2016.
- [5] A.E. Hubbard, **N. Hejazi**, W. Cai, and A. Decker. Targeted learning for high-dimensional variable importance. Montréal, QC, Canada, July 2016. Statistical Causal Inference and Applications to Genetics, Centre de Recherches Mathématiques.

### *Posters — Invited and Contributed*

- [1] **N. Hejazi** and A.E. Hubbard. Super Learner for optimal prediction. Montréal, QC, Canada, July 2016. Statistical Causal Inference and Applications to Genetics, Centre de Recherches Mathématiques.
- [2] **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. Chapel Hill, NC, USA, May 2017. Atlantic Causal Inference Conference.