

# Nima Hejazi

✉ [nhejazi@berkeley.edu](mailto:nhejazi@berkeley.edu)

🌐 [nimahejazi.org](http://nimahejazi.org)

in [nshejazi](#)

🐦 [nshejazi](#)

🌐 [nhejazi](#)

## Curriculum Vitæ

---

updated: 28 April 2021

### Current Positions

Ph.D. Candidate, Graduate Group in Biostatistics, University of California, Berkeley  
Advisors: Mark van der Laan and Alan Hubbard

### Research Interests

Methodology causal inference; non/semi-parametric inference; statistical machine learning; high-dimensional inference; targeted loss-based estimation; statistical computing; reproducible research.

Applications vaccine efficacy trials and immunology; computational biology and bioinformatics; precision medicine; electronic health records; epidemiology; neurology and psychiatry.

### Education

2017–2021 **Doctor of Philosophy in Biostatistics**, *University of California, Berkeley*.  
Designated emphasis (doctoral minor) in Computational and Genomic Biology  
Committee: Mark van der Laan, co-chair; Alan Hubbard, co-chair; Nicholas Jewell; Fernando Pérez  
Thesis: *Nonparametric Causal Inference with Stochastic Treatment Regimes*

2016–2017 **Master of Arts 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**, *University of California, Berkeley*.  
Triple major in Molecular & Cell Biology (em. Neurobiology), Psychology, and Public Health

### Selected Experience

09/2021–2023 **NSF Postdoctoral Research Fellow**, *Weill Cornell Medicine*, New York, NY.  
Sponsor: Prof. Iván Díaz, Division of Biostatistics, Department of Population Health Sciences

- Co-leading the formulation of novel methodology for causal mediation analysis with flexible multiple time point interventions and complex confounding, including effect bridging via causal transport.
- Developing techniques for efficient estimation under outcome-dependent two-phase sampling and with time-to-event variables, motivated by applications in vaccine efficacy trials and epidemiology.
- Collaborating in the planning and execution of statistical analyses for vaccine efficacy trials of malaria, HIV-1, and COVID-19, including the evaluation of immune correlates of risk and protection.

2017–2021 **Graduate Student Researcher**, *University of California, Berkeley*, Berkeley, CA.  
Advisor: Prof. Mark J. van der Laan, Division of Biostatistics, School of Public Health

- Extended and developed novel statistical methods for causal inference with stochastic interventions, including robust estimation in two-phase sampling designs and flexible effects for mediation analysis.
- Constructed and evaluated sieve-like strategies for nonparametric inverse probability weighting and flexible conditional density estimation based on undersmoothing the highly adaptive lasso estimator.
- Served as a founding core developer of the [tlverse project](#), a centralized ecosystem of R packages for causal machine learning in the Targeted Learning framework, including co-leading a new book.

- 2016–2021 **Graduate Student Researcher**, *University of California, Berkeley*, Berkeley, CA.  
Advisor: Prof. Alan E. Hubbard, Division of Biostatistics, School of Public Health
- Developed robust methods for stable statistical inference in high-dimensional biology, including data-adaptive hypothesis testing and variance moderation for non/semi-parametric estimators.
  - Performed statistical data analysis of high-dimensional genomics and epigenetics experiments, and provided bioinformatics support for molecular, environmental, and toxicological health sciences.
- 2018–present **Statistical Consultant**.
- Private consultation for industrial and academic clients on issues of causal inference, experimental design, observational studies, statistical machine learning, targeted learning, and applied statistics.
  - Consultation for researchers in the biological and social sciences through the Statistical Consulting Service, Department of Statistics, University of California, Berkeley (Spring 2020 and 2021 terms).
- 05–08/2021 **Data Scientist Intern**, *Netflix*, Los Gatos, CA.  
Supervisor: Dr. Wenjing Zheng, Data Science and Engineering
- Developed and implemented novel statistical methodology for causal machine learning of complex parameters, tailored for large-scale studies with millions of units and complex intervention schedules.
- 10–12/2020 **Data Scientist Intern**, *SiriusXM+Pandora*, Oakland, CA.  
Supervisor: Dr. Samuel D. Lendle, Shared Science Foundation
- Designed and implemented inverse propensity score weighting methods for improving contextual bandits action evaluation and selection, with applications to personalization of online radio services.
  - Evaluated doubly robust targeted estimators in settings with logged decision probabilities and non-stationary rewards, in support of offline analysis of data from contextual bandits A/B experiments.
- 06–09/2020 **Data Scientist Intern**, *Netflix*, Los Gatos, CA.  
Supervisor: Dr. Wenjing Zheng, Data Science and Engineering
- Development of novel methodology for the construction of robust, efficient, and flexible estimators of the causal impacts of longitudinal interventions, leveraging the highly adaptive lasso algorithm.
  - Methodological improvement of dynamic linear modeling strategies for assessing causal impacts of multiple time point interventions in time-series data, with application in quasi-experimental designs.
  - Core software development for and initial open source release of the `causalMethods` R meta-package.
- 2018–2019 **Research Biostatistician Intern**, *Kaiser Permanente Division of Research*, Oakland, CA.  
Supervisor: Dr. Romain S. Neugebauer, Biostatistics Core
- Development and software implementation of prediction methodology combining ensemble machine learning with inverse probability weighting, applied to massive electronic health records-based studies.
  - Core software development of an R package designed to reformat and interface with a proprietary electronic health records database in support of building domain-specific datasets for analytic studies.
- 2014–2015 **Undergraduate Student Researcher**, *University of California, Berkeley*, Berkeley, CA.  
Lab of Prof. Mark T. D'Esposito, Helen Wills Neuroscience Institute & Department of Psychology
- Study of the properties of resting-state voxel connectivity in the visual system using fMRI data.
- 2012–2014 **Undergraduate Student Researcher**, *UCSF Medical Center*, San Francisco, CA.  
Supervisor: Dr. Paul S. Larson, Department of Neurological Surgery
- Analysis of patient outcomes from deep brain stimulation surgery using interventional MR imaging.
- 2012–2014 **Undergraduate Research Assistant**, *University of California, Berkeley*, Berkeley, CA.  
Lab of Prof. Robert T. Knight, Helen Wills Neuroscience Institute & Department of Psychology
- Study of disruptions in lexical selection from neurological disease based on human EEG recordings.

---

## Awards & Honors

- 2021 Mathematical Sciences Postdoctoral Research Fellowship, National Science Foundation

- April 2021 Registration Scholarship, Summer Institute in Statistics and Modeling in Infectious Diseases, Department of Biostatistics, University of Washington
- June 2020 The Wallace Lowe Fellowship, School of Public Health, University of California, Berkeley
- May 2020 Student Travel Award for the Joint Statistical Meetings, San Francisco Bay Area Chapter of the American Statistical Association
- May 2019 Tom Ten Have Memorial Award (for “exceptionally creative or skillful research in causal inference”) for “Nonparametric-efficient causal mediation analysis for stochastic interventions”, Atlantic Causal Inference Conference
- May 2019 Student Travel Award for the Joint Statistical Meetings, San Francisco Bay Area Chapter of the American Statistical Association
- April 2019 The Eki & Nobuta Akahoshi and Seiko Baba Brodbeck Endowed Fund Scholarship, School of Public Health, University of California, Berkeley
- April 2019 Travel Scholarship, BioC 2019: Where Software and Biology Connect, Bioconductor: Open Source Software for Bioinformatics
- June 2018 Travel Scholarship, BioC 2018: Where Software and Biology Connect, Bioconductor: Open Source Software for Bioinformatics
- April 2018 The Wellness Scholarship in Honor of Chin Long Chiang, School of Public Health, University of California, Berkeley
- 2017 NIH-NLM “Big Data to Knowledge” (BD2K) Training Grant Fellowship, Biomedical Big Data Training Program, University of California, Berkeley
- April 2017 Registration and Travel Scholarships, Summer Institute in Statistics for Big Data, Department of Biostatistics, University of Washington
- May 2017 Honorable Mention for the Tom Ten Have Memorial Award (for “exceptionally creative or skillful research in causal inference”) for “Nonparametric variable importance for continuous exposures with applications in high-dimensional biology”, Atlantic Causal Inference Conference
- June 2015 NIH-NHLBI Training Grant Recipient, Summer Institute in Biostatistics, Department of Biostatistics & Medical Informatics, University of Wisconsin — Madison

---

## Publications

†: denotes equal contribution or co-first authorship; authors (usually) listed alphabetically.

### Statistical Methodology (refereed)

- [1] I. Díaz and **N. S. Hejazi**, “Causal mediation analysis for stochastic interventions,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 82, no. 3, pp. 661–683, 2020. DOI: [10.1111/rssb.12362](https://doi.org/10.1111/rssb.12362).
- [2] P. Boileau, **N. S. Hejazi**, and S. Dudoit, “Exploring high-dimensional biological data with sparse contrastive principal component analysis,” *Bioinformatics*, vol. 36, no. 11, pp. 3422–3430, 2020. DOI: [10.1093/bioinformatics/btaa176](https://doi.org/10.1093/bioinformatics/btaa176).
- [3] **N. S. Hejazi**, M. J. van der Laan, H. E. Janes, P. B. Gilbert, and D. C. Benkeser, “Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials,” *Biometrics*, 2020. DOI: [10.1111/biom.13375](https://doi.org/10.1111/biom.13375).
- [4] I. Díaz, **N. S. Hejazi**, K. E. Rudolph, and M. J. van der Laan, “Non-parametric efficient causal mediation with intermediate confounders,” *Biometrika*, 2020. DOI: [10.1093/biomet/asaa085](https://doi.org/10.1093/biomet/asaa085).

- [5] A. Ertefaie, **N. S. Hejazi**, and M. J. van der Laan, “Nonparametric inverse probability weighted estimators based on the highly adaptive lasso,” revision invited at *Biometrics*, 2021. [Online]. Available: <https://arxiv.org/abs/2005.11303>.
- [6] J. R. Coyle<sup>†</sup>, **N. S. Hejazi**<sup>†</sup>, I. Malenica<sup>†</sup>, R. V. Phillips<sup>†</sup>, B. F. Arnold, A. Mertens, J. Benjamin-Chung, W. Cai, S. Dayal, J. M. Colford Jr., A. E. Hubbard, and M. J. van der Laan, “Targeting Learning: Robust statistics for reproducible research,” under review at *Nature*, 2021. [Online]. Available: <https://arxiv.org/abs/2006.07333>.
- [7] **N. S. Hejazi**, K. E. Rudolph, M. J. van der Laan, and I. Díaz, “Nonparametric causal mediation analysis for stochastic interventional (in)direct effects,” under review at *Journal of the American Statistical Association*, 2021. [Online]. Available: <https://arxiv.org/abs/2009.06203>.
- [8] P. Boileau, **N. S. Hejazi**, M. J. van der Laan, and S. Dudoit, “Cross-validated loss-based covariance matrix estimator selection in high dimensions,” under review at *Journal of the American Statistical Association*, 2021. [Online]. Available: <https://arxiv.org/abs/2102.09715>.
- [9] **N. S. Hejazi**, M. J. van der Laan, and A. E. Hubbard, “A generalization of moderated statistics to data adaptive semiparametric estimation in high-dimensional biology,” submitted to *Statistical Methods in Medical Research*, 2021. [Online]. Available: <https://arxiv.org/abs/1710.05451>.
- Scientific Applications (refereed)*
- [10] **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](https://doi.org/10.14293/s2199-1006.1.sor-stat.afhtwc.v1.rxidzs).
- [11] A. Sobh, A. Loguinov, G. N. Yazici, R. S. Zeidan, A. Tagmount, **N. S. Hejazi**, A. E. Hubbard, L. Zhang, and C. D. Vulpe, “Functional profiling identifies determinants of arsenic trioxide cellular toxicity,” *Toxicological Sciences*, vol. 169, no. 1, pp. 108–121, 2019. DOI: [10.1093/toxsci/kfz024](https://doi.org/10.1093/toxsci/kfz024).
- [12] K. E. Rudolph, I. Díaz, **N. S. Hejazi**, M. J. van der Laan, S. X. Luo, M. Shulman, A. Campbell, J. Rotrosen, and E. V. Nunes, “Explaining differential effects of medication for opioid use disorder using a novel approach incorporating mediating variables,” *Addiction*, 2020. DOI: [10.1111/add.15377](https://doi.org/10.1111/add.15377).
- [13] G. E. Gray, L.-G. Bekker, F. Laher, M. Malahleh, M. Allen, Z. Moodie, N. Grunenberg, Y. Huang, D. Grove, B. Prigmore, J. J. Kee, D. Benkeser, J. Hural, C. Innes, E. Lazarus, G. Meintjes, N. Naicker, D. Kalonji, M. Nchabeleng, M. Sebe, N. Singh, P. Kotze, S. Kassim, T. Dubula, V. Naicker, W. Brumskine, C. N. Ncayiya, A. M. Ward, N. Garrett, G. Kistnasami, Z. Gaffoor, P. Selepe, P. B. Makhoba, M. P. Mathebula, P. Mda, T. Adonis, K. S. Mapetla, B. Modibedi, T. Philip, G. Kobane, C. Bentley, S. Ramirez, S. Takuva, M. Jones, M. Sikhosana, M. Atujuna, M. Andrasik, **N. S. Hejazi**, A. Puren, L. Wiesner, S. Phogat, C. Diaz Granados, M. Koutsoukos, O. Van Der Meeren, S. W. Barnett, N. Kanesa-thasan, J. G. Kublin, M. J. McElrath, P. B. Gilbert, H. Janes, and L. Corey, “Vaccine efficacy of ALVAC-HIV and bivalent subtype C gp120–MF59 in adults,” *New England Journal of Medicine*, vol. 384, no. 12, pp. 1089–1100, 2021. DOI: [10.1056/NEJMoa2031499](https://doi.org/10.1056/NEJMoa2031499).

- [14] J. Benjamin-Chung, A. Mertens, J. M. Colford Jr., A. E. Hubbard, M. J. van der Laan, J. Coyle, O. Sofrygin, W. Cai, A. Nguyen, N. N. Pokpongkiat, S. Djajadi, A. Seth, W. Jilek, E. O. Chung, S. Dayal, **N. Hejazi**, I. Malenica, H. Li, R. Hafen, V. Subramoney, J. Häggström, T. Norman, K. H. Brown, P. Christian, B. F. Arnold, and the ki Child Growth Consortium, “Early childhood linear growth failure in low- and middle-income countries,” revision invited at *Nature*, 2021. DOI: [10.1101/2020.06.09.20127001](https://doi.org/10.1101/2020.06.09.20127001).
- [15] A. Mertens, J. Benjamin-Chung, J. M. Colford Jr., A. E. Hubbard, M. van der Laan, J. Coyle, O. Sofrygin, W. Cai, W. Jilek, S. Dayal, A. Nguyen, N. Pokpongkiat, S. Djajadi, A. Seth, E. O. Chung, I. Malenica, **N. Hejazi**, H. Li, R. Hafen, V. Subramoney, J. Häggström, T. Norman, P. Christian, K. H. Brown, B. F. Arnold, and the ki Child Growth Consortium, “Child wasting and concurrent stunting in low- and middle-income countries,” revision invited at *Nature*, 2021. DOI: [10.1101/2020.06.09.20126979](https://doi.org/10.1101/2020.06.09.20126979).
- [16] A. Mertens, J. Benjamin-Chung, J. M. Colford Jr., J. Coyle, M. J. van der Laan, A. E. Hubbard, S. Dayal, I. Malenica, **N. Hejazi**, O. Sofrygin, W. Cai, H. Li, A. Nguyen, N. N. Pokpongkiat, S. Djajadi, A. Seth, E. O. Chung, W. Jilek, V. Subramoney, R. Hafen, J. Häggström, T. Norman, K. H. Brown, P. Christian, B. F. Arnold, and the ki Child Growth Consortium, “Causes and consequences of child growth failure in low- and middle-income countries,” revision invited at *Nature*, 2021. DOI: [10.1101/2020.06.09.20127100](https://doi.org/10.1101/2020.06.09.20127100).
- [17] H. K. Amato, C. Hemlock, K. L. Andrejko, A. R. Smith, **N. S. Hejazi**, A. E. Hubbard, S. C. Verma, R. K. Adhikari, D. Pokhrel, K. R. Smith, J. P. Graham, and A. Pokhrel, “Effect of biodigester cookstove interventions on child diarrhea in semi-rural Nepal: a causal analysis of daily observations,” under review at *Environmental Health Perspectives*, 2021.

#### Statistical Software (refereed)

- [18] **N. S. Hejazi**, W. Cai, and A. E. Hubbard, “biotml: Targeted Learning for biomarker discovery,” *Journal of Open Source Software*, vol. 2, no. 15, p. 295, Jul. 2017. DOI: [10.21105/joss.00295](https://doi.org/10.21105/joss.00295).
- [19] J. R. Coyle and **N. S. Hejazi**, “origami: A generalized framework for cross-validation in R,” *Journal of Open Source Software*, vol. 3, no. 21, p. 512, Jan. 2018. DOI: [10.21105/joss.00512](https://doi.org/10.21105/joss.00512).
- [20] **N. S. Hejazi**, R. V. Phillips, A. E. Hubbard, and M. J. van der Laan, “methyvim: Targeted, robust, and model-free differential methylation analysis in R,” *F1000Research*, vol. 7, p. 1424, Sep. 2018. DOI: [10.12688/f1000research.16047.1](https://doi.org/10.12688/f1000research.16047.1).
- [21] W. Cai, A. E. Hubbard, and **N. S. Hejazi**, “adaptest: Data-adaptive statistics for high-dimensional testing in R,” *Journal of Open Source Software*, vol. 3, no. 30, p. 161, Oct. 2018. DOI: [10.21105/joss.00161](https://doi.org/10.21105/joss.00161).
- [22] P. Boileau, **N. S. Hejazi**, and S. Dudoit, “scPCA: A toolbox for sparse contrastive principal component analysis in R,” *Journal of Open Source Software*, vol. 5, no. 46, p. 2079, Feb. 2020. DOI: [10.21105/joss.02079](https://doi.org/10.21105/joss.02079).
- [23] **N. S. Hejazi**, J. R. Coyle, and M. J. van der Laan, “hal9001: Scalable highly adaptive lasso regression in R,” *Journal of Open Source Software*, vol. 5, no. 53, p. 2526, Sep. 2020. DOI: [10.21105/joss.02526](https://doi.org/10.21105/joss.02526).
- [24] **N. S. Hejazi** and D. C. Benkeser, “txshift: Efficient estimation of the causal effects of stochastic interventions in R,” *Journal of Open Source Software*, vol. 5, no. 54, p. 2447, Oct. 2020. DOI: [10.21105/joss.02447](https://doi.org/10.21105/joss.02447).

- [25] P. Boileau, **N. S. Hejazi**, B. Collica, M. J. van der Laan, and S. Dudoit, “cvCovEst: Cross-validated covariance matrix estimation in R,” under review at *Journal of Open Source Software*, 2021.

#### *Books and Book Chapters*

- [26] M. J. van der Laan, J. R. Coyle, **N. S. Hejazi**, I. Malenica, R. V. Phillips, and A. E. Hubbard, *Targeted Learning in R: Causal Data Science with the tLverse Software Ecosystem*. CRC Press, 2022, in preparation. [Online]. Available: <https://tlverse.org/tlverse-handbook>.
- [27] **N. S. Hejazi**, J. R. Coyle, and M. J. van der Laan, “Stochastic treatment regimes,” in *Targeted Learning in R: Causal Data Science with the tLverse Software Ecosystem*, in preparation, CRC Press, 2022. [Online]. Available: <https://tlverse.org/tlverse-handbook/stochastic-treatment-regimes>.
- [28] **N. S. Hejazi**, J. P. Duncan, D. B. McCoy, and M. J. van der Laan, “Causal mediation analysis,” in *Targeted Learning in R: Causal Data Science with the tLverse Software Ecosystem*, in preparation, CRC Press, 2022. [Online]. Available: <https://tlverse.org/tlverse-handbook/causal-mediation-analysis>.

#### *Statistical Analysis Plans*

- [29] P. B. Gilbert, Y. Fong, D. Benkeser, J. Andriesen, B. Borate, M. Carone, L. N. Carpp, I. Díaz, M. P. Fay, A. Fiore-Gartland, **N. S. Hejazi**, Y. Huang, Y. Huang, O. Hyrien, H. E. Janes, M. Juraska, K. Li, A. Luedtke, M. Nason, A. K. Randhawa, L. van der Laan, B. D. Williamson, W. Zhang, and D. Follmann, “CoVPN COVID-19 Vaccine Efficacy Trial Immune Correlates Statistical Analysis Plan,” 2021. DOI: [10.6084/m9.figshare.13198595](https://doi.org/10.6084/m9.figshare.13198595).

#### *Dissertations and Theses*

- [30] **N. S. Hejazi**, “Generalized application of empirical Bayes statistics to asymptotically linear parameters,” Master’s thesis, University of California, Berkeley, 2017. [Online]. Available: <https://www.stat.berkeley.edu/~nhejazi/publications/thesis-ma-biostat.pdf>.
- [31] **N. S. Hejazi**, “Nonparametric and robust methods for causal inference with stochastic interventions,” PhD thesis, University of California, Berkeley, 2021. [Online]. Available: <https://www.stat.berkeley.edu/~nhejazi/publications/thesis-phd-biostat.pdf>.

#### *Technical Reports (non-refereed)*

- [32] W. Cai, **N. S. Hejazi**, and A. E. Hubbard, “Data-adaptive statistics for multiple hypothesis testing in high-dimensional settings,” University of California, Berkeley, Tech. Rep., 2017. [Online]. Available: <https://arxiv.org/abs/1704.07008>.

#### *Currently in Preparation*

- [33] **N. S. Hejazi**, D. C. Benkeser, I. Díaz, and M. J. van der Laan, “An efficient estimator of the generalized propensity score for the causal effects of modified treatment policies,” Target: *Biometrics*, 2021.
- [34] **N. S. Hejazi**, P. B. Gilbert, and D. C. Benkeser, “Evaluating causal vaccine efficacy with stochastic interventional shifts of immunologic markers in COVID-19 vaccine efficacy trials,” Target: *Statistics in Medicine*, 2021.
- [35] **N. S. Hejazi**, M. J. van der Laan, and W. Zheng, “Robust and data-adaptive inverse probability weighted estimation of longitudinal causal effects,” Target: *Journal of the American Statistical Association*, 2021.

- [36] **N. S. Hejazi**, I. Díaz, and M. J. van der Laan, “Direct estimation of robust inverse probability weights for nonparametric causal inference,” *Target: Journal of the American Statistical Association*, 2021.

---

## Presentations

### *Conference and Seminar Talks*

- [1] **N. S. Hejazi**, M. J. van der Laan, S. Anand, and W. Zheng, “Efficient inverse probability weighted estimation of longitudinal causal effects in quasi-experimental studies,” Joint Statistical Meetings, Seattle, WA, USA, Aug. 2021, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2021\\_jsm\\_ipwhal\\_longit\\_nflx.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2021_jsm_ipwhal_longit_nflx.pdf).
- [2] **N. S. Hejazi**, P. B. Gilbert, and D. C. Benkeser, “Evaluation of causal vaccine efficacy with stochastic interventional shifts of immunologic markers in COVID-19 vaccine efficacy trials,” Society for Epidemiologic Research Meeting, San Diego, CA, USA, Jun. 2021, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2021\\_ser\\_txshift\\_covpn.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2021_ser_txshift_covpn.pdf).
- [3] **N. S. Hejazi**, D. C. Benkeser, I. Díaz, and M. J. van der Laan, “Efficient estimation of the generalized propensity score in evaluating modified treatment policies,” European Causal Inference Meeting, Oslo, Norway, May 2021, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2021\\_eurocim\\_haldensify.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2021_eurocim_haldensify.pdf).
- [4] **N. Hejazi**, A. Ertefaie, and M. van der Laan, “Nonparametric inverse probability weighted estimators based on the highly adaptive lasso,” Causal Inference Research Seminar, Department of Statistics, UC Berkeley, Berkeley, CA, USA, Mar. 2021, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2021\\_berkeley\\_ipwhal.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2021_berkeley_ipwhal.pdf).
- [5] **N. Hejazi**, M. van der Laan, H. Janes, P. Gilbert, and D. Benkeser, “Evaluating the causal impacts of vaccine-induced immune responses in two-phase vaccine efficacy trials,” Biostatistics Research Seminar, Public Health Sciences and Vaccine & Infectious Disease Divisions, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, Jan. 2021, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2021\\_fredhutch\\_txshift.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2021_fredhutch_txshift.pdf).
- [6] **N. S. Hejazi**, M. J. van der Laan, D. C. Benkeser, H. E. Janes, and P. B. Gilbert, “Leveraging the causal effects of stochastic interventions to evaluate vaccine efficacy in two-phase trials,” Society for Epidemiologic Research Meeting, Boston, MA, USA, Dec. 2020, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2020\\_ser\\_txshift.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2020_ser_txshift.pdf).
- [7] M. J. van der Laan, A. Ertefaie, and **N. S. Hejazi**, “Nonparametric inverse probability weighted estimators based on the highly adaptive lasso,” 13th International Conference of the ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on Computational and Methodological Statistics (CMStatistics 2020), London, Eng., UK, Dec. 2020.
- [8] **N. Hejazi**, M. van der Laan, H. Janes, P. Gilbert, and D. Benkeser, “Evaluating the causal impacts of vaccine-induced immune responses in two-phase vaccine efficacy trials,” Causal Inference Research Seminar, Department of Statistics, UC Berkeley, Berkeley, CA, USA, Nov. 2020, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2020\\_berkeley\\_txshift.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2020_berkeley_txshift.pdf).

- [9] **N. Hejazi**, M. van der Laan, H. Janes, P. Gilbert, and D. Benkeser, "Evaluating the causal impacts of vaccine-induced immune responses in two-phase vaccine efficacy trials," Biomedical Big Data Research Seminar, Division of Biostatistics, UC Berkeley, Berkeley, CA, USA, Oct. 2020, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2020\\_bbd\\_txshift.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2020_bbd_txshift.pdf).
- [10] **N. S. Hejazi**, M. J. van der Laan, and I. Díaz, "Efficient estimation of stochastic intervention effects in causal mediation analysis," Joint Statistical Meetings, Philadelphia, PA, USA, Aug. 2020, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2020\\_jsm\\_medshift.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2020_jsm_medshift.pdf).
- [11] **N. S. Hejazi**, A. Ertefaie, and M. J. van der Laan, "Nonparametric-efficient inverse probability weighted estimators based on the highly adaptive lasso," American Causal Inference Conference (canceled due to COVID-19), Austin, TX, USA, May 2020.
- [12] K. E. Rudolph, I. Díaz, **N. S. Hejazi**, S. X. Luo, M. Shulman, J. Rotrosen, and E. V. Nunes, "Using a flexible, efficient, nonparametric estimator of direct and indirect effects to explain differences in opioid treatment responsivity," American Causal Inference Conference (canceled due to COVID-19), Austin, TX, USA, May 2020.
- [13] **N. Hejazi**, M. van der Laan, and I. Díaz, "Nonparametric causal mediation analysis for stochastic interventions," Causal Inference Research Seminar, Department of Statistics, UC Berkeley, Berkeley, CA, USA, Apr. 2020, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2020\\_berkeley\\_medshift.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2020_berkeley_medshift.pdf).
- [14] **N. S. Hejazi**, M. J. van der Laan, and A. E. Hubbard, "Generalized variance moderation for locally efficient estimation in high-dimensional biology," BioC 2019: Where Software and Biology Connect (contributed), New York, NY, USA, Jun. 2019, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2019\\_bioc\\_modtmle.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2019_bioc_modtmle.pdf).
- [15] **N. S. Hejazi**, D. C. Benkeser, and M. J. van der Laan, "Vaccine efficacy assessment under two-phase sampling based on the causal effects of stochastic interventions," Berkeley Statistics Annual Research Symposium, Department of Statistics, UC Berkeley (contributed), Berkeley, CA, USA, Mar. 2019, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2019\\_bstars\\_shift.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2019_bstars_shift.pdf).
- [16] **N. S. Hejazi** and M. J. van der Laan, "Fair inference through semiparametric-efficient estimation over constraint-specific paths," Joint Statistical Meetings (contributed), Vancouver, BC, Canada, Aug. 2018, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2018\\_jsm\\_fairtmle.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2018_jsm_fairtmle.pdf).
- [17] **N. S. Hejazi**, A. E. Hubbard, and M. J. van der Laan, "Data-adaptive estimation and inference for differential methylation analysis," BioC 2018: Where Software and Biology Connect (contributed), Toronto, ON, Canada, Jul. 2018, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2018\\_bioc\\_methyvim/2018\\_bioc\\_methyvim.html](https://www.stat.berkeley.edu/~nhejazi/present/2018_bioc_methyvim/2018_bioc_methyvim.html).
- [18] **N. Hejazi**, M. J. van der Laan, and D. C. Benkeser, "Robust nonparametric inference for stochastic interventions under multi-stage sampling," Biostatistics Seminar Series, Division of Biostatistics, UC Berkeley (contributed), Berkeley, CA, USA, Apr. 2018, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2018\\_berkeley\\_txshift.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2018_berkeley_txshift.pdf).



- [19] **N. Hejazi**, K. Benac, and N. P. Jewell, “Efficient estimation of survival prognosis under immortal time bias,” Berkeley Statistics Annual Research Symposium, Department of Statistics, UC Berkeley (contributed), Berkeley, CA, USA, Mar. 2018, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2018\\_bstars\\_itb.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2018_bstars_itb.pdf).
- [20] **N. Hejazi**, A. E. Hubbard, and M. J. van der Laan, “Data-Adaptive estimation and inference in the analysis of differential methylation,” Computational & Genomic Biology Retreat, Center for Computational Biology, UC Berkeley (contributed), Berkeley, CA, USA, Nov. 2017, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2017\\_berkeley\\_methyvim.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2017_berkeley_methyvim.pdf).
- [21] **N. Hejazi** and A. E. Hubbard, “Moderated finite-sample inference for asymptotically linear parameters,” Biostatistics Seminar Series, Division of Biostatistics, UC Berkeley (contributed), Berkeley, CA, USA, Mar. 2017, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2017\\_berkeley\\_biotmle.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2017_berkeley_biotmle.pdf).
- [22] A. E. Hubbard, **N. Hejazi**, W. Cai, and A. Decker, “Targeted learning for high-dimensional variable importance,” Statistical Causal Inference and Applications to Genetics, Centre de Recherches Mathématiques (invited), Montréal, QC, Canada, Jul. 2016, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2016\\_crm\\_limmatmle.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2016_crm_limmatmle.pdf).

#### *Conference and Symposia Posters*

- [23] K. E. Rudolph, I. Díaz, **N. Hejazi**, M. J. van der Laan, S. X. Luo, M. Shulman, A. Campbell, J. Rotrosen, and E. V. Nunes, “Using mediation to explain differences in opioid treatment responsivity,” Society for Epidemiologic Research Meeting, Boston, MA, USA, Dec. 2020.
- [24] **N. S. Hejazi**, M. J. van der Laan, M. T. Smith, and A. E. Hubbard, “Semiparametric estimation with robust empirical Bayes inference and supervised clustering in high-dimensional biological exposure studies,” NIEHS Superfund Research Program Annual Meeting, Seattle, WA, USA, Nov. 2019, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/posters/2019\\_srp.pdf](https://www.stat.berkeley.edu/~nhejazi/posters/2019_srp.pdf).
- [25] **N. S. Hejazi**, D. C. Benkeser, and M. J. van der Laan, “Robust inference on the causal effects of stochastic interventions under two-phase sampling, with applications in vaccine efficacy trials,” Joint Statistical Meetings, Denver, CO, USA, Jul. 2019, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/posters/2019\\_jsm.pdf](https://www.stat.berkeley.edu/~nhejazi/posters/2019_jsm.pdf).
- [26] **N. S. Hejazi**, M. J. van der Laan, and I. Díaz, “Nonparametric-efficient causal mediation analysis for stochastic interventions,” Atlantic Causal Inference Conference, Montréal, QC, Canada, May 2019, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/posters/2019\\_acic.pdf](https://www.stat.berkeley.edu/~nhejazi/posters/2019_acic.pdf).
- [27] **N. S. Hejazi**, D. C. Benkeser, and M. J. van der Laan, “Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling,” Berkeley Statistics Annual Research Symposium, Berkeley, CA, USA, Mar. 2019, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/posters/2019\\_bstars.pdf](https://www.stat.berkeley.edu/~nhejazi/posters/2019_bstars.pdf).
- [28] **N. S. Hejazi**, M. J. van der Laan, and A. E. Hubbard, “Variance moderation of locally efficient estimators and supervised clustering with applications in high-dimensional biology,” Computational & Genomic Biology Retreat, Center for Computational Biology, UC Berkeley, Berkeley, CA, USA, Nov. 2018, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/posters/2018\\_ccb.pdf](https://www.stat.berkeley.edu/~nhejazi/posters/2018_ccb.pdf).

- [29] **N. S. Hejazi**, M. J. van der Laan, and D. C. Benkeser, “Robust nonparametric inference for stochastic interventions under multi-stage sampling,” Atlantic Causal Inference Conference, Pittsburgh, PA, USA, May 2018, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/posters/2018\\_acic.pdf](https://www.stat.berkeley.edu/~nhejazi/posters/2018_acic.pdf).
- [30] **N. S. Hejazi**, K. Benac, and N. P. Jewell, “Efficient estimation of survival prognosis under immortal time bias,” Berkeley Statistics Annual Research Symposium, Berkeley, CA, USA, Mar. 2018, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/posters/2018\\_bstars.pdf](https://www.stat.berkeley.edu/~nhejazi/posters/2018_bstars.pdf).
- [31] **N. S. Hejazi**, I. Malenica, A. K. Waschka, A. E. Hubbard, and M. J. van der Laan, “Non-parametric variable importance for continuous exposures with applications in high-dimensional biology,” Atlantic Causal Inference Conference, Chapel Hill, NC, USA, May 2017, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/posters/2017\\_acic.pdf](https://www.stat.berkeley.edu/~nhejazi/posters/2017_acic.pdf).
- [32] **N. S. Hejazi** and A. E. Hubbard, “Super Learner for optimal prediction,” Statistical Causal Inference and Applications to Genetics, Centre de Recherches Mathématiques, Montréal, QC, Canada, Jul. 2016, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/posters/2016\\_crm-comp.pdf](https://www.stat.berkeley.edu/~nhejazi/posters/2016_crm-comp.pdf).
- Other Talks and Presentations*
- [33] **N. Hejazi** and S. Lendle, “On the radio: reweighting and efficient estimation with bandits,” “Science for Lunch” Forum, Science and Music Analysis, SiriusXM+Pandora, Dec. 2020.
- [34] **N. Hejazi** and W. Zheng, “Efficient estimation of causal impacts using machine learning: the highly adaptive lasso,” Machine Learning Forum, Data Science and Engineering, Netflix, Sep. 2020.
- [35] **N. Hejazi** and W. Zheng, “Flexible and efficient inverse probability weighted estimation,” Experimentation Forum, Data Science and Engineering, Netflix, Sep. 2020.
- [36] **N. S. Hejazi**, I. Díaz, and M. van der Laan, “Efficient estimation of stochastic intervention effects in causal mediation analysis,” Student Travel Award Presentation, San Francisco Bay Area Chapter of the American Statistical Association, Jun. 2020, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2020\\_sfasa\\_jsm.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2020_sfasa_jsm.pdf).
- [37] **N. Hejazi**, M. van der Laan, H. Janes, P. Gilbert, and D. Benkeser, “Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials,” Student Paper Competition, Western North American Region of the International Biometric Society, Jun. 2020, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2020\\_wnar\\_txshift.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2020_wnar_txshift.pdf).
- [38] **N. Hejazi**, “Building R packages: “Good enough” practices for applied statistics,” Guest Lecture in Public Health 292: Biostatistics Master’s Thesis Seminar, Feb. 2020, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2020\\_rpkgs\\_intro/rpkgs\\_intro.html](https://www.stat.berkeley.edu/~nhejazi/present/2020_rpkgs_intro/rpkgs_intro.html).
- [39] **N. Hejazi**, “Causal inference in biomedical research,” California Department of Public Health, Sep. 2019, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2019\\_cdph\\_causal.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2019_cdph_causal.pdf).

- [40] **N. Hejazi**, D. Benkeser, and M. van der Laan, “Robust inference on the causal effects of stochastic interventions under two-phase sampling, with applications in vaccine efficacy trials,” Student Travel Award Presentation, San Francisco Bay Area Chapter of the American Statistical Association, Jun. 2019, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2019\\_sfasa\\_jsm.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2019_sfasa_jsm.pdf).
- [41] **N. Hejazi**, “Towards the realistic, robust, and efficient assessment of causal effects with stochastic shift interventions,” Ph.D. Qualifying Examination, Graduate Group in Biostatistics, UC Berkeley, Sep. 2018, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2018\\_phd\\_qualis.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2018_phd_qualis.pdf).
- [42] **N. Hejazi**, “Targeted biomarker discovery,” Statistics Undergraduate Student Association, UC Berkeley, Apr. 2018, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2018\\_susa\\_berkeley/2018\\_susa\\_berkeley.html](https://www.stat.berkeley.edu/~nhejazi/present/2018_susa_berkeley/2018_susa_berkeley.html).
- [43] **N. Hejazi**, “Targeted biomarker discovery,” Graduate Group in Biostatistics, UC Berkeley, Mar. 2017, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2017\\_admitday\\_berkeley/2017\\_admitday\\_berkeley.html](https://www.stat.berkeley.edu/~nhejazi/present/2017_admitday_berkeley/2017_admitday_berkeley.html).
- [44] **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, Dec. 2016, [Online]. Available: [https://www.stat.berkeley.edu/~nhejazi/present/2016\\_talk\\_h2oSL\\_THW.pdf](https://www.stat.berkeley.edu/~nhejazi/present/2016_talk_h2oSL_THW.pdf).

---

## Software

- [1] **N. S. Hejazi**, M. J. van der Laan, and A. E. Hubbard, *biotmle: Targeted Learning with moderated statistics for biomarker discovery*, version 1.12.0, R package. DOI: [10.18129/B9.bioc.biotmle](https://doi.org/10.18129/B9.bioc.biotmle). [Online]. Available: <https://bioconductor.org/packages/biotmle>.
- [2] **N. S. Hejazi** and M. J. van der Laan, *methyvim: Targeted, robust, and model-free differential methylation analysis*, version 1.10.0, R package. DOI: [10.18129/B9.bioc.methyvim](https://doi.org/10.18129/B9.bioc.methyvim). [Online]. Available: <https://bioconductor.org/packages/methyvim>.
- [3] D. C. Benkeser and **N. S. Hejazi**, *survtmle: Compute targeted minimum loss-based estimates in right-censored survival settings*, version 1.1.1, R package. DOI: [10.5281/zenodo.835867](https://doi.org/10.5281/zenodo.835867). [Online]. Available: <https://CRAN.R-project.org/package=survtmle>.
- [4] J. R. Coyle and **N. S. Hejazi**, *origami: Generalized framework for cross-validation*, version 1.0.3, R package. DOI: [10.5281/zenodo.835602](https://doi.org/10.5281/zenodo.835602). [Online]. Available: <https://CRAN.R-project.org/package=origami>.
- [5] J. R. Coyle, **N. S. Hejazi**, I. Malenica, and O. Sofrygin, *sl3: Modern pipelines for machine learning and Super Learning*, version 1.3.7, R package. DOI: [10.5281/zenodo.1342293](https://doi.org/10.5281/zenodo.1342293). [Online]. Available: <https://github.com/tlverse/sl3>.
- [6] W. Cai and **N. S. Hejazi**, *adapttest: Data-adaptive statistics for high-dimensional multiple testing*, version 1.7.1, R package. DOI: [10.18129/B9.bioc.adapttest](https://doi.org/10.18129/B9.bioc.adapttest). [Online]. Available: <https://bioconductor.org/packages/adapttest>.
- [7] J. R. Coyle, **N. S. Hejazi**, and M. J. van der Laan, *hal9001: The scalable highly adaptive lasso*, version 0.2.7, R package. DOI: [10.5281/zenodo.3558313](https://doi.org/10.5281/zenodo.3558313). [Online]. Available: <https://CRAN.R-project.org/package=hal9001>.

- [8] **N. S. Hejazi** and D. C. Benkeser, *txshift: Efficient estimation of the causal effects of stochastic interventions*, version 0.3.4, R package. [Online]. Available: <https://CRAN.R-project.org/package=txshift>.
- [9] **N. S. Hejazi**, J. R. Coyle, and M. J. van der Laan, *tmle3shift: Targeted Learning of the causal effects of stochastic interventions*, version 0.1.9, R package. [Online]. Available: <https://github.com/tlverse/tmle3shift>.
- [10] **N. S. Hejazi** and I. Díaz, *medshift: Causal mediation analysis for stochastic interventions*, version 0.1.1, R package. [Online]. Available: <https://github.com/nhejazi/medshift>.
- [11] **N. S. Hejazi**, D. C. Benkeser, and M. J. van der Laan, *haldensify: Highly adaptive lasso conditional density estimation*, version 0.0.6, R package. DOI: [10.5281/zenodo.3698329](https://doi.org/10.5281/zenodo.3698329). [Online]. Available: <https://CRAN.R-project.org/package=haldensify>.
- [12] P. Boileau, **N. S. Hejazi**, and S. Dudoit, *scPCA: Sparse contrastive principal component analysis*, version 1.2.0, R package. DOI: [10.18129/B9.bioc.sc pca](https://doi.org/10.18129/B9.bioc.sc pca). [Online]. Available: <https://bioconductor.org/packages/scPCA>.
- [13] **N. S. Hejazi**, I. Díaz, and K. E. Rudolph, *medoutcon: Efficient causal mediation analysis with intermediate confounders*, version 0.1.0, R package. [Online]. Available: <https://github.com/nhejazi/medoutcon>.
- [14] P. Boileau, B. Collica, and **N. S. Hejazi**, *cvCovEst: Cross-validated covariance matrix estimation*, version 0.3.1, R package. [Online]. Available: <https://CRAN.R-project.org/package=cvCovEst>.
- [15] **N. S. Hejazi**, W. Zheng, and S. Lane, *causalMethods: Causal machine learning methods for quasi-experimental studies*, version 0.1.0, R package. [Online]. Available: <https://github.com/Netflix/causalMethods>.

## Research Support & Funding

- 09/2021–08/2023 Nonparametric Causal Inference in Complex Randomized and Observational Studies
- Source: National Science Foundation DMS–2102840
  - Role: Postdoctoral research fellow (PI: N.S. Hejazi)
- 06/2017–05/2021 Causal Effects of Endogenous and Exogenous Risk Factors for Wasting
- Source: The Bill & Melinda Gates Foundation OPP–1165144
  - Role: Graduate student researcher (PI: M.J. van der Laan)
- 08/2018–09/2019 Predicting Conversion from Prediabetes to Diabetes
- Source: Kaiser Permanente Division of Research
  - Role: Biostatistical research analyst (PI: J.A. Schmittdiel)
- 06/2017–05/2018 UC Berkeley Biomedical Big Data Training Program
- Source: National Library of Medicine T32–LM012417
  - Role: Graduate student research fellow (PI: M.J. van der Laan)
- 11/2016–05/2017 Molecular Mechanisms of Obesity in Children Exposed to Phthalates in Utero
- Source: National Institute of Environmental Health Sciences R01–ES021369
  - Role: Graduate student researcher (PI: N.T. Holland)

- 03/2016– Toxic Substances in the Environment  
 05/2017
- Source: National Institute of Environmental Health Sciences P42–ES004705
  - Role: Graduate student researcher (PI: M.T. Smith)

## Service & Affiliations

### Ad-Hoc Reviewing

American Journal of Epidemiology

Biometrika

IEEE/ACM Transactions on Computational Biology and Bioinformatics

Journal of Business & Economic Statistics

Journal of Machine Learning Research

Journal of Open Source Software

Journal of the Royal Statistical Society: Series B (Statistical Methodology)

Journal of the Royal Statistical Society: Series C (Applied Statistics)

Statistical Methods in Medical Research

### Education and Outreach

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

### University and Departmental Committees

- 2018–2021 Member of the Executive Board, Biostatistics Graduate Student Association, Graduate Group in Biostatistics, University of California, Berkeley; Berkeley, CA, USA
- Feb. 2018 Invited Member, Admissions Committee for Graduate Studies (M.A./Ph.D.), Graduate Group in Biostatistics, University of California, Berkeley; Berkeley, CA, USA

### Affiliations

- 2021–present Society for Causal Inference
- 2020–present Society for Epidemiologic Research
- 2017–present International Biometric Society (Western North American Region)
- 2016–present American Statistical Association
- 2016–2020 Berkeley Institute for Data Science

## Advising & Mentoring

### Master's Students

- 06/2020– Brian Collica, Department of Statistics, University of California, Berkeley, through a research  
 05/2021 mentorship program, jointly with Phillippe Boileau and Sandrine Dudoit (faculty sponsor).  
 Project: “Developing an R package for cross-validated covariance matrix estimator selection.”

### Undergraduate Projects

- 09–12/2020 Jamarcus Liu, Department of Statistics, University of California, Berkeley, through a research  
 mentorship program, jointly with Phillippe Boileau and Sandrine Dudoit (faculty sponsor).  
 Project: “Developing an R package for cross-validated covariance matrix estimator selection.”

09–12/2020 Dylan Chima-Sanchez, Department of Statistics, University of California, Berkeley, through a research mentorship program, with Mark van der Laan (faculty sponsor). Project: “Evaluating efficient nonparametric inverse probability weighted estimators.”

## Teaching

### *Short Course and Workshop Instructorship*

- 2021–present “Causal Mediation: Modern Methods for Path Analysis”, with Iván Díaz and Kara E. Rudolph
- May 2021: Authored and revised instructional materials and provided one-on-one instruction for a 1-day short course offered at the annual meeting of the Society for Epidemiologic Research, held in ~~San Diego, CA, USA~~ virtually (due to COVID-19).
    - Workshop materials: [https://code.nimahejazi.org/ser2021\\_mediation\\_workshop](https://code.nimahejazi.org/ser2021_mediation_workshop)
- 2019–present “Targeted Learning in the t1verse: Causal Inference Meets Machine Learning”, with Mark J. van der Laan, Alan E. Hubbard, Jeremy R. Coyle, Ivana Malenica, and Rachael V. Phillips
- Aug. 2021: Authored and revised instructional materials and provided one-on-one instruction for a 4-day short course offered for the Programa de Actualización en Salud Pública y Epidemiología of the Instituto Nacional de Salud Pública, held in CDMX, MX.
    - Workshop materials: <https://tlverse.org/paspe2021-workshop>
  - Jun. 2021: Authored and revised instructional materials and provided one-on-one instruction for a 1-day short course offered at the annual meeting of the Society for Epidemiologic Research, held in ~~San Diego, CA, USA~~ virtually (due to COVID-19).
    - Workshop materials: <https://tlverse.org/ser2021-workshop>
  - Mar. 2021: Authored and revised instructional materials and provided one-on-one instruction for a 1-day short course offered at the annual meeting of the Eastern North American Region of the International Biometric Society, held in ~~Baltimore, MD, USA~~ virtually (due to COVID-19).
    - Workshop materials: <https://tlverse.org/enar2021-workshop>
  - Feb. 2020: Authored and revised instructional materials and provided one-on-one instruction for a 1-day short course offered at the Conference on Statistical Practice, held in Sacramento, CA, USA.
    - Workshop materials: <https://tlverse.org/csp2020-workshop>
  - Nov. 2019: Co-presented module on targeted maximum likelihood estimation with the `tmle3` R package as part of a 3-hour invited short course for the Bill & Melinda Gates Foundation, held in Seattle, WA, USA.
    - Workshop materials: <https://tlverse.org/gates2019-workshop>
  - Oct. 2019: Presented modules on ensemble machine learning and estimating causal effects of stochastic interventions with the `s13` and `tmle3shift` R packages, and provided one-on-one instruction for a 3-day short course for the group of Prof. Ashley I. Naimi, Dept. of Epidemiology, Univ. of Pittsburgh.
    - Workshop materials: <https://tlverse.org/pitt2019-workshop>
  - May 2019: Presented a module on estimating causal effects of stochastic treatment regimes with the `tmle3shift` R package, authored and revised instructional materials, and provided one-on-one instruction, for a 1-day short course offered at the Atlantic Causal Inference Conference, held in Montréal, QC, Canada.
    - Workshop materials: <https://tlverse.org/acic2019-workshop>

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.

- Jan. 2019: “Scientific Computing with Bash, Git, and R”; workshop instructor (co-instructors: S. Peterson, N. Varoquaux); Berkeley Institute for Data Science; website: <https://bids.github.io/2019-01-17-bids/>
- 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>
- 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.

- 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>
- 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>

2017–2019 The R Bootcamp, module instructor with Dr. Christopher J. Paciorek, Department of Statistics, University of California, Berkeley

- Course description: <https://www.stat.berkeley.edu/computing/r-bootcamp>
- Aug. 2019: Presented modules on programming concepts (e.g., iteration, flow control, writing functions) and statistical data analysis and modeling; revised instructional materials; and provided one-on-one instruction.
  - Workshop materials: <https://github.com/berkeley-scf/r-bootcamp-fall-2019>
- Jan. 2019: Presented modules on data tidying and modern data visualization, focusing on the tidyverse ecosystem of R packages; revised instructional materials; and provided one-on-one instruction.
  - Workshop materials: <https://github.com/berkeley-scf/r-bootcamp-winter-2019>
- Aug. 2018: Presented modules on modern data manipulation and visualization, with a focus on the tidyverse ecosystem of R packages; contributed new instructional materials; and provided one-on-one instruction.
  - Workshop materials: <https://github.com/berkeley-scf/r-bootcamp-2018>
- Aug. 2017: Presented a module on programming concepts in R (e.g., iteration, flow control, writing functions), contributed new materials on modern functional programming, and provided one-on-one instruction.
  - Workshop materials: <https://github.com/berkeley-scf/r-bootcamp-2017>

### Guest Lectures Delivered

Spring 2020 “Building R Packages: ‘Good Enough’ Practices for Applied Statistics”, Biostatistics Master’s Thesis Seminar (PBHLTH 292), *Instructor*: Corinne A. Riddell, Division of Biostatistics, School of Public Health, University of California, Berkeley

### Teaching Assistantships

Spring 2021 “Biomedical Big Data Training Capstone Seminar” (PBHLTH 290), with Prof. Mark J. van der Laan, Division of Biostatistics, School of Public Health, University of California, Berkeley

- Newly designed advanced graduate-level seminar course of 31 masters and doctoral students.
- Graduate student instructor overall effectiveness rating: TBD/7.00 (TBA departmental average).
- Co-designed all course content (based on our in-development book *Targeted Learning in R: Causal Data Science with the tidyverse Software Ecosystem*), and jointly supervised collaborative research projects.

- Fall 2020 “Survival Analysis and Causality” (PBHLTH 240B & STAT 245B), with Prof. Mark J. van der Laan, Division of Biostatistics, School of Public Health, University of California, Berkeley
- Advanced graduate-level lecture course of 24 masters and doctoral students.
  - Graduate student instructor overall effectiveness rating: 6.67/7.00 (6.34 departmental average).
  - Designed and delivered weekly computer laboratory instruction, and supervised course research projects.
- Spring 2020 “Biomedical Big Data Training Capstone Seminar” (PBHLTH 290), with Prof. Alan E. Hubbard, Division of Biostatistics, School of Public Health, University of California, Berkeley
- Advanced graduate-level seminar course of 8 doctoral students.
  - Designed and delivered course lectures and jointly supervised collaborative research projects.
- Fall 2019 “Longitudinal Data Analysis” (PBHLTH 242C & STAT 247C), with Prof. Alan E. Hubbard, Division of Biostatistics, School of Public Health, University of California, Berkeley
- Introductory graduate-level lecture course of 33 masters and doctoral students.
  - Graduate student instructor overall effectiveness rating: 5.56/7.00 (6.23 departmental average).
  - Delivered weekly computer laboratory instruction, and provided support for course data analysis projects.
- Spring 2018 “Targeted Learning in Biomedical Big Data” (PBHLTH 290), with Prof. Mark J. van der Laan, Division of Biostatistics, School of Public Health, University of California, Berkeley
- Advanced graduate-level seminar course of 11 masters and doctoral students.
    - Course website: <https://vanderlaan-group.github.io/tlbbd-sp2018>
    - Laboratory materials: <https://github.com/tlbbd-sp2018>
  - Developed all laboratory materials and assessments, delivered weekly computer laboratory instruction, and supervised course projects.

### *Miscellaneous Teaching*

- 2017–2019 Private Tutor in Statistics (introductory 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: Linux (Ubuntu, Debian, Arch, Red Hat), macOS / OS X, Windows  
 Languages: R, PYTHON, JULIA, shell scripting, SQL, L<sup>A</sup>T<sub>E</sub>X, markdown, HTML  
 Tools: Git, GitHub, Jupyter, Google Colaboratory, R Markdown, Amazon Web Services (EC2), Google Compute Engine, Google Workspace, Microsoft Office

## Accreditation & Training

- 2021 Summer Institute in Statistics and Modeling in Infectious Diseases, University of Washington  
 2018 Biomedical Big Data Training Program, University of California, Berkeley  
 2017 Instructor Training Program Certification, Software Carpentry  
 2017 Summer Institute in Statistics for Big Data, University of Washington  
 2016 Genomic Data Science Specialization, Johns Hopkins University via Coursera  
 2015 Data Science Specialization, Johns Hopkins School of Public Health via Coursera



---

## Selected Courses (credit or audit)

- 2016–2021 **Berkeley (graduate):** Theoretical Stat. I (W. Fithian), Theoretical Stat. II (M.I. Jordan), Applied Stat. I (P.B. Stark), Applied Stat. II (J.D. McAuliffe), Stat. Genomics (S. Dudoit), Machine Learning in Bio. (J. Listgarten), Prob. Models in Genomics (Y.S. Song), Intro. Causal Inference (M.L. Petersen), Modern Causal Inference (M.L. Petersen), Observational Studies & Sensitivity Analysis (S.D. Pimentel), Experimental Design (S.D. Pimentel), Causal Inference (P. Ding), Adaptive Designs (M.J. van der Laan), Targeted Learning (M.J. van der Laan), Fairness in Machine Learning (M. Hardt), Deep Reinforcement Learning (S. Levine), Deep Time-Series Learning (L. El Ghaoui), Stat. Consulting (P.B. Stark, L.R. Goldberg, J.D. McAuliffe)
- 2011–2015 **Berkeley (undergraduate):** Modern Biostat. Theory (M.J. van der Laan and A.E. Hubbard), Survival Analysis (M.J. van der Laan), Comp. Stat. in Biology (S. Dudoit), Repro. Stat. Data Sci. (K.J. Millman), Longit. Data Analysis (N.P. Jewell), Multivar. Stat. (L. Li), Big Data in Public Health (L. Li), Fourier Analysis (J.A. Strain), Mathematical Bio. (L. Pachter), Chemical Bio. (M.B. Francis), Biophysical Neurobio. (H.E. Lecar), Molecular Neurobio. (N. Caporale), Mammalian Neuroanatomy (H. Roelink), Human Neuropsych. (R.T. Knight), Functional MRI (M.T. D'Esposito), Genetics and Genomics (E. Ünal), Intro. Epidemiology (L.F. Barcellos)

---

## Distractions

Trail and distance running (recreational)

- Urban/trail running: ~60–120 miles per month.
- Cross Country (2007, 2008, 2010)