

# Salvador Balkus

## Curriculum Vitae

✉ [sbalkus@g.harvard.edu](mailto:sbalkus@g.harvard.edu)  
🌐 [salbalkus.github.io](https://salbalkus.github.io)  
🐙 [salbalkus](#)  
in [salvador-balkus](#)  
🆔 0000-0003-4695-833X

### Research Interests

Statistical Causal inference, debiased machine learning, spatial and network data  
Scientific Environmental health, pollution, public policy

### Education

2022—2027 **Doctorate of Philosophy in Biostatistics**, *Harvard University*  
Advised by Professor Nima Hejazi  
2018—2022 **Bachelor of Science in Data Science**, *University of Massachusetts Dartmouth*  
Minors: Mathematics, Economics. GPA: 4.0/4.0; Summa Cum Laude

### Experience

2022—Present **Graduate Student Researcher**, *NSH Lab + National Studies on Air Pollution and Health*,  
Harvard University, Cambridge, MA  
Developed new machine learning methodology and open-source software packages for causal inference  
in spatial and network data with continuous treatments.  
2020—2022 **Research Assistant**, *Computational Statistics & Data Science Lab*,  
University of Massachusetts Dartmouth, North Dartmouth, MA  
Harmonized dietary data in SAS, developed R package for COVID-19 modeling.  
Jun—Aug 2021 **Research Project Manager**, *Research in Industrial Projects for Students*,  
Institute for Pure and Applied Mathematics, Los Angeles, CA  
Led team to develop object tracking algorithms and visualizations for The Aerospace Corporation.  
Jun—Aug 2020 **Student Researcher**, *Ecological Modeling REU*,  
University of Wisconsin, La Crosse, Remote  
Led team to develop forest cover classification model for U.S. Geological Survey.  
2019—2020 **Research Assistant**, *Public Policy Center*,  
University of Massachusetts Dartmouth, North Dartmouth, MA  
Created reports and infographics communicating data to local governments and organizations.

### Publications

- [1] **S. V. Balkus** and N. S. Hejazi, “CausalTables.jl: Simulating and storing data for statistical causal inference in Julia”, *Journal of Open Source Software*, vol. 10, no. 106, p. 7580, 2025. DOI: 10.21105/joss.07580.
- [2] **S. V. Balkus**, S. W. Delaney, and N. S. Hejazi, “The causal effects of modified treatment policies under network interference”, *arXiv:2412.02105*, 2024.
- [3] **S. V. Balkus** and D. Yan, “Improving short text classification with augmented data using GPT-3”, *Natural Language Engineering*, pp. 1–30, 2023. DOI: 10.1017/s1351324923000438.

- [4] **S. V. Balkus**, H. Fang, and H. Wang, “Federated fuzzy clustering for longitudinal health data”, *2022 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pp. 128–132, 2022.
- [5] **S. V. Balkus**, H. Wang, B. D. Cornet, C. Mahabal, H. Ngo, and H. Fang, “A survey of collaborative machine learning using 5G vehicular communications”, *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1280–1303, 2022. DOI: 10.1109/comst.2022.3149714.
- [6] V. S. Gurugubelli, H. Fang, J. M. Shikany, **S. V. Balkus**, J. Rumbut, H. Ngo, H. Wang, J. J. Allison, and L. M. Steffen, “A review of harmonization methods for studying dietary patterns”, *Smart Health*, vol. 23, p. 100263, 2022. DOI: 10.1016/j.smhl.2021.100263.
- [7] **S. V. Balkus**, H. Fang, J. Rumbut, A. Moormann, and E. Boyer, “A multi-level biosensor-based epidemic simulation model for COVID-19”, *IEEE Internet of Things Journal*, pp. 1–1, 2021. DOI: 10.1109/jiot.2021.3127804.
- [8] **S. V. Balkus**, J. Rumbut, H. Wang, and H. Fang, “An adaptive and dynamic biosensor epidemic model for COVID-19”, in *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, IEEE, Aug. 2020. DOI: 10.1109/iri49571.2020.00051.

## Presentations

- [1] **S. V. Balkus**, “Speeding up the Highly Adaptive Lasso”, in *Harvard Biostatistics Student Seminar*, Mar. 2025.
- [2] **S. V. Balkus**, “Statistics in Julia: Is it right for you?”, in *Harvard Biostatistics Student Seminar*, Dec. 2024.
- [3] **S. V. Balkus**, “Nonparametric network causal inference for continuous exposures in mobile source air pollution”, in *American Causal Inference Conference*, May 2024.
- [4] **S. V. Balkus**, “Assumption-lean causal inference for mobile source air pollution”, in *ASA Boston Chapter Student Research Symposium on Statistics and Data Science*, Apr. 2024.
- [5] **S. V. Balkus**, “Language models that teach themselves: Augmenting training data for topic classification using GPT-3”, in *ASA Boston Chapter Student Research Symposium on Statistics and Data Science*, Apr. 2022.
- [6] N. Pai, **S. V. Balkus**, and T. Zeng, “Multi-hypothesis tracking of space objects and targets”, in *AMS Joint Mathematics Meetings (JMM) Poster Session*, Apr. 2022.
- [7] **S. V. Balkus**, “Multi-level biosensor-based epidemic forecasting in small areas”, in *Joint Statistical Meetings*, American Statistical Association, Aug. 2021.
- [8] **S. V. Balkus**, “Lunchtime Computing: Basics of AWS Sagemaker”, in *Center for Scientific Computing and Visualization Research*, University of Massachusetts Dartmouth, Feb. 2021.
- [9] **S. V. Balkus**, M. McDevitt, and N. Dean, “A classification system for characterizing diversity across floodplain forests of the Upper Mississippi River System”, University of Wisconsin La Crosse, Aug. 2020.
- [10] **S. V. Balkus**, “Lunchtime Computing: Getting started with Git and GitHub”, in *Center for Scientific Computing and Visualization Research*, University of Massachusetts Dartmouth, Feb. 2020.

## Awards

- 2025 **Early Career Travel Award**, *Joint Statistical Meetings*, Nashville, TN
- 2024—2025 **Certificate of Distinction in Teaching**, *BST231: Methods I*, Harvard University
- 2022 **Graduate Research Fellowship**, *National Science Foundation*
- 2022 **Best Analysis and Visualization**, *SEMASS DataFest*, American Statistical Association
- 2021 **Travel Award**, *Joint Mathematics Meetings*, American Mathematical Society

---

## Teaching

2024—2025	<b>Qualifying Exam Tutor</b> , <i>Topics: Regression, Optimization</i> , Harvard University
Summer 2025	<b>Project Mentor</b> , <i>StatStart</i> , Harvard University
Summer 2024	<b>Workshop Instructor</b> , <i>Replicathon, IQ BIO REU</i> , University of Puerto Rico Río Piedras
Summer 2024	<b>Curriculum Fellow</b> , <i>BST 232: Methods I</i> , Harvard University
2023—2024	<b>Teaching Fellow</b> , <i>BST 232: Methods I</i> , Harvard University

---

## Service and Extracurricular

### Ad-hoc Reviewer

Journal of Causal Inference (3x), Journal of Open Source Software (3x), Natural Language Processing

2024 —Present	<b>Co-Chair</b> , <i>Biostatistics Student Committee</i> , Harvard University Led peer mentoring program, qualifying exam tutoring program, social event planning, and resource organization for 70+ PhD students.
2022 —2025	<b>Graphic Designer and Blog Contributor</b> , <i>Science in the News</i> , Harvard University Designed infographics and wrote blog posts explaining scientific topics to a lay audience.
2023 —2024	<b>Committee Chair</b> , <i>Biostatistics Peer Mentoring Program</i> , Harvard University Supported class of 12 incoming students by serving as peer mentor, creating instructional material, acquiring event funding, and planning seminar series.
2022	<b>Biostatistics Session Chair</b> , <i>Student Research Symposium on Statistics and Data Science</i> , ASA Boston Chapter
2020—2022	<b>President</b> , <i>Big Data Club</i> , University of Massachusetts Dartmouth Organized data science workshops and networked with local clients for consulting projects

---

## Scientific Blog Contributions

- [1] A. Bogatch (Author) and **S. V. Balkus (Illustrator)**, “Pregnancy causes astonishing brain changes why haven’t we noticed this before?”, *Harvard Science in the News*, May 2025.
- [2] S. Renauld (Author) and **S. V. Balkus (Illustrator)**, “Gene editing to treat disease: A success story in sickle cell disease”, *Harvard Science in the News*, May 2024.
- [3] H. Blume (Author) and **S. V. Balkus (Illustrator)**, “Liquid-within-liquid compartments: A ground-breaking discovery in biology”, *Harvard Science in the News*, Jun. 2023.
- [4] **S. V. Balkus (Author and Illustrator)**, “How do scientists know whether to trust their results?”, *Harvard Science in the News*, Jan. 2023.

---

## Skills

Systems	Linux (Fedora), Windows
Languages	Julia, R, Python, $\text{\LaTeX}$ , markdown, HTML/CSS
Tools	Git, GitHub, Quarto, RStudio, VSCode, AWS, SLURM

---

## Open-Source Contributor To...

- The Book of Statistical Proofs, <https://statproofbook.github.io/>
- TMLE.jl
- DensityRatioEstimation.jl

*last updated July 10, 2025*