

Education

- 2019 **PhD, Biostatistics**, *University of Washington*, Seattle, WA
Dissertation: "Statistical miscellany: causality, networks, and bandits"
Advisor: Ali Shojaie, PhD
- 2014 **BMath, Statistics & Pure Mathematics**, *University of Waterloo*, Waterloo, ON

Work Experience

- 2019-
present **Senior Data Scientist**, *Flatiron Health*, New York, NY
Research and develop novel statistical methods for real world EHR-derived data to support oncology research. Develop machine learning models and analytics tools for clinical decision support.
- 2014-
2019 **Research & Teaching Assistant**, *University of Washington*, Seattle, WA
Developed statistical methods for large network-structured data, with applications to genomics and metabolomics. Taught introductory biostatistics and machine learning courses to public health graduate students.
- 2018 **Research Scientist Intern**, *Facebook*, Menlo Park, CA
Developed novel method for improved counterfactual policy evaluation in contextual bandit and reinforcement learning settings.
- 2017 **Data Scientist Intern**, *Google*, New York, NY
Implemented machine learning methods to improve causal ad attribution product.

Publications

- 2023 Arjun Sondhi, Alexander S Rich, Siruo Wang, and Jeffrey T Leek. "Postprediction Inference for Clinical Characteristics Extracted With Machine Learning on Electronic Health Records". Published in *JCO: Clinical Cancer Informatics*.
- 2023 Corey M Benedum, Arjun Sondhi, Erin Fidyk, Aaron B Cohen, Sheila Nemeth, Blythe Adamson, Melissa Estévez, and Selen Bozkurt. "Replication of Real-World Evidence in Oncology Using Electronic Health Record Data Extracted by Machine Learning". Published in *Cancers*.
- 2023 Arjun Sondhi, Janick Weberpals, Chengsheng Jiang, Prakirthi Yerram, Michael Taylor, Meghna Samant, and Sarah T Cherng. "A systematic approach towards missing lab data in electronic health records: a case study in non-small cell lung cancer and multiple myeloma". Published in *CPT: Pharmacometrics & Systems Pharmacology*.
- 2022 Arjun Sondhi "Estimating survival parameters under conditionally independent left truncation". Published in *Pharmaceutical Statistics*.
- 2021 Arjun Sondhi, Brian D Segal, Jeremy Snider, Olivier Humblet, and Margaret McCusker. "Bayesian additional evidence for decision making under small sample uncertainty". Published in *BMC Medical Research Methodology*.

- 2021 Jean Feng, Arjun Sondhi, Jessica Perry, and Noah Simon. "Selective prediction-set models with coverage guarantees". Published in *Biometrics*.
- 2021 David Arbour, Drew Dimmery, and Arjun Sondhi. "Permutation Weighting". Published in *International Conference on Machine Learning (ICML 2021)*.
- 2020 Arjun Sondhi, David Arbour, and Drew Dimmery. "Balanced off-policy evaluation in general action spaces". Published in *International Conference on Artificial Intelligence and Statistics (AISTATS 2020)*.
- 2019 Arjun Sondhi and Ali Shojaie. "The Reduced PC-Algorithm: Improved Causal Structure Learning in Large Random Networks". Published in *Journal of Machine Learning Research*.
- 2017 Arjun Sondhi and Kenneth M. Rice. "Fast permutation tests and related methods, for association between rare variants and binary outcomes". Published in *Annals of Human Genetics*.
- 2016 Arjun Sondhi and Ali Shojaie. "Causal structure learning with reduced partial correlation thresholding". Published in *IEEE Conference on Data Science and Advanced Analytics (DSAA 2016)*.

Abstracts

- 2023 Xavier Orcutt, Ronac Mamtani, Arjun Sondhi, Aaron B Cohen, Ravi B Parikh. "Evaluating generalizability of practice-changing randomized clinical trials in non-small cell lung cancer using machine learning-based in-silico trials." *ASCO Annual Meeting 2023*.
- 2023 Maneet Kaur, Filip Frahm, Mustafa Ascha, Jenny Guadamuz, Adam Gottesman, Barry Leybovich, Arjun Sondhi, Yihua Zhao, Neal J Meropol, Trevor J Royce. "Trial eligibility criteria (EC) and diversity among patients with advanced non small cell lung cancer (advNSCLC)." *ASCO Annual Meeting 2023*.
- 2022 Corey Benedum, Blythe Adamson, Aaron B Cohen, Melissa Estevez, Arjun Sondhi, Erin Fidyk, Sheila Nemeth, Selen Bozkurt. "Machine learning-accelerated outcomes research: a real-world case study of biomarker-associated overall survival in oncology." *ISPOR Europe 2022*.
- 2022 Arjun Sondhi, Corey Benedum, Aaron B Cohen, Sheila Nemeth, Selen Bozkurt. "Can ML-extracted variables reproduce real world comparative effectiveness results from expert-abstracted data? A case study in metastatic non-small cell lung cancer treatment." *ISPOR Europe 2022*.
- 2021 Jaron Lee, Melissa Estevez, Brian D Segal, Arjun Sondhi, Aaron B Cohen, Sarah T Cherng. "Quantifying bias in ML-extracted variables for inference in clinical oncology" *ISPOR 2021*.
- 2021 Akshay Swaminathan, Jeremy Snider, Arjun Sondhi, Meghna Samant, Olivier Humblet. "Statistical methods for pantumor analysis: models to account for tumor-level heterogeneity" *ISPOR 2021*.

- 2021 Felipe Batalini, Russell Madison, Dean C Pavlick, Ethan Sokol, Tamara Snow, Arjun Sondhi, Garrett M Frampton, Colby Jenkins, Judy Ellen Garber, Gerburg M Wulf, Jeffrey M Venstrom, Nadine M Tung, Emily Castellanos, Alexa B Schrock, Kimberly McGregor. "Analysis of real-world (RW) data for metastatic breast cancer (mBC) patients (pts) with somatic BRCA1/2 (sBRCA) or other homologous recombination (HR)-pathway gene mutations (mut) treated with PARP inhibitors (PARPi)." *ASCO Annual Meeting 2021*.

Presentations

- 2023 *Postprediction Inference for Clinical Characteristics Extracted With Machine Learning on Electronic Health Records*. Joint Statistical Meetings 2023.
- 2019 *Balanced off-policy evaluation*. WNAR IBS Meeting 2019.
- 2018 *The Reduced PC-Algorithm: Improved Causal Structure Learning in Large Random Networks*. Joint Statistical Meetings 2018.
- 2017 *Association testing for rare genetic variants*. American Society for Human Genetics Meeting 2017.
- 2017 *Causal structure learning in high dimensions*. Joint Statistical Meetings 2017.
- 2016 *Association testing for rare genetic variants*. International Biometric Conference 2016.

Teaching Experience

- 2023- **Data Science Instructor**, *BrainStation*
- present Teach 8-week intensive data science course for working professionals
- 2014- **Graduate Teaching Assistant**, *University of Washington*
- 2019 Statistical Inference, Advanced Regression Methods, Applied Biostatistics, Introduction to Machine Learning
- 2010- **Peer Tutor/Undergraduate Teaching Assistant**, *University of Waterloo*
- 2014 Calculus I-II, Linear Algebra I-II, Linear Optimization, Introduction to Statistics

Honours and Awards

- 2018 ASA Biopharmaceutical Section Student Paper Award
- 2017 UW Graduate School Fund for Excellence and Innovation Award
- 2016 WNAR-IBS Conference Travel Award
- 2013 NSERC Undergraduate Student Research Award

Software

- glmfunk R package with C++ backend implementing generalized linear models for two-way network-structured data.
Available on GitHub: <https://github.com/asondhi/glmfunk>
- AUtests R package implementing a variety of association tests for rare genetic variants.
Available on CRAN: <https://cran.rstudio.com/web/packages/AUtests/>