

Aaron J. Fisher

Biostatistics & Machine Learning Researcher

🌐 aaronjfisher.com | ✉ afishe27@alumni.jh.edu | 🐦 @ajfishr | 📍 Boston, MA

Education

PhD in Biostatistics (2016), *Johns Hopkins Bloomberg School of Public Health*, Baltimore, MD

- Advisors: Vadim Zipunnikov & Brian Caffo
- Awards & Scholarships:
 - Margaret Merrell Award for outstanding research by a Biostatistics doctoral student
 - June B. Culley Award for outstanding achievement on an oral examination paper
 - Doctoral Training Grant in Environmental Biostatistics

BA in Economics (2010), *University of Rochester*, Rochester, NY

- Summa cum laude

Professional experience

Principal Statistician (2019-Present), *Takeda Pharmaceuticals, Statistics and Quantitative Sciences*, Boston, MA

- Analysis of wearable devices in early-stage clinical trials (with Dmitri Volfson)

Postdoctoral Research Fellow (2016-2019), *Department of Biostatistics at the Harvard T.H. Chan School of Public Health*, Boston, MA

- Research on interpretability for machine learning models (with Francesca Dominici & Cynthia Rudin)

Statistical Consultant (2016), *Pfizer*, Boston, MA

- Analysis of wearable devices and temperature probes in human sleep studies

Skills

Statistics & Machine Learning: Causal inference, matrix decompositions, regression in a RKHS, Bayesian regression trees, random forests, neural networks, finite sample bounds, adaptive clinical trials, non-convex quadratic programming, functional data analysis

Computing: R package development, git, Python, PyTorch, MATLAB, Stata, shell scripting, \LaTeX

Summary of selected projects

Fast bootstrap PCA for millions of covariates: In order to quantify uncertainty for the dominant patterns in brain MRI data, with millions of voxels measured for each patient, we developed a fast bootstrap principal component analysis procedure. Our approach reduces computation time from the order of days to the order of minutes with no loss in accuracy. This work was published in *JASA T&M*, and my associated R package (bootSVD) has been downloaded over 23,000 times.

Interpretability for machine learning models: Black-box, proprietary prediction models can provoke distrust and produce predictions that are difficult to combine with supplementary information in order to make decisions. In response, we proposed a method for estimating

how much unknown proprietary models rely on different covariates, which we applied to the criminal recidivism model COMPAS. The technical aspects of our work combined approaches from finite sample theory, U-statistics, covering numbers, and non-convex quadratic programming. Our paper was published in *JMLR* and has been cited over 169 times ([link](#)), including citations to previous arXiv versions.

Visual intuition for influence functions used in causal inference: Influence functions have emerged as a popular, though highly technical, framework for combining machine learning with statistical inference, especially within the field of real world (observational) data analysis. In order to make influence functions more approachable, our educational paper, published in *The American Statistician*, builds intuition based on rigorous, visual illustrations.

Fast prognostic scores for prostate cancer screening: We implemented an importance sampling approach to obtain fast, in-clinic, Bayesian risk updates in response to new patient information.

[Selected academic papers \(see Google Scholar for full list\)](#)

Submitted

A. J. Fisher (2020). Treatment effect bias from sample snooping: Blinding outcomes is neither necessary nor sufficient. ([link](#).)

Peer-reviewed publications

A. J. Fisher & E. H. Kennedy (2020). Visually communicating and teaching intuition for influence functions. *The American Statistician*. ([link](#).)

A. J. Fisher, C. Rudin, F. Dominici (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *The Journal of Machine Learning Research*. ([paper link](#); [169 citations](#) as of July 16, 2020, including citations to previous arXiv versions)

A. J. Fisher & M. Rosenblum (2018). Stochastic optimization of adaptive enrichment designs for two subpopulations. *Journal of Biopharmaceutical Statistics*. ([link](#).)

R. Y. Coley, **A. J. Fisher**, M. Mamawala, H. B. Carter, K. J. Pienta, S. L. Zeger (2017). A Bayesian hierarchical model for prediction of latent health states from multiple data sources with application to active surveillance of prostate cancer. *Biometrics*. ([link](#).)

A. J. Fisher, B. Caffo, B. Schwartz, V. Zipunnikov (2016). Fast, exact bootstrap principal component analysis for $p > 1$ million. *Journal of the American Statistical Association* TM. ([link](#).)

A. J. Fisher, G. B. Anderson, R. Peng, J. Leek (2014). A randomized trial in a massive online open course shows people don't know what a statistically significant relationship looks like, but they can learn. *PeerJ*. ([link](#); 10,610 unique visitors as of July 16, 2020.)

Reviewer service

Journal of the American Statistical Association; Journal of Machine Learning Research; Journal of Computational and Graphical Statistics; Computational Statistics and Data Science; Risk Analysis.