# Aaron Fisher

*Statistics & Machine Learning Researcher*
*Boston, MA*
✉ *afishe27@alumni.jh.edu*
`https://aaronjfisher.github.io`

---

## Education

| | |
|---|---|
| 2016 | **PhD in Biostatistics**, *Johns Hopkins Bloomberg School of Public Health*, Baltimore, MD.<br>Advisors: Vadim Zipunnikov & Brian Caffo<br>Dissertation: Methods for High Dimensional Analysis, Multiple Testing, and Visual Exploration |
| 2010 | **BA in Economics**, *University of Rochester*, Rochester, NY.<br>Summa cum laude |

## Professional Experience

| | |
|---|---|
| 2019-Present | **Principal Statistician**, *Takeda Pharmaceuticals, Statistics and Quantitative Sciences*, Boston, MA.<br>Analysis of wearable devices in early-stage clinical trials (with Dmitri Volfson) |
| 2016-2019 | **Postdoctoral Research Fellow**, *Harvard T.H. Chan School of Public Health, Dept of Biostatistics*, Boston, MA.<br>Research on interpretability for machine learning models (with Francesca Dominici & Cynthia Rudin) |
| 2016 | **Statistical Consultant**, *Pfizer*, Boston, MA.<br>Analysis of wearable devices and temperature probes in human sleep studies |

## Academic Papers

| | |
|---|---|
| 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.)

T. Qian, E. Colantuoni, **A. J. Fisher**, M. Rosenblum, for the Alzheimer's Disease Neuroimaging Initiative (2017). Sensitivity of adaptive enrichment trial designs to accrual rates, time to outcome measurement, and prognostic variables. *Contemporary Clinical Trials Communications.* (link.)

Y. Webb-Vargas, S. Chen, **A. J. Fisher,** A. Mejia, Y. Xu, C. Crainiceanu, B. Caffo, M. A. Lindquist (2017). Big data and neuroimaging. *Statistics in Biosciences.* (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.)

M. Rosenblum, T. Qian, Y. Du, H. Qiu, **A. J. Fisher** (2016). Multiple testing procedures for adaptive enrichment designs: Combining group sequential and reallocation approaches. *Biostatistics.* (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.)

Technical Reports  **A. J. Fisher,** R. Y. Coley, S. L. Zeger (2015). Fast out-of-sample predictions for Bayesian hierarchical models of latent health states. (link.)

**A. J. Fisher**, H Jaffee, M Rosenblum (2014). interAdapt – An interactive tool for designing and evaluating randomized trials with adaptive enrollment criteria. (link.)

---

## Awards and Scholarships

2016  **Margaret Merrell Award (co-winner with Amanda Mejia):** Departmental award recognizing outstanding research by a Biostatistics doctoral student (link).

2014  **June B. Culley Award:** Honors outstanding achievement by a Biostatistics student on his or her school-wide oral examination paper (link).

2012-2015  **Doctoral Training Grant in Environmental Biostatistics:** Provides funding for at least three years.

2006-2010  **Undergraduate Awards:** Phi Beta Kappa; John Dows Mairs Prize (University of Rochester Economics Dept); Omicron Delta Epsilon International Honor Society for Economics; Theta Chi Long, Walter, Ott Award; Theta Chi Valentine H. Zahn Fund.

---

## Software

**bootSVD**  (23,000 downloads as of July 16, 2020) an R package for implementing fast, exact bootstrap principal component analysis and singular value decompositions for high dimensional data (i.e. $> 1$ million covariates). Matrices too large for memory can be entered as class `ff` objects, with contents stored on disk (CRAN link; GitHub link).

| | |
|---|---|
| **ggBrain** | An R package for beautiful brain image figures with ggplot. This packages allows color to be mapped to both (1) tissue intensities of the template image, and (2) values of a voxel-wise test statistic (<u>GitHub link</u>). |
| **interAdapt** | An interactive tool for designing and evaluating randomized trials with adaptive enrollment criteria (<u>Shiny App link</u>; <u>CRAN link</u>; <u>Github link</u>). |

## 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 |

## Reviewer

Journal of the American Statistical Association (1); Journal of Machine Learning Research (1); Journal of Computational and Graphical Statistics (1); Computational Statistics and Data Science (1); Risk Analysis (2).

## Teaching

### Co-Instructor

| | |
|---|---|
| 2015 | Statistical Reasoning I and II: My role included teaching independently for 13 hours of lectures *(MPH Level Course, JHSPH Summer Institute of Epidemiology and Biostatistics)* |

### Guest Lecturer

| | |
|---|---|
| 2013 | Essentials of Probability and Statistical Inference I-II *(Biostatistics ScM Level, JHSPH)* |

### Lab Lecturer with Content Design

| | |
|---|---|
| 2012-2014 | Essentials of Probability and Statistical Inference I-IV: Designed and administered a weekly 1-hour lab lecture. In the second year of this course, we reduced this lab to a 1-hour session every two weeks. *(Biostatistics ScM Level, JHSPH)* |

### Lab Lecturer without Content Design

| | |
|---|---|
| 2014-2015 | Statistical Methods in Public Health II: Administered approximately 16 hours of lab lecture in each year of the course. *(MPH Level, JHSPH)* |

### Educational Presentations

| | |
|---|---|
| 2013-2015 | JHU Biostatistics Journal Club: I have given talks on high dimensional asymptotics, adaptive clinical trials, and the Bayesian Bootstrap |
| 2013-2015 | JHU Biostatistics Computing Club: I have given talks on environments in R, and on LaTeX |

### General TA Roles

2014-2015   Statistical Methods in Public Health I and IV *(MPH Level, JHSPH)*

2012   Statistical Reasoning I and II, *(MPH Level, JHSPH Summer Institute of Epidemiology and Biostatistics)*

## Conference Presentations

2018   "Model Class Reliance: Variable Importance when all Models are Wrong, but *Many* are Useful." JSM, Vancouver. *Contributed Speed Talk & Poster.*

2018   "Model Class Reliance: Variable Importance when all Models are Wrong, but *Many* are Useful." Conference on Statistical Learning and Data Science / Nonparametric Statistics, New York City, NY. *Invited Talk.*

2018   "Model Class Reliance: Variable Importance when all Models are Wrong, but *Many* are Useful." Atlantic Causal Inference Conference, Pittsburgh, PA. *Contributed Poster.*

2018   "Model Class Reliance: Variable Importance when all Models are Wrong, but *Many* are Useful." Harvard Chan Poster Day, Boston, MA. *Contributed Poster.*

2016   "Optimizing Adaptive Enrichment Designs." JSM, Chicago IL. *Contributed Poster.*

2015   "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." JSM, Seattle WA. *Contributed Speed Session & Poster.*

2015   "Fast Exact Bootstrap Principal Component Analysis for $p > 1$ million." ENAR, Miami, FL. *Contributed Talk.*

2014   "Fast, Exact Bootstrap Principal Component Analysis for $p > 1$ Million." 4th Annual Hopkins Imaging Conference (link). Baltimore, MD, *Invited Short Talk & Poster.*

2014   "Fast Exact Bootstrap Principal Component Analysis for $p > 1$ million: Leveraging Low-Dimensional Structure Across High-Dimensional Bootstrap Samples." JSM, Boston, MA. *Contributed Speed Session & Poster.*

2014   "People Can't See Statistical Significance: A Massive Randomized Trial on the Visual Perception of Relationships." ENAR Spring Meeting, Baltimore, MD, *Contributed Talk.*

## Other Leadership & Service Roles

2020   Compiled list of educational resources for the Takeda Machine Learning and Artificial Intelligence Community of Practice.

2017   Invited panel member for HSPH Biostatistics departmental discussion on preparing for the job market

2016   Organizer for HSPH Biostatistics faculty panel on preparing for the job market

2015   Volunteer with Thread: Thread is a mentorship and tutoring program that enrolls underperforming high school students who face significant barriers outside of the classroom.

| | |
|---|---|
| 2015 | Facilitator at JHU Data Science Hackathon: Assisted a team through the process of scraping web data and building a shiny app (3-day event) |
| 2014-2015 | JHU Biostatistics Meat Chili Champion |
| 2013-2014 | JHU Biostatistics Vegetarian Chili Champion |
| 2012-2013 | Co-organizer of JHU Biostatistics Computing Club, with Prasad Patil (speaker schedule link) |

Last updated: July 19, 2020