Shira Mitchell
Statistician
sam942@mail.harvard.edu

Positions

2020-Present Blue Rose Research, http://BlueRoseResearch.Org/

2019-2020 Civis Analytics, Political Team

2018-2019 New York City Mayor's Office of Data Analytics (MODA)

2016-2018 Mathematica Policy Research

2014-2016 Postdoctoral Fellow, Columbia University

Statistical consultant

2011-Present Human Rights Data Analysis Group (HRDAG)

2018-Present Data for Progress

Education

2009-2014 PhD in Biostatistics, Harvard University

2005-2009 BA in Mathematics (minor in Computer Science), Harvard University

Teaching

2023 The Nelson Mandela African Institution of Science and Technology (NM-AIST)

- Instructor
    * Statistical Workflow: Experimental Design and Model Building

2021 New York University

- Instructor
    * Survey Research Methods

2007-2014 Harvard University

- Teaching Assistant
  Lectured for recitations, developed and graded homework and exams, held office hours.
    * Graph Theory & Combinatorics
    * Efficient Algorithms
    * Applied Longitudinal Analysis
    * Regression and Analysis of Variance in Experimental Research
- Tutor
  Led group review sessions for Health Care Policy PhD students.
    * Introduction to Probability
    * Statistical Inference
    * Econometric Methods in Impact Evaluation
    * Statistical Methods for Evaluating Causal Effects

Experience

2020-Present Blue Rose Research, http://BlueRoseResearch.Org/

2019-2020 Civis Analytics, Political Team

- Developing Bayesian models to reduce survey non-response bias in polls.
- Estimating causal effects of political advertising across time (considering decay and diminishing returns) and campaigns. Used to give guidance on where and when to spend resources.
- Estimating conditional average treatment effects (CATE) of messages across strata of the electorate, modeling data from hundreds of randomized message tests.

2011-Present Human Rights Data Analysis Group (HRDAG)
Developed Bayesian models for population size estimation using multiple systems estimation (also called capture-recapture). Applied methods to estimate the number of deaths in armed conflicts.

2018-2019 New York City Mayor's Office of Data Analytics (MODA)

- NYC Automated Decision Systems (ADS) Task Force
  Researched for the task force, an interdisciplinary group appointed by the Mayor to study how City agencies make decisions using data.
- NYC Department of Housing Preservation and Development (HPD)
  Modeled tenant harassment to help target building inspections.
- NYC Open Data Week
  Taught groups across the city about the City's public urban data.
- NYC Department of City Planning (DCP)
  Supported DCP's analysis of census mail return rates across neighborhoods.

2017-2019 Airwars, a journalistic transparency project, London, UK
Produced statistical graphics of reported civilian casualties from Coalition airstrikes in Iraq and Syria.

2014-2018 Millennium Villages Project (MVP)
Designed, led, and implemented a 10-country impact evaluation of the MVP, a 10-year economic development project in Africa.

2016-2018 Mathematica Policy Research

- US Department of Labor (DOL)
  Produced guidance (text, videos, tools) on survey sampling and analysis for child labor research.
- US Food and Nutrition Service (FNS)
  Used a matched-comparison method to estimate impact of direct certification with Medicaid for free / reduced-price meals.
- US Department of Defense (DOD)
  Developed multilevel regression and poststratification (MRP) to estimate TRICARE acceptance rates in hospital service areas.
- US Internal Revenue Service (IRS)
  Validated methods to identify low-income communities for designation of Qualified Opportunity Zones, using data from the American Community Survey (ACS).
- US Center of Medicare & Medicaid Services (CMS)
  Analyzed clinician and staff survey data from the Comprehensive Primary Care (CPC) initiative, accounting for sampling design and non-response. Developed a Bayesian heterogeneous-treatment-effects model for CPC's impact on Medicare costs.
- US Center of Medicare & Medicaid Services (CMS)
  Wrote a brief on small area estimation (SAE) for health care quality measures for Medicaid enrollees.
- US Institute of Education Sciences (IES)
  Wrote a brief on (mis)use of p-values in impact evaluations.

2009, 2012-2013 Management and Development for Health - Dar es Salaam, Tanzania
Managed data and analysis for HIV patient care, including estimating the impact of quality of care on maternal mortality among HIV-positive pregnant women.

2010 Ministry of Health - Lilongwe, Malawi
Developed lot quality assurance sampling methods to assess data quality in HIV clinics.

Academic Papers

2021 Mitchell, S., E. Potash, S. Barocas, A. D'Amour, K. Lum. "Prediction-Based Decisions and Fairness: A Catalogue of Choices, Assumptions, and Definitions." *Annual Review of Statistics and Its Application (2021)*, https://arxiv.org/abs/1811.07867

2018 Mitchell, S., A. Gelman, R. Ross, J. Chen, S. Bari, U. Kim Huynh, M. Harris, S.E. Sachs, E.A. Stuart, A. Feller, S. Makela, A.M. Zaslavsky, L. McClellan, S. Ohemeng-Dapaah, P. Namakula, C. Palm, and J.D. Sachs. "The Millennium Villages Project: A Retrospective Observational End-Line Evaluation." *The Lancet Global Health*, May 2018, Volume 6, Issue 5, e500 - e513.

2015 Mitchell, S., A. Ozonoff, A. M. Zaslavsky, K. Lum, and B.A. Coull. "Population Size Estimation with Inactive Lists: Hierarchical Mixture Models and Missing Data with Application to Armed Conflict Data." *Annals of Applied Statistics*, revision requested.

2015 Exner, N., S. Mitchell, and M. Pagano. "The Use of the Finite Population Correction in Survey Design for National Disease Surveillance." Working paper.

2013 Mitchell, S.A. Ozonoff, A.M. Zaslavsky, B. Hedt-Gauthier, K. Lum, and B.A. Coull. "A Comparison of Marginal and Conditional Models for Capture-Recapture Data with Application to Human Rights Violations Data." *Biometrics*, vol. 69, no. 4, December 2013, pp. 1022-1032.

2012 Hedt-Gauthier, B.L., L. Tenthani, S. Mitchell, F.M. Chimbwandira, S. Makombe, Z. Chirwa, E.J. Schouten, M. Pagano, and A. Jahn. "Improving Data Quality and Supervision of Antiretroviral Therapy Sites in Malawi: An Application of Lot Quality Assurance Sampling." *BMC Health Services Research*, vol. 12, no. 196, 2012.

2012 Mitchell, S., and M. Pagano. "Pooled Testing for Effective Estimation of the Prevalence of Schistosoma Mansoni." *American Journal of Tropical Medicine and Hygiene*, vol. 87, no. 5, 2012, pp. 850-861.

2012 Mitchell, S., and M. Pagano. "Effective Classification of the Prevalence of Schistosoma Mansoni." *Tropical Medicine International Health*, vol. 17, no. 12, December 2012, pp. 1470-1477.

2007 Gravano, A., S. Bensus, J. Hirschberg, S. Mitchell, and I. Vovsha. "Classification of Discourse Functions of Affirmative Words in Spoken Dialogue." *Proceedings of Interspeech*, August 2007, pp. 1613-1616.

Authored Media Articles

2018 Mitchell, S. "Worries Grow Census Data Could Be Used to Target the Undocumented." *The Indypendent*, May 2018, https://indypendent.org/2018/05/census-data-could-be-used-to-target-the-undocumented/

Invited Talks

2020 Department of Public Policy, Tel Aviv University, Tel Aviv, Israel (zoom)
The Millennium Villages Project: a retrospective, observational, endline evaluation

2019 Data Council NYC 2019, New York, NY
Prediction-based decisions and fairness: choices, assumptions, and definitions

2019 Statistics Department Seminar, Columbia University, New York, NY
Prediction-based decisions and fairness: choices, assumptions, and definitions

2018 Rosenkranz Global Health Policy Symposium at Stanford University, Stanford, CA
The Millennium Villages Project: a retrospective, observational, endline evaluation

2018 The Math and Democracy Seminar, New York University (NYU), New York, NY
Prediction-based decisions and fairness: choices, assumptions, and definitions

2018 Google AI, Cambridge, MA
Prediction-based decisions and fairness: choices, assumptions, and definitions

2018 Joint Statistical Meetings, Vancouver, Canada
The Millennium Villages Project: a retrospective, observational, endline evaluation

2018 Data Analytics Networking Night, New York, NY
The Millennium Villages Project: a retrospective, observational, endline evaluation

2018 Program in Global Surgery and Social Change, Harvard Medical School, Boston, MA
The Millennium Villages Project: a retrospective, observational, endline evaluation

2018 Bayesian Statistics for the Social Sciences, Columbia University, New York, NY
The Millennium Villages Project: a retrospective, observational, endline evaluation

2018 Keynote Address at the Global Health & Innovation Conference, Yale University, New Haven, CT
with S.E. Sachs and J.D. Sachs
Lessons from the Millennium Villages Health System

2018 Department of Computer Science, Princeton University, Princeton, NJ
Fairness and Causality

2017 Goldman School of Public Policy at University of California, Berkeley, Berkeley, CA
The Error of Our Ways: Making Policy Decisions Under Uncertainty

2017 Department of Public Policy, Tel Aviv University, Tel Aviv, Israel
Evaluating the Fight Against Extreme Poverty: The Millennium Villages Project

2016 Department of Mathematics and Statistics, Georgetown University, Washington D.C.
Causal Inference with Small Samples and Incomplete Baseline for the Millennium Villages Project

2016 Department of Biostatistics, Johns Hopkins, Baltimore, MD
Causal Inference with Small Samples and Incomplete Baseline for the Millennium Villages Project

2015 Department of Mathematics, Bucknell University, Lewisburg, PA
Causal Inference with Small Samples and Incomplete Baseline for the Millennium Villages Project

2015 Seminar for the Study of Development Strategies, Columbia University, New York, NY
Causal Inference with Small Samples and Incomplete Baseline for the Millennium Villages Project

2014 Joint Statistical Meetings, Boston, MA
Population Size Estimation with Inactive Lists: Hierarchical Mixture Models and Missing Data with Application to Armed Conflict Data

2014 Conference of the ASA Section on Statistical Learning and Data Mining, Durham, NC
Population Size Estimation with Inactive Lists: Hierarchical Mixture Models and Missing Data with Application to Armed Conflict Data

2013 Global Maternal Health Conference, Arusha, Tanzania
Does the Quality of Clinical Care Impact Maternal Mortality Among HIV-Positive Pregnant Women in HIV Care and Treatment Clinics in Dar es Salaam?

Skills

Proficient in R, Stan, SQL, Git/GitHub, LaTeX, Markdown, command line
Experience with Python, STATA, SAS
Languages: English (fluent), Hebrew (fluent), Spanish (basic)

Academic Service

Referee/Reviewer: *Statistical Modeling, Biometrics, Annals of Applied Statistics, American Journal of Public Health, Statistics in Medicine, Health Policy and Planning, BMC Medical Research Methodology, Journal of the Royal Statistical Society: Series A, Journal of the Royal Statistical Society: Series C, ICML Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML)*

Program Committee member: *ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT), ICLR Workshop on Debugging Machine Learning Models*

Panel member: *NeurIPS Queer in AI*

Workshop co-organizer: *NeurIPS Workshop on Consequential Decision Making in Dynamic Environments*

Academic Awards and Fellowships

2011   Robert Reed Prize (highest score on the PhD qualifying exam)

2010   National Science Foundation Graduate Research Fellowship

2009   Gertrude M. Cox Scholarship Honorable Mention

2009   Carl and Lily Pforzheimer Foundation Public Service Fellowship

2008,2009   Harvard Teaching Award

2008   Harvard University Phi Beta Kappa Junior 24 (highest GPA)

References

Andrew Gelman
Professor of Statistics
Columbia University

Joseph K Blitzstein
Professor of the Practice in Statistics
Harvard University

Jeffrey D Sachs
Professor of Sustainable Development
Columbia University

Alan M Zaslavsky
Professor of Health Care Policy (Statistics)
Harvard University

Mariel Finucane
Senior Statistician
Mathematica Policy Research

Deena Patel
Data Scientist
NYC Mayor's Office of Data Analytics

Solon Barocas
Principal Researcher and Professor of Information Science
Microsoft Research and Cornell University