Sofia Barragan

sofb@uw.edu · portfolio · (she/her)

EDUCATION

PhD, Biostatistics, University of Washington, Seattle — **September 2022 - Present**

BA, Statistics, Macalester College, Saint Paul, *Magna Cum Laude* — **August 2018 - May 2022**

> **Departmental Honor Thesis:** *Statistical Genetics for Pediatric Leukemia: Characterizing racial disparities in pediatric acute lymphoblastic leukemia*

WORK EXPERIENCE

Statistics & Machine Learning Researcher

> **Graduate Research Assistant,** University of Washington , Seattle, WA — **September 2022 - Present**
> - Researching gradient-boosting techniques to develop a fairness-aware, intersectional boosting algorithm and minimize racial bias of suicide risk prediction models with Kaiser Permanente Washington Health Research Institute's Biostatistics Division.
>
> **Research Consultant,** Rose International | Kaiser Permanente, Seattle, WA — **March 2023 - September 2023**
> - Developing proprietary simulation functions to model and test the consequences of imbalanced outcomes upon racial bias and outcome misclassification of tree-based suicide risk models with R.

Researcher 1, University of Minnesota, Minneapolis, MN — **May 2022 - September 2022**
- Identified unique treatment and survival disparities by race/ethnicity in young, adult, and elderly patients with Osteosarcoma from National Cancer Database (NCDB) data in R.
- Integrated differential gene expression, penalized regression, and survival modeling to study sex-based disparities in acute lymphoblastic leukemia and acute myeloid leukemia.
- Routine technical manuscript writing and preparation.

Research Intern, University of Minnesota, Minneapolis, MN — **June 2020 - May 2022**
- Independently created an internal database of massive pediatric genomic data scraped from multiple public repositories for researcher use with Python, R, and BASH.
- Performed demographic and differential gene expression analysis in 5 major pediatric cancers using TARGET & PeCan datasets via linear regression, empirical Bayes, and random forest classification in R.
- Conducted survival analyses of intersectional survival disparities in 17 pediatric cancers using the National Cancer Institute's SEER program in R. Confirmed & published evidence of survival disparities by race & sex in 11 different pediatric cancers.

Mann-Hill Research Fellow, Macalester College, St. Paul, MN — **June 2021 - August 2021**
- Developed bioinformatic and statistical pipelines with R, BASH, & Python to conduct first study of major survival disparities by race/ethnicity in pediatric leukemia using local and global genetic ancestry.

TEACHING EXPERIENCE

Teaching Assistant, University of Washington, Seattle, WA — **January 2024 - March 2024**
- Assisted instruction of Biostatistics for Health Sciences (BIOST 310) for Dr. William Brown
- Hosted weekly discussion sections and office hours, graded, and independently designed exam review material to develop students' statistical literacy and competency of biostatistical fundamentals.

Preceptor, Macalester College, St. Paul, MN — **August 2019 - May 2022**
- Assisted instruction of Applied Multivariable Calculus (MATH 135), Introduction to Statistical Modeling (STAT 155), and Statistical Machine Learning (STAT 253) for Drs. Kristin Heysse, Kelsey Grinde, Brianna Heggeseth, Alicia Johnson, and Leslie Myint.

- Hosted twice-weekly office hours, graded, and worked intimately with students to develop their fluency with advanced supervised and unsupervised ML techniques in R for semester-long projects.

PROJECTS IN PROGRESS

Intersectional Statistical Methods and Intimate Partner Violence Epidemiology: Characterizing IPV reporting delay behaviors in Los Angeles, California (2010-2023)
- Developed a novel extension of multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA) techniques to time-to-event settings to characterize intimate partner violence reporting delays across intersectionally-defined subgroups.

PROJECTS

Narcotic Arrests in Seattle: A Spatial Analysis
- Identified neighborhood-level risk factors of narcotic arrests and opioid use in Seattle, WA across the second wave of the U.S. Opioid Epidemic (2011-2013) using spatial regression in R & ArcGIS.

Neighborhood Deserts: Transportation Access & Housing Disparities in NYC
- Mapped transportation (in)access and its relationships to NYC rental prices, demographic change, and eviction rates using hierarchical Bayesian regression with R & ArcGIS.
- Harmonized census data and multiple from the Departments of Transportation, Health, Education, and Housing to construct a spatial dataset of relevant neighborhood and county-level features.

AWARDS, GRANTS, AND FELLOWSHIPS

- **Bill & Melinda Gates Foundation ARCS Endowed Fellowship**, ARCS Foundation, Seattle   **2022**
- **Excellence Award**, University of Washington, Department of Biostatistics   **2022**
- **Pat Wahl Award**, University of Washington, Department of Biostatistics   **2022**
- **Bressoud Prize**, Macalester College, Department of MSCS   **2022**
- **Research Supplements to Promote Diversity in Health-Related Research (PA-21-071)**, NCI   **2022**
  Parent Grant: *Admixture Analysis of Acute Lymphoblastic Leukemia in African American Children: the ADMIRAL Study (NIH R01 CA239701-01A1S1)*
- **Mann-Hill Research Fellowship**, Macalester College   **2021**
- **Best Video Poster**, Electronic Undergraduate Statistics Research Conference   **2020**
- **Catharine-Lealtad Scholarship**, Macalester College   **2018**

PUBLICATIONS

† = Indicates joint first-authorship

1. **Barragan, F.**, Mills, L., Raduski, A., Marcotte, E., Grinde, K., Spector, L., Williams, L., "Genetic Ancestry, differential gene expression, and survival in pediatric B-cell acute lymphoblastic leukemia", *Cancer Medicine* 00:1-12 (2022): 10.1002/cam4.5266
2. Moore, K.[†], **Barragan, F.**[†], Williams, L., "Survival disparities for childhood cancers exist when defined by race/ethnicity and sex." *Cancer Epidemiology* 81 (2022): 102262.

MANUSCRIPTS IN REVIEW

1. Williams, L., **Barragan, S.**, Lu, Z., Weigel, B., Spector, L., "Osteosarcoma survival is worse for males compared to females in most age groups: a National Cancer Database analysis (2004-2016)." (*International Journal of Cancer*)

RESEARCH PRESENTATIONS

1. **Barragan, F.**, Mills, L., Raduski, A., Marcotte, E., Spector, L., Grinde, K., Williams, L., "Genetic Ancestry, gene expression, and survival in children with B-cell acute lymphoblastic leukemia". Contributed poster at 2022 Pediatric Research, Education, & Scholarship Symposium, Virtual.
2. **Barragan, F.**, Mills, L., Raduski, A., Marcotte, E., Spector, L., Grinde, K., Williams, L., "Gene Expression Differences by Race and Genetic Ancestry in in B-Cell Acute Lymphoblastic Leukemia". Contributed poster at American Society for Human Genetics 2021 Annual Meeting, Virtual.
3. **Barragan, F.**, Mills, L., Raduski, A., Marcotte, E., Spector, L., Grinde, K., Williams, L., "Characterizing Racial Disparities in Pediatric Cancer: Ancestry, Gene Expression, and Survival Disparities in B-Cell Acute Lymphoblastic Leukemia". Contributed poster at Underrepresented Students in STEM Symposium 2021, Minneapolis, MN.
4. **Barragan, F.**, Mills, L., Raduski, A., Marcotte, E., Spector, L., Grinde, K., Williams, L., "Statistical Methods for Pediatric Leukemia: Gene Expression & Ancestry in B-Cell Acute Lymphoblastic Leukemia". Contributed poster at Macalester Summer Research Showcase 2021, Saint Paul, MN.
5. **Barragan, F.**, Moore, K., Williams, L., "Survival disparities for some childhood brain tumors exist when defined by race/ethnicity and sex". Contributed poster at Neuro-Oncology Symposium Conference 2021, Minneapolis, MN.
6. **Barragan, F.**, Mills, L., Spector, L., Williams, L., "Gene Expression & Clinical Differences in Pediatric Neuroblastoma by Sex". Video presentation at Electronic Undergraduate Statistics Research Conference 2020.

SERVICE & LEADERSHIP

CGH Steering Committee Member, Macalester College, St. Paul, MN   September 2020 - April 2022
- Senior student representative for 59 Community & Global Health (CGH) concentrators that advocated for student needs to the academic board.
- Organized and hosted academic panels and webinars and community outreach events to connect potential students with other CGH concentrators and CGH faculty.
- Presented at community outreach events, lectures, and new student orientations to help guide CGH concentrators through the internship process.

WMCN 91.7 FM Station Staff, Macalester College, St. Paul, MN   January 2019 - December 2021
- Organize event-programming and trained shows dedicated to international music, as a senior staff member of a FM radio station.
- Independently designed and implemented new training programs in September 2019 for beginning DJs.

MACCESS Coordinator, MPIRG, St. Paul, MN   October 2018 - November 2020
- Lead coordinator of an independent admissions and college counseling event for 50 first-generation, low-income, high school students of color from the Twin Cities' Public School District. 2019-2021 Sessions cancelled due to COVID-19.
- Collaborated with faculty to provide an introductory Computational Linear Algebra lecture on facial recognition software for 24 attendees interested in Computer Science and Mathematics.

SKILLS

Technical: Highly Proficient in R; Proficient in Git, SQL, and Excel; Familiar with Python, BASH, & ArcGIS

Spoken Languages: English (Native), Spanish (Bilingual Proficiency)

CERTIFICATIONS AND COURSES

- **Computational Pipeline for WGS Data**, SISG, University of Washington   **2021**
- **Bayesian Statistics for Genetics**, SISG, University of Washington   **2021**
- **Association Mapping: GWAS and Sequencing Data**, SISG, University of Washington   **2021**

PROFESSIONAL MEMBERSHIPS

- **American Statistical Association** (ASA)
- **International Society for Bayesian Analysis** (ISBA)
- **Society for Advancement of Chicanos/Hispanics and Native Americans in Science** (SACNAS)