

Sofia Barragan

barraganfia@gmail.com · [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

Research Assistant, University of Washington, Seattle, WA

September 2022 - Present

- Developing computational pipelines to evaluate clustering procedures for random forest estimation from longitudinal EHR data in R.
- Designing weighting schemes to account for heterogeneity of sample sizes across subgroups when predicting suicide risk with tree-based machine learning models in R.
- Conducted literature reviews of fair machine learning methods to develop novel risk prediction models in collaboration with Kaiser Permanente Washington Health Research Institute's Biostatistics Division.

Research Consultant, Rose International | Kaiser Permanente, Seattle, WA

March 2023 - September 2023

- Devised proprietary simulation functions to model and test the consequences of imbalanced outcomes upon racial bias and outcome misclassification of tree-based suicide risk models on EHR data with R.

Researcher 1, University of Minnesota, Minneapolis, MN

May 2022 - September 2022

- Led statistical and bioinformatic data analyses across multiple projects conducted in collaboration with medical practitioners and epidemiologists with R, Python, and BASH
- 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.
- Routine technical manuscript writing and report preparation.

Research Intern, University of Minnesota, Minneapolis, MN

June 2020 - May 2022

- Independently created and managed an internal database of massive pediatric genomic data scraped from multiple public repositories for researcher use with Python, R, and BASH.
- Identified genetic risk factors of sex-disparities across five major pediatric cancers by integrating linear mixed models, differential gene expression, and machine learning algorithms in R.
- Confirmed evidence of survival disparities by intersectional identities in 11 different pediatric cancers using survival methods trained on clinical data from the National Cancer Institute's SEER program in R

Mann-Hill Research Fellow, Macalester College, St. Paul, MN

June 2021 - August 2021

- Independently developed bioinformatic and statistical pipelines with R, BASH, & Python to conduct first study of underlying ancestral factors of racial/ethnic survival disparities in pediatric leukemia (B-ALL).
- Built an extended pipeline to conduct ancestral inference quality control and identify inferentially-weak chromosomal locations with R, RFMix, BASH and Python.

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 121 students' statistical literacy and competency in epidemiological fundamentals.

Preceptor, Macalester College, St. Paul, MN

August 2019 - May 2022

- Received endowed departmental awards for excellence and accessibility of teaching.
- Assisted instruction of Applied Multivariable Calculus (MATH 135), Introduction to Statistical Modeling (STAT 155), and Statistical Machine Learning (STAT 253) for 200+ students.
- Hosted twice-weekly office hours, graded, and worked intimately with students to develop their fluency in study design, statistical models, and machine learning in R.

PERSONAL PROJECTS

Neighborhood Deserts: Transportation Access & Housing Disparities in NYC

- Utilized hierarchical Bayesian techniques to evaluate relationships between transportation inaccess and rental prices, demographic change, and eviction rates in NYC with R & ArcGIS.
- Harmonized census data and multiple datasets from the Departments of Transportation, Health, Education, and Housing to construct a spatial dataset of relevant neighborhood and county-level features.

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.

Intersectional Statistical Methods and Intimate Partner Violence Epidemiology: Characterizing IPV reporting delay behaviors in Los Angeles, California (2010-2023)

- Independently developed novel extension of intersectional statistical methods using hierarchical survival techniques and approximate Bayesian inference via INLA in R.
- Harmonized census data with LAPD report records, precinct shapefiles, and reporting codes to conduct first analysis of intimate partner violence reporting delays across intersectionally-defined subgroups in R.

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 CA239701-01A1SI)*
- **Mann-Hill Research Fellowship**, Macalester College **2021**
- **Best Video Poster**, Electronic Undergraduate Statistics Research Conference **2022**
- **Catharine-Lealtad Scholarship**, Macalester College **2018**

PUBLICATIONS

† = Indicates joint first-authorship

1. Williams L., **Barragan, S.**, Lu Z., Weigel B., Spector L., “Sex differences in osteosarcoma survival across the age spectrum: A National Cancer Database analysis (2004-2016).” *Cancer Epidemiology*, 102565. 3 Apr. 2024, doi:10.1016/j.canep.2024.102565
2. **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
3. Moore, K.[†], **Barragan, F.**[†], Williams, L., “Survival disparities for childhood cancers exist when defined by race/ethnicity and sex.” *Cancer Epidemiology* 81 (2022): 102262.

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 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**

- Acted as the student representative for 59 Community & Global Health (CGH) concentrators, advocating for student needs to the academic board.
- Organized and facilitated academic panels, webinars, and community outreach events to foster connections between CGH concentrators, faculty, and external organizations.
- Presented at various events, including community outreach events, lectures, and new student orientations, guiding CGH concentrators through the internship process.

WMCN 91.7 FM Station Staff, Macalester College, St. Paul, MN **January 2019 - December 2021**

- Organized event-programming and led training sessions for shows dedicated to international music, as a supervising staff member of the FM radio station.
- Designed and implemented new training programs for beginning DJs in September 2019, streamlining station operations, equity training, and DJ skills.

MACCESS Coordinator, MPIRG, St. Paul, MN **October 2018 - November 2020**

- Lead coordinator of an independent admissions and college counseling event aimed at 50 first-generation, low-income, high school students of color from the Twin Cities’ Public School District. Sessions for 2019-2021 canceled due to COVID-19.
- Collaborated with faculty to provide an introductory Computational Linear Algebra lecture on facial recognition software, attracting 24 attendees interested in Computer Science and Mathematics.

SKILLS

Technical: Expert in R; Proficient in SQL, Excel, and Tableau; Familiar with Python, BASH, ArcGIS, and SAS

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)**