

University of California, Irvine
Distinguished Statistics Seminar

The Data Struggle of the Unseen

Bhramar Mukherjee
Distinguished Professor of Biostatistics
Professor of Epidemiology
Professor of Global Public Health
University of Michigan

4 p.m., Thursday, February 29, 2024
6011 DBH

Despite several proposed roadmaps to increase diversity in scientific research, most of the world's research data are collected on people of European ancestry. We rely on summary statistics from historically privileged populations and then devise clever statistical methods to transfer/transport them for cross-ancestry use. In this talk, I would first argue the obvious: for building fair algorithms we need fair training datasets. However, till we have reached the dream of equitable big data at a global scale, statisticians have an important role to play. In fact we have the perfect tools to study the "unobserved" through modeling of missing data, selection bias and alike. I will share examples from my personal journey as a statistician where doing good and timely statistical work with imperfect data quantified important disparity in health outcomes and led to policy impact. I will conclude the talk with a call to arms for statisticians to lead efforts for creating, curating, collecting data and pioneering new scientific studies, not just remain on the design and analytic fringes. As public health statisticians, our job is not just to predict, but to prevent. The talk is based on years of work with my students and colleagues at the Department of Biostatistics, University of Michigan and inspired by the transformative experience we shared as a statistical team working on the COVID-19 pandemic.