Methods for missing outcomes in healthcare and public health

Emma Pierson

**Assistant professor of computer science at Cornell Tech,
with a secondary joint appointment as an Assistant Professor of
Population Health Sciences at Weill Cornell Medical College**

February 08, 2024
1:30PM-2:50PM
MSOB X303

Abstract:

Missing data problems are pervasive in biomedicine, and profoundly entwined with health inequality: for example, health data is often less comprehensively collected for underserved populations. In this talk, we consider a frequently occurring missing data problem: *selective labels problems*, which occur when we observe outcomes only for a subset of the population determined by historical decision-making. For example, in a medical testing setting, we observe test outcomes only for the subset of patients who were historically tested. We show that selective labels problems arise when assessing racial discrimination in police traffic stops, develop scalable Bayesian methods to address them, show how these methods helped inform policing policy, and provide evidence of racial discrimination in traffic stops across the United States. We then show how similar methods can be applied to assess disparities in medical testing, provide evidence of disparities in COVID-19 testing, and show how to leverage plausible constraints in medical settings to improve model estimation.

Reading list:

1. https://arxiv.org/abs/1702.08536
2. https://www.nature.com/articles/s41562-020-0858-1