Estimating Racial Disparities when Race is Not Observed

Kosuke Imai

Harvard University

Keynote Talk
Promises and Limits of Inferring Protected-Class Data
for Disparate Impact Testing of AI Systems
September 8, 2023

Joint work with Cory McCartan, Jacob Goldin, and Daniel E. Ho

Motivation

- Importance of racial disparity estimation in many fields:
  public health, employment, voting, criminal justice, taxation, housing, lending, and internet technology
- But, often individual race is not available
  - law may prohibit collection of information about race (e.g., Equal Credit Opportunity Act)
  - agencies and companies may not wish to collect such information
- How should we estimate racial disparities when race is not observed?
  - standard methods use BISG (Bayesian Improved Surname Geocoding)
  - but, it has been shown that they are likely to yield biased estimates
- Can we improve the standard methods and eliminate their bias?

Executive Order 13985: Advancing Racial Equity and Support for Underserved Communities through the Federal Government

- Sec. 4. Identifying Methods to Assess Equity.
  The Director of the Office of Management and Budget (OMB) shall, in partnership with the heads of agencies, study methods for assessing whether agency policies and actions create or exacerbate barriers to full and equal participation by all eligible individuals. The study should aim to identify the best methods, consistent with applicable law, to assist agencies in assessing equity with respect to race, ethnicity, religion, income, geography, gender identity, sexual orientation, and disability.

- Sec. 5. Conducting an Equity Assessment in Federal Agencies.
  The head of each agency, or designee, shall, in consultation with the Director of OMB, select certain of the agency's programs and policies for a review that will assess whether underserved communities and their members face systemic barriers in accessing benefits and opportunities available pursuant to those policies and programs.

The Setup

- Data
  - Y_i: outcome of interest
  - R_i: (unobserved) race
  - S_i: surname
  - G_i: residence location
  - X_i: other Census variables (optional)
  - W_i: covariates of interest
- Census data
  - $\mathbb{P}(G_i = g, R_i = r, X_i = x)$
  - $\mathbb{P}(R_i = r, S_i = s)$ for frequently occurring surnames
- Regression estimands
  - short regression: $\mathbb{P}(Y_i = y \mid R_i = r)$
  - long regression: $\mathbb{P}(Y_i = y \mid R_i = r, W_i = w)$
- Racial disparity estimands
  - $\mathbb{P}(Y_i = y \mid R_i = r) - \mathbb{P}(Y_i = y \mid R_i = r')$ for $r \neq r'$
  - $\mathbb{P}(Y_i = y \mid R_i = r, W_i = w) - \mathbb{P}(Y_i = y \mid R_i = r', W_i = w)$

Standard Estimation Methods

1. Predict race via BISG (or its variant)
   - Assumption: $G_i \perp\!\!\!\perp S_i \mid R_i$
   - Bayes rule:

$$\begin{aligned} \hat{P}_{ir} &= \mathbb{P}(R_i = r \mid G_i = g, S_i = s) \\ &= \frac{\mathbb{P}(S_i = s \mid R_i = r)\,\mathbb{P}(G_i = g, R_i = r)}{\sum_{r'} \mathbb{P}(S_i = s \mid R_i = r')\,\mathbb{P}(G_i = g, R_i = r')} \end{aligned}$$

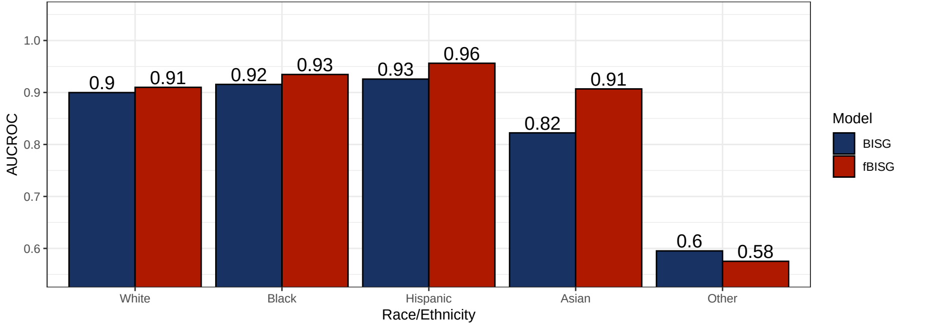
   - WRU software package (Imai and Kahna 2016)

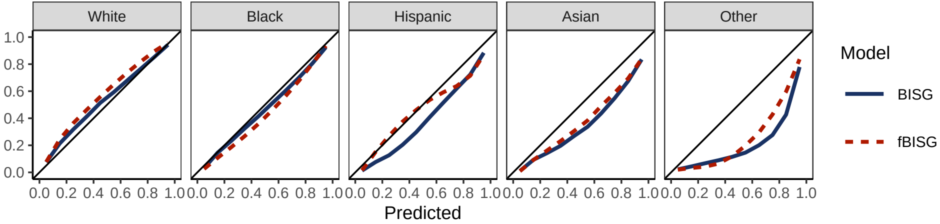
2. Estimate racial disparities $\mu_{Y|R}(y \mid r) = \mathbb{P}(Y_i = y \mid R_i = r)$
   - weighting:

$$\hat{\mu}^{\text{wtd}}_{Y|R}(y \mid r) = \frac{\sum_i \mathbf{1}\{Y_i = y\} \hat{P}_{ir}}{\sum_i \hat{P}_{ir}}$$

   - thresholding: use the racial group with the largest probability as imputed race

BISG Prediction Works Reasonably Well (Imai et al. 2022. *Sci. Adv.*)





Good Race Prediction Can Bias Racial Disparity Estimates

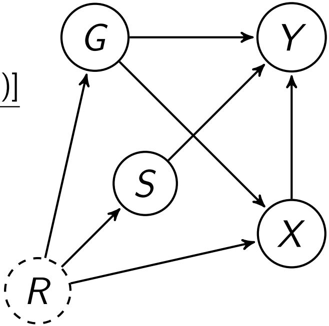
- Bias of the weighted estimator (Chen *et al.* 2019)

$$\hat{\mu}^{\text{wtd}}_{Y|R}(y \mid r) - \mathbb{P}(Y_i = y \mid R_i = r)$$
$$= -\frac{\mathbb{E}[\text{Cov}(\mathbf{1}\{Y_i = y\}, \mathbf{1}\{R_i = r\} \mid G_i, X_i, S_i)]}{\mathbb{P}(R_i = r)}$$

  - bias tends to be large for minority groups
  - racial disparity tends to be underestimated

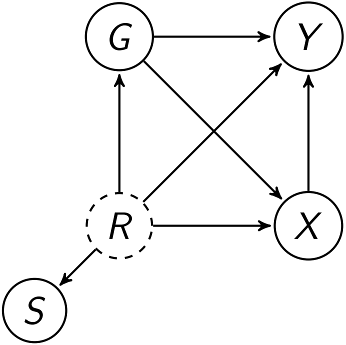
- Required assumption:

$$Y_i \perp\!\!\!\perp R_i \mid G_i, S_i, X_i$$



Problem: race affects many aspects of the society

BIRDiE (Bayesian Instrumental Regression for Disparity Estimation)



- Required assumption:

$$Y_i \perp\!\!\!\perp S_i \mid G_i, R_i, X_i$$

- Race can directly or indirectly affects the outcome
- Potential violations:
  - name-based discrimination within a racial group
  - racial categories being too coarse
- The assumption holds for:
  - anonymous applications
  - algorithmic decisions without use of names

- BIRDiE incorporates BISG and its variable, estimates racial disparity, and produces improved race probabilities

Incorporating Finer Racial Categories

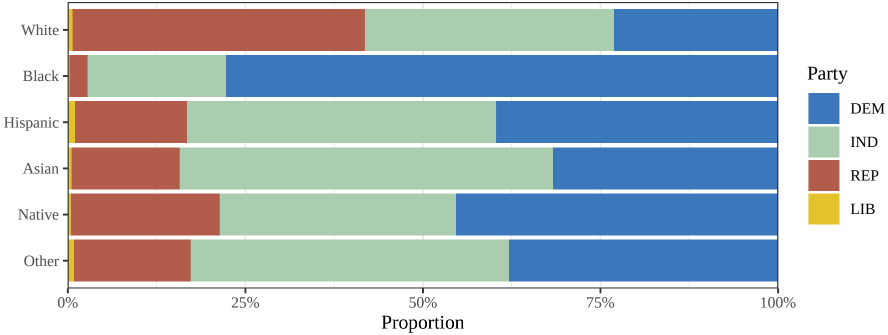
- Suppose we can have information about finer ethnic groups f
  - $f(\text{Imai}) = \text{Japanese}$, $f(\text{McCartan}) = \text{Irish}$, etc.
  - Assume instead
    $$Y_i \perp\!\!\!\perp S_i \mid f(S_i), R_i, G_i, X_i$$
  - Can include $f(S_i)$ as a covariate in BIRDiE

- 1930 Census provides 22 groups
  - Anglosphere and Black surname (third-or-more generation Whites and Blacks): Smith, Williams, Brown, ...
  - First wave European immigration (German, Nordic, and Irish): Burns, Olson, Wagner, ...
  - East Asian (Chinese, Japanese, Korean), South Asian (Indian, Southwest Asian), Southeast Asian and Pacific (Vietnamese, Filipino)
  - Non-Cuban Hispanic (Mexican, Latin American), Cuban

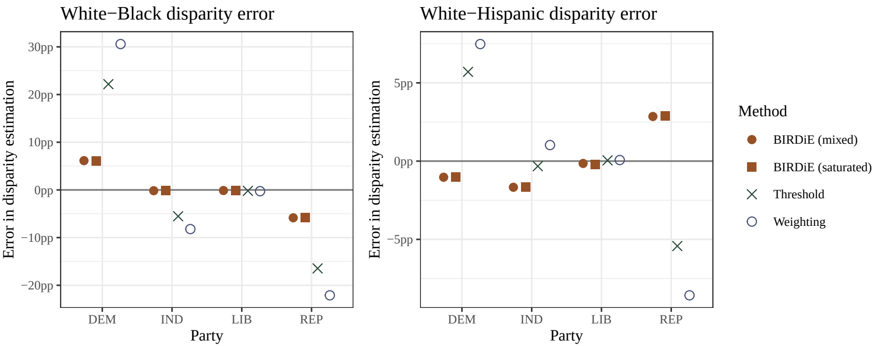
Empirical Validation

- 2022 North Carolina voter file: 5.8M voters with self-reported race
- Subset 1M voters $\rightsquigarrow$ negligible sampling uncertainty

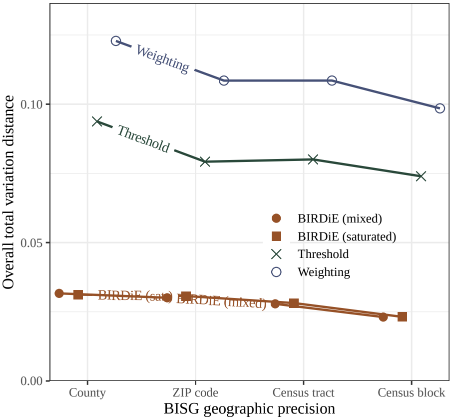
- Focus on party registration

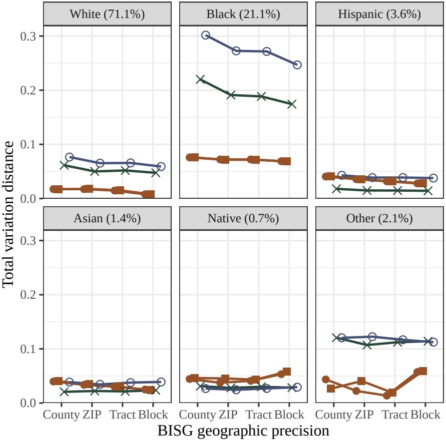


Estimates of Racial Disparity in Party Registration

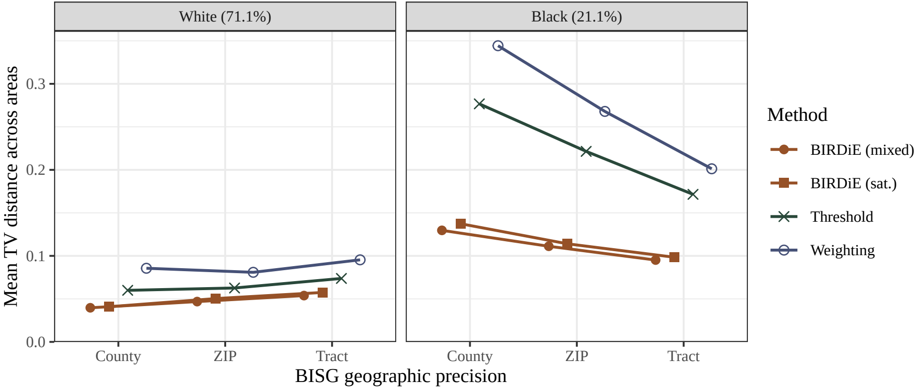


Total Variation Distance

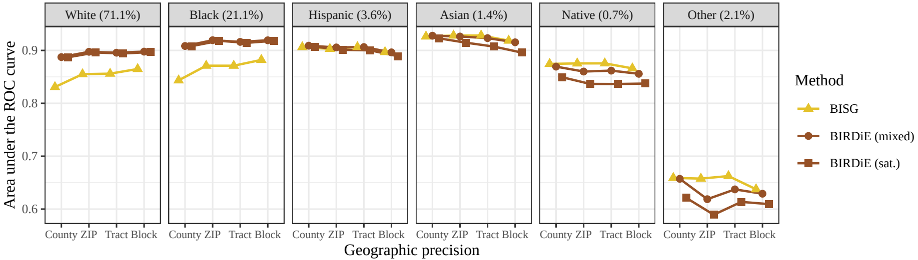




Small Area Estimation



Improved Race Probabilities



Concluding Remarks

- BIRDiE
  - new identification assumption
  - flexible modeling with scalable estimation
  - improved BISG race probabilities
  - sensitivity analysis with finer racial categories
- Future work
  - collaboration with IRS: racial disparity in tax system
  - additional empirical validations: understanding bias
  - generalization to record linkage and data combination
  - better use of auxiliary information in sensitivity analysis

The paper is available at
https://imai.fas.harvard.edu/research/birdie.html
The software is available at
https://corymccartan.com/birdie/