# Algorithmic Selection of Patients for Case Management: Alternative Proxies to Healthcare Costs

Moninder Singh

IBM Research, Yorktown Heights, NY, USA

## Motivation

- Racial and ethnic disparities in access to healthcare in the United States are well-known and documented[1].
- Healthcare expense is used as a proxy for health in algorithms that drive healthcare systems, and this exacerbates the existing bias.
- Prior work[2] showed significant racial bias in a widely used algorithm - black patients with highest predicted health risk have significantly more chronic illnesses than white patients with the same risk.
- Similar phenomenon observed[3] using the publicly available, and nationally representative, Medical Expenditure Panel Survey (MEPS) dataset[4].
- Question: Are there alternative proxies for health that are fairer?

## MEPS data

- Collection of surveys of families/individuals, medical providers, and employers across the US by the US department of HHS.
- Datasets (from 1996) contain two major components: household and insurance.
- We use the household component data which contains detailed information on demographics, health conditions, utilization, access to care, insurance, income, employment, charges, and payment sources.
- A single panel consists of unique individuals interviewed in five rounds over two calendar years.
- In each dataset, each sample is weighted so that the total weight in a panel sums to the entire US civilian, non-institutionalized population.

## Racial bias in predicted high expense individuals using MEPS data[3]

- Predicted top healthcare expense individuals had high racial disparity
- 10.7% whites selected as opposed to 6.8% blacks
- Blacks sicker than whites, especially top decile.

| Metric | Race | |
|---|---|---|
| | White | Black |
| Average number of priority conditions | 4.89 | 5.18 |
| Average perceived physical health status | 3.55 | 3.90 |
| Average perceived mental health status | 2.53 | 2.99 |

First year health indicator metrics for predicted high total expenditure individuals.

## Analysis

- Analyzed 2-yr longitudinal data comprising MEPS panel 20, 2015-16, as well as panel 19, 2014-15.
- Logistic regression model.
- For health conditions, considered # of priority conditions, and self-assessed health status, cognitive/activity/social/work limitations, smoking history, vision/hearing problems, etc.

## Predicting individuals with high expected utilization ( IP nights and/or >= 2 ER visits) using the raw MEPS data

| Metric | Race | |
|---|---|---|
| | White | Black |
| % of race predicted to be high-utilization | 9.7 | 11.2 |
| Average expected costs | $16.9K | $14.5K |

Racial differences between predicted high utilization individuals

| Metric | Predicted High Utilizers | | Overall Population | |
|---|---|---|---|---|
| | White | Black | White | Black |
| Average number of priority conditions | 4.76 | 4.31 | 1.97 | 1.8 |
| Average perceived physical health status | 3.57 | 3.59 | 2.08 | 2.23 |
| Average perceived mental health status | 2.67 | 2.79 | 1.82 | 1.85 |

First year health indicator metrics for predicted high utilization individuals.

| Diagnosis | Predicted High Utilizers | | Overall Population |
|---|---|---|---|
| | White | Black | |
| Angina | 0.14 | 0.08 | 0.02 |
| Arthritis | 0.74 | 0.70 | 0.28 |
| Asthma | 0.21 | 0.18 | 0.11 |
| Cancer | 0.33 | 0.18 | 0.13 |
| Chronic Bronchitis | 0.08 | 0.07 | 0.02 |
| High Cholesterol | 0.66 | 0.58 | 0.32 |
| Diabetes | 0.31 | 0.42 | 0.10 |
| Emphysema | 0.11 | 0.05 | 0.02 |
| High Blood Pressure | 0.79 | 0.85 | 0.35 |
| Joint Pain | 0.75 | 0.67 | 0.39 |
| Coronary Heart Disease | 0.20 | 0.14 | 0.05 |
| Myocardial Infarction | 0.20 | 0.14 | 0.04 |
| Stroke | 0.26 | 0.26 | 0.04 |
| Other Heart Disease | 0.36 | 0.23 | 0.12 |

First year average rate of prevalence of Priority Conditions (chronic diseases) for predicted high utilization individuals

- Built model to predict 2nd year utilization using demographics / health conditions from 1st year.
- Did not use expenditure, income, or employment status.
- Binary classification task: predict if person in top decile.
- Racial disparity was substantially lower than total healthcare expenditure case
- Patients were much sicker than overall population; racially too were more balanced.
- Blacks have higher rates of diabetes & BP
- Whites have higher cancer, cardiac diseases, etc.
- Data itself may reflect bias due to lack of access, since access to medical care is needed for diagnosis

## Discussion and Conclusion

- Future healthcare utilization needing inpatient nights or at least two ER visits results in fairer racial outcomes than total healthcare costs
- Pool data across multiple MEPS panels to get larger dataset.
- Address people who are chronically very sick but don't utilize IP/ER

1. Agency for Healthcare Research and Quality. National Healthcare Quality and Disparities Reports. https://www.ahrq.gov/research/findings/nhqrdr/index.html.
2. Z. Obermeyer and S. Mullainathan. Dissecting racial bias in an algorithm that guides health decisions for 70 million people. In Proceedings of FAT'19, pages 89–89. ACM, 2019.
3. M. Singh and K. N. Ramamurthy. Understanding racial bias in health using the Medical Expenditure Panel Survey data. NeurIPS Fair Machine Learning for Health Workshop, 2019.
4. Agency for Healthcare Research and Quality. Medical Expenditure Panel Survey (MEPS). https://www.ahrq.gov/data/meps.html.