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

Moninder Singh
IBM T. J. Watson Research Center
Yorktown Heights, NY 10598, USA
moninder@us.ibm.com

Abstract

Expected healthcare costs are commonly used in the United States as a proxy for health to select patients for case management. However, several recent studies have shown that AI algorithms for predicting costs exacerbate underlying disparities in the healthcare system and result in substantial bias against blacks, who have to be much sicker than whites to be chosen by these algorithms. We look at an alternative proxy based on emergency-room and inpatient utilization, and show that it results in more fair outcomes, reducing racial disparity while choosing patients truly in need for such services. We evaluate the effectiveness of this approach using the publicly available and nationally representative Medical Expenditure Panel Survey data collected annually by the U.S. Department of Health and Human Services.

Introduction

Healthcare costs, as a proxy for health conditions, are being increasingly used to identify patients for care management (Fleishman and Cohen 2009; Wherry, Burns, and Leininger 2014). By providing enhanced support to such patients, the aim is to improve patient outcome as well as reduce unnecessary healthcare expenses. However, several recent studies have demonstrated that algorithms used to predict future healthcare costs accentuate underlying biases in the healthcare system and result in substantial racial bias. Obermeyer *et al.* (Obermeyer and Mullainathan 2019; Obermeyer et al. 2019b; 2019a) analyzed a commercially-used algorithm used to select patients in the United States for care-management using risk-scores based on expected healthcare costs and found that it resulted in significant bias against blacks - blacks had substantially lower risk scores than whites who were equally sick, resulting in the selection of far fewer blacks for care management than would result from health considerations alone. Singh *et al.* (Singh and Ramamurthy 2019) showed that models for predicting healthcare costs learnt from the publicly available, and nationally representative, Medical Expenditure Panel Survey (MEPS) dataset (Agency for Healthcare Research and Quality 2018) resulted in similar racial bias against blacks.

To address this bias, alternative proxies for health need to be identified that are better related to the health of the patient than healthcare expenditure, and as such do not lead to racially-biased decisions for care management. Healthcare utilization (Ash and Ellis 2012; Obermeyer et al. 2019a), number of chronic health conditions (Fleishman and Cohen 2009), and self health assessments (Ash and Ellis 2012; Wherry, Burns, and Leininger 2014; Fleishman and Cohen 2009) have all been suggested as such alternative for future health needs. In this work, we study a certain type of health utilization - specifically, whether a patient will need any kind of inpatient visit, or at least two emergency-room visits (to eliminate a one-oft visit due to an acute health event) for identifying patients for care management using the MEPS data and show that the outcomes are much more desirable - the racial disparity is almost eliminated and the patients selected are substantially sicker than average as measured by a variety of metrics, and thus would benefit from additional care.

Description of MEPS data

Collected annually by the US Department of Health and Human Services, the Medical Expenditure Panel Survey (MEPS) dataset is a collection of surveys of families of individuals, medical providers, and employers across the United States. Each year, a new cohort is started and interviewed over five rounds over the next two calendar years. In any given dataset, each sample is weighted so that the total weight sums to the entire US civilian, non-institutionalized population. For this analysis, we use the household component piece of this dataset which contains detailed information on demographic characteristics, health conditions, healthcare utilization, access to care, health insurance coverage, income, employment, and payments.

Predicting high expected healthcare utilization

Since Singh *et al.*(Singh and Ramamurthy 2019) studied bias resulting from models to predict future healthcare costs learned from the MEPS data, we followed the same methodology, and used the same data, in our study to facilitate an easy comparison between the two approaches. We used 2-year longitudinal data for the cohort initiated in 2014 (panel

19) as well as 2015 (panel 20), and restricted the population to individuals who provided data during all five rounds, and indicated their ethnicity/race as non-Hispanic white, or non-Hispanic black.

We built a logistic regression model to predict second year healthcare utilization of individuals, based only on their demographics (age, race, gender) as well as health-related attributes in the first year. Health attributes consisted of diagnoses received for various chronic conditions, the count of these chronic conditions, physical and mental health assessments, and cognitive/hearing/vision limitations. Prior studies have shown that the total number of chronic conditions (Fleishman and Cohen 2009) as well as self health assessments and limitations (Ash and Ellis 2012; Wherry, Burns, and Leininger 2014; Fleishman and Cohen 2009) are strongly related to future healthcare expenditures and utilization.

The chronic conditions considered consisted of the set of *priority conditions*, a set of conditions (heart disease, stroke, high cholesterol, emphysema, chronic bronchitis, diabetes, cancer, arthritis, asthma, attention deficit/hyperactivity disorder (ADHD or ADD), and high blood pressure/hypertension.) that have been marked as such due to their frequency, expense, as well as importance to policy (Agency for Healthcare Research and Quality 2019). Four different heart related conditions (coronary heart disease, angina, myocardial infarction, and other unspecified heart disease) were combined into a single condition for the sake of our analysis. For each such feature, only records with diagnosis = yes/no were retained. All other records where any diagnostic values were not available (refused or don't know) were dropped. The chronic condition count feature was created by simply adding up the number of such diagnoses for a patient.

We also considered two different self assessed health status measures, one for perceived health status and one for perceived mental health status. Both measures were rated from 1 (excellent) to 5 (poor). In addition, numerical features representing various measures of health status, including Kessler Index of non-specific psychological distress, the Patient Health Questionnaire score, the Physical Component Summary Score, and the Mental Component Summary Score were also used. Additional features included the presence/absence of cognitive limitations, activity limitations, social limitations, work limitations, vision difficulty, hearing difficulty, pregnancy, and smoking history.

Since most of the features were measured multiple times during the panel (all 5 rounds or 2 rounds or annually), the feature representing the first time the value was collected was used.

Features such as prior year healthcare expenditure, income, and employment status, that are known to be strong predictors of future healthcare expenditure (Wherry, Burns, and Leininger 2014; Morid et al. 2017; Sushmita et al. 2015), and thus utilization, were, however, not used. This was primarily to ensure that predictions made by the model would be based on health-related attributes on which actions could be taken (e.g. manage a chronic disease), rather than features that were highly predictive but more reflective of access and affordability, and also non actionable (such as prior year expenditure).

Table 1: Racial differences between predicted high utilization individuals in panel 20, 2015-2016 cohort.

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

We modeled this problem as a binary classification task - the objective being to predict whether an individual would need at least one inpatient visit or at least two emergency room visits in the future (second year).

The training data for the model consisted of the 2014-2015 Panel 19 data, whereby the model was learned to predict patients who would incur utilization in 2015, based on 2014 demographic and health features. The learned model was then applied to the 2015-2016 Panel 20 data to predict the top 2016 utilization individuals in the cohort based on their 2015 features. After restricting the population to individuals who provided data during all five rounds, and indicated their ethnicity/race as non-Hispanic white, or non-Hispanic black, the panel 19 data consisted of 8136 records, representing appx. 214M people (179M or 84% White, 35M or 16% Black). The panel 20 data consisted of 8737 records, representing appx. 215M people (179.5M or 83.5% White, 35.5M or 16.5% Black). The data was further processed to one-hot encode categorical features. Feature standardization was done by centering (removing mean) and scaling (to unit variance). The sklearn Python library's implementation of Logistic Regression was used to construct the model. The sample weights were used to fit the resultant model.

Table 2: First year (2015) health indicator metrics for predicted high expense individuals in panel 20, 2015-2016 cohort (Singh and Ramamurthy 2019)

| | Race | |
Metric	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

The balanced accuracy of the model on the test set was 68.2% (using a threshold obtained from the training data using 5-fold cross validation) and the AUC was 0.74 . However, to enable a fair comparison with the results of Singh *et. al* (Singh and Ramamurthy 2019), the predicted model scores were sorted and only the individuals with scores in the top 10% were predicted to be the future high healthcare utilizers that would be flagged for care-management.

Table 3: First year (2015) health indicator metrics for predicted high utilization individuals in panel 20, 2015-2016 cohort.

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

As shown in Table 1, the top decile consisted of 11.2% blacks and 9.7% whites. The racial disparity in this group was much smaller, compared to what was observed in the group of patients with top decile of predicted healthcare costs (Singh and Ramamurthy 2019) - 6.8% blacks and 10.7% whites. The expected costs of the high-utilization patients is only slightly lower than those for patients chosen using total healthcare costs.

More importantly, in the study by Singh *et. al* (Singh and Ramamurthy 2019), using healthcare costs as a proxy for health to select patients for care management resulted in blacks having to be far sicker than whites to be included in that population (Table 2) on all metrics.

However, as shown in Table 3, while the individuals chosen using predicted utilization are substantially sicker than the average individual, they are quite similar in terms of sickness by race. Moreover, while whites are slightly sicker on average than blacks in terms of the number of priority conditions, blacks are sicker in terms of both perceived physical as well as mental health. Looking at individual chronic conditions (Table 4), the people selected for case-management display, on average, a much higher rate of occurrence of the priority conditions than the overall population. Blacks in the targeted population have higher rates than whites, on average, of incidence of diabetes and high blood pressure, whites are generally sicker than blacks in terms of the other diagnoses. Moreover, since chronic condition diagnoses in themselves probably reflect racial bias due to lack of access, a case could be made that blacks are slightly more sicker than indicated by the number of chronic conditions, since one has to have access to medical care to have a diagnosis. As such, the selection of slightly more blacks is expected, and consistent with the overall population as well (Table 4).

Discussion and Conclusion

Recent studies have shown that predictive algorithms that use expected medical expenditure as a proxy for health to identify patients for care management result in substantial racial bias. With such algorithms, blacks have to be substantially sicker than whites to be selected for enhanced support. Using publicly available, nationally representative, Medical Expenditure Panel Survey data, we show that an alternative proxy for health results in fairer algorithmic risk predictions.

Table 4: Average rate of prevalence of Priority Conditions (chronic diseases) in patients at time of prediction (2015) of second year (2016) utilization in panel 20, 2015-2016 cohort.

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

Specifically, future healthcare utilization requiring inpatient nights or at least two emergency room nights as a proxy for patient health in predictive models reduces the disparity between blacks and whites, while selecting similarly sick patients as chosen when future total healthcare costs are used.

Although this metric results in fairer predictions than when expenditure is used, the outcomes are now slightly biased towards whites. This could possibly be due to the fact that blacks are higher utilizers of inpatient/ER services as they have lower access to normal healthcare (Obermeyer et al. 2019a), although the differences vary by race when specific chronic conditions are examined. However, this may also be due to the fact that the chronic conditions diagnoses themselves are biased due to lack of medical access for blacks - whites get diagnosed with more chronic conditions since they have better access to medical care. This is supported by the fact that blacks appear sicker than whites when compared on the basis of self-assessments of physical/mental health. In that case, the aforementioned disparity in predictions would be expected.

Going forward, one planned enhancement is to pool together data from multiple MEPS panels to get larger, more robust data samples, and validate these findings on that data. Second, further detailed information on medical conditions, such as prescriptions, as well as event level details, such as diagnoses received during an ER visit or IP stay, are also available in the MEPS data in two components that have not been used: the medical conditions and event files. This detailed level data may help in better understanding, and modeling, of the racial disparities that exist across the entire healthcare space. Finally, it is likely that the proposed metric leaves out people who are relatively sick and would benefit from care management, but are not selected simply because they have not had any ER or IP night utilization. We would like to identify/explore this population and consider alternative approaches to identify them for care management as well.

References

Agency for Healthcare Research and Quality. 2018. Medical Expenditure Panel Survey (MEPS). https://www.ahrq.gov/data/meps.html.

Agency for Healthcare Research and Quality. 2019. Priority Conditions. https://meps.ahrq.gov/data_stats/MEPS_topics.jsp?topicid=41Z-1.

Ash, A. S., and Ellis, R. P. 2012. Risk-adjusted payment and performance assessment for primary care. *Med Care.* 50(8):643–653.

Fleishman, J. A., and Cohen, J. W. 2009. Using information on clinical conditions to predict high-cost patients. *Health Serv Res.* 45(2):532–553.

Morid, M. A.; Kawamoto, K.; Ault, T.; Dorius, J.; and Abdelrahman, S. 2017. Supervised learning methods for predicting healthcare costs: Systematic literature review and empirical evaluation. In *AMIA Annu Symp Proc.*, 1312–1321.

Obermeyer, Z., and Mullainathan, S. 2019. Dissecting racial bias in an algorithm that guides health decisions for 70 million people. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 89–89. ACM.

Obermeyer, Z.; Powers, B.; Vogeli, C.; and Mullainathan, S. 2019a. Algorithmic bias in health care: A path forward. *Health Affairs Blog*.

Obermeyer, Z.; Powers, B.; Vogeli, C.; and Mullainathan, S. 2019b. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366(6464):447–453.

Singh, M., and Ramamurthy, K. N. 2019. Understanding racial bias in health using the medical expenditure panel survey data. In *NeurIPS Fair Machine Learning for Health Workshop*.

Sushmita, S.; Newman, S.; Marquardt, J.; Ram, P.; Prasad, V.; Cock, M. D.; and Teredesai, A. 2015. Population cost prediction on public healthcare datasets. In *Proceedings of the 5th International Conference on Digital Health 2015*, 87–94.

Wherry, L. R.; Burns, M. E.; and Leininger, L. J. 2014. Using self-reported health measures to predict high-need cases among medicaid-eligible adults. *Health Serv Res.* 49(2):2147–2172.