David Strube - dstrube3@gatech.edu
CS6603 - AI, Ethics, & Society
Final Exam

Task 1:

- **Public Artifact**:
  - **Title** - "AI Bias - What Is It and How to Avoid It?"
  - **Released** - 2022-11-16
  - **Link** - https://levity.ai/blog/ai-bias-how-to-avoid
- **Application/Scenario/Domain of Misuse**: Healthcare resource allocation
- **Regulated Domain/Protected Class**: Healthcare / Race
- **Evidence: Research publication -** https://www.science.org/doi/full/10.1126/science.aax2342

Task 2:

In order to reduce healthcare costs, many hospitals worked with health insurers to develop and implement a high-risk care management program. The aim of this program was to provide additional resources to medically complex patients before their health deteriorated, given that earlier access to care for these high-risk patients would prevent burdensome and costly complications like emergency visits and hospitalizations. This program required precise targeting of patients, using commercial algorithms to predict which patients would have the most significant and complex healthcare needs.
Researchers found in October 2019 that an algorithm used to predict which patients would likely need extra medical care heavily favored white patients over black patients. More than 200 million people in US hospitals were subject to this algorithm. While race itself wasn't a variable used in this algorithm, another variable highly correlated to race - healthcare cost history - was. The rationale was that this cost summarizes how many healthcare needs a particular person has had in the past and thus would indicate how much healthcare this person would need in the future. For various reasons, black patients incurred lower health-care costs than white patients with the same conditions on average.

Task 3:

- The algorithm is given a dataframe with two elements:

(where subscript i is presumably an indicator of the patient's identity)

1. C_{it} (label): Total medical expenditures in year t
2. $X_{i,t-1}$ (features): Records of care from the patient's insurer over the year t-1:
   a. Demographics (e.g., age and sex, but specifically excluding race),
   b. Insurance type,
   c. ICD-9 diagnosis and procedure codes,
   d. Prescribed medications,
   e. Encounters, categorized by type of service (e.g., surgical, radiology, etc.),
   f. Billed amounts, categorized by type (e.g., outpatient specialists, dialysis, etc.)

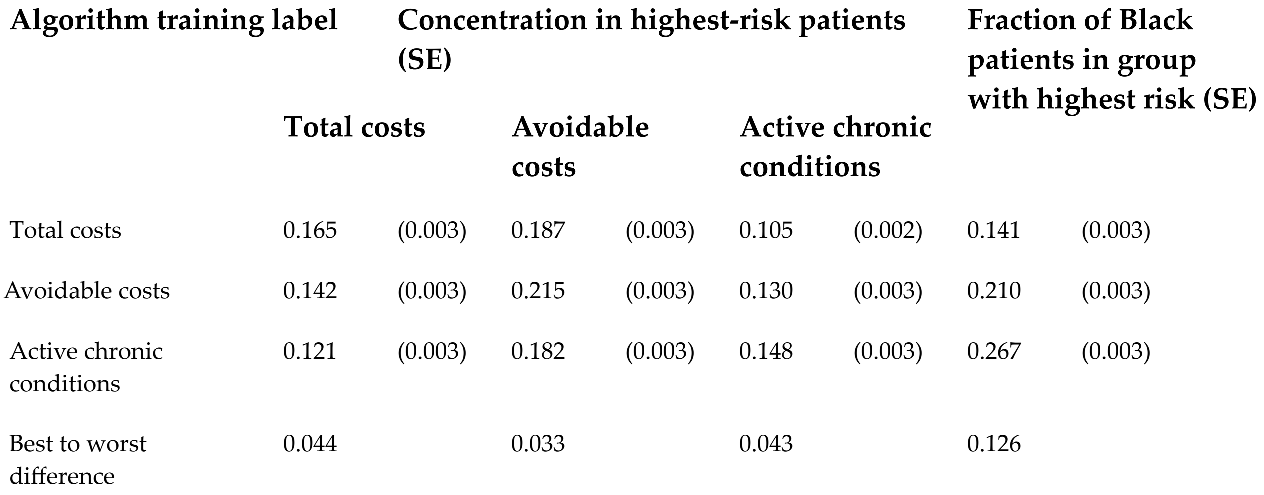
- These data for all eligible patients (those enrolled in risk-based insurance contracts) for a given year t-1 are fed into the commercial software, which delivers back a risk score for year t.
- The algorithm is run three times per year, during the enrollment period for the program. Patients whose scores exceed a critical threshold, approximately the 97th percentile of all available data, are auto-identified for enrollment in the program.

Table 1—Descriptive statistics on the study's sample, by race. BP, blood pressure; LDL, low-density lipoprotein.

	White	**Black**
n (patient-years)	88,080	11,929
n (patients)	43,539	6,079
Demographics		
Age	51.3	48.6
Female (%)	62	69
Care management program		
Algorithm score (percentile)	50	52
Race composition of program (%)	81.8	18.2
Care utilization		
Actual cost	$7,540	$8,442
Hospitalizations	0.09	0.13
Hospital days	0.50	0.78
Emergency visits	0.19	0.35
Outpatient visits	4.94	4.31
Mean biomarker values		
HbA1c (%)	5.9	6.4
Systolic BP (mmHg)	126.6	130.3
Diastolic BP (mmHg)	75.5	75.7
Creatinine (mg/dl)	0.89	0.98

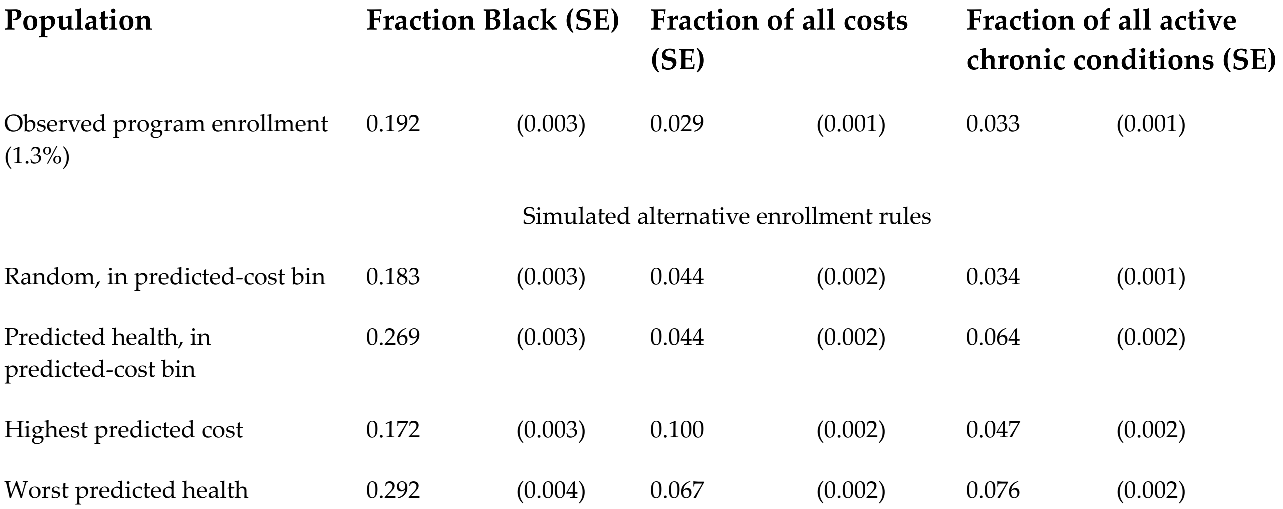
Hematocrit (%)	40.7	37.8
LDL (mg/dl)	103.4	103.0
Active chronic illnesses (comorbidities)		
Total number of active illnesses	1.20	1.90
Hypertension	0.29	0.44
Diabetes, uncomplicated	0.08	0.22
Arrhythmia	0.09	0.08
Hypothyroid	0.09	0.05
Obesity	0.07	0.18
Pulmonary disease	0.07	0.11
Cancer	0.07	0.06
Depression	0.06	0.08
Anemia	0.05	0.10
Arthritis	0.04	0.04
Renal failure	0.03	0.07
Electrolyte disorder	0.03	0.05
Heart failure	0.03	0.05
Psychosis	0.03	0.05
Valvular disease	0.03	0.02
Stroke	0.02	0.02
Peripheral vascular disease	0.02	0.02
Diabetes, complicated	0.02	0.07
Heart attack	0.01	0.02
Liver disease	0.01	0.02

Table 2—Performance of predictors trained on alternative labels.



Algorithm training label	Concentration in highest-risk patients (SE)						Fraction of Black patients in group with highest risk (SE)	
	Total costs		Avoidable costs		Active chronic conditions			
Total costs	0.165	(0.003)	0.187	(0.003)	0.105	(0.002)	0.141	(0.003)
Avoidable costs	0.142	(0.003)	0.215	(0.003)	0.130	(0.003)	0.210	(0.003)
Active chronic conditions	0.121	(0.003)	0.182	(0.003)	0.148	(0.003)	0.267	(0.003)
Best to worst difference	0.044		0.033		0.043		0.126	

Table 3—Doctors’ decisions versus algorithmic predictions.



Population	Fraction Black (SE)		Fraction of all costs (SE)		Fraction of all active chronic conditions (SE)	
Observed program enrollment (1.3%)	0.192	(0.003)	0.029	(0.001)	0.033	(0.001)
			Simulated alternative enrollment rules			
Random, in predicted-cost bin	0.183	(0.003)	0.044	(0.002)	0.034	(0.001)
Predicted health, in predicted-cost bin	0.269	(0.003)	0.044	(0.002)	0.064	(0.002)
Highest predicted cost	0.172	(0.003)	0.100	(0.002)	0.047	(0.002)
Worst predicted health	0.292	(0.004)	0.067	(0.002)	0.076	(0.002)

- **Privileged group:** white individuals
- **Unprivileged group:** black individuals
- **Any misleading graphs?**: In the research publication, the horizontal axis of Figure 1B didn’t start at 0; the same is

true for the vertical axes of Figure 2C, 2E, 3A, & 3B. Furthermore, the proportions of the vertical axes in Figures 3A & 3B were unequal.

- **Sources of Data Bias:** Black individuals with chronic illnesses spend less on healthcare than white individuals with the same conditions.
- **Sources of Sampling Bias:** The proportion of white individuals to black individuals in the study wasn't precisely the proportion of those affected by the program (e.g., 19.2% Blacks in the study versus 11.9% Blacks in the entire sample), but it was close enough that one could say that there was a negligible sampling bias.
- **Sampling Methods Used to Collect Data:** First, the researchers calculated the realized program enrollment rate (RPER) within each percentile of individuals within the predicted risk bins of an original algorithm and then randomly sampled patients in each bin for enrollment. Second, they sample those individuals with the highest predicted number of active chronic illnesses within a risk bin. Finally, they compared this to simply assigning those with the highest predicted costs, or the highest number of active chronic illnesses.
- **Correlations found in the data:** Healthcare expenses and healthcare needs are closely related, because sicker patients (on average) require and receive more care. However, a disparity appeared between needing healthcare and receiving healthcare, and the researchers found this disparity to be correlated with race. At a given level of health, Blacks incur less healthcare expenses than Whites.
- **Outcome measures:** Percentile of Algorithm Risk Score of Blacks compared to Whites along several chronic illnesses, including: Hypertension, Diabetes, Bad cholesterol, Renal failure, & Anemia severity

Task 4:

Issue related to one of the quantifiable metrics listed above (Task 3) that, if addressed, might help mitigate bias and/or unfairness:

1- Healthcare cost history summarizes how many healthcare needs a particular person has had in the past and thus would indicate how much healthcare this person would need in the future.

2- Black individuals with chronic illnesses spend less on healthcare than white individuals with the same conditions.

Method to help address the issue identified:

1- My first method is a little outside-the-box as it is a policy change and thus doesn't lend itself to pseudo-code. My idea is to revise one of the criteria used by the algorithm - healthcare cost history (as it has been shown to be correlated with race) - and replace it with another criterion like number of incidents of a given chronic illness. If the number is greater than one, then the patient qualifies for the healthcare resources. The anticipated change of outcome would be more healthcare resource allocation for all qualifying participants.

2- My second method is more generic and less bold, but possibly as much a step in the right direction as my first method. Given that black individuals with chronic illnesses spend less on healthcare than white individuals with the same conditions *on average*, the solution to this problem may lie in other statistical measures, namely mode and median. The mean, median, and mode of incomes of families in the statistics example in Module 2 were all wildly different. Likewise, finding the median and mode values of healthcare cost history of individuals, both without regard to race and with regard to race, could shed light on how better to allocate healthcare resources. With this information, the anticipated change of outcome would be a better informed and more fair allocation of healthcare resources.

3- My third method would be to use what was described in Module 4 as demographic parity. If each demographic was set to have a certain percentage of approval, then that would ensure fairness among demographics and no racial discrimination, inadvertent or otherwise. This could be incorporated into the original algorithm fairly easily so as to not have to completely scrap the whole program and start from scratch. For example:

PSEUDOCODE:

- For each demographic, the algorithm is given a dataframe with two elements:

1. C_{it} (label): Total medical expenditures in year t
2. $X_{i,t-1}$ (features): Records of care from the patient’s insurer over the year t-1
    a. (see Task 3 for details of the features)

- Next, the algorithm finds the midpoint (M) of healthcare resource allocation between all demographics from the previous runs (i.e., prior to October 2019). M is used as the target rate of healthcare resource allocation in each demographic cycle.

The anticipated change of outcome would be an allocation of healthcare resources that is not racially biased.