


Lab 8: Bias in Algorithms

Gregory Bruich, Ph.D

HARVARD UNIVERSITY

OPPORTUNITY INSIGHTS

Lab 8: Bias in Algorithms

- In the last lab, we developed two methods for prediction that address overfit:
  1. **Cross validation** as a general-purpose tool
  2. **Random forests** as a specific improvement upon decision trees
- In today's lab, we will dive deeper into the issue of **bias in algorithms**, using random forests to predict patient costs and patient health
- We will show that getting the right target—**label choice**—is central
  - A major source of bias in health algorithms
  - And other fields e.g., crime vs. arrests

Impact Pro®

Better understand individual and population health needs

Impact Pro leverages the power of Optum Symmetry®, an industry leading foundational analytics suite for assessing cost, risk and quality, to help you identify, profile and stratify your members.

With Impact Pro, you can determine which individuals are in need of specialized intervention programs and which intervention programs are likely to have an impact on the quality of individuals' health.

The advantage of Impact Pro

- **Clinician-driven identification and stratification,** helping you flag individuals early, before their health becomes catastrophic
- **Easy-to-use strategic dashboards and opportunity reports,** with data displayed to highlight key opportunities to enhance care or reduce unnecessary costs

To learn more about how Impact Pro can help you target health care services to those individuals who will benefit the most, please contact us at **1-800-765-6807** or **empower@optum.com.**

Examples of Label Choice Bias in Health Care

Algorithm	Ideal Target
Care Management Prioritization: Identifying patients for additional services	Health needs, benefit from high-risk care management programs
Emergency Severity Index (ESI): emergency triage	Medical condition needing immediate attention
6-Clicks Mobility Score: Decisions about discharge destination	Inability to care for self and live independently at home without help
"No-show" prediction: Clinic scheduling	Voluntary no-show to appointment
Predicting Disease Onset: Targeting preventative care	New disease onset (e.g., heart failure, kidney failure)
Kellgren-Lawrence Grade: Osteoarthritis on knee x-rays	Severity of knee osteoarthritis

Source: Mullainathan and Obermeyer (2021)

Examples of Label Choice Bias in Health Care

Algorithm	Ideal Target	Actual Target
Care Management Prioritization: Identifying patients for additional services	Health needs, benefit from high-risk care management programs	Total costs of care
Emergency Severity Index (ESI): emergency triage	Medical condition needing immediate attention	Nurse-rated acuity, "resources patient is expected to consume"
6-Clicks Mobility Score: Decisions about discharge destination	Inability to care for self and live independently at home without help	Physical measures of mobility and daily activities
"No-show" prediction: Clinic scheduling	Voluntary no-show to appointment	Any no-show to prior appointment
Predicting Disease Onset: Targeting preventative care	New disease onset (e.g., heart failure, kidney failure)	Provider–insurer transaction with ICD code for disease
Kellgren-Lawrence Grade: Osteoarthritis on knee x-rays	Severity of knee osteoarthritis	Severity of osteoarthritis seen by radiologist on knee x-rays

Source: Mullainathan and Obermeyer (2021)

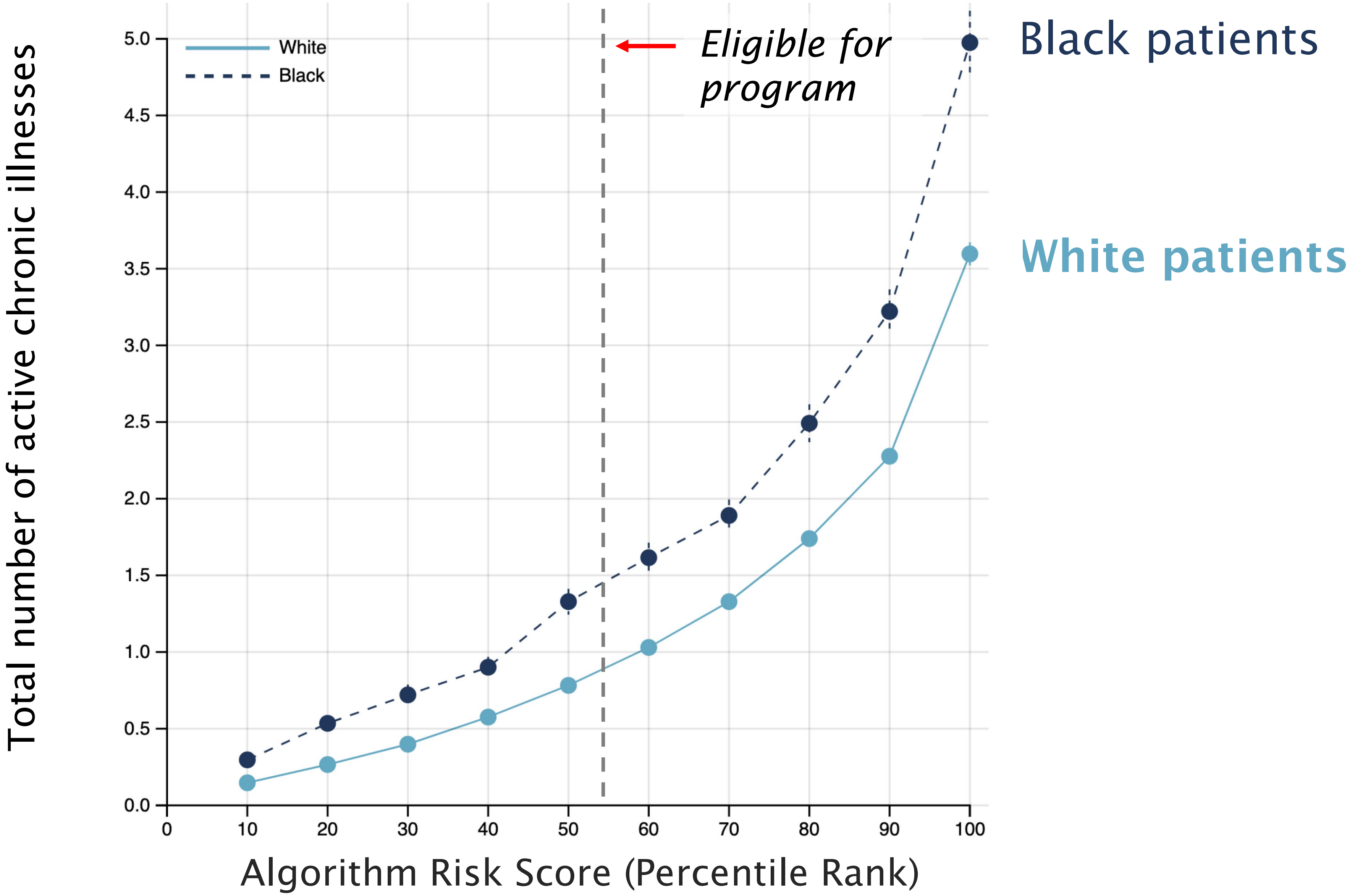
Examples of Label Choice Bias in Health Care

Algorithm	Ideal Target	Actual Target	Risk of Bias
Care Management Prioritization: Identifying patients for additional services	Health needs, benefit from high-risk care management programs	Total costs of care	**High.** Less money is spent on Black patients who have the same level of need
Emergency Severity Index (ESI): emergency triage	Medical condition needing immediate attention	Nurse-rated acuity, "resources patient is expected to consume"	**High.** Resource consumption varies by race and insurance for any given acuity
6-Clicks Mobility Score: Decisions about discharge destination	Inability to care for self and live independently at home without help	Physical measures of mobility and daily activities	**High.** Similar physical mobility scores have larger impact on those lacking income
"No-show" prediction: Clinic scheduling	Voluntary no-show to appointment	Any no-show to prior appointment	**High.** No shows relate to access: barriers are unequally distributed
Predicting Disease Onset: Targeting preventative care	New disease onset (e.g., heart failure, kidney failure)	Provider–insurer transaction with ICD code for disease	**High.** Probability of being coded varies by physician quality, hospital billing, insurance, etc.
Kellgren-Lawrence Grade: Osteoarthritis on knee x-rays	Severity of knee osteoarthritis	Severity of osteoarthritis seen by radiologist on knee x-rays	**High.** Radiologists miss causes of knee pain affecting underserved groups

Source: Mullainathan and Obermeyer (2021)

Health vs. Risk Score from Commercial Algorithm Currently Used in Healthcare

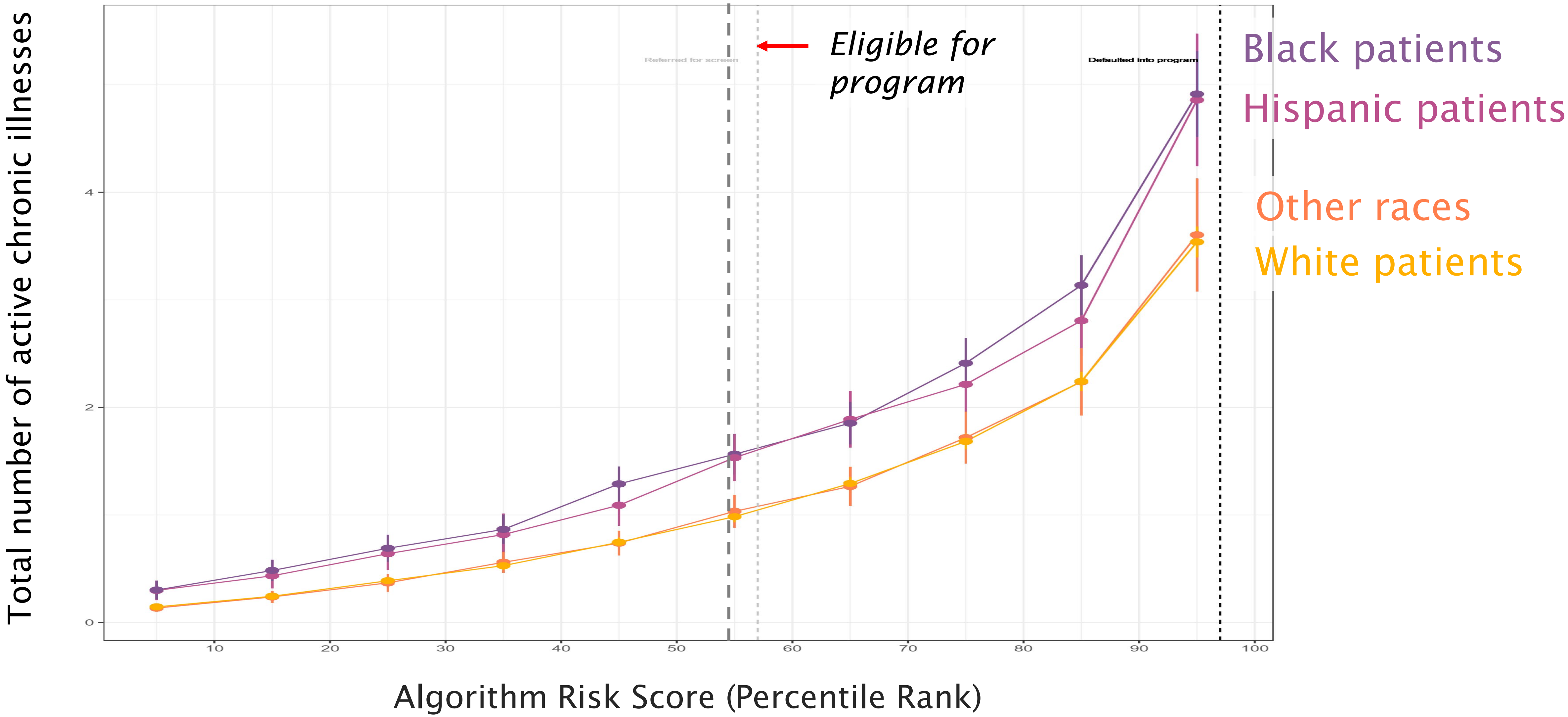
Patients with the Same Risk Score should have Same Health Needs



Source: Obermeyer, Powers, Vogeli, and Mullainathan (2019)

Health vs. Risk Score from Commercial Algorithm Currently Used in Healthcare

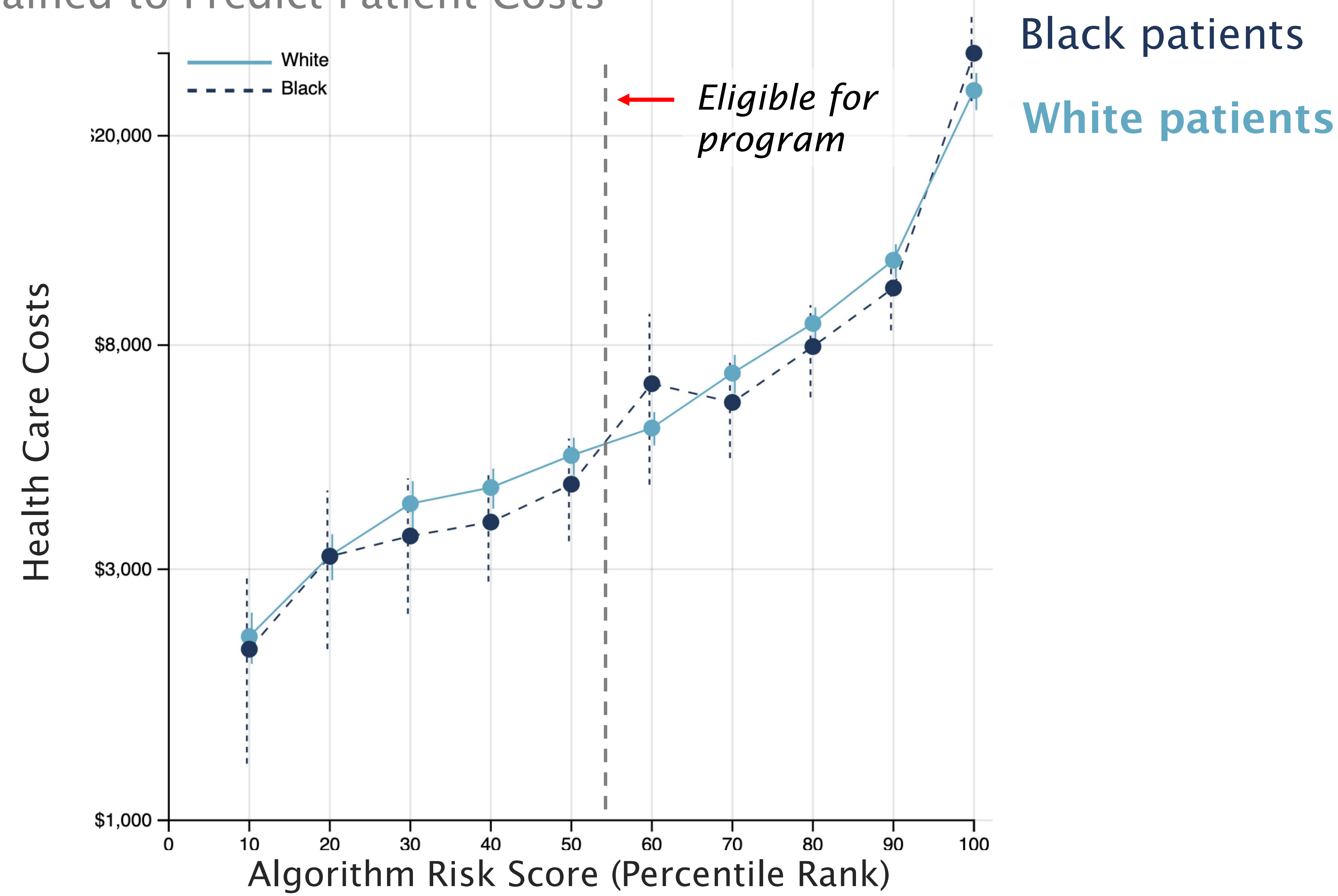
Patients with the Same Risk Score should have Same Health Needs



Source: Obermeyer, Powers, Vogeli, and Mullainathan (2019)

Comparing the Actual Target for Black versus White Patients

Algorithm was Trained to Predict Patient Costs



Source: Obermeyer, Powers, Vogeli, and Mullainathan (2019)

What is the source of bias in algorithms? Labels versus predictors

- Principle: patients with the same risk score should have same health needs
- In this lab, we will use binned scatter plots to compare the bias of four models:
  - **Random forest model 1:** trained to predict **patient costs** and does **not** use patient race as a predictor
  - **Random forest model 2:** trained to predict **patient costs** and **uses patient race** as a predictor variable
  - **Random forest model 3:** trained to predict **patient health** and does **not** use patient race as a predictor
  - **Random forest model 4:** trained to predict **patient health** and **uses patient race** as a predictor variable