CPSC 464: Fairness in Predicting the Risk of Stillbirth and Preterm Pregnancy

Bradley Yam, Cove Geary, Ivy Fan

Final Presentation

Background / Introduction

Introduction

- Machine learning is increasingly being used to predict health outcomes and diagnose patients at a theoretical level and in clinical practice. (Akbulut et al)
- Medical data can reflect disparities in care across subgroups which encode hidden variables such as inequities in housing, education, employment and criminal justice that affect healthcare access, utilization and quality, which subsequently create disparities in predictions. (Bailey et al) Further compounded by underrepresentation of minority groups in clinical trials and warped financial incentives of healthcare (Pfohl et al). Self-fulfilling predictions.
- Typical formulations of algorithmic fairness such as group fairness criteria are unaware of the complex sociotechnical context and thus do not capture the ideal state. Model performance does not equate accrual of benefit (Pfohl et al).
- Our paper focuses on assessing disparities in risk prediction for fetal-health risk and birth outcomes, and assessing the implications of such disparities in clinical practice.

Context

- Stillbirth occurs 1 out of 160 births, roughly 24,000 stillbirths per year
- Preterm births occurs in 1 out of 8 births, have higher chances of death and disability, accounting for 17% of infant deaths.
- According to the CDC: risk of stillbirth and preterm birth is increased if you are **of the black race**, 35 or older, low socioeconomic status, smoked cigarettes during pregnancy, have multiple medical conditions, or have a previous pregnancy loss. In 2019, rates of preterm birth were 50% higher amongst African American women than White women.
- **There is no surefire test for stillbirth, explanations are usually post-hoc.**
- Interventions include hormone therapy, cervical cleavage and mental health counselling for when the baby cannot be saved.

Related Works

Pfhol et al, 2021 - An empirical characterization of fair machine learning for clinical risk prediction

Bjarnadottir and Anderson, 2020 - Machine Learning in Healthcare: Fairness, Issues, and Challenges (are we doing as well as we can for everyone that we can?)

Davidson and Boland, 2021 - Towards deep phenotyping pregnancy: a systematic review on artificial intelligence and machine learning methods to improve pregnancy outcomes

Akbulut et al, 2018 - Fetal health status prediction based on maternal clinical history using machine learning techniques

Replicating Research → Koivu and Sairanen, 2020 - Predicting risk of stillbirth and preterm pregnancies with machine learning

Incorporating Methods → Seyyed-Kalantari et al., 2003 - CheXclusion: Fairness gaps in deep chest X-ray classifiers

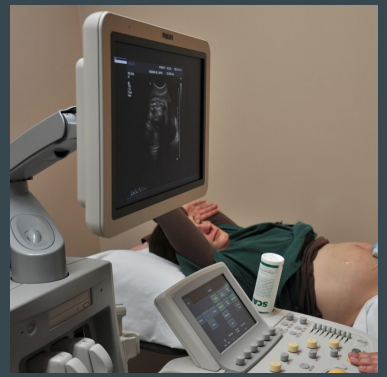
Goal: assess fairness in models trained to predict birth outcomes based on pregnancy data

Implications:

- Disparity in TPR can lead to undertreatment of stillbirth and preterm cases
- Disparity in FPR can lead to costly and invasive treatments, including hormone treatment and cervical cleavage. This is especially dangerous when we don't have other means of assessing risk (there is no surefire test)
- Disparity in TPR and FPR can also lead to mis-weighted cost outcomes in paying more out of pocket for tests and increased insurance premiums.

Roleplay Healthcare Professional

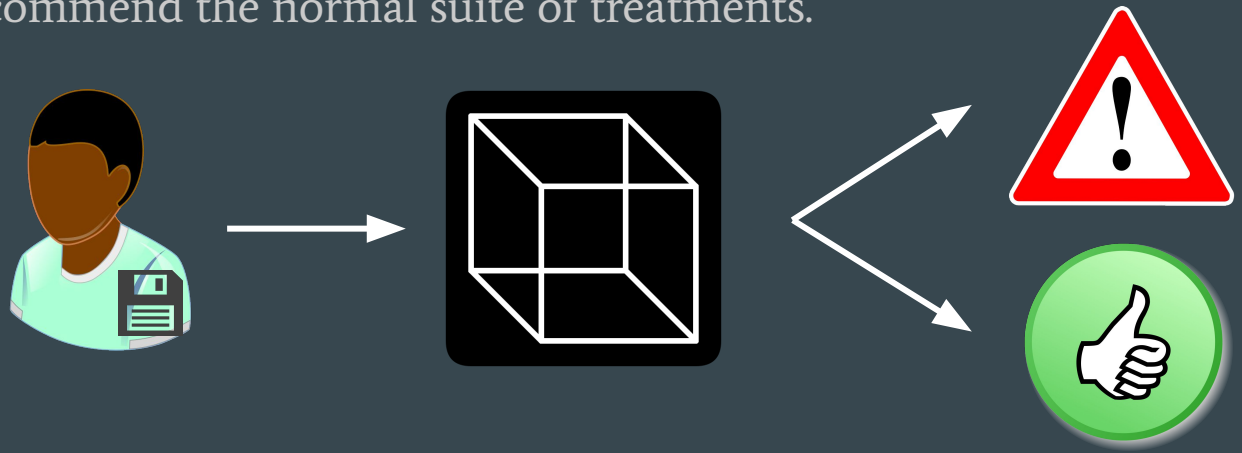
- You are a practicing OBGYN and you want to know what tests and treatments to order for your patients. There is blood work, cervical exams, ultrasounds etc.
- How do you know which medical services to prescribe?



Check Box to Prescribe

Solution: Training a Machine Learning Model

- You train a ML model to determine if your patient is at a higher risk of stillbirth or preterm pregnancy based on electronic health record (EHR) data.
- If your patient is “at risk”, you prescribe the maximum suite of tests and treatments, and you might opt for more aggressive medical interventions such as C-sections, cervical cleavage or hormone therapy. Otherwise, you recommend the normal suite of treatments.

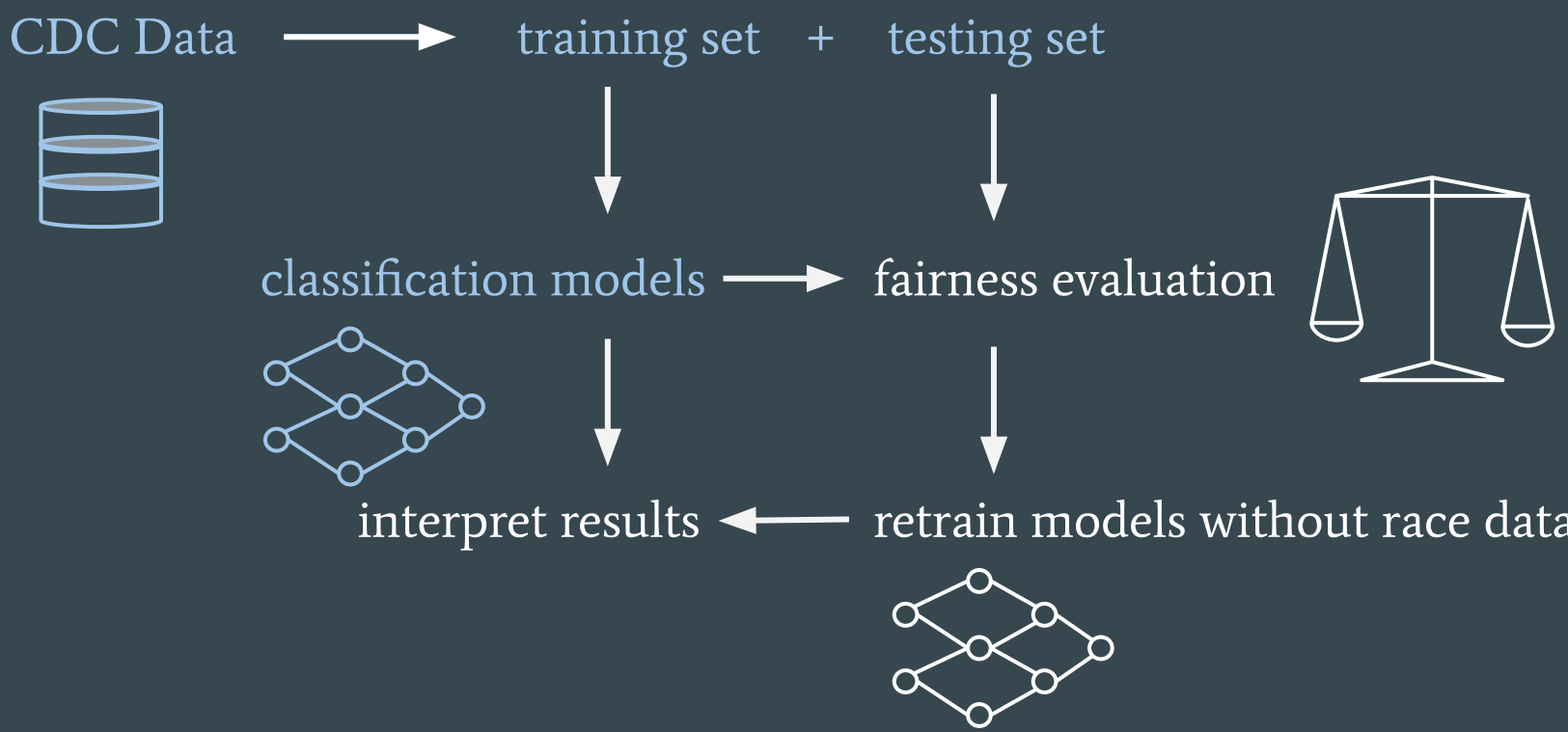


Follow-Up: Bias

- You observe that the model predicts that the number of black persons at risk are 7x the number of white persons.
- Most of the black persons that are identified proceed to have normal pregnancies.
- You wonder if the treatments might be causing more stress for the patients.
- You also wonder about how insurance companies are calculating the premiums for black persons given extra costs.
- You decide to investigate if this model is allocatively fair -- is it distributing the limited resources that your hospital has in the optimal way?

7:1

Approach: replicate, evaluate, explicate/mitigate



Specific Methods

CDC Data → Classification Models → Fairness Evaluation → Retrain Models

- 2013-2016 fetal birth and death data sets, following Koivu & Sairanen
- 26 feature variables, demographic, previous birth history

- 12 models total: 3 classification tasks x 4 algorithms
- Logistic regression, gradient boosted trees, NN w/ leaky ReLu, NN w/ scaled exponential linear unit activation fxn

- Evaluate each model for each outcome by comparing disparities in TPR, FPR, accuracy across protected attribute (race)

- Attempt to mitigate bias in models by dropping the race variable from data set and retraining > fairness through unawareness

Approach

Replication

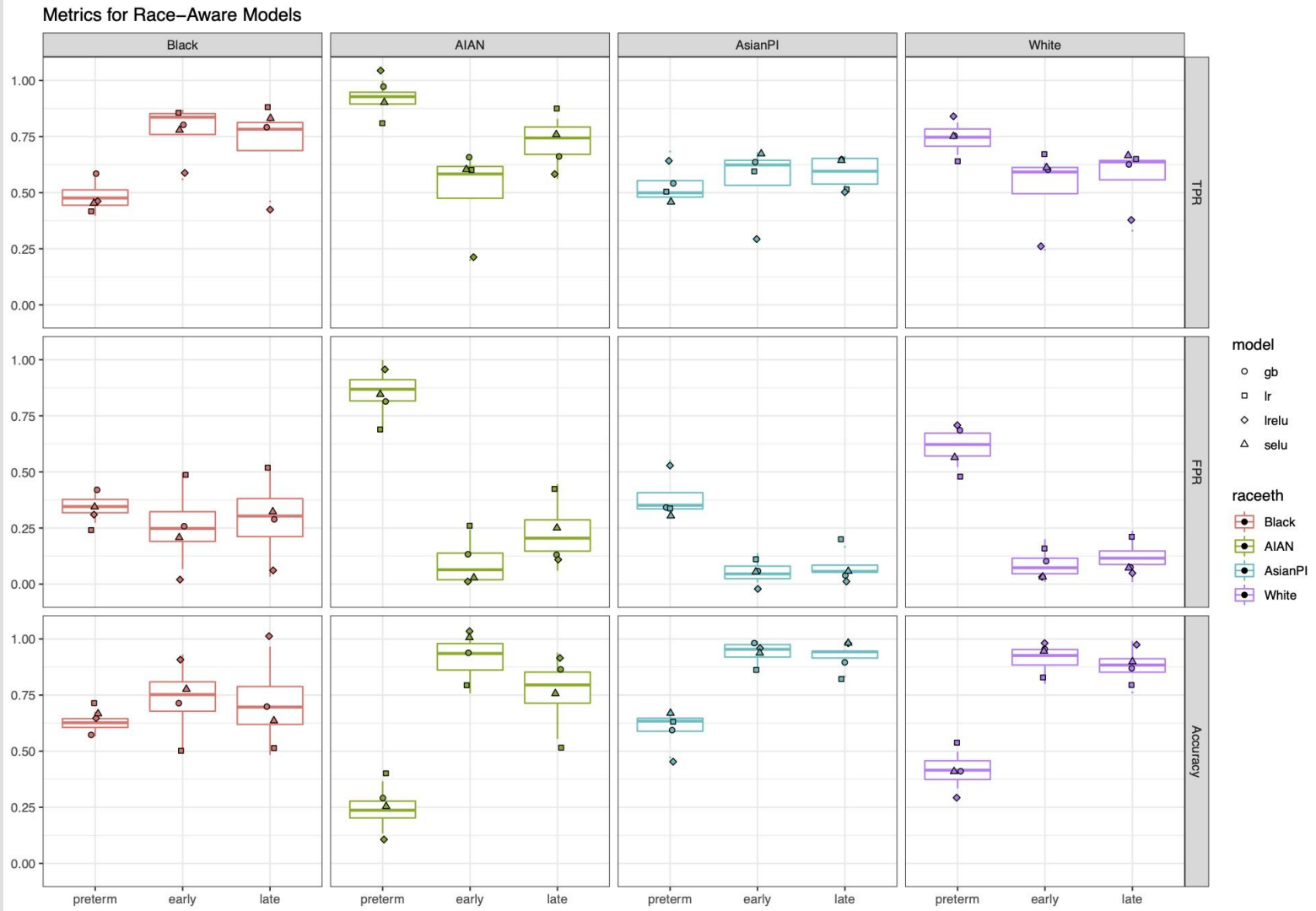
CDC Data → Classification Models → Fairness Evaluation

- Same data as Koivu & Sairanen:
- 2013-2016 infant birth and fetal death data based on birth and death certificates; nearly 16 million pregnancies
- Select 26 feature variables including demographic data, pregnancy-related medical history, sexually transmitted infections

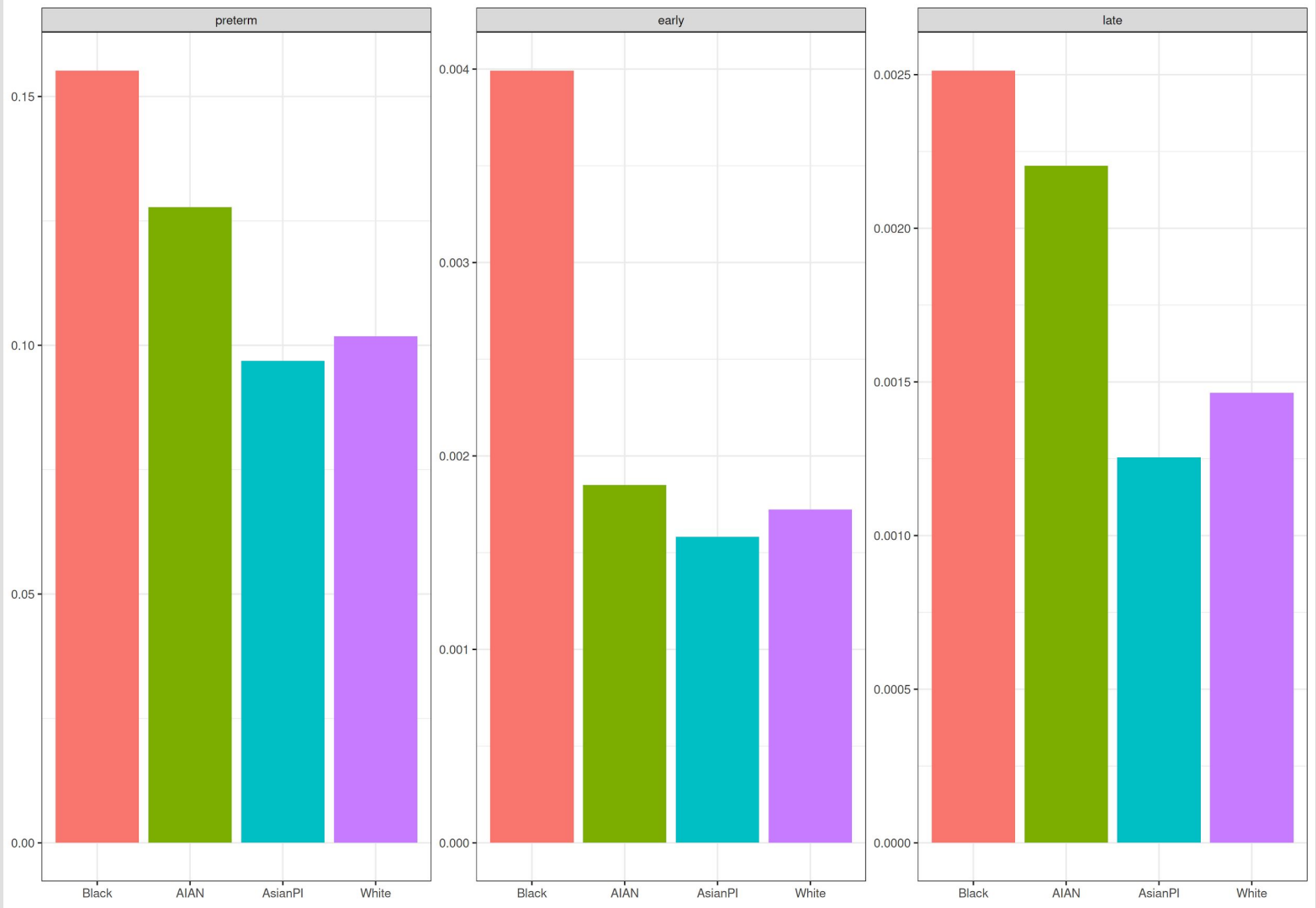
- Three binary classification tasks: early stillbirth, late stillbirth, preterm birth
- Four models: logistic regression, gradient boosted trees, two deep neural networks

- Evaluate each model for each outcome by comparing disparities in TPR, FPR, accuracy
- Also calculate AUC and TPR@10%FPR

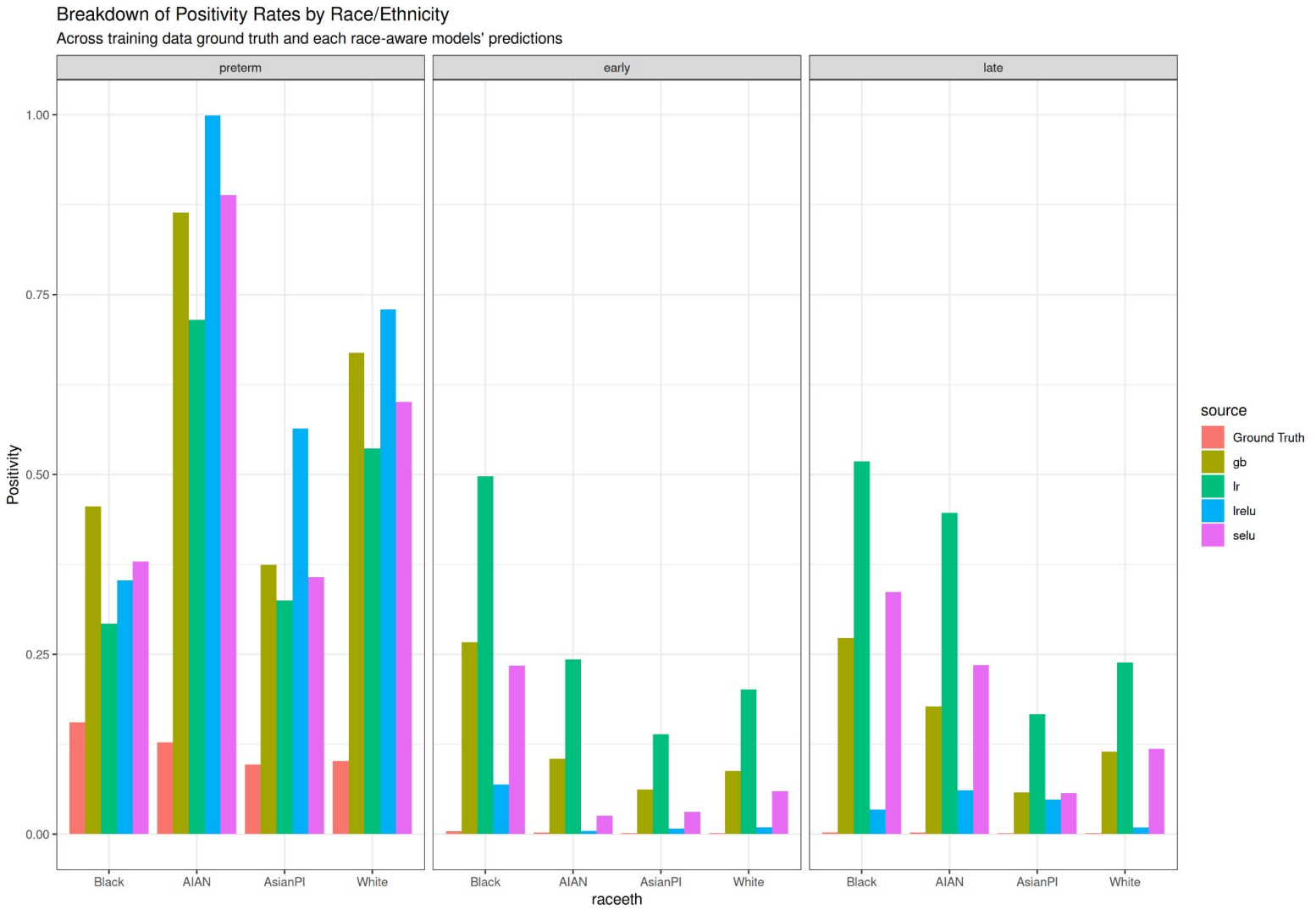
Results: Model Metrics (Aware)



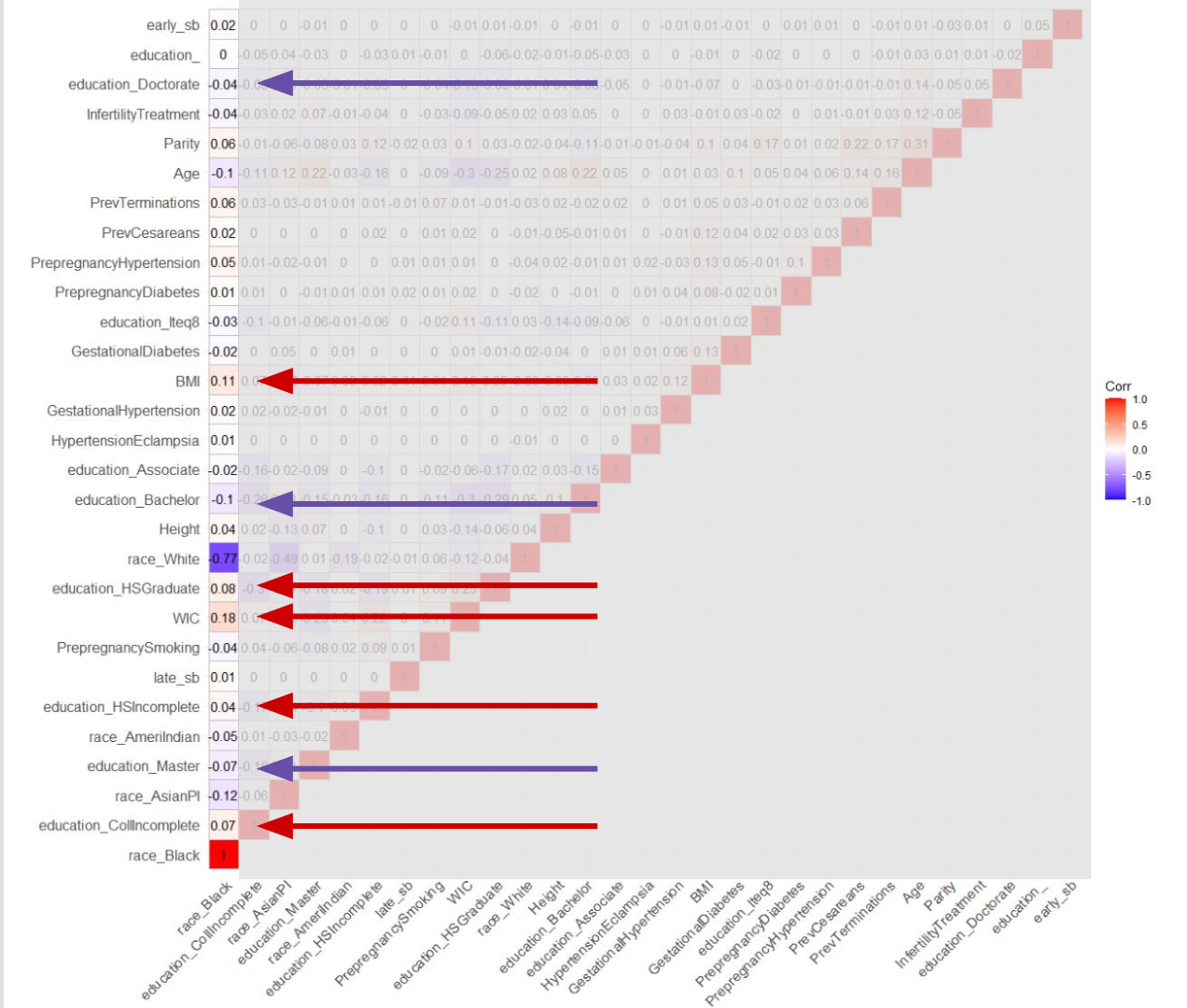
Risk Burden: Data Breakdown



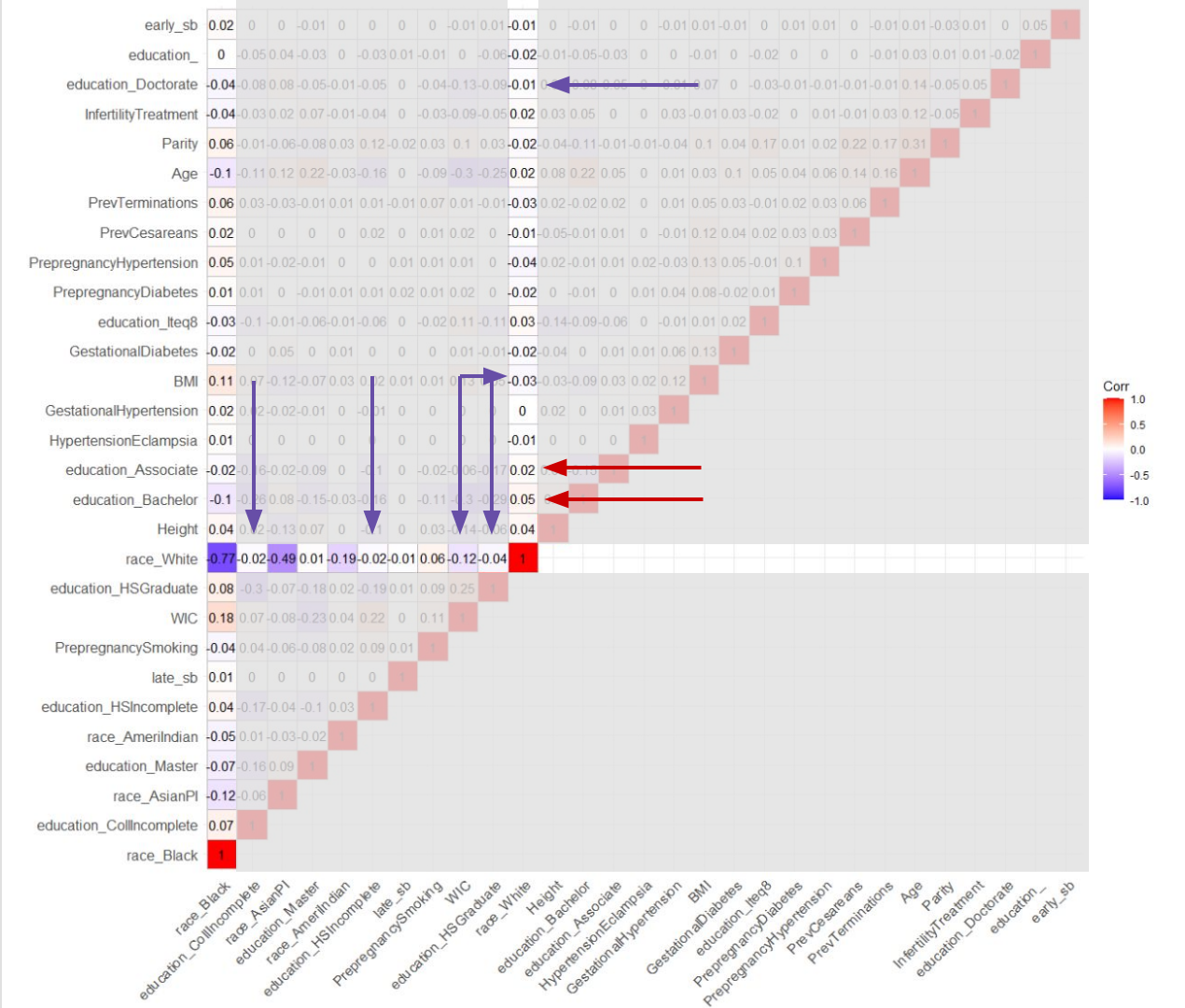
Risk Burden: Data+Models



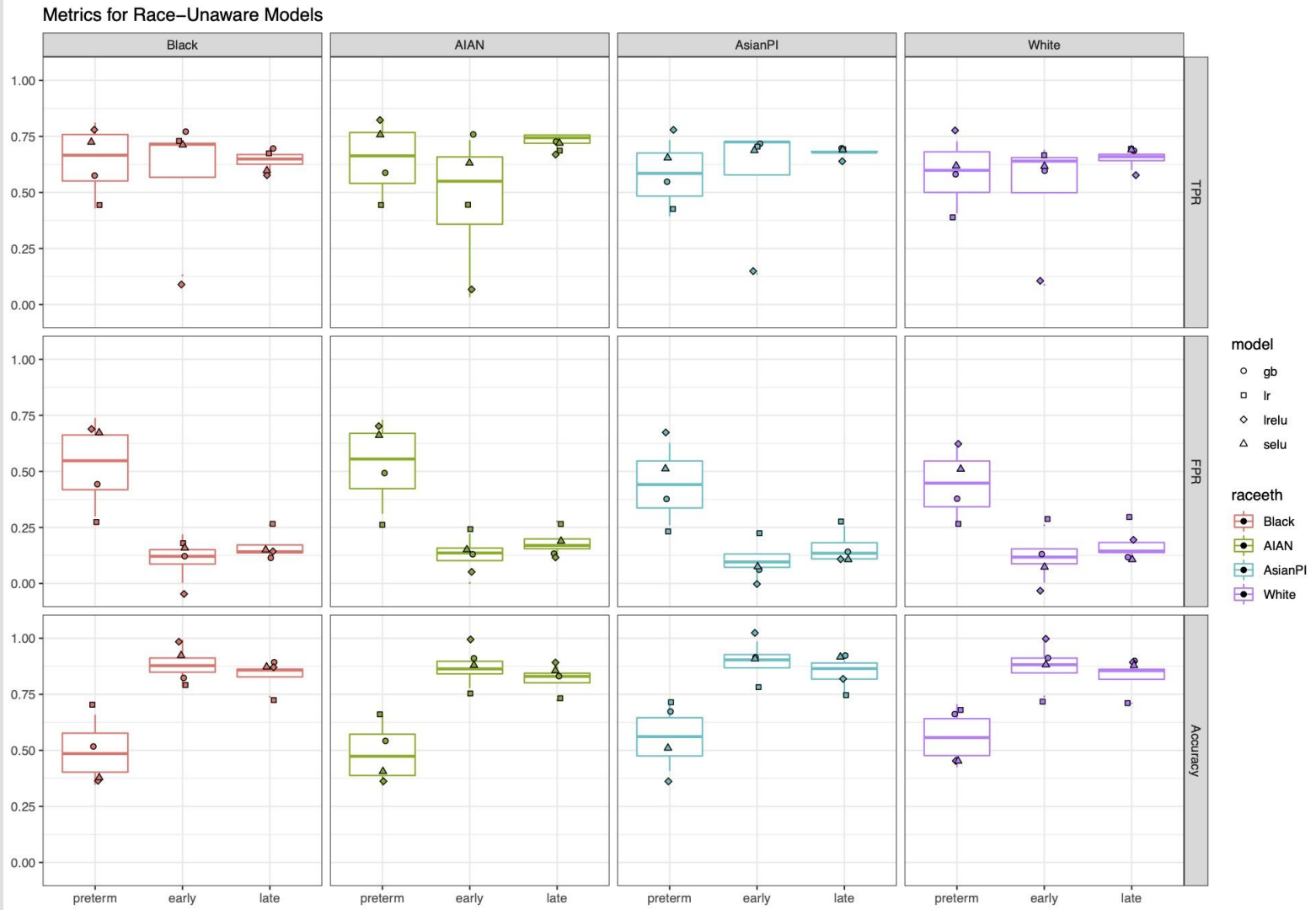
Feature Correlation: Data Breakdown



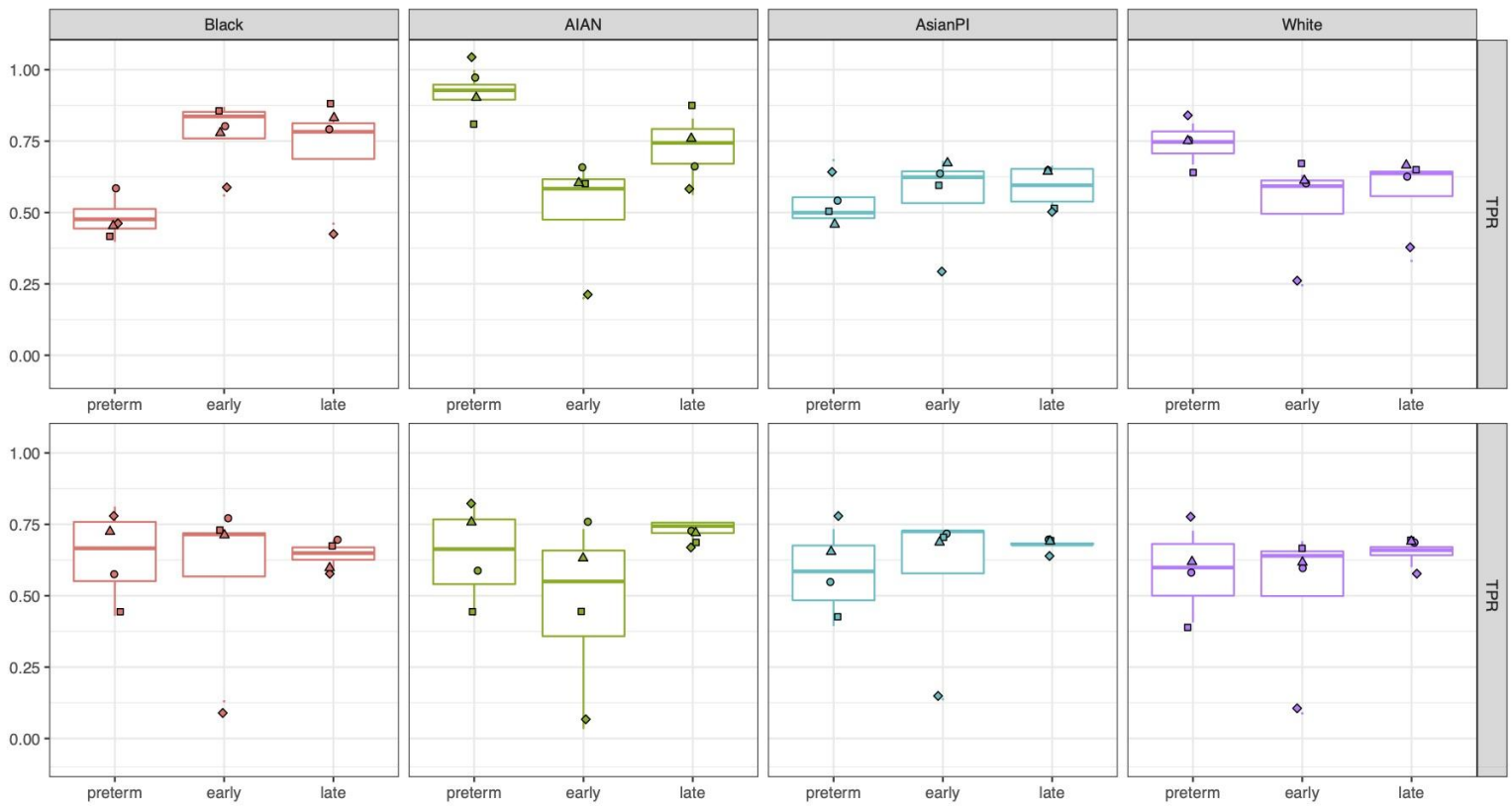
Feature Correlation: Data Breakdown



Results: Model Metrics (Unaware)



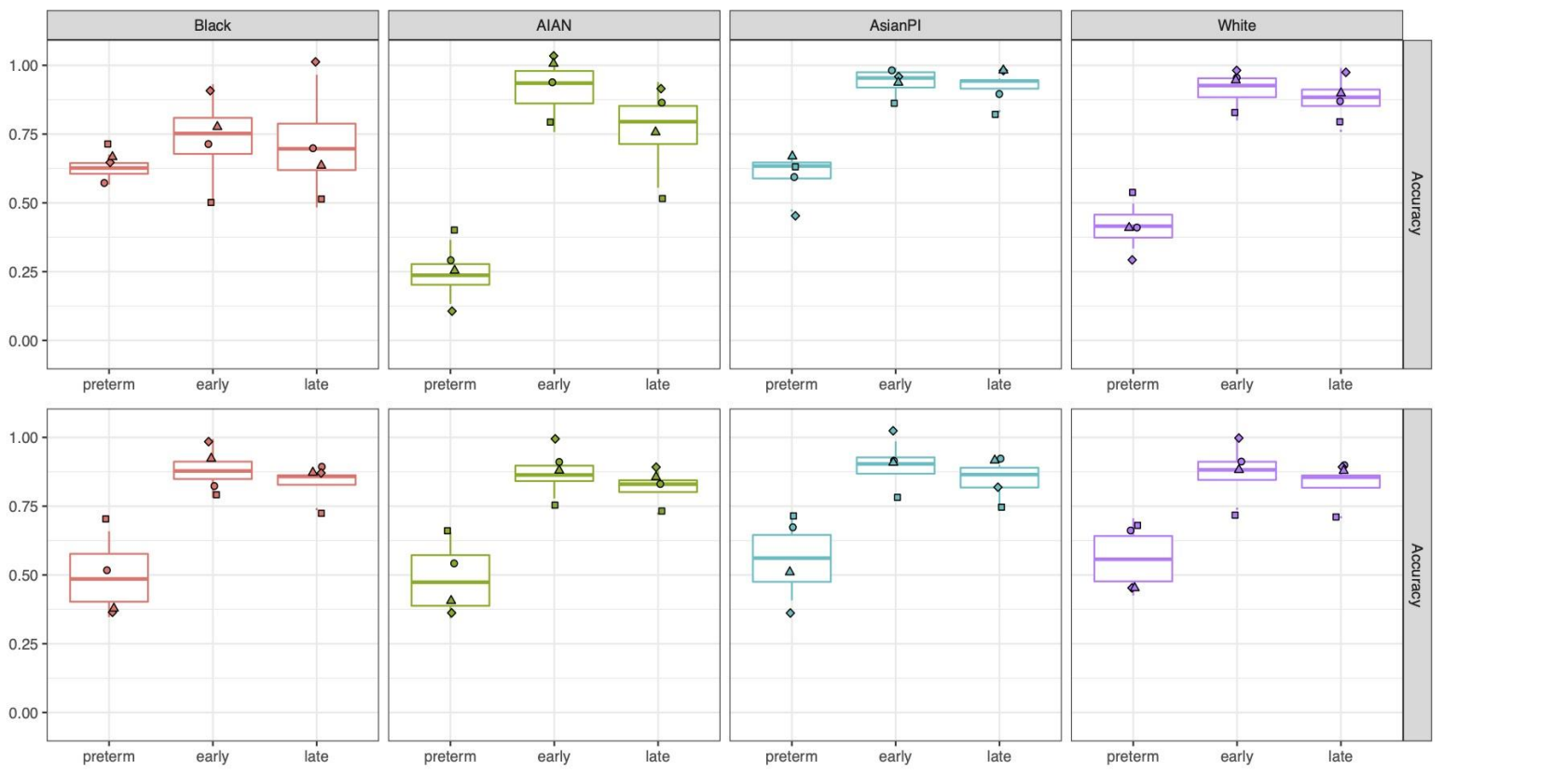
Aware versus Unaware: TPR



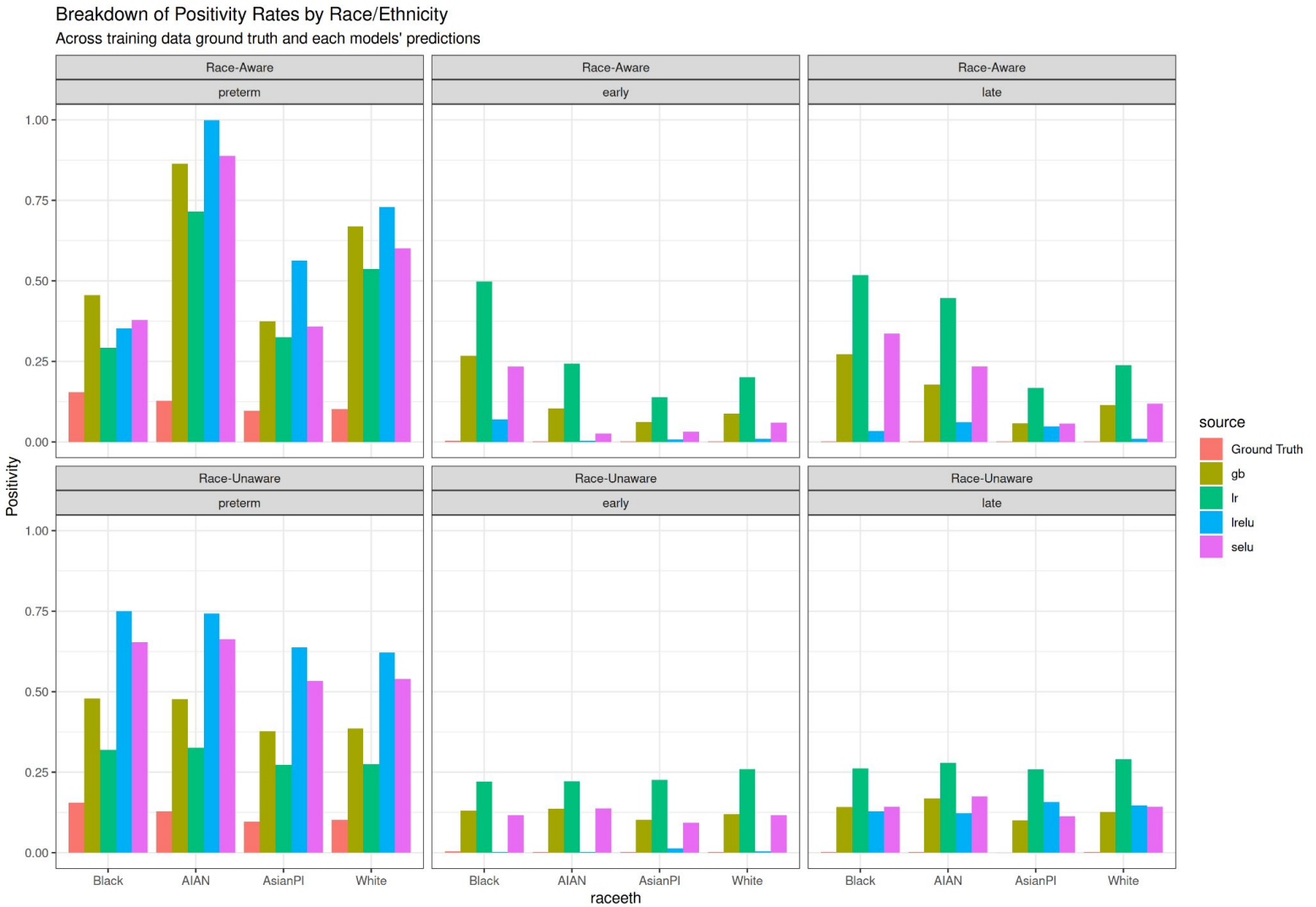
Aware

Unaware

Aware versus Unaware: Accuracy



Aware vs Unaware: Positivity



Policy Implications

- It matters how the model is embedded in decision-making and how it enables the healthcare professional to allocate resources. For e.g. how do we weight and incorporate new information from tests?
- We want to pursue a policy of ***equal allocation***: *allocation of resources as decided by the model is equal across groups, possibly after controlling for all relevant factors.*
- "All relevant factors" may include the disease burden given to each race. It may that we have to be even more granular. E.g. low-income black persons vs high-income black persons.
- Since treatment can be costly, we also want to ensure that the actual ***accrual of benefit*** is roughly equal across groups, therefore, we must also strive toward ***equal performance***: *the model performs equally well across groups for such metrics as accuracy, sensitivity, specificity, and positive predictive value*

Takeaways

- Models trained without using the race variable show less disparities in TPR and accuracy across racial cross-sections
- Collecting more biologically relevant data during pregnancy might provide features with better predictive power
- **However, the performance of models should be contextualized with impacts the predictions will have on patients through policy**

Limitations and Future Work

- Disparities in dataset suggest that our analysis of the models may not accurately reflect the models that we attempted to replicate
- Explore causal relationships in the models to explore strength of feature variables
- Evaluate fairness with respect to other protected attributes

CPSC 464: Fairness in Predicting the Risk of Stillbirth and Preterm Pregnancy

•••

Bradley Yam, Cove Geary, Ivy Fan

https://github.com/bradleyyam/fair-child