Probing Automated Treatment of Urinary Tract Infections for Bias:
A Case-Study Where Machine Learning Perpetuates Structural Differences and Racial Disparities

Garrett Yoon BS, Vincent J. Major PhD

NYU Grossman School of Medicine, Department of Population Health

NYU Langone Health

INTRODUCTION

Urinary tract infections (UTI) are a common indication for antibiotic treatment but some widely prescribed antibiotics are 'second-line' treatments and contribute antibiotic resistance. A recent paper[1] described a machine learning system to recommend the narrowest appropriate antibiotic for an individual's UTI. Data-driven methods integrated with clinical decision support may improve antibiotic stewardship and slow the onset of resistance.

Machine learning models may inadvertently contain bias and should be vetted before implementation. Unintended discriminatory practices have been found in widely used healthcare algorithms.[2]

MATERIALS AND METHODS

The openly-available AMR-UTI database[3] was used including patients from two Boston hospital systems. The system's performance is measured by two metrics of prescriber behavior:

1. the proportion of inappropriate antibiotic therapy (IAT) and
2. the proportion of second-line usage (SLU).

The authors use these metrics to select optimal thresholds and compare the system to human prescribers.

IAT and SLU were computed by race groups (white vs. non-white) to assess unfairness. Unfairness by race would be characterized by non-white patients receiving very different treatment than white patients. To explore potential causes of racial bias, the underlying data was probed for differences. The models were then re-trained without race as a feature as an attempt to reduce the racial disparity.

RESULTS

Prior resistance to one of four commonly prescribed antibiotics was much higher in non-white patients (Figure 1; 21.6% vs 11.0%, p<0.001), likely a result of higher antibiotic exposure (51.9% vs 41.8%, p<0.001). Algorithms can learn these systemic patterns of inequity present in data to reproduce and perpetuate racial disparities. For example, non-white patients were predicted to have much higher probability of resistance (Figure 2).

Among test set specimens (n=3,941), the balance of IAT vs. SLU was dramatically different by race (Figure 3 left). Non-white patients were recommended more second-line therapies (20.1% vs 6.8%, p<0.001) and more often received ineffective therapies (11.3% vs 9.2%, p=0.04).
Re-training the models without the race feature reduced the second-line usage disparity, but it remained significant (15.1% vs 11.7%, p<0.001). Given the racial differences in prior exposure and resistance, masking patient race is insufficient likely due to other associated features acting as proxies for race.

CONCLUSIONS

A consistent trend of higher exposure and resistance to antibiotics is observed for non-white patients. Systemic disparities in which antibiotics are prescribed would be perpetuated by this system. Historical patterns cannot be changed but the risk scores may be debiased to improve recommendations in the future.

REFERENCES

1. Kanjilal, S. *et al.* A decision algorithm to promote outpatient antimicrobial stewardship for uncomplicated urinary tract infection. *Sci. Transl. Med.* **12**, eaay5067 (2020).
2. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).
3. Oberst, M., Boominathan,S., Zhou, H., Kanjilal, S. & Sontag, D.. AMR-UTI: Antimicrobial Resistance in Urinary Tract Infections (version 1.0.0). PhysioNet (2020). https://doi.org/10.13026/se6w-f455

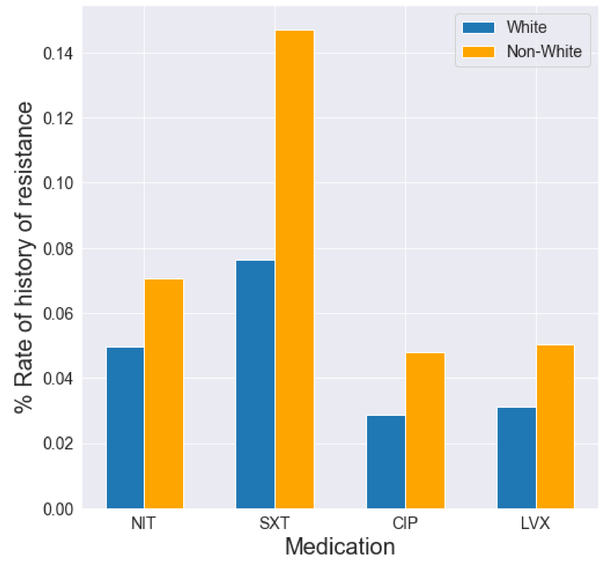

Figure 1: Rates of prior resistance to one of the four prescribed antibiotics conditional on racial group.
NIT = nitrofurantoin, SXT = trimethoprim-sulfamethoxazole, CIP = ciprofloxacin, and LVX = levofloxacin.

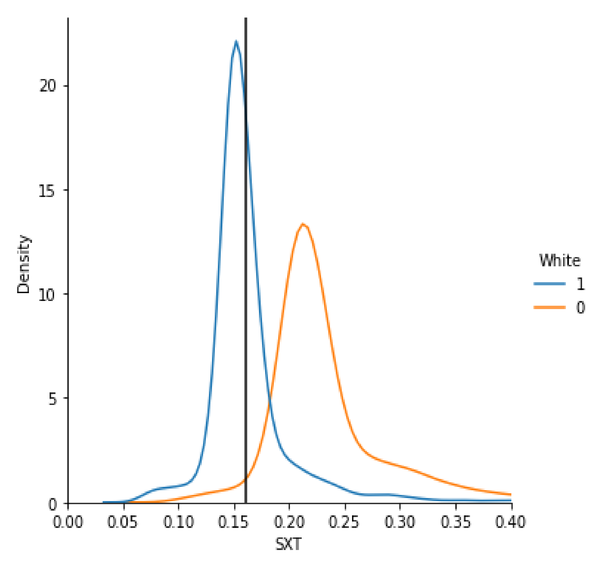

Figure 2: Model output risk distribution for non-susceptibility to SXT for white (blue) and non-white (orange) patients. The vertical black line is the chosen threshold for resistance (higher).

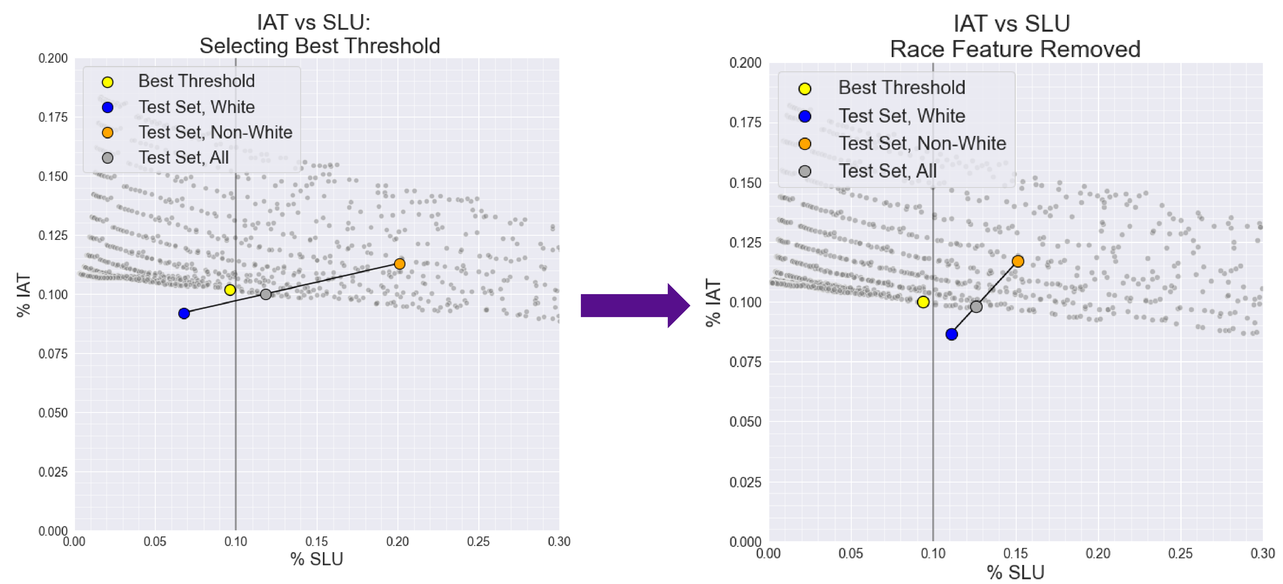

Figure 3: Thresholding sensitivity analysis. Left: the best threshold (indicated by the yellow dot selected from the small grey dots) was selected by choosing the set of thresholds that resulted in a SLU rate of <10% while minimizing IAT rate. When applied to the test set, the metrics generalize to higher SLU (grey dot) but separate drastically when stratified into white and non-white patients (blue and orange dots). Right: after removal of race as a feature, the SLU disparity is reduced (at the cost of increasing SLU for white patients) but a significant disparity remains between IAT-SLU.