

On the Validity of Arrest as a Proxy for Offense: Race and the Likelihood of Arrest for Violent Crimes

Riccardo Fogliato[†], Alice Xiang[‡], Zachary Lipton[†], Daniel Nagin[†], Alexandra Chouldechova[†]

[†]Carnegie Mellon University, [‡] Sony AI

On the Validity of Arrest as a Proxy for Offense: Race and the Likelihood of Arrest for Violent Crimes

Riccardo Fogliato[†], Alice Xiang[‡], Zachary Lipton[†], Daniel Nagin[†], Alexandra Chouldechova[†]

[†]Carnegie Mellon University, [‡] Sony AI

Re-arrest vs. re-offense

Re-offense risk is considered in decision-making at many stages of the criminal justice system. To aid decision-makers in their assessments, institutions increasingly rely on algorithmic risk assessment instruments (RAIs). However, since not all crimes result in arrests, RAIs do not directly assess the risk of re-offense. Disparities in the risk of re-arrest can lead to biases in the risk scores.

Arrest for violent crimes as a proxy?

Past work has argued that arrests for violent offenses, unlike other types of crimes, represent (racially-)unbiased proxies of offending and thus justified their use in the study of algorithmic fairness in RAIs. The key argument: Black and White offenders are arrested, on average, at similar rates. However, this does not imply that the likelihood of arrest is equal for all incidents.

Our study: racial disparities in NIBRS data

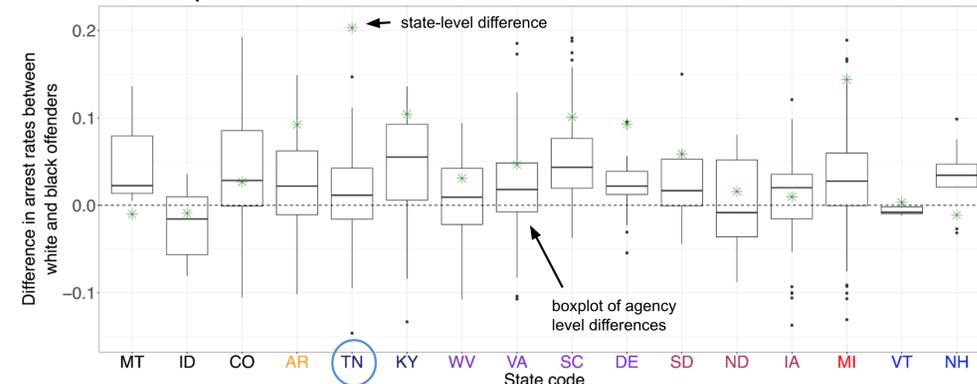
In this work, we study racial disparities in the likelihood of arrest for violent offenses. Our study focuses on 2007–2016 incident-level data of violent offenses with lone victims and offenders that were not cleared by exceptional means in 16 US states as recorded in NIBRS, a national crime data collection program. Our analysis is close to that of D'Alessio and Stolzenberg [2003]. Our dataset consists of more than 3 million offenses comprising simple assault, aggravated assault, forcible rape, robbery, and murder.

Macro-level variations (part I)

Black offenders are arrested at lower rates than White offenders in case of assaults and robbery, and at similar rates for murder and rape. However, arrest rates substantially vary across
 → offense types
 → states
 → law enforcement agencies
 These variations largely drive the observed disparities.

Macro-level variations (part II)

Consider simple assault. Overall arrest rate is 57% for White offenders and 43% for Black offenders.



Tennessee:

- Memphis:
 - arrest rate = 26%
 - % Black pop. = 89%
- Nashville:
 - arrest rate = 54%
 - % Black pop. = 57%

Agencies with larger shares of Black offenders tend to be characterized by lower arrest rates.

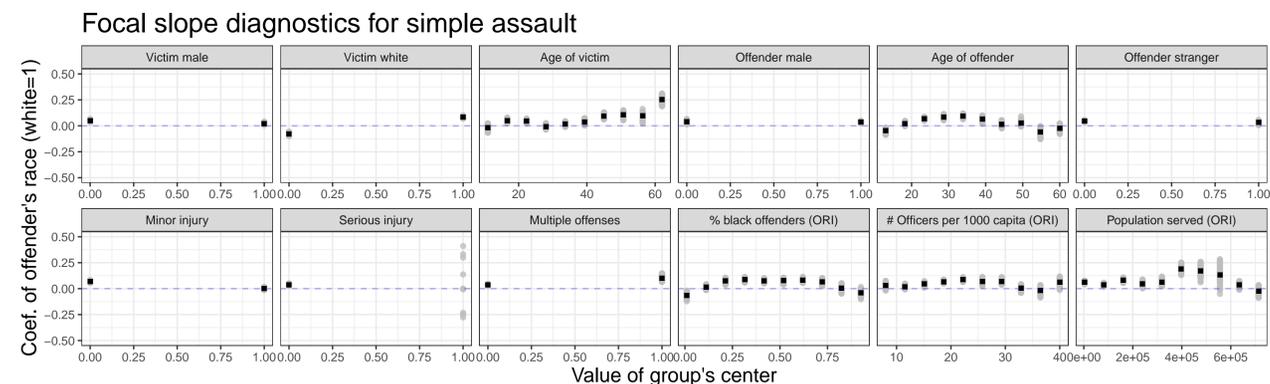
Regression analysis

We use logistic regression models to analyze racial disparities in the likelihood of arrest, controlling for the characteristics of the crime, e.g., state, year, presence of a weapon. The relevant coefficients estimates (and corresponding standard errors):

Variable	Murder/n.n.m.	Forcible rape	Robbery	Aggravated assault	Simple assault
Age of offender	-0.01 (0.00)***	0.01 (0.00)***	0.01 (0.00)***	0.00 (0.00)***	-0.01 (0.00)***
Offender male	-0.18 (0.09)*	2.74 (1.02)**	-0.11 (0.03)**	-0.11 (0.01)***	0.01 (0.00)***
Offender white	-0.20 (0.09)*	0.04 (0.02).	0.24 (0.02)***	0.03 (0.01)***	0.04 (0.00)***

Thus, according to our regression model, White offenders appear to be *more* likely to be arrested than Black offenders for offenses of robbery and assaults, but not for forcible rape and murder.

- Past work has generally assumed that the model is well specified. But is it really?
- We test whether the model is correctly specified using the *focal slope* model diagnostics tool proposed by Buja et al. [2019]. The idea: Even if the distribution of the predictors varies, the estimate of the coefficient of the offender's race should not change.



For simple assault, the sign and size of the estimate of the (conditional) association between the offender's race and arrest vary with the (reweighting of the) distribution of the predictors, indicating that the postulated model is *not* well specified. We find similar variations in the coefficient's estimates for aggravated assault and forcible rape.

Takeaways

- ★ (Potential) explanations of macro-level disparities:
 - differential allocation of policing resources;
 - differential cooperation of the community.
- ★ Researchers should call into question the assumption of well-specification underlying the regression model by examining whether and how it affects their conclusions.
- ★ Given the observed variations in arrest rates and the presence of racial disparities in the data, arrest does not represent a good proxy for offending, even in case of violent offenses.

Implications for RAIs

Given the same risk of re-offense, the RAIs will estimate the risk of re-offense to be lower for
 → the types of defendants that are more prevalent in areas where arrest rates are lower. RAI will also not be calibrated across areas. Fix? Account for these differences or train jurisdiction-specific tools.
 → the defendants that commit crimes of rape and robbery will be judged, compared to those that commit assaults. Fix? Train crime-specific RAIs, e.g., as done for IPV.

Debiasing cannot be based (only) on NIBRS data

Can we “debias” RAIs based on our findings? No. Why?
 → effect of specific types of bias cannot be estimated
 → NIBRS only contains data of offenses known to law enforcement
 → model misspecification threatens estimation
 → longitudinal estimates are needed, but they are not available in NIBRS
 Fix? Leverage NIBRS together with other data sources, such as self-reported data of offending behavior.

References

- Buja, A., Brown, L., Kuchibhotla, A. K., Berk, R., George, E., Zhao, L., et al. (2019). Models as approximations ii: A model-free theory of parametric regression. *Statistical Science*, 34(4):545–565.
- D'Alessio, S. J. and Stolzenberg, L. (2003). Race and the probability of arrest. *Social forces*, 81(4):1381–1397.