Implications of race and ethnicity (mis)measurement in the U.S. criminal justice system*

Keith Finlay
U.S. Census Bureau

Elizabeth Luh
University of Michigan

Michael Mueller-Smith
University of Michigan

March 16, 2023

Abstract

In this extended abstract, we explore several related analyses regarding the (mis)measurement of race and ethnicity in statistical reporting on the U.S. criminal justice system. First, we explore how the common reliance on administratively reported race and ethnicity in criminal justice data collection creates non-trivial bias regarding racial and ethnic disparities in justice system contact. To accomplish this, we link individual criminal histories contained in the Criminal Justice Administrative Records System (CJARS), which include race and ethnicity information as recorded by arresting officers and other actors in the justice system, to self-reported race and ethnicity from the 2010 Decennial Census. We find discrepancies between self-reported and agency-reported racial and ethnic information on 14.3% of individuals, with the largest gaps for self-identified American Indian and Asian individuals (but present for all groups). When we apply these preliminary results to extant federal incarceration statistical reports, we find that American Indian populations actually exhibit the highest imprisonment rates in the U.S. population, exceeding that of Black, non-Hispanic individuals. Second, we delve into the impact of discordance between self-reported and perceived race in the justice system, which we view as a potential novel source of variation to measure racial discrimination in the U.S. justice system. In these exercises, we examine how discordant racial and ethnic information impact a range of court outcomes, including types of charges (felony, violent, drug, etc.), disposition outcomes (convictions), and sanctions (incarceration). Across these exercises, we find strong impacts of "labeling" information; in fact, labeling appears to dominate self-reported racial information in the determination of adjudication outcomes. Those mislabeled as Black experience outcomes as if they were Black; those mislabeled as White experience the same phenomenon in reverse. Our evidence indicates that this pattern is not the simple consequence of errors in the record linkage process. Third, in a final planned exercise, we seek to explore why racial and ethnic discordance exists in the U.S. justice system. To accomplish this exercise, we will document when and where discordance is more pronounced, and examine whether criminal justice policies or professional incentives (e.g., electoral cycles) play a role in generating this behavior.

JEL Codes: J15, K42

*Any conclusions expressed herein are those of the authors and do not reflect the views of the U.S. Census Bureau. All results were approved for release by the Disclose Review Board of the U.S. Census Bureau (Data Management System number: P-7500378, Disclosure Review Board (DRB) approval numbers: CBDRB-FY23-0178.

1 Introduction

One of the most important lines of research over the last half century on the U.S. justice system has focused on racial disparities. A wealth of research has been developed on topics ranging from policing behavior (i.e., use of force, stop and frisk, traffic ticketing, highway searches) (Hoekstra and Sloan 2022; Goncalves and Mello 2021; Horn 2020; Coviello and Persico 2015), federal sentencing (Tuttle 2019), exoneration (Alesina and La Ferrara 2014), and bail decisions (Arnold, Dobbie, and Yang 2018).[1] Yet, at the core of the majority of this work is the underlying assumption that racial and ethnic information is collected by the justice system without error. In contrast with many other data collections and facets of government, racial and ethnic information in the justice system is often populated by justice agency personnel, reflecting their perception of race and ethnicity. From an operational perspective, this may be efficient since officers may need to identify a suspect or a defendant. But the implications for statistical reporting and research are serious considering that whether this accurately reflects an individual's own self-identified race or ethnicity remains an open question.

In this proposal, we leverage a novel linkage between agency-recorded race and ethnicity contained in data from the Criminal Justice Administrative Records System (CJARS) and self-reported race and ethnicity data from the U.S. Census Bureau to quantify the accuracy of racial and ethnic information in the criminal justice system. A major advantage of these partnered data systems is that we can accurately link an individual's self-identified race using responses from a multitude of surveys available on the Federal Statistical Research Data Center (FSRDC) (e.g., the 2010 Decennial Census) to an individual's agency-recorded race. Using this information, we conduct three exercises (two with preliminary evidence; a third planned but not executed at this time). First, we quantify the degree of discordance by reporting the confusion matrix on self-identified versus agency-labeled racial and ethnic categories. This helps us answer questions about what share of agency-recorded data agree with self-reports and how are they mislabeled. These estimates are then applied to existing federal statistical series on incarceration rates to yield new bias-corrected estimates of the incarceration rate per capita by self-identified racial and ethnic groups. Second, we examine how these "mistakes" in racial and ethnic labeling by justice agents themselves influence outcomes in the justice system, a unique and previously unexplored source of variation to investigate potential racial bias in the justice system. Finally, we explore why discordance between self-identification and agency-perception exists, tracing out both geographic and temporal variation and mapping these into differences in IT infrastructure (e.g., insufficient default options at data entry), policy environments, and professional incentives (e.g., electoral cycles), given prior

[1] See Doleac (2021), ACLU and Sentencing Project (2022), Hinton, Henderson, and Reed (2018), Sawyer (2020), Wiley (2022), and Western (2006) for a review of past work on racial disparities in the criminal justice system.

research provides local evidence of intentional manipulation of racial and ethnic information by justice officials (Luh 2022).

Our preliminary evidence suggests that the use of agency-reported race greatly limits our understanding of racial disparities in the justice system. For example, we find that 14.3% of justice-involved individuals contained in CJARS, that we can link to their 2010 Decennial Census response, self-identify their racial and ethnic information inconsistent with justice agency labeling.[2] Agency labeled racial subgroups with the largest degree of misclassification are American Indians and Alaska Natives (28.1%) and Asian/Pacific Islanders (26.4%) justice-involved individuals, although a non-trivial share of White, Black, and Hispanic individuals are misclassified as well (12.2%, 17.4%, 10.5%, respectively). The consequences of such misclassification are not inconsequential. Our bias-corrected estimates of Bureau of Justice Statistics-reported imprisonment rates show that in fact American Indians experience the highest rate of incarceration per capita in the U.S. population, and that Black imprisonment has been overstated by 14.1%, a substantial error that might contribute to harmful negative stereotypes about this minority population.

At the micro-level, agency mislabeling is consequential for individual outcomes in the justice system. Black self-identified justice-involved individuals who are labeled by agencies as White are -6.3 percentage points (p.p.) less likely to be charged at the felony level, -2.2 p.p. (-3.6 p.p.) less likely to be charged with violent (drug) charges, -1.7 p.p. less likely to be convicted, and -3.4 p.p. less likely to receive an incarceration sentence. These effects are so strong that they almost fully neutralize the correlation of court outcomes with self-identifying as Black, a relationship the literature often described as combining both potential discrimination and omitted socioeconomic characteristics that make Black individuals different ex-ante compared to White defendants. Remarkably, we observe a symmetric relationship where defendants who self-identify as White but are labeled as Black experience worse court outcomes just as if they self-identified as Black. Ascribing a specific motivation or mechanism to this relationship remains an area for future work, but we can rule out one mechanical explanation relating to failures in the record linkage process through looking at whether pre-existing income levels are different for mislabeled versus non-mislabeled individuals. If what we call mislabeled was simply the byproduct of improperly matching CJARS records to Decennial Census records, we would expect pre-existing income levels to be noticeably higher for mislabeled individuals since they would not be drawn from the justice-

[2]The Decennial Census supersedes the other surveys and data on self-reported race available on the FSRDC because of its coverage. Due to federal mandate, the Census is designed to enumerate every single individual in the U.S. We exclude all Group Quarters responses in this exercise since, for many jurisdictions, the Decennial Census relies on racial and ethnic information recorded by the facility rather than self-reported by the individual. Such practice actually extends the potential scope of mismeasurement in federal statistics beyond just the criminal justice system, but to population enumeration in general. While there is a more recent 2020 Decennial Census, we chose the 2010 Decennial Census because the COVID-19 pandemic interrupted data collection efforts of the 2020 Decennial Census. We have plans on leveraging other surveys (e.g., ACS) and data available on the FSRDC).

involved population.[3] This is not what we find empirically, instead finding precise null results on pre-existing income differences.

The final section of the project is currently planned but not executed. In this final section of the paper, we plan to explore when, where, and why records are being mislabeled. Is this the product of antiquated data entry system that provide insufficient options when logging racial and ethnic information? Is it the intentional manipulation of data entry to minimize potential signs of racial discrimination in a local population? Could (unconscious) bias in racial perceptions play a role, and would places with more racial segregation or less diversity in their police force exhibit higher rates of mislabeling? Understanding why such patterns exist in the data will be a critical step in improving the accuracy of information on the justice system, as well as the fairness of its administration.

The remainder of this extended abstract is split into four sections. In Section 2, we describe the aggregate rates of mislabeling and the resulting implications for federal statistical series. In Section 3, we delve into the analysis of the microdata on how mislabeling impacts individual-level outcomes for defendants in the criminal justice system. In Section 4, we talk about our planned analysis to document variation in mislabeling rates and potential correlates of this phenomenon. Our final section offers some concluding thoughts on the implications of this research agenda.

2 Differences between administrative and self-reported measures of race and ethnicity among the justice-involved population

Since 1997, the Office of Management and Budget (OMB) has required federal agencies, including the Census Bureau and the Bureau of Justice Statistics, to use the following minimum categories for race: American Indian or Alaska Native (AIAN), Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White (Office of Management and Budget 1997). The same standard requires a separate identification for ethnicity, using the following minimum categories: "Hispanic or Latino" and "Not Hispanic or Latino." The OMB regulation also states that "self-identification is the preferred means of obtaining information about an individual's race and ethnicity, except in instances where observer identification is more practical (e.g., completing a death certificate)". The Bureau of Justice Statistics combines the Asian and Native Hawaiian or Other Pacific Islander categories and reports statistics only for the combined group (see, e.g., Harrell 2009).

[3] Even if the records did match to someone with contact with the justice system, it has been observed that those facing criminal charges exhibit an Ashenfelter-type dip prior to facing criminal charges. So, one would expect differences in pre-existing income for mismatched individuals even if they were justice-involved, since the temporal pattern of when charges were filed would be highly unlikely to line up with the record in CJARS.

While race and ethnicity categories are standardized at the federal level, criminal justice agencies may use different race and ethnicity categories or may not collect information on some categories; additionally, the categories collected may change over time. This leads to heterogeneity in race/ethnicity data quality across criminal justice agencies. For example, departments of corrections (DOCs) in Minnesota, Missouri, Montana, Ohio, and Vermont only report race and not ethnicity.[4] Thus, these states do not report any incarceration of Hispanics, which could lead to underreporting of Hispanic individuals in the justice system within those states. Since federal statistics of incarceration, such as those published by the Bureau of Justice Statistics, depend on data provided by states, non-identification of Hispanics in one state can cause underestimates of incarcerated Hispanics nationally. Thus, low estimates of the incarceration of Hispanics could be driven by poor race measurement, rather than actual low incarceration rates of Hispanic individuals.

Table 1 depicts the different race categories reported by each state's DOC covered by the CJARS data set. Of these listed state agencies, only Michigan DOC reports race and ethnicity separately. Furthermore, there is significant heterogeneity in what race categories are recorded. While each state's DOC consistently reports White, Black, and Asian, Hispanic and AIAN are not reported in 6 and 5 out of the 23 states included, respectively. Ten states also include an 'Other' race category or race/ethnicity categories that are not included in data from other states (e.g., Middle Eastern), while the rest do not allow for an additional 'Other' race.

To estimate how much administrative and self-reported measures of race and ethnicity differ in the justice-involved population, we link individuals ever incarcerated in the CJARS data to their 2010 Decennial Census response. We harmonize the Census Bureau race/ethnicity groups to match those of BJS, as shown in Table 2, and measure the amount of misreporting within each race group, based on the agency-reported race, which we report in Table 3.[5] Each cell reports the percentage of individuals within each race combination within the total sample (e.g., *White, White* = 41.5% of 2.8 million observations self-identified and were recorded as White). We also present the total percentage of racial identity responses that matched the administratively recorded racial identity in Column 7. For example, 28.1% of all individuals who self-identified as AIAN in the 2010 Decennial Census were correctly recorded as AIAN by the justice system. Similarly, we present the total percentage of observations where administratively recorded racial identity matched with the self-reported response in the second to last row.

We demonstrate how the compounding of errors in data collection on race/ethnicity in the justice system impacts federal published reports of criminal justice statistics by race using the annual

[4] We determined whether a state collected ethnicity data based on if we observed any records of Hispanic or Latino ethnicity in the raw data in the CJARS system.

[5] See Appendix A for details on sample restrictions.

report on prisoners published by the Bureau of Justice Statistics (BJS). BJS creates this report based on data and information provided by state departments of corrections (DOC) as part of the National Prisoner Statistics (NPS) program.[6] BJS reports are largely dependent on agency responses and therefore will reflect the same limitations in how race and ethnicity were measured.[7] Taken together, the annual statistics on prisoners across race may be inaccurate due to the compounding of these measurement issues.

To adjust published BJS trends to reflect self-identification, we create weights based on the proportion of individuals within each cell within each agency-recorded race/ethnicity group (sum of the column) and use those percentages to reweight the annual imprisonment rates by race and ethnicity reported by BJS where the imprisonment rate is defined as number of "individuals sentenced to more than 1 year under state/federal jurisdiction per 100,000 U.S. residents" (Carson 2021). Figure 1 Panel A shows the original BJS reported imprisonment rates (dashed line) and the reweighted rates (solid lines) with the weights used in the table below (Panel B).[8]

As seen in the figure, the ranking of imprisonment rate changes with the reweighting. When BJS uses agency-reported race and ethnicity, Blacks have the highest imprisonment rate of about 1,200 per 100,000 residents, followed by AIANs (900), then Hispanics (600), Whites (200), and lastly Asian/PIs (100). After reweighting to reflect self-reported race and ethnicity, AIANs have the highest imprisonment rate at about 1,500 per 100,000 residents, which outranks the adjusted Black rate of 1,000. Reweighting increases the imprisonment rate for AIANs by 60%, while decreasing the imprisonment rate for Blacks 12%. Hispanic, White, and Asian/Pacific Islander imprisonment rates also rise (by 2.6%, 8%, and 18%, respectively) but not substantially enough to change the relative ranking of imprisonment rates across race/ethnicity.

3 Adjudication outcomes of people with discordant administratively recorded versus self-reported race and ethnicity

In this section, we shift gears to examine the individual-level implications of race/ethnicity (mis)labeling by criminal justice agencies in our linked microdata. In particular, we are interested in whether courts are more responsive to self-identified racial/ethnic identity or how the justice system has labeled them. While there are examples of intentional manipulation of racial and ethnic information in the literature (Luh 2022), the working example we have in mind are individuals with ambiguous

[6] Another reporting mechanism on the U.S. prison system is the National Corrections Reporting Program (NCRP). This statistical program also relies on data that originates within state and local entities.

[7] For example, according to BJS, Michigan's DOC reports Hispanic as an ethnicity, rather than a race. Thus, "the numbers for Hispanics were significantly underreported" (Carson 2021). Furthermore, if the state chooses to not submit a report, for example, Oregon in 2019 and 2020, then BJS must impute statistics based on other data (e.g., the NCRP).

[8] See the figure notes of Figure 1 for more details on the reweighting and BJS measurement of imprisonment rates.

racial identity either due to their dress or appearance, their mixed race heritage, or their ambiguous skin tone such that mislabeling might be due to random chance.

We test the relative importance of self-identified race versus agency-labeled race directly by measuring how agreement and mismatch between administratively recorded race and self-identified race impacts sentencing outcomes during adjudication, allowing for asymmetric effects based on whether individuals are mislabeled to be minorities or not. In this preliminary evidence, we focus on the simple exercise of examining individuals who either self-identify as Black or White *and* are labeled as Black or White by criminal justice agencies.[9] Thus, individuals will fall into one of the four mutually exclusive categories: self-identify as White, labeled as White; self-identify as White, labeled as Black; self-identify as Black, labeled as Black; self-identify as Black, labeled as White.[10]

To understand the impact of race labeling for an individual i with charge j on sentencing outcomes, we estimate the following:

$$\begin{aligned} Y_{i,j} = &\beta_0 + \beta_1 I(Race_i^{Self-Report} = \text{Black}) + \beta_2 \underbrace{I(Race_i^{Self-Report} = \text{Black}) \times I(Race_i^{Agency} = \text{White})}_{\text{Mislabeled White}} \\ &+ \beta_3 \underbrace{I(Race_i^{Self-Report} = \text{White}) \times I(Race_i^{Agency} = \text{Black})}_{\text{Mislabeled Black}} + X_i\delta + \varepsilon_{i,j} \end{aligned} \tag{1}$$

where court outcome Y for individual i facing charge j is a function of self-identified race ($Race_i^{Self-Report}$) as recorded in the 2010 Decennial Census and agency-recorded race ($Race_i^{Agency}$) and other individual and charge characteristics $X_{i,j}$, which include fixed effects for the county of adjudication, year of birth fixed effects, year of adjudication fixed effects, and an indicator if the individual had any prior convictions in the 3 years preceding disposition. For this regression and all subsequent regressions, standard errors are clustered at the individual level as captured by the Protected Identification Key (PIK).

We include an indicator for self-identifying as Black as well as indicators for the two potentially mislabeled statuses (Black $\rightarrow$ White and White $\rightarrow$ Black). The excluded category is individuals who self-identify as White and are also labeled by agencies as White. Thus, β_1, β_2, and β_3 describe the impacts relative to this excluded group. β_1 measures the correlation of self-identifying as Black with court outcomes, regardless of agency label. β_2 describes the marginal impact of being labeled as White conditional on self-identifying as Black; the total effect of being Black and mislabeled as White would be $\beta_1 + \beta_2$. β_3 captures the marginal impact of being mislabeled as Black, conditional on self-identifying as White.

[9] See Appendix A for the full details on sample restrictions used in this analysis.

[10] Here, we mean White to refer to as White, not Hispanic and Black to mean Black, non-Hispanic.

We examine a range of outcomes reflecting various (potentially discretionary) decisions in the adjudication process, including whether the case was charged at the felony level, whether the case included violent or drug charges, if the case resulted in a conviction, and whether the case generated an incarceration sentence.[11] Such outcomes involve a range of actors (e.g., prosecutors, defense attorneys, investigating officers, judges, etc.) and we remain agnostic on who might react to mislabeled information and how that might translate into differences in court outcomes. While labeled information could potentially be counteracted by physical proximity and visual observation, there are several reasons why we believe labels may continue to have an impact. First, because of their skin tone or appearance some individuals may be ethnically ambiguous in the eyes of court or law enforcement officials and labels may be relied on in the face of uncertainty (Ludwig and Mullainathan 2022). Second, labels may prime subjects in specific subconscious ways even if the information conflicts with visual confirmation of the defendant. Third, many actions occur in criminal charges with limited or no physical interaction with the defendant; for example, a large majority of criminal cases are disposed through plea deals in which the judge may only review a summary sheet (including the mislabeled information) before coming to a ruling on the case.

We show our results in Figures 2–3 with a corresponding table in Table 4. For these figures, the red line denotes the constant estimate, β_0. We present estimates of β_1–β_3 using the arrows to indicate the coefficient size and gray bars to indicate 95% confidence intervals. Since β_2, the impact of mislabeling as White when self-identifying as Black, is measured relative to β_1, we present β_2 relative to β_1, using a blue, dashed line to visually depict the connection of the two estimates.

Figure 2, Panels A–C shows the estimates of β_1–β_3 on agencies' charging behavior. We find that individuals who self-identify as Black are 7 p.p. more likely to be charged with a felony, 4.2 p.p. and 2.7 more likely to be charged with a drug or violent offense, respectively compared to those who self-identify as White. This difference could reflect population differences, such as income, contact with the justice system, or other factors. Interestingly, we find that labeling of race plays an important role in sentencing outcomes as well. For all charges in the figure, being mislabeled as White, despite identifying as Black, nearly completely reverses the increases in charging behavior due to self-identifying as Black. Furthermore, we observe the reverse with the opposite case: self-identifying as White but being mislabeled as Black. Individuals in this category have a 4.7 p.p. significantly higher risk of being charged with a felony, and 1.3 p.p. (2.6 p.p.) higher likelihood for a violent (drug) offense compared to individuals who are labeled as White, regardless of self-identification.

We also find that the differential treatment based on perception of race impacts sentencing out-

[11] In Table 4 we include other offense types: likelihood for property/public order/unknown charge and traffic/DUI charge.

comes. Being labeled as Black regardless of self-identification increases likelihood of conviction by 1.4 p.p. if the individual self-identifies as Black and 0.5 p.p. if an individual self-identifies as White relative to the excluded group. We also find that this affects likelihood of being sentenced to incarceration. Here again, being labeled as Black significantly increases likelihood of being sentenced to incarceration by 4 p.p. and 3 p.p. if the individual self identifies as Black or White, respectively. For both of these outcomes, we find that being labeled as White, regardless of self-identification, leads to better sentencing outcomes compared to those labeled as Black.

Taken together, these coefficients suggest that racial prejudice rather than population differences across race may play a role in the racial disparities we observe in the adjudication stage. Furthermore, we provide the first evidence that perceived race, rather than an individual's true race, plays an important role in criminal case outcomes. We plan on fully exploring the causes and mechanisms for this observed relationship in future work. For instance, it would be interesting to revisit these results when restricting to just those with two parents of different racial/ethnic groups to see whether impacts are concentrated among those who might be racially ambiguous or whether these relationships are observed for all criminal defendants.

3.1 Testing correlation of data linkage errors with racial misalignment

An alternative explanation for these results is that linkage quality of the CJARS data to the Decennial Census varies across the misalignment of race between the two data sources. Thus, observations where race is aligned are more likely to be matching the same person, while observations where race is misaligned are actually matching records from two different persons together. The consequence of this would be that the labeled race should be more predictive of court outcomes because it is a more accurate measure of race for the defendant.

To test this hypothesis, we examine the correlation of prior earnings, an outcome preceding the focal charge, with an indicator variable for if the labeled race matches the perceived race. If the linkage quality varies by race alignment, then wrongly linked observations are statistically more likely to match a justice involved individual with a non-justice involved individual. As past research, such as Bushway, Stoll, and Weiman (2007), Holzer, Raphael, and Stoll (2003), and Fairlie (1999), has noted, justice involved individuals are less attached to the formal labor market and thus report less earnings compared to individuals not involved. Thus, observations where racial identity is misaligned are more likely to have higher prior income. We measure the correlation of individual i's prior earnings at the time of disposition for charge j on race agreement using the following specification:

$$Prior\ Earnings_{i,j} = \alpha_0 + \alpha_1 I(Race_i^{Self-report} = Race_i^{Agency}) + \alpha_2 I(Race_i^{Agency} = \text{White}) + \varepsilon_{i,j} \quad (2)$$

where $I(Race_i^{Self-report} = Race_i^{Agency})$ is an indicator variable for if the agency recorded race matches the self-identified race from the 2010 Decennial Census. We also include controls for recorded race, $I(Race_i^{Agency} = \text{White})$, to control for underlying population differences across the two race groups. We measure the outcome variable, $Prior\ Earnings_{i,j}$, using total earnings reported on IRS 1040 tax filings in the 3 years prior to the charge disposition date.[12]

We present our results in Table 5. If differential linkage quality across race and race match rate are driving our findings in Figures 2–3, we would expect to observe an economically meaningful and statistically significant negative coefficient on $I(Race_i^{Self-report} = Race_i^{Agency})$ given ample research that income is lower among, and especially declines right before a criminal charge filed for, justice involved individuals (Finlay, Mueller-Smith, and Street 2021; Agan et al. 2022) We do not find any evidence of this occurring in our data. In fact, the estimate suggests that when race matches, individuals earn approximately $100 less per year. This estimate is insignificant and also relatively small in comparison to the constant estimate of almost $15,000 income per year. While these results are preliminary and not conclusive, they provide confidence that the relationships we have documented are not mechanical in nature relating to the specific circumstances of our data environment.

4 Determinants of discordance in the measurement of race and ethnicity

In this section, we outline our ongoing and future work for this line of research. As demonstrated in this paper, collection of race and ethnicity data in the justice system has created systematic errors in the measurement of race and ethnicity and racial disparities. These errors may vary based on environmental characteristics, location, and time, which we plan on exploring in detail. Given the vast amount of records in CJARS and the coverage of the survey data and administrative data available on the FSRDC, we have enough observations to accurately measure the amount of racial discordance at finer geographic levels, such as county level or agency, using the same strategy in Table 3. Furthermore, as shown in this proposal, CJARS has sufficient coverage of data over time to allow us to also measure contemporaneous changes in the racial/ethnic misalignment rate. Thus, we plan on creating a panel data set of racial/ethnic discordance at the chosen geographic level over time.

This analysis is crucial for understanding potential drivers of the discordance we observe in Table 3. Past research and journalists have uncovered systematic under-reporting of minorities in administrative records on justice contact that greatly affect measures of racial disparities (Luh 2022; Webster 2021). Numerous reasons could explain these errors: antiquated data systems,

[12]Since the IRS 1040 tax filings are only available on an annual basis from 1998 onwards, we focus on charges adjudicated from 2001–2019.

human error (e.g., Texas Department of Public Safety allegedly accidentally overwriting minorities as White in highway stops (Luh 2022)), outdated racial/ethnic categorization practices (e.g., defaulting to White when race is unknown), subconscious biases, and varying institutional incentives. Understanding the source for the misalignment will be important for designing policy to improve accuracy in racial/ethnic data in the criminal justice system. Preliminary evidence by Luh (2022) finds that significant public scrutiny and backlash against systematic misreporting of race reduces misreporting in Texas. Our research will build on this limited existing work to understand other correlations of race/ethnicity mislabeling, which will help the accuracy of justice records and reducing institutional disparities.

We have a wealth of administrative data and surveys identified that we plan on linking to the planned panel data set to help conduct this analysis (e.g., annual county-level measures of gross domestic product, local, state, or national electoral outcomes, changing demographic composition over time) that will help grow understanding on local characteristics have the strongest correlations with the racial discordance. Furthermore, since agencies themselves may change their own data collection over time (e.g., Texas added Hispanic as a race category in 2010 (Luh 2022)), we also plan on measuring how these data infrastructural changes impact the rate of mislabeling.

5 Conclusion

This research identifies significant shortcomings in how race and ethnicity are measured in the criminal justice system. We show how reliance on agency-recorded race distorts measurements of racial disparities in federal reports, leading to significant misperceptions on imprisonment rates within each race and ethnic group. Notably, the imprisonment rate for AIANs is underreported on average by 60%. This work is the first step in providing more accurate measurement of racial or ethnic identity in the justice system through linking in self-identified race from the Decennial Census to the recorded race.

We also find that the labeling of Black or White racial identity impacts charging and sentencing outcomes in the justice system. Individuals labeled as Black face worse outcomes in adjudication compared to those labeled as White; specifically, those labeled as Black are significantly more likely to be charged with a felony, drug, or violent offense and are more likely to be convicted and sentenced to incarceration. These results suggest that racial prejudice, rather than underlying differences between Black and White individuals, may play a role in the disparate outcomes we observe across race.

This work will help improve the existing research on measuring racial disparities and racial bias in the criminal justice system. In recent years, events such as the death of George Floyd at the hands of police and the growing Black Lives Matter movement, have made clear that these

disparities have significant consequences on the general public. But, without addressing the flaws in the underlying data, targeted policy in ameliorating these disparities is significantly harder.

References

ACLU and Sentencing Project. 2022. *Report to the United Nations on Racial Disparities in the U.S. Criminal Justice System* [in en-US], July. Accessed February 23, 2023. https://www.sentencingproject.org/reports/report-to-the-united-nations-on-racial-disparities-in-the-u-s-criminal-justice-system/.

Agan, Amanda, Andrew Garin, Dmitri Koustsas, Alex Mas, and Crystal Yang. 2022. Removing the Mark: Labor Market Impacts of Criminal Record Remediation. *working paper.*

Alesina, Alberto, and Eliana La Ferrara. 2014. A Test of Racial Bias in Capital Sentencing. *American Economic Review* 104 (11): 3397–433. Accessed February 28, 2023. https://www.aeaweb.org/articles?id=10.1257/aer.104.11.3397.

Arnold, David, Will Dobbie, and Crystal S Yang. 2018. Racial Bias in Bail Decisions. *The Quarterly Journal of Economics* 133 (4): 1885–932. Accessed February 28, 2023. https://doi.org/10.1093/qje/qjy012.

Bushway, Shawn D., Michael A. Stoll, and Weiman. 2007. *Barriers to Reentry? The Labor Market for Released Prisoners in Post-Industrial America.* Russell Sage Foundation, June.

Carson, E. Ann. 2021. *Prisoners in 2020 – Statistical Tables,* December. Accessed February 28, 2023. https://bjs.ojp.gov/library/publications/prisoners-2020-statistical-tables.

Coviello, Decio, and Nicola Persico. 2015. An Economic Analysis of Black-White Disparities in the New York Police Department's Stop-and-Frisk Program. Publisher: The University of Chicago Press, *The Journal of Legal Studies* 44 (2): 315–60. Accessed February 28, 2023. https://www.journals.uchicago.edu/doi/full/10.1086/684292.

Doleac, Jennifer. 2021. Racial Bias in the Criminal Justice System. In *A Modern Guide to Economics of Crime.* Edward Elgar Publishing, June.

Fairlie, Robert W. 1999. The Absence of the African-American Owned Business: An Analysis of the Dynamics of Self-Employment [in en]. *Journal of Labor Economics* 17 (1): 29.

Finlay, Keith, Michael Mueller-Smith, and Brittany Street. 2021. Convictions, Incarceration, and Earnings in an Event Studies Framework. *working paper.*

Goncalves, Felipe, and Steven Mello. 2021. A Few Bad Apples? Racial Bias in Policing. *American Economic Review* 111 (5): 1406–41. Accessed February 28, 2023. https://www.aeaweb.org/articles?id=10.1257/aer.20181607.

Harrell, Erika. 2009. Asian, Native Hawaiian, and Pacific Islander Victims of Crime. Washington, DC: Bureau of Justice Statistics. NCJ 225037. https://bjs.ojp.gov/content/pub/pdf/anhpivc.pdf.

Hinton, Elizabeth, Leshae Henderson, and Cindy Reed. 2018. An Unjust Burden: The Disparate Treatment of Black Americans in the Criminal Justice System. *Vera Evidence Brief.*

Hoekstra, Mark, and CarlyWill Sloan. 2022. Does Race Matter for Police Use of Force? Evidence from 911 Calls. *American Economic Review* 112 (3): 827–60. Accessed February 28, 2023. https://www.aeaweb.org/articles?id=10.1257/aer.20201292.

Holzer, Harry, Steven Raphael, and Michael A Stoll. 2003. Employer Demand for Ex-Offenders [in en]. *The Urban Institute,* 42.

Horn, Chris. 2020. *Racial disparities revealed in massive traffic stop dataset,* June. Accessed February 28, 2023. https://www.sc.edu/uofsc/posts/2020/06/racial_disparities_traffic_stops.php.

Ludwig, Jens, and Sendhil Mullainathan. 2022. *Algorithmic Behavioral Science: Machine Learning as a Tool for Scientific Discovery* [in en]. SSRN Scholarly Paper. Rochester, NY, July. Accessed March 16, 2023. https://papers.ssrn.com/abstract=4164272.

Luh, Elizabeth. 2022. *Not So Black and White: Uncovering Racial Bias from Systematically Misreported Trooper Reports.* SSRN Scholarly Paper. Rochester, NY, July. Accessed February 28, 2023. https://papers.ssrn.com/abstract=3357063.

Office of Management and Budget. 1997. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. *Federal Register* 62 (210): 58782–90. https://www.federalregister.gov/documents/1997/10/30/97-28653/revisions-to-the-standards-for-the-classification-of-federal-data-on-race-and-ethnicity-.

Sawyer, Wendy. 2020. Visualizing the racial disparities in mass incarceration. *Prison Policy Initiative,* accessed February 28, 2023. https://www.prisonpolicy.org/blog/2020/07/27/disparities/.

Tuttle, Cody. 2019. Racial Disparities in Federal Sentencing: Evidence from Drug Mandatory Minimums. *SSRN Electronic Journal,* accessed February 28, 2023. https://www.ssrn.com/abstract=3080463.

U.S. Census Bureau. 2021. *2020 Census Self-Response Rate Map FAQs,* March. https://www2.census.gov/programs-surveys/decennial/2020/data/tracking-response-rates/response-rate-map-faqs.pdf.

Webster, Richard A. 2021. *"If Everybody's White, There Can't Be Any Racial Bias": The Disappearance of Hispanic Drivers From Traffic Records,* November. Accessed February 28, 2023. https://www.propublica.org/article/if-everybodys-white-there-cant-be-any-racial-bias-the-disappearance-of-hispanic-drivers-from-traffic-records.

Western, Bruce. 2006. *Punishment and Inequality in America.* Google-Books-ID: ZfUWAwAAQBAJ. Russell Sage Foundation, May.

Wiley, Ella. 2022. *How Racism in the Courtroom Produces Wrongful Convictions and Mass Incarceration* [in en-US], July. Accessed February 28, 2023. https://www.naacpldf.org/judicial-process-failures/.

6 Tables

Table 1: Race and ethnicity categorizations in CJARS data received from state departments of corrections

State	White	Black	Asian	Hispanic	AIAN	Other
AR	✓	✓	✓	✓	✓	✓
AZ	✓	✓	✓	✓	✓	✓
FL	✓	✓	✓	✓	*	-
GA	✓	✓	✓	*	*	-
IA	✓	✓	✓	✓	✓	-
IL	✓	✓	✓	✓	✓	✓
IN	✓	✓	✓	✓	*	-
KS	✓	✓	✓	✓	✓	-
MI	✓	✓	✓	✓	✓	-
MN	✓	✓	✓	-	✓	-
MO	✓	✓	✓	-	✓	-
MS	✓	✓	✓	✓	✓	-
MT	✓	✓	✓	-	✓	-
NE	✓	✓	✓	✓	✓	✓
NJ	✓	✓	✓	✓	*	-
NV	✓	✓	✓	✓	✓	-
NY	✓	✓	✓	✓	✓	✓
OH	✓	✓	✓	-	✓	✓
OK	✓	✓	✓	✓	✓	✓
OR	✓	✓	✓	✓	✓	-
PA	✓	✓	✓	✓	*	✓
TX	✓	✓	✓	✓	✓	✓
VT	✓	✓	✓	-	✓	✓

Source: Authors' calculations from criminal justice histories using data from the CJARS 2022Q4 vintage. Calculations were made from CJARS data at the University of Michigan, not from Census Bureau data protected by 13 USC §9(a).

Note: "✓" indicates a state DOC records that race category. "-" indicates that the state DOC does not record that race category. "*" indicates that the state reports less than .5% proportion of that race/ethnicity. Only state wide DOC that provide any race information in their incarceration data and are covered by CJARS are included in this table. Race categories are modeled off of race categories used in Bureau of Justice Statistics annual reports of prisoners in the United States.

Table 2: Harmonizing Census Race to BJS/CJARS Race

BJS/CJARS race	Census Race	Census Hispanic
White	White	No
Black	Black	No
Asian/Pacific Islander	Asian Indian, Japanese, Native Hawaiian, Chinese, Korean, Guamanian or Chamorro, Filipino, Vietnamese, Samoan, Other Asian, Other Pacific Islander	No
Hispanic	Any race	Yes
AIAN	AIAN	No
Other	Some other race	No

Table 3: Race and ethnicity as recorded by agencies versus self reported, total table percents

Self-reported 2010 Decennial Census race/ethnicity	**Administratively recorded race/ethnicity** White* (1)	Black* (2)	AIAN* (3)	Asian or Pac. Isl.* (4)	Other (5)	Hispanic (6)	% of admin.-recorded that concord (7)	Row total (8)
White*	41.5%	4.5%	0.1%	0.1%	0.0%	1.0%	87.9%	1,328,000
Black*	1.6%	29.6%	0.0%	0.0%	0.3%	0.3%	92.8%	896,000
AIAN*	1.1%	0.4%	0.6%	0.0%	0.0%	0.1%	28.1%	64,000
Asian or Pac. Isl.*	0.3%	0.2%	0.0%	0.3%	0.0%	0.1%	31.9%	25,500
Other*	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	1.0%	4,600
Hispanic	2.7%	1.1%	0.1%	0.0%	0.0%	13.7%	78.0%	492,000
% of self-reported that concord	87.8%	82.6%	73.4%	71.9%	0.4%	89.5%		
Column total	1,329,000	1,006,000	24,500	11,500	11,000	429,000		2,810,000

Source: Authors' calculations from the 2010 Decennial Census and criminal justice histories from the CJARS 2022Q4 vintage.

Note: This table presents the proportion of the individuals that self-identify as a particular race group within the total sample. For example, *White, White* indicates 41.5% of the 2.8 million in the sample are labeled as White also self-identified as White in the 2010 Decennial Census. We also present the summation of the proportions for each row and column in the second to last row and column, respectively. See Appendix A for details on sample restrictions. * indicates that the race group excludes Hispanic origin (e.g., White* means White, Non-Hispanic). Estimates and sample sizes have been rounded according to Census Bureau DRB rules. All results were approved for release by the U.S. Census Bureau, Data Management System number: P-7500378 and approval number CBDRB-FY23-0178.

Table 4: Impact of race labeling and own racial identity for Black/White individuals on sentencing outcomes

Panel A: Charging outcomes	Felony grade charge	Violent charge	Drug charge	Property+ charge	Traffic/DUI charge
Constant	0.295***	0.125***	0.181***	0.454***	0.241***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$I(Race_i^{Self-Report} = \text{Black})$	0.071***	0.027***	0.042***	-0.031***	-0.038***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$I(Race_i^{Self-Report} = \text{Black}) \times I(Race_i^{Agency} = \text{White})$	-0.063***	-0.022***	-0.036***	0.041***	0.017***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
$I(Race_i^{Self-report} = \text{White}) \times I(Race_i^{Agency} = \text{Black})$	0.047***	0.013***	0.026***	0.008***	-0.047***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Total observations	22,480,000	22,480,000	22,480,000	22,480,000	22,480,000

Panel B: Sentencing outcomes	Convicted	Sentenced to incarceration
Constant	0.524***	0.351***
	(0.000)	(0.000)
$I(Race_i^{Self-Report} = \text{Black})$	0.014***	0.040***
	(0.000)	(0.000)
$I(Race_i^{Self-Report} = \text{Black}) \times I(Race_i^{Agency} = \text{White})$	-0.017***	-0.034***
	(0.001)	(0.002)
$I(Race_i^{Self-report} = \text{White}) \times I(Race_i^{Agency} = \text{Black})$	0.005***	0.030***
	(0.001)	(0.002)
Total observations	22,480,000	11,910,000

Source: Authors' calculations from the 2010 Decennial Census and criminal justice histories from the CJARS 2022Q4 vintage.

Note: This table presents the impacts of being perceived as White or Black and self-identifying as Black or White in the 2010 Decennial Census. Each column is a single regression. Fixed effects for year of adjudication, year of birth, county of adjudication, and controls for any prior criminal history in the 3 years preceding the adjudicating event are included. Standard errors are clustered at the PIK (individual) level and are in parentheses. "Property+" includes property, public order, or unknown charge. Sample is restricted to males adjudicated from 2000–2019 who either self-identify as Black/White in the 2010 Decennial Census and are labeled as Black/White in the CJARS data. See Section 3 on full regression specifications. $*$ p<0.1, $**$ p<0.05, $***$ p<0.01. Estimates and sample sizes have been rounded according to Census Bureau DRB rules. All results were approved for release by the U.S. Census Bureau, Data Management System number: P-7500378 and approval number CBDRB-FY23-0178.

Table 5: Correlation of match rate between agency recorded race and self-identified race on prior earnings

	Prior earnings
$I(Race_i^{Self-report} = Race_i^{Agency})$	-302.9
	(55,660)
$I(Race_i^{Agency} = \text{White})$	19,870
	(18,470)
Constant	44,180
	(55,240)
Total observations	21,810,000

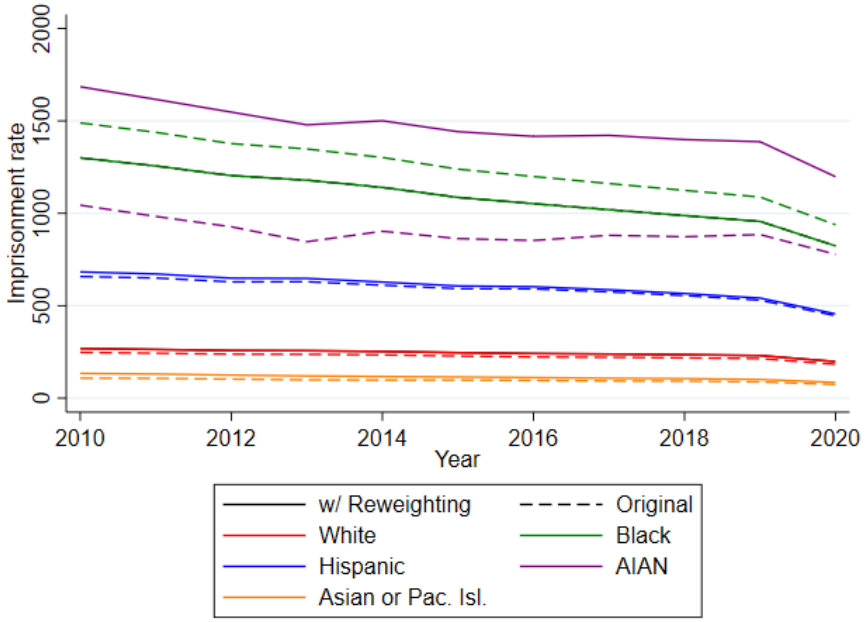
Source: Authors' calculations from the 2010 Decennial Census, criminal justice histories from the CJARS 2022Q4 vintage, and total earnings reported on IRS 1040 tax filings from 1998–2018.

Note: This table presents the correlation of likelihood of race match between agency recorded race and self-identified race from the 2010 Decennial Census, $I(Race\ Agree)$, with prior earnings. Prior earnings are measured using total earnings reported on 1040 tax filings in the 3 years prior to charge disposition. $I(Agency/White)$ is an indicator variable for if the individual is labeled as White by law enforcement agency. Standard errors are clustered at the PIK (individual) level and are in parentheses. Sample is restricted to males adjudicated from 2001–2019 who either self-identify as Black, Non-Hispanic or White, Non-Hispanic in the 2010 Decennial Census and are labeled as Black, Non-Hispanic or White, Non-Hispanic in the CJARS data. See Section 3 on full regression specifications. $*$ p<0.1, $**$ p<0.05, $***$ p<0.01. Estimates and sample sizes have been rounded according to Census Bureau DRB rules. All results were approved for release by the U.S. Census Bureau, Data Management System number: P-7500378 and approval number CBDRB-FY23-0178.

7 Figures

Figure 1: Demonstration of the difference between administratively and self-reported race and ethnicity in incarceration

Panel A: Bureau of Justice Statistics' published imprisonment rate by race, original (dashed) and reweighted (solid)



Panel B: Weights for reweighting imprisonment rates

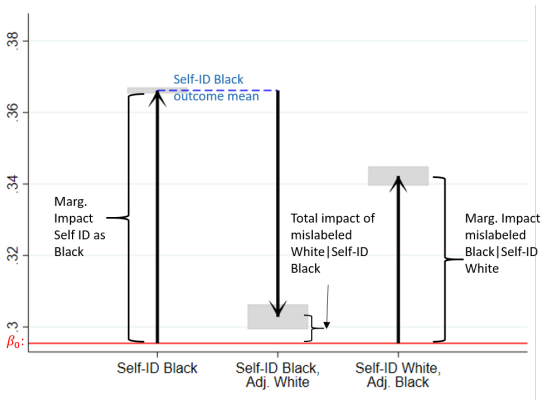
2010 Census	**Administratively reported race**				
race response	White*	Black*	AIAN*	Asian or Pac. Isl.*	Hispanic
White*	87.9%	12.6%	14.1%	15.7%	6.6%
Black*	3.5%	82.8%	3.3%	5.0%	2.1%
AIAN*	2.3%	1.1%	73.5%	1.2%	0.9%
Asian or Pac. Isl.*	0.7%	0.5%	0.9%	72.6%	0.6%
Hispanic	5.6%	3.0%	8.2%	5.5%	89.7%

Source: Authors' calculations from the 2010 Decennial Census and criminal justice histories from the CJARS 2022Q4 vintage.

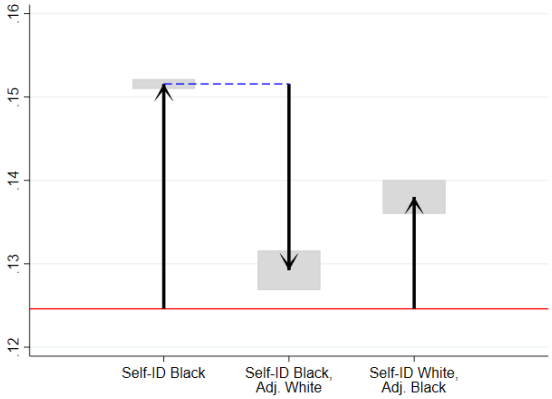
Note: This figure presents the imprisonment rates by race or ethnicity presented in Figure 2 of the Bureau of Justice Statistics (BJS) report, Prisoners of 2020 (Carson 2021). The dashed line represents the imprisonment rates using BJS data tables and the solid line represents the imprisonment rate reweighted using the table below the figure. BJS defines imprisonment rates as: the number of prisoners sentenced to more than 1 year under state/federal jurisdiction per 100,000 U.S. residents. To reweight imprisonment rates, we applied the weights to the total number of sentenced prisoners by race reported in Table 3 of the same report and converted it to the imprisonment rate by multiplying the weighted counts with the ratio of total prisoners by race divided by the original imprisonment rate. The weights are from Table 3 and are calculated by dividing each of the cells by the summation of each column (e.g., *White,White* = 87.9 = 41.5/(41.5 + 1.6 + 1.1 + 0.3 + 2.7)). We follow BJS race categorizations and exclude "Other, Non-Hispanic" from the reweighting. Estimates and sample sizes have been rounded according to Census Bureau DRB rules. All results were approved for release by the U.S. Census Bureau, Data Management System number: P-7500378 and approval number CBDRB-FY23-0178.

Figure 2: Impact of race/ethnicity discordance on charge outcomes

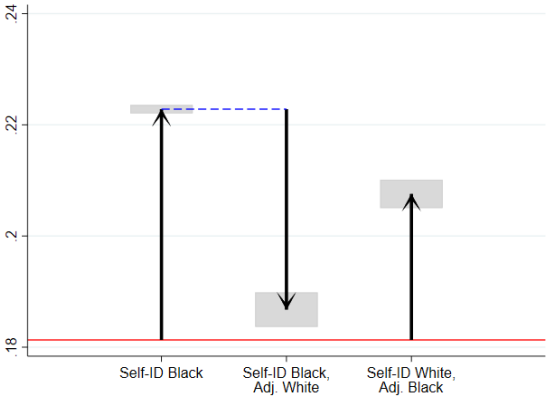
Panel A: Felony charge



Panel B: Violent charge



Panel C: Drug charge

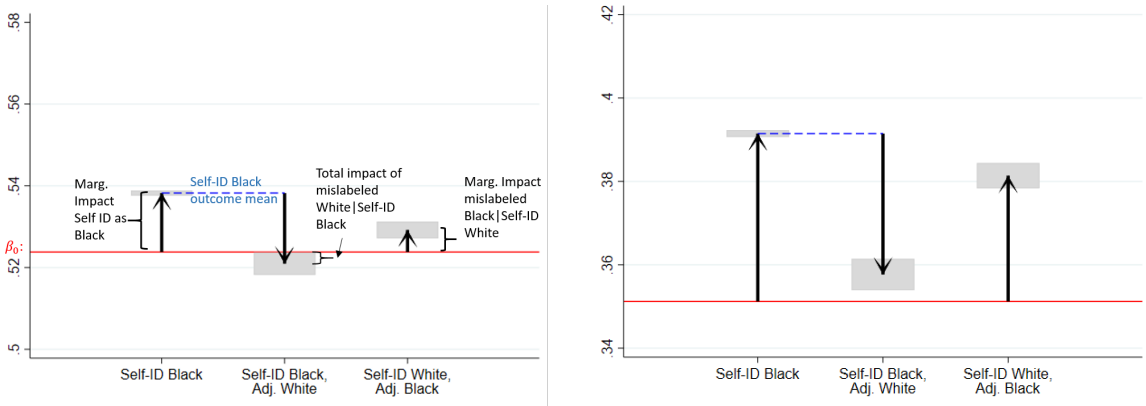


Source: Authors' calculations from the 2010 Decennial Census and criminal justice histories from the CJARS 2022Q4 vintage.

Note: This table presents the coefficient estimates of β_{0-3} from Equation 1 with standard errors clustered at the PIK (individual) level. Regressions also include fixed effects for year of birth, year of adjudication, county of adjudication, and controls for prior criminal history. See Panel A for details on graph exposition. See Section A for more details on sample restrictions. Estimates and sample sizes have been rounded according to Census Bureau DRB rules. All results were approved for release by the U.S. Census Bureau, Data Management System number: P-7500378 and approval number CBDRB-FY23-0178.

Figure 3: Impact of race/ethnicity discordance on disposition and sentencing outcomes

Panel A: Conviction *Panel B: Incarceration sentence conditional on guilt*



Source: Authors' calculations from the 2010 Decennial Census and criminal justice histories from the CJARS 2022Q4 vintage.

Note: This table presents the coefficient estimates of β_{0-3} from Equation 1 with standard errors clustered at the PIK (individual) level. Regressions also include fixed effects for year of birth, year of adjudication, county of adjudication, and controls for prior criminal history. See Panel A for details on graph exposition. See Section A for more details on sample restrictions. Estimates and sample sizes have been rounded according to Census Bureau DRB rules. All results were approved for release by the U.S. Census Bureau, Data Management System number: P-7500378 and approval number CBDRB-FY23-0178.

A Data Appendix

A.1 Incarceration sample data restrictions

We combine the administrative criminal justice data from CJARS with the 2010 Decennial Census individual responses. These data are linked at the person level using the Protected Identifier Key (PIK). PIKs are generated for all individuals in the Social Security Administration Numident file or have an Individual Taxpayer Identification Number. Thus, individuals in the CJARS data set who are undocumented will not be in the final linked data. For this exercise, we focus on the subset of individuals who were ever incarcerated a state department of corrections in the CJARS data.

We also exclude individuals who were enumerated in Group Quarters in 2010 Decennial Census responses, since the responses may have been recorded by someone other than the respondent (U.S. Census Bureau 2021). In order to best match the BJS sample, we also limit individuals in the CJARS sample to responses from a state department of corrections. We are able to link approximately a third of the 8.4 million eligible individuals in the CJARS data to a 2010 Decennial Census response. The remaining unlinked individuals either were not assigned a PIK, were not enumerated in the 2010 Decennial Census, or were enumerated in Group Quarters.

A.2 Adjudication sample data restrictions

Starting with the combined data of CJARS with the 2010 Decennial Census, we restrict our sample to the criminal charges of male individuals, who were charged with a criminal offense in the CJARS data set starting from 2000 to 2019, and who either reported a Black, Non-Hispanic or White, Non-Hispanic race in the 2010 Census. Thus, all individuals in our sample are PIK'ed and have responded to the 2010 Decennial Census. Out of the 47 million criminal charges of a Black or White individuals in this sample, 48% were linked to a 2010 Decennial Census response where the individual also self-identified as White or Black. The unlinked charges were not included because the individual: 1) was not assigned a PIK, 2) did not respond to the 2010 Decennial Census, or 3) did not choose Black or White as their race or chose Hispanic as their ethnicity in the 2010 Decennial Census.