RACE AND ETHNICITY (MIS)MEASUREMENT IN THE U.S. CRIMINAL JUSTICE SYSTEM

Keith Finlay
Elizabeth Luh
Michael G. Mueller-Smith

Race and Ethnicity (Mis)measurement in the U.S. Criminal Justice System
Keith Finlay, Elizabeth Luh, and Michael G. Mueller-Smith
NBER Working Paper No. 32657
July 2024
JEL No. J15,K42

ABSTRACT

The United States criminal justice system is characterized by substantial disparities in outcomes across racial and ethnic groups. Understanding these disparities requires accurate measures of race and ethnicity of people involved in the justice system. We document how race and ethnicity are recorded by administrative agents and how operational concerns limit corrections to misreported race and ethnicity. To understand the impacts of these administrative processes, this paper uses novel linkages between person-level microdata from the Criminal Justice Administrative Records System (CJARS) and race and ethnicity composites from U.S. Census Bureau census and administrative records, mostly composed of self-reported or family-reported race/ethnicity, to quantify mismeasurement of race and ethnicity in the justice system. We find that 17 percent of misdemeanor and felony defendants and 10 percent of prison inmates have an agency-recorded label that does not concord with the composite measure, largely driven by justice agencies poorly measuring people identified in Census Bureau data as Hispanic, Asian, Pacific Islander, or American Indian and Alaska Native (AIAN). Using estimated correspondences between agency-recorded and the composite race and ethnicity, we reweight federal series on imprisonment rates and show that those series, which currently rely on small survey samples to impute racial and ethnic population shares, have substantially underestimated the incarceration rates of Whites, Blacks, and AIANs.

Keith Finlay
U.S. Census Bureau
4600 Silver Hill Road
Washington, DC 20233
United States
keith.ferguson.finlay@census.gov

Elizabeth Luh
University of Michigan
eluh@uh.edu

Michael G. Mueller-Smith
Department of Economics
University of Michigan
365C Lorch Hall
611 Tappan Ave.
Ann Arbor, MI 48109-1220
and NBER
mgms@umich.edu

1 Introduction

Investigating the causes and consequences of racial and ethnic disparities has been one of the most important lines of criminal justice research over the last half century. For example, Blacks in the United States are 5 times more likely to be incarcerated compared to Whites, while Hispanics are 3 times as likely to be incarcerated as Whites (American Civil Liberties Union and The Sentencing Project 2022). Beyond incarceration, both qualitative and quantitative research shows evidence of racial disparities in stop-and-frisks (Goel, Rao, and Shroff 2016), charging decisions (Tuttle 2019), and sentencing outcomes (Alesina and La Ferrara 2014). Identifying these types of patterns is critically important given the well-documented broader impacts of justice involvement on individuals' lives (Pager 2003; Mueller-Smith and Schnepel 2021; Decker et al. 2014) and their family members (Rose and Clear 1998; Finlay, Mueller-Smith, and Street 2023).

A fundamental requirement for studying racial and ethnic disparities in criminal justice outcomes is high-quality measurement of race and ethnicity. As research increasingly relies on criminal justice administrative data or derivative series produced by statistical agencies, researchers must generally rely on the quality of race and ethnicity data originally collected for operational purposes, such as the physical identification of suspects. In contrast to databases for public programs such as Medicaid or household surveys, where race and ethnicity may be self-reported or reported by family members, racial and ethnic information in the justice system is often determined by justice agency personnel—reflecting their perceptions of another's race and ethnicity. Thus, researchers and policymakers are dependent on the ability of officers to accurately and reliably identify race, which may be problematic if officers perceive race and ethnicity differently than does the subject. Even more troubling are cases where some police agencies have been accused of deliberately misreporting racial and ethnic information with the goal of subverting efforts to curtail racial profiling and discrimination (Webster 2021; Collister 2016; Luh 2022).

Some researchers turn to another data source: group quarters (GQ) responses in decennial censuses or the American Community Survey (e.g., Neal and Rick 2014; Bayer and Charles 2018).

These data track people who reside in group living arrangements like dorms, healthcare facilities, jails, or prisons. The Census Bureau increasingly relies on adult correctional facilities to electronically report facility-wide data on behalf of their residents (Belton and Durante 2021; National Academies of Sciences and Medicine 2023). These data are ultimately sourced from agency records systems, which allows flawed data that originated in the justice system to propagate through to survey data. The American Community Survey faces the additional challenge of a two-stage sampling procedure wherein correctional facilities are first sampled and then a subset of respondents are selected within those facilities. In order to cover non-sampled facilities, full responses are imputed, drawing from the bank of observed GQ responses with weights assigned to reproduce the expected population based on the last previously observed enumeration (U.S. Census Bureau 2022). Consequently, misreports present in justice administrative data may both directly and indirectly impact Census Bureau survey efforts to measure the incarcerated population.

As we document below, race and ethnicity data collection in the justice system is inconsistent with best practices, such as those defined at the federal level. The Office of Management and Budget (1997) revision to Statistical Policy Directive No. 15 states that "self-identification is the preferred means of obtaining information about an individual's race and ethnicity," although other means of collection are permitted when deemed more practical. An example of a permissible deviation from self-identification given in the guidelines is a coroner recording information about an autopsy where the individual in question can no longer self-report since they are deceased.

In this paper, we leverage a novel linkage between agency-recorded race and ethnicity from the Criminal Justice Administrative Records System (CJARS) and race and ethnicity composites from Census Bureau census and administrative records to quantify the accuracy of race and ethnicity information recorded in the criminal justice system.[1] A major advantage of integrating these data systems is that we can link an individual's agency-reported race and ethnicity to their race and ethnicity from census or administrative data that are primarily based on a self-report or family report, utilize paradata to exclude observations that might reflect agency perceptions (e.g. group quarters),

---

[1]CJARS is a longitudinal, harmonized data system of criminal justice events and a partnership between the University of Michigan and the Census Bureau (Finlay and Mueller-Smith 2022).

and these data are available to researchers through the Federal Statistical Research Data Centers (FSRDCs).[2] The longitudinal structure of the CJARS data also enables us to follow individuals through the justice system (e.g., from adjudication to incarceration). Thus, we can measure the amount of racial or ethnic mismeasurement in the criminal justice system and identify the stages of the justice system where misreporting is more likely to occur. Using these linked data, we explore several related questions around the accuracy of race and ethnicity information in the criminal justice system and the implications for extant federal statistical series.

First, we document overall levels of concordance in racial and ethnic information, distinguishing between court records and state prison records. Court records typically rely on police incident reports to populate race/ethnicity fields, and these data are often not used for operational purposes. Prison systems are more likely to have intake processes that provide the inmate the opportunity to self-identify and may use that information for security and housing decisions (e.g., to prevent gang violence). Consistent with these features, we observe lower concordance rates in court data compared to prison data. Between 2000 and 2019, we find that 83 percent of felony and misdemeanor defendants and 90 percent of prison inmates have agency-recorded labels that concord with their Census Bureau race and ethnicity composites. While White and Black justice-involved individuals are typically accurately recorded by agencies as such, the justice system often struggles to accurately classify American Indians and Alaska Natives, Asians, Pacific Islanders, and Hispanics.[3]

Second, we consider how race and ethnicity data quality has evolved over time and how it varies across the United States. We find that there have been only modest changes in the accuracy of this information over the last two decades. There is substantial variation by geography, however. While all jurisdictions are likely to accurately label White and Black justice-involved individuals,

---

[2]See Charles and Guryan (2011) for an excellent discussion of the important theoretical implications of self-identified versus perceived race and ethnicity, the potential mutability of racial identity, as well as a review of the relevant empirical literature grappling with the taxonomical challenge of how to measure race. Snipp (2003) reviews the history of racial data collected the federal government, such as the changes in racial categories, and its impact on social science research.

[3]Similar observations for the American Indian and Alaska Native population have also been made regarding health and mortality data (Small-Rodriguez and Akee 2021; Mays et al. 2022).

locations with higher shares of residents who are Hispanic or American Indian and Alaska Native generally tend to more accurately identify those populations.

Finally, we apply our findings to re-estimate the imprisonment rate of racial and ethnic groups in the United States since 2000. We use data from the Bureau of Justice Statistics (BJS) National Prisoner Statistics (NPS) program together with estimated annual correspondences between agency-recorded race/ethnicity and Census Bureau race/ethnicity composites for the population of prisoners observed in CJARS to produce new aggregate counts and per capita rates that should better reflect how inmates self-identify. We find that current reporting by BJS may underestimate the size of the White and Black inmate populations in 2018 by 20 percent and 17 percent, respectively. For American Indians and Alaska Natives, the incarceration rate since 2010 may have been underestimated by 46 percent. These findings suggest that improved data linkage infrastructure can provide agencies like BJS with new opportunities to improve the accuracy of race and ethnicity measurement of prisoner populations.

The remainder of this paper is organized as follows. Section 2 provides background on the standards for, practices of, and sources of misreporting in race and ethnicity measurement. Section 3 presents our findings on the discordance between race and ethnicity recorded by criminal justice agencies relative to census and administrative data, including comparisons across stages of the justice system, across geography, and over time. Section 4 applies these insights to federal statistical series published by BJS to quantify how misreporting impacts the measurement of racial and ethnic disparities in the justice system. Section 5 concludes.

2   Standards, practices, and challenges of collecting race and ethnicity data

2.1   Federal guidelines and state practices

In 1997, the Office of Management and Budget (OMB) revised Statistical Policy Directive No. 15 (SPD15) to require that federal agencies, including the Census Bureau and the Bureau of Justice Statistics, use the following minimum categories for race: American Indian and Alaska Native

5

(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 method for measuring an individual's race and ethnicity, except where identification by an observer is more practical. The Bureau of Justice Statistics often reports a category that combines the Asian and Native Hawaiian or Other Pacific Islander categories (e.g., Harrell 2009). In 2024, the Office of Management and Budget revised SPD15 again, calling for the collection of race and ethnicity in a single question that allows multiple responses, and for the addition of Middle Eastern or North African (MENA) as a new minimum category. Our research was conducted before these revisions. Justice records management systems generally do not allow for multiple races or ethnicities to be selected, but we have found records systems that could support a new MENA category.

State and local criminal justice agencies are not obligated to collect race and ethnicity data according to OMB guidelines. Their categories may differ and change over time. Table 1 identifies the different race categories reported by each state's department of corrections (DOC) covered by the CJARS data set. Of the state agencies listed, only the Michigan DOC reports race and ethnicity separately. Furthermore, there is significant heterogeneity in what race or ethnicity categories are recorded. While each state's DOC consistently reports for the White, Black, and Asian categories, Hispanic categories are not reported in 6 out of the 23 states included and AIAN categories are not reported in 5 out of the 23 states. Ten states also include an Other race/ethnicity category or race/ethnicity categories that are not included in the 1997 OMB minimum categories (e.g., Middle Eastern), while the rest do not allow for an additional Other race.

Different collection practices across states causes heterogeneity in race/ethnicity data quality across criminal justice agencies. Although the CJARS project requests race and ethnicity variables from all agencies, DOCs in Minnesota, Missouri, Montana, Ohio, and Vermont only report race and not ethnicity.[4] These states do not report any incarceration of Hispanics, which could lead

---

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

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 (Carson 2021). Thus, low estimates of the incarceration of Hispanics could be driven by poor ethnicity measurement, rather than actually low incarceration rates of Hispanic individuals.

These findings are consistent with a report from the National Conference of State Legislatures that found that out of 13 states that reallocated prisoner residences during the 2020 redistricting cycle, only one state's department of corrections recorded race and ethnicity in a way consistent with federal guidelines (Williams 2023). The report recommended that states considering reallocation adopt policies to require the federal standards, but recognized the substantial costs in recanvassing their prisoner populations.

2.2 Mismeasurement of race and ethnicity in criminal justice data

Our perceptions of someone else's race or ethnicity may differ from how that person self identifies for a number of reasons. Characteristics such as names (Gaddis 2017; Fryer and Levitt 2004; Luh 2022; Bertrand and Mullainathan 2004), clothing (Freeman et al. 2011), occupation (Freeman et al. 2011), skin tone (Monk 2021; Gyimah-Brempong and Price 2006; Naik and Farrukh 2022; Noe-Bustamante et al. 2021; Davenport 2020; Dixon and Telles 2017), and other physical features (Blair et al. 2002) can impact the perception of an individual's race or ethnicity. These physical or socioeconomic characteristics are also associated with worse criminal justice outcomes like conviction and arrest rates (Gyimah-Brempong and Price 2006; Finkeldey and Demuth 2021; King and Johnson 2016). Furthermore, the characteristics of the recorder can also impact their ability to identify an individual's race (Harris 2002). For example, Luh (2022) finds that non-White troopers are better at recording motorist race in highway searches compared to their White peers.

In addition, work by Eberhardt et al. (2004) and Bordalo et al. (2016) shows that people associate Black racial identity when primed with words associated with crime. For law enforcement

officers, this may lead to over-labeling of racial minorities. While not necessarily a cause of inaccurate data, it is interesting to note that Eberhardt et al. (2004) finds that causality flows in both directions, showing that seeing Black faces also primes people to detect crime related images. These conclusions are consistent with work demonstrating statistical discrimination among employers against minorities in response to the adoption of ban-the-box policies. Agan and Starr (2018) and Doleac and Hansen (2020) find that withholding criminal records from job applications increases the Black-White gap in employment, suggesting that employers infer criminal history using race in the absence of criminal records in employment applications.

It is important to acknowledge that racial and ethnic identity need not remain static. While empirical research frequently assumes that racial identity is a fixed, mutually exclusive characteristic (Holland 2003), there is research to the contrary which shows that self-identification of racial and ethnic identity may change over an individual's lifetime (Saperstein and Penner 2012; Liebler et al. 2017; Harris and Sim 2002; Doyle and Kao 2007).[5] Seminal work by Saperstein and Penner (2012) using the National Longitudinal Survey of Youth (NLSY) showed that an individual's racial identity evolves over time. This research highlights that racial identity may not be solely based on one's parents' identities, but can be influenced by other social and environmental experiences. Other research has shown that self-identified race can also change across contexts and in response to events (Mason 2017; Telles and Paschel 2014). These findings are consistent with labeling theory, which would suggest that individuals may change their self-perceptions based on criminal justice system involvement, altering their self-identify to match beliefs about characteristics that are associated with criminality (Besemer, Farrington, and Bijleveld 2017). These factors may contribute to our estimates of racial and ethnic discordance and the rate of disagreement between race/ethnicity measures across censuses that we describe below, but not necessarily bias statistical reporting or research on disparities.

---

[5]See Darity, Mason, and Stewart (2006), Roberts (2012), Nelson (2016), Rose (2023), and Morning (2007) for additional discussion of the definition and construction of race in society. Rose (2023) provides excellent discussion of the implications of a constructivist perspective of race and ethnicity for economic theory and discrimination research.

2.3 The life cycle of race and ethnicity data in the justice system

To understand how race and ethnicity are recorded in the criminal justice system, how that information persists, and what opportunities there are to update it, we conducted qualitative research by interviewing 7 personnel with knowledge of records management who have worked in multiple jurisdictions and across different stages of the justice system. To supplement these interviews, we also reviewed 5 publicly available procedural guidelines that document official policies and practices within state and local criminal justice agencies.[6] A summary of these sources, which cover procedures in 8 states, is available in Table 2.

Several common themes emerged from this investigation. First, information frictions arise due to the decentralization of information between agencies. Records tend to pass in one direction only, which allows potentially flawed data to propagate through multiple agencies. Second, the reliability of race and ethnicity data varies tremendously, and depends on whether agencies view this information as operationally relevant to their missions. Third, correcting misreported race and ethnicity is seen as legally or bureaucratically costly, which discourages updates. Together, these factors create an environment where flawed information may persist, which ultimately impacts the statistical reporting that relies on these data.

Figure 1 documents the typical flow of information and interactions in the pre-trial and adjudication phases of the criminal justice system, indicating where race and ethnicity data might be electronically recorded and where direct observation could feasibly identify misreports in the record. Based on our interviews, the police incident report (PIR) appears to be the critical point of data collection for race and ethnicity data. Typically, the information contained in the PIR reflects what the arresting officer recorded about the suspect in the field shortly after arrest, and is supplemented with victim statements and fingerprints. The PIR is then transmitted to jails, prosecutors, and courts, and tends to receive significant deference as the official record of a criminal incident

---

[6]While our conclusions in this section are drawn from a convenience sample and may not be nationally representative, the sources touched on a breadth of processes that impact race and ethnicity measurement. We could not find a systematic review of procedures to validate the findings. But see, for example, Rodriguez and Tublitz (2023) for an agency survey that measures whether Hispanic ethnicity is coded separately from race in data management systems.

9

and arrest.

Several sources emphasized that, even though a PIR regularly contains racial and ethnic descriptions of a suspect, officers within the court system do not view race and ethnicity information as operationally relevant to their jobs since using that information during court proceedings could be illegal and a violation of civil rights law. Since it is considered improper to use this information, these fields are not monitored for data quality issues. Consequently, race and ethnicity in court records are viewed with some degree of doubt by those involved in records management. In contrast, measurement of a suspect's sex in court records is considered to be higher quality.

If misreports are identified, staff do not have an incentive to correct the record, especially for fields viewed as secondary to operational needs. For example, fixing flawed race and ethnicity information typically requires filing formal paperwork and dealing with a records management bureaucracy, work that generally would fall on top of existing job responsibilities. In addition, prosecutors and investigating officers may be hesitant to contradict information contained in the original PIR as those filings could be used by defense counsel to undermine the prosecution's case, perhaps by suggesting that law enforcement officers may have misidentified the suspect.

If a felony defendant is convicted and sentenced to incarceration in prison, all states require an intake process to assess security risk and make housing assignments. This intake process will generally leverage some combination of the inmate's criminal history, the court documentation associated with their conviction, and a risk assessment form completed after a case manager interviews the inmate. Some court documentation may propagate into correctional records systems and thereby continue the use of any misreporting in the PIR. For example, most states that we engaged with use the Ohio Risk Assessment System (Latessa et al. 2009), which does not explicitly ask inmates about their race and ethnicity. But information on known or suspected gang affiliation, which often depends on a specific racial or ethnic identity, is taken into account by the case manager and added to the prisoner's record (York 2021). Because of the security risk associated with gang conflict, greater effort is made to accurately record and verify racial and ethnic information about inmates, since this is operationally relevant to improve safety within facilities. And because

departments of corrections typically maintain their own independent data management systems, updating flawed information is significantly less burdensome on staff and does not require formal court filings.

Based on our interviews, it appears that the information recorded at the time of arrest carries significant weight in the official record. Few incentives exist to correct the official record within the adjudication phase, and this potentially flawed information may get passed onto corrections data systems. Due to the operational need and opportunity to update data management systems through the intake process at the start of an incarceration spell, race and ethnicity data within the prison context likely is of higher quality than within the arrest and adjudication phases.

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

3.1 Data and methods

We identify people who have been involved in the justice system using data from the Criminal Justice Administrative Records System (CJARS), a repository of criminal history information covering all stages of the criminal justice system, including both adjudication events and incarceration spells (Finlay, Mueller-Smith, and Papp 2022).[7] CJARS is constructed from criminal history data collected from state and local agencies, then harmonized into a unified national repository. In addition to harmonizing criminal justice event data, CJARS also collects agency-recorded race and ethnicity (Finlay and Mueller-Smith 2022). We use these administrative records to assign the most common agency-recorded race and ethnicity measures to justice-involved individuals. Importantly, we allow agency-recorded measures of race and ethnicity to vary across the stages of the justice system (e.g., adjudication, incarceration) to capture dynamics in the accuracy and persistence of information as individuals are processed through the justice system.

From the universe of available CJARS data, we focus on two common events in the criminal

---

[7]States for which CJARS has longitudinal, statewide coverage in at least one criminal justice domain constitute 66 percent of the U.S. population.

justice system—adjudication and incarceration events. For adjudication, we focus on defendants who were charged with misdemeanor or felony offenses in a given calendar year. Court events in the CJARS data come from Arizona, Florida, Maryland, Michigan, North Carolina, Oregon, Pennsylvania, Texas, Virginia, Wisconsin. For incarceration, we consider three measures of the incarcerated population: entry cohorts, exit cohorts, and the population of incarcerated individuals. We define an entry cohort as the population of individuals who began an incarceration spell in a given calendar year, exit cohorts as the population of individuals whose incarceration spell ends in a given calendar year, and the population as the incarcerated population as of year-end. Incarceration events in the CJARS data come from Arizona, Florida, Illinois, Indiana, Kansas, Michigan, Minnesota, Montana, Nebraska, New York, North Carolina, Oklahoma, Pennsylvania, Texas.

To construct a Census Bureau composite race/ethnicity, we use data from the 2000 and 2010 Censuses and from the Census Numident. In the 2000 and 2010 Census data, we exclude any race or ethnicity responses that were imputed or edited, and any responses from group quarters such as prisons or jails, since those may have been based on race or ethnicity measures recorded by justice personnel. The remaining race and ethnicity responses will primarily reflect self or family reports. The Census Numident is based on the Social Security Administration (SSA) Numident and serves as the backbone of record linkage at the Census Bureau. Race and ethnicity data in the Census Numident will also primarily reflect self or family reports.[8]

We link CJARS records at the person level with records from the 2000 and 2010 Censuses and the Census Numident using Protected Identification Keys (PIK), which are anonymous person identifiers assigned by the Census Bureau using the Person Identification Validation System (PVS; Wagner and Lane 2014). CJARS records without PIKs and CJARS records that could not be linked to at least one Census Bureau race/ethnicity measure were excluded from the analysis.[9]

---

[8] The Census Numident contains SSA-collected race and ethnicity data (Scott 1999). Prior to 1980, White, Negro, and Other were the only options. In 1980, SSA began allowing White, Black, Hispanic, Asian or Pacific Islander, and American Indian and Alaska Native as options in a combined race/ethnicity field. Beginning in 1989, SSA began receiving SS-5 applications for Social Security numbers directly from hospitals at birth, but these filings often exclude race/ethnicity because it can be considered protected health data. For people born after 1989, most of these records would received race/ethnicity data when the SSA records was updated later in life.

[9]The reference files used for PVS only include people with Social Security numbers and Individual Tax Identification Numbers. Many undocumented immigrants, who are disproportionately Hispanic, are therefore not able to be

Using data from these three sources, we can construct a composite race/ethnicity that gets closer to an individual's own perception of their race and ethnicity, while covering as much of the justice-involved population as possible. Although the Census Numident data is administrative, like the census data, it will have mostly been submitted by an individual or a member of an individual's family or household. The Census Bureau composite will be much closer to the best practice in race/ethnicity measurement than those from justice sources. Future research could explore more of the survey and administrative sources of race and ethnicity available at the Census Bureau to understand how they vary for the justice-involved population.

Because of the limitations in how race and ethnicity are recorded in justice records management systems, we use the following combined race/ethnicity categories to assess the accuracy of justice-assigned labels: Hispanic, any race; White, non-Hispanic (NH); Black, NH; Asian/Pacific Islander, NH; American Indian and Alaska Native (AIAN), NH; and Some Other Race or Two or More Races/Multiracial, NH.[10] These categories, shown in detail in Table 3, were chosen because they are mutually exclusive and align with the BJS statistics we study below.

For each justice-involved person in the CJARS data, the Census Bureau race and ethnicity sources are recoded into a single race/ethnicity. If there is a modal race/ethnicity across the three sources, that category is chosen. If there is not a single modal race/ethnicity, we use a race/ethnicity ordering that prioritizes Hispanic ethnicity then less populous racial groups to resolve disagreements across the three sources. Hispanic ethnicity is selected first; then American Indian and Alaska Native, NH; then Some Other Race and Two or More Races/Multiracial, NH; then Asian/Pacific Islander, NH; then Black, NH; then White, NH. This algorithm is analogous to those used in Ennis et al. (2018) and Limburg, Kurczewski, and Udalova (2023).

Table 4 documents how the resulting Census Bureau race and ethnicity composites agree with the component data sources. Most composite values are constructed from at least one census response. For felony and misdemeanor defendants, 79 to 89 percent have one or more Decennial

---

linked and this could cause an undercount of the true share of the Hispanic population.

[10]Unless otherwise noted, we use Hispanic, White, Black, Asian/PI, AIAN, and Other interchangeably with these labels.

responses. Prison inmates have somewhat lower linkage rates (58 to 79 percent), which is expected since we exclude group quarters responses. When we observe at least one census response, there is substantial concordance between the census responses and the composite race/ethnicity. At least one of the census responses aligns with the composite in 77 to 100 percent of records. In the minority of cases where we do not have a link to a census response, we rely on the Census Numident. While we cannot validate the information in cases where we only have links to the Numident, among records with links to at least one census response and the Numident, we see very high rates of agreement for Whites and Blacks (>98 percent), with lower rates for Hispanics, Asian/PIs, and AIANs (60–83 percent).[11]

With a large set of agency-recorded race/ethnicity measures matched to high-quality race/ethnicity composites from Census Bureau data, we construct criminal justice event-by-year empirical correspondence functions that map agency-recorded to Census Bureau composite race and ethnicity over the period 2000–2019, with events defined as above (e.g., individuals with felony charges, incarcerated population). We average the correspondence tables into a single matrix by event type by taking the unweighted average across all years.

3.2 Concordance of agency-recorded race/ethnicity versus Census Bureau race/ethnicity composites

We consider two perspectives when trying to evaluate the quality of race and ethnicity information in criminal justice administrative records. First, we measure the percent of agency-recorded labels that are consistent with the Census Bureau race/ethnicity composites. Second, we identify the percent of Census Bureau race/ethnicity composites that concord with the racial/ethnic categories assigned by criminal justice agencies. These statistics respectively correspond to the concepts of *precision* and *recall* typically used to evaluate the performance of classifiers. While closely related, the two approaches are differentiated by who appears in the denominators—subgroups defined

---

[11]Numident agreement rates for the Other race category are low because this was a default category for older Numident records where White and Black where the only options and then was removed as an option for newer records (Scott 1999).

by Census Bureau or agency-recorded race and ethnicity. For example, precision and recall in assessing the identification of Whites would be defined as:

$$\text{precision} = \frac{\text{White in Census Bureau composite and agency data}}{\text{White in agency data}}, \text{ and}$$

$$\text{recall} = \frac{\text{White in Census Bureau composite and agency data}}{\text{White in Census Bureau composite}}.$$

Figure 2 presents our findings. Panel A shows that the majority of agency-recorded labels are consistent with the Census Bureau composites, meaning they generally have high precision. For example, 80–87 percent of justice-involved individuals recorded by agencies as White we also identify as White in the composite. Black and Hispanic labels fare even better with an even higher share (>92 percent, >94 percent) agreeing with the composite race/ethnicity. Two populations show relatively lower rates of precision: Asians/Pacific Islanders (61–78 percent) and American Indians and Alaska Natives (76–81 percent).

Panel B shows divergent results when examining the share of Census Bureau race/ethnicity composites that are accurately identified by agencies, a measurement concept similar to recall. White and Black populations have very high concordance, meaning that the vast majority (>94 percent, >96 percent) of those identified as White or Black are recorded as such by the justice system. Asian/Pacific Islanders, Hispanics, and AIANs, however, show substantially lower concordance. For example, whether looking at court records or prison records, slightly over one quarter of those who self-identify as Asian/Pacific Islander are recorded as such by agencies. Hispanic individuals are rarely recorded as Hispanic by criminal court systems (15–16 percent), but have somewhat better rates within the prison context (67 percent). AIANs show a similar dynamic as Hispanics, albeit with better performance in courts (38 percent) and worse performance in prison data (50 percent).

One conclusion from these exercises is that the justice system tends to default to labeling individuals as White or Black, the largest justice-involved racial groups. As has long been recognized by the Bureau of Justice Statistics, many agencies vary in whether information on Hispanic iden-

tity is recorded in electronic records (e.g., Carson 2021, p. 42). We also see evidence that race and ethnicity records from prison systems tend to be higher quality than those in court data. This is consistent with what we know about how some departments of correction rely on race and ethnicity information for operational purposes, which provides an incentive for higher data standards for these fields.

This latter point regarding varying quality across procedural stages of the justice system is also shown in the full empirical correspondences shown in Tables 5 and 6. These tables capture the complete set of transition probabilities for each agency recorded and Census Bureau composite pair, first for court defendants then for prison caseloads. Overall, the concordance rate of the Census Bureau composites with agency-recorded race and ethnicity is only 83 percent in adjudication records, which increases to 90 percent in prison records. This means that 17 percent and 10 percent of these respective caseloads have some disagreement between agency-recorded and Census composite labels.

It is also interesting to consider the most common types of misreports made in these systems. For example, the most common misalignment we observe in both court and prison systems is that individuals who we identify as Hispanic in the Census Bureau composites are recorded as White in administrative records. The next most common off-diagonal pairing are individuals identified as Hispanic in the composites but labeled as Black by justice agencies. These patterns are not surprising given that some agencies do not systematically collect information about ethnicity. Three additional racial groups, AIAN, Asian/Pacific Islander, and Other, which includes Multiracial identification, also show systematic sorting to White and Black agency-recorded racial codes. While these groups constitute a smaller share of the justice-involved population overall, the misclassifications are significant enough that the majority of agency records do not reflect how individuals are identified in Census Bureau data.

The second to last rows and columns from Tables 5 and 6 show the implied racial and ethnic composition of the justice system based on agency-recorded versus Census Bureau composite identity. We observe that 38 percent of inmates are identified as White, 40 percent as Black,

18 percent as Hispanic, 2 percent as American Indian and Alaska Native, and less than 1 percent identify as Asian/Pacific Islander or Other. The caseload demographics for misdemeanor and felony defendants are similar to those for prison inmates but with relatively more White defendants (48 percent) and a smaller share of Black (35 percent) and Hispanic (13 percent) defendants. While interpreting this exercise comes with the caveat that the estimates are limited to only jurisdictions contained within CJARS, Finlay, Mueller-Smith, and Papp (2022) show that CJARS coverage is generally representative of the U.S. as a whole on a wide range of characteristics.

Delving further, we can disaggregate our measures of race and ethnic data quality by more detailed types of interaction with the justice system, explore how they change over time, and measure how they vary across states. Figures 3 and 4 show the annual rates of precision and recall between 2000 and 2019 for felony and misdemeanor defendants as well as incarceration entries, exits, and populations. Overall, we generally see a fair degree of stability in these quality measures over time. There is a notable increase between 2000 and 2011 in the percent of agency-recorded Asian/Pacific Islander records for misdemeanor defendants that concord with the composite race/ethnicity, but other measures tend to remain relatively constant. Since incarceration entries and exits are constructed from smaller populations, they do show higher variance year-to-year even if the running average roughly matches the more stable incarcerated population measures. Focusing on the share of Census Bureau composites accurately labeled by agencies, there do appear to be signs of improvements in data quality over time. For example, an increasing share of Hispanic inmates are labeled as such by prison systems up through 2012; unfortunately data quality on this dimension returns to the lower values from the early 2000s by the end of our statistical series in 2019.

Across states, justice agencies vary in how they code race and ethnicity. Figures 5 and 6 show the geographic variation in the share of criminal defendants and prison inmates identified in the Census Bureau composites in the same way as agency-recorded labels by race and ethnicity. Nationwide, the concordance for White and Black justice-involved populations is consistently high. With the exception of Maryland court records where only 56 percent of defendants who are identified in the composite as White are recorded as White, agreement rates often are in excess of 97

percent. We find somewhat lower recall for: Black defendants in Arizona (95 percent), Oregon (90 percent), and Virginia (94 percent); White defendants in Virginia (95 percent); Black inmates in Montana (92 percent) and New York (94 percent); and White inmates in Arizona (95 percent) and New York (92 percent).

In contrast, few jurisdictions record race and ethnicity accurately for people who we identify in Census Bureau records as American Indian and Alaska Native, Asian, Pacific Islander, or Hispanic. For AIAN and Hispanic populations, it does appear that jurisdictions with larger resident populations from these groups tend to fare better. The label agreement of AIAN justice-involved individuals is highest in Montana, Minnesota, Arizona, and Nebraska. Likewise the coverage of Hispanic inmates is highest in Arizona, Illinois, and Texas. That said, some jurisdictions break this pattern. A minority of Hispanic defendants and inmates in Florida, for example, are recorded as such by the criminal justice agencies they interact with in spite of Florida's large Hispanic population. Across the map, there is low data quality for Asian/Pacific Islander populations, perhaps related to this population's infrequent contact with the justice system in general.

3.3   Evaluating alternative hypotheses related to record linkage errors

These exercises rely on the assumption that the Census Bureau composite race/ethnicity linked with agency-recorded race/ethnicity from administrative records correspond to the same underlying person. One reason that discordance could arise between the Census Bureau composite and agency-recorded race and ethnicity is if this assumption regarding accurate record linkage fails, which may occur since the data integration relies on probabilistic record linkage. This would be even more problematic if linkage errors disproportionately affected specific minority populations.

To explore how much bias might be introduced due to record linkage errors, we explore two exercises. First, we examine the degree of concordance between agency-recorded sex and sex as recorded in the Census Numident. As discussed above, sex information in criminal justice administrative data is viewed by records administrators as high quality. Consequently, high levels of sex discordance would be warning sign for potential flaws in the record linkage procedure

18

that may result in an overestimate of race/ethnicity misreporting. For this exercise, we use an analogous algorithm to select a Census Bureau composite sex from the 2000 and 2010 Censuses and the Census Numident.

Figure 7 reports the sex concordance results. We find high agreement rates between agency-recorded and the composite sex. Over 99 percent of individuals identified as male are recorded as such by agencies. Similarly, over 98 percent of individuals identified as female are recorded that way by agencies. Sex is used as an input in the Census Bureau PVS record linkage process, whereas race and ethnicity are not, so this exercise may overstate the degree of concordance between the sex measures. Given that some minor degree of discordance may be attributable to measurement error in the Census Bureau data, differences in measurement concepts, and differences in reporting over time, these results provide evidence that record linkage is working as intended.[12]

The second exercise considers whether linked Medicaid administrative records exhibit the same degree of racial and ethnic discordance observed in criminal justice administrative records. Three factors make this a particularly interesting comparison exercise. First, the Medicaid data are compiled and linked from state agency records in a process similar to that used to harmonize the CJARS data. Second, Medicaid records should capture self-reported or family-reported race and ethnicity submitted by applicants on their enrollment paperwork. Finally, the income-targeted nature of Medicaid means the data should draw from a population more closely resembling the racial and ethnic distribution of justice-involved individuals. There are documented limitations in the quality of race and ethnicity data in Medicaid records, which might generate discordance with Census Bureau survey records in ways distinct from divergences between the composite race/ethnicity and ones based on agency perception (Limburg, Kurczewski, and Udalova 2023). Consequently, to the extent that Medicaid data show stronger concordance relative to the criminal justice data, this would be a lower bound on the relative impact of agency-perception.

Figure 8 shows the findings from the exercise. For the overall population as well as specific

---

[12]Concordance of race and ethnicity measures may vary by sex since factors like surname change at marriage may disproportionately impact women and may influence perceptions of race and ethnicity. Understanding the intersection of these characteristics should be an important area for future research.

racial/ethnic subgroup, we show the share of Census Bureau composite labels that agree with the agency label. This is broken out for four groups: felony and misdemeanor defendants, prison inmates, justice-involved Medicaid recipients (either in court or prison records), and all Medicaid recipients. The first two columns reproduce results from above as a point of comparison. Overall, the Medicaid data, whether looking at justice-involved Medicaid recipients or all Medicaid recipients, have higher concordance rates (91–92 percent) than those observed in the CJARS data (83–90 percent), especially relative to the court records. Aligning with prior findings, the improvements come from better coverage of Asian, Pacific Islander, Hispanic, and American Indian and Alaska Native populations.

These two exercises show that the degree of discordance between agency-recorded race/ethnicity and the Census Bureau race/ethnicity composite is uniquely poor. Sex, as measured by justice agencies, compares well with sex measured by Census Bureau sources, and race/ethnicity is measured more accurately in Medicaid data than justice data, even for the Medicaid-enrolled population that is justice involved.

4   Using Census Bureau composite race and ethnicity microdata to adjust federal statistics on incarceration

Since 1980, the Bureau of Justice Statistics (BJS) has published regular reports about the size and composition of state and federal prisoner populations in the United States. To do this, they rely on data from the National Prisoner Statistics (NPS) program, which solicits aggregate information on year-end prison counts from state and federal corrections agencies. Jurisdictions report back data based on information available in their records management systems.

Because not all NPS data providers record information in a manner consistent with OMB guidelines, BJS relies on supplementary survey data as a benchmark for the racial and ethnic composition of the prisoner population in the United States. For state records prior to 2016, BJS adjusted annual prisoner counts to match the distribution of self-reported race and ethnicity from the 2004

20

Survey of Inmates in State and Federal Correctional Facilities (SISFCF). Starting in 2016, BJS also incorporated equivalent data from the 2016 Survey of Prison Inmates (SPI) to account for potential changes in the demographic profile of inmates. Prisoner counts for prior reporting years over the preceding decade were re-released using a new set of weights that smoothed the transition between the 2004 and 2016 survey distributions. For federal prisoners, a similar process was applied using corresponding federal survey responses. Table 7 shows the difference between the 2004 SISFCF and 2016 SPI race distributions, and Figure 9 shows how the weight adjustment affected the distribution of BJS-reported imprisonment per capita by race/ethnicity subgroup.

The race/ethnicity reweighting strategy employed by BJS to meet OMB guidelines creates a dependence on the accuracy of costly and therefore infrequently administered survey instruments. For example, the re-reported 2015 prisoner estimates following the incorporation of 2016 SPI data into the reweighting methodology reduced the number of White inmates by 48,800 (9.8 percent), reduced Black inmates by 26,600 (5.1 percent), and increased Hispanic inmates by 16,400 (5.1 percent).[13] The net decline in total inmates falling into the three largest racial/ethnic categories is attributable to a high rate of respondents in the 2016 SPI reporting Two or More Races, a category for which BJS does not report prison population counts because almost no department of corrections has data to support this category. From the 2004 to 2016 survey instruments, the percent of state inmates self-reporting as Multiracial grew by 242 percent from 3.3 percent to 11.3 percent.[14]

The growth in multiracial identification is consistent with broader demographic trends in the United States. That said, it is also possible that the differences observed between these two survey waves is a byproduct of the surveys' stratified two-stage sample design, wherein a set of prison facilities are first sampled for inclusion in the survey (301 [364] out of 1,801 [2,001] unique facilities in the 2004 SISFCF [2016 SPI]) and then inmates within just that subset of facilities are

---

[13] Authors' calculations based on data underlying Figure 9. In 2015, BJS reported 2015 prisoner counts for Whites, Blacks, and Hispanics as 499,400, 523,000, and 319,400. In 2022, the 2015 prisoner counts, which incorporated the 2016 SPI data, were reported as 450,600, 496,400, and 335,800.

[14] The net reduction of prison inmates reported to be in one of the three largest racial and ethnic subgroups following the incorporation of the 2016 SPI is not the consequence of retroactive revisions to the number of inmates reported to BJS in the NPS; Panels D and E of Figure 9 show that reported imprisonment rates for men and women line up perfectly across reporting years, regardless of how racial and ethnic reweighting was conducted.

sampled for the survey.[15] Clustered sampling strategies, while cost-effective, are known to generate higher variance in their estimates. Additionally, inmate non-response (10 percent [30 percent] in the 2004 SISFCF [2016 SPI]) may limit the representativeness of the survey estimates to the prisoner population as a whole.

With the luxury of person-level data linkages, we can take an alternative approach to reweight prison populations. We leverage the annual NPS reports by state and federal agencies of their raw, unadjusted counts by agency-recorded race/ethnicity. Those raw counts can be reweighted by our estimated year-specific concordance matrices for incarcerated populations to generate estimates that adjust for the racial and ethnic distribution of the prisoner population based on the Census Bureau race/ethnicity composite. This permits us to reallocate subpopulations within agency-recorded bins to different Census Bureau race/ethnicity composites. For example, we observe that 8 percent (4 percent) of agency-recorded White (Black) inmates in 2019 are identified in the composite measure as Hispanic. An equivalent strategy could be used with the 2004 SISFCF and 2016 SPI if those survey responses were linked to raw agency records, but the CJARS data provide a much larger sample to estimate the reweighting matrix.

There are strengths and weaknesses of this approach. On the positive side, we are able to incorporate the original NPS-reported aggregate counts by agency-recorded race and ethnicity for the entire United States. While imperfectly suited to meet OMB guidelines, this ground-truth agency data is high quality annual information about the racial and ethnic distribution of the prisoner population. Applying a concordance matrix to these counts translates agency-recorded labels to labels closer in principle to something self reported. In the context of the population of prisoners, fewer than 10 percent of inmates on average between 2000 and 2019 are relabeled using this method. We estimate concordance matrices for these populations on an annual basis so they can adjust as the relationship between agency-record race and ethnicity and self-identification evolves over

---

[15]In practice, only a subset of facilities are sampled at random and randomly selected facilities are not selected with equal probability. The facilities with the largest inmate populations were "sampled" with certainty thereby automatically entering the sample, while the remaining were stratified by census region and given a sampling probability proportional to the size of their inmate population. As a result, selected prisons are larger than non-selected prisons and selected inmates may not be nationally representative of the overall prisoner population.

time, whether due to improvements in data quality, changes in records policies and procedures, or changes in how justice-involved people perceive their own races and ethnicities.[16]

The drawback is that CJARS does not include data from all state prisons. As a result, we have to assume that the estimated concordance matrices generalize to all states. While important, we view this as an assumption that is secondary in nature since the majority of information regarding the racial and ethnic composition of the prisoner population will come from the NPS counts, which do not face the same coverage limitations as the linked CJARS microdata.

Figure 10 shows the results of reweighting using the CJARS/Census Bureau correspondence matrices. Results are shown separately by the five race/ethnicity groups included in BJS reporting as well as the non-reported residual category: White, Black, AIAN, Asian/Pacific Islander, Hispanic, and Other. Each panel includes the raw unadjusted NPS imprisonment rate per 100,000 adults (dot-dash line), the adjusted BJS-reported imprisonment rate using the 2004 SISFCF and 2016 SPI (dash line), and the CJARS-Census Bureau matrix-adjusted imprisonment rate (solid line). To construct the BJS-adjusted rate for the Other series, we take the total overall count of prisoners reported by BJS and subtracted the reported counts from the five major racial and ethnic groups.

Several important findings emerge. The CJARS/Census Bureau-adjusted imprisonment rates almost always exceed the BJS-adjusted statistical series. The two exceptions are the Hispanic imprisonment rate, where the two series are very close to each other, and the Other imprisonment rate, where the implied BJS-adjusted rate is significantly higher than the BJS unadjusted rate or the CJARS/Census Bureau-adjusted rate. This pattern arises because a disproportionate share of 2004 SISFCF and 2016 SPI respondents identify as multiracial compared to the agency-record race codes in the NPS or the Census Bureau composite labels linked to CJARS data.[17] These gaps become more pronounced over time between 2005 and 2016, especially for White and Black pris-

---

[16]Because of lags in data production, we apply the 2019 correspondence matrix to the 2020 and 2021 NPS counts to get CJARS-adjusted estimates in those years.

[17]As can be seen in Table 6, Column 8, only 0.39% of prisoners in the CJARS data between 2000 and 2019 are identified as Other/Multiracial in the Census Bureau race/ethnicity composite. This is 88 percent and 97 percent smaller than the shares reported as Multiracial in the state prison estimates from the 2004 SISFCF (3.3 percent) and the 2016 SPI (11.3 percent).

oners, as the BJS reweighting strategy shifts towards placing more weight on the 2016 SPI implied race and ethnicity distributions. For 2018, CJARS/Census Bureau-adjusted estimates place the imprisonment rate of White and Blacks 20 percent and 17 percent higher, respectively, compared to BJS-reported rates. In levels, this corresponds to approximately 86,000 previously unreported additional White inmates and 80,000 previously unreported additional Black inmates in 2018.[18]

If the BJS adjustments are accurate, it would imply that the imprisonment rate for Other/Multiracial individuals is the highest of any demographic group in the U.S., more than 200 percent higher than for Blacks. In contrast, the BJS unadjusted and CJARS/Census Bureau-adjusted rates for the Other population place them somewhere between rates for Hispanics and Whites. Table 8 uses data from the American Community Survey (ACS) between 2005 and 2019 to show a variety of socioeconomic characteristics in the overall population for these racial and ethnic groups, including educational attainment, employment and earnings, and household structure.[19] On each of these measures, the Other adults appear much more similar to White and Hispanic adults, suggesting the CJARS-adjusted statistical series are more consistent with other observed population differences across racial and ethnic groups. Stated differently, it is not the case that Other adults are more economically disadvantaged than Black adults on any measure, as should be the case if the imprisonment rate in this population was more than 200 percent higher than Black adults. The treatment of the Other/Multiracial category is a potential shortcoming in the BJS imputation strategy that could be improved with the kind of record linkage used in this paper.

Another finding from Figure 10 is that the CJARS/Census Bureau-adjusted imprisonment rate for the AIAN population is on average 45 percent larger than the BJS-adjusted imprisonment rate since 2010. In 2015, the CJARS/Census Bureau-adjusted series is 62 percent larger than the BJS-reported imprisonment rate. The increase reflected in the CJARS/Census Bureau-adjusted series is so substantial that by the latter half of the 2010s, the imprisonment rates of Blacks and AIANs

---

[18]Authors' calculations based on data underlying Figure 10. In 2018, the adjusted BJS prisoner counts for Whites and Blacks were 430,500 and 461,500; the equivalent prisoner counts using the CJARS/Census Bureau-adjustment were 516,381 and 541,163.

[19]Using publicly available 2005–2019 5-year ACS IPUMS files (Ruggles et al. 2024), we restrict to a sample of respondents aged between 18 and 65 years and stratify by the mutually exclusive race/ethnicity categories used throughout this paper. For each group, we calculate person-weighted means for the characteristics listed in Table 8.

are approximately equivalent. This potential major revision to our understanding of the prevalence of prison incarceration within this group reflects the challenge of accurately capturing information about less populous racial and ethnic groups with small samples from surveys like the SISFCF and the SPI.

5 Conclusion

In this paper, we consider the challenge of trying to analyze race and ethnicity information in the context of the U.S. criminal justice system, where race and ethnicity date are often assigned by law enforcement officers and other justice officials. Operational priorities, limited resources, and other bureaucratic incentives limit the improvement of race and ethnicity records even when misreports have been identified.

We link agency-recorded race and ethnicity data from CJARS at the person level to a composite race and ethnicity measure constructed from 2000 and 2010 Census and Census Numident data. By being composed primarily of self- or family-reported data, the Census Bureau race/ethnicity composite better reflects best practices in race and ethnicity measurement. These matched data allow us to quantify both the degree of discordance between the justice agency-reported and Census Bureau composite sources and the types of misreports that occur in practice. Overall, we find that 17 percent of felony and misdemeanor defendants and 10 percent of prison inmates have a Census Bureau composite race/ethnicity that differs from their agency-recorded label. The most common types of discordance are where individuals are identified as Hispanic in the Census Bureau composite but are labeled by agencies as White or Black (and not Hispanic). While agency labels typically concord for the vast majority of Whites and Blacks in their caseloads, agencies are much less successful at identifying American Indians and Alaska Natives, Asians, Pacific Islanders, and Hispanics.

We apply the estimated empirical correspondences between agency-recorded and Census Bureau composite race and ethnicity to the raw, unadjusted annual National Prisoner Statistics data to re-estimate imprisonment rates by race and ethnicity. Because of concerns about the quality of

race and ethnicity information reported in the NPS and its adherence to OMB guidelines, BJS's current practice is to reweight the racial and ethnic distribution of imprisonment using two national prisoner surveys from 2004 and 2016. We find large discrepancies between the BJS approach and our estimates that incorporate the original NPS data with our estimated annual correspondence matrices. For example, we find that there were approximately 85,000 additional White inmates and 80,000 Black inmates in 2018 in 2018 than BJS reported and that the imprisonment rate of AIANs was 46 percent higher on average since 2010.

These results have implications for Census Bureau collections and the downstream uses of those data to study the justice system. As decennial census and ACS group quarters responses are increasingly collected in bulk from administrators, census and ACS race and ethnicity data will more greatly reflect the limitations of justice data systems. Race/ethnicity data collected from jails are likely to be of particularly poor quality, since those data are often collocated with court data, which we find misidentify people who are not White or Black. The 2024 SPD15 revision requires federal agencies to transparently describe how race and ethnicity data have been collected (Office of Management and Budget 2024). Working with data-providing justice agencies to better document their own collection processes would allow the Census Bureau to further inform the public about any limitations of resulting census or survey data. This could be part of a broader research agenda that measures how justice data can be mapped to the collection of race and ethnicity in a single question that allows multiple responses, and for the addition of Middle Eastern or North African as a minimum category—changes required under the 2024 SPD15 revision (Office of Management and Budget 2024).

Research on crime and justice in the U.S. increasingly relies on administrative data with agency-recorded race and ethnicity. The race/ethnicity misreporting we measure raises questions about research that relies on agency data without further linkage or adjustments. Systematic mismeasurement may mean that existing estimates on racial and ethnic disparities in criminal justice outcomes are currently underestimated. This paper shows the value of data infrastructure efforts like CJARS that support the linkage of individual-level microdata across a range of sources. We demonstrate

how leveraging the full depth of available information in integrated data environments like the Census Bureau's Data Linkage Infrastructure can strengthen our knowledge about—and meaningfully improve the monitoring of—the criminal justice system in the United States.

References

Agan, Amanda, and Sonja Starr. 2018. Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment. *Quarterly Journal of Economics* 133 (1): 191–235. https://doi.org/10.1093/qje/qjx028.

Alesina, Alberto, and Eliana La Ferrara. 2014. A Test of Racial Bias in Capital Sentencing. *American Economic Review* 104 (11): 3397–433. https://doi.org/10.1257/aer.104.11.3397.

American Civil Liberties Union and The Sentencing Project. 2022. Racial Disparities in Sentencing in the United States. Shadow Report to the 107th Session of the UN Committee on the Elimination of Racial Discrimination. https://www.sentencingproject.org/app/uploads/2022/10/07-14-2022_CERD-Shadow-Report-Draft_with-endnotes.pdf.

Bayer, Patrick, and Kerwin Kofi Charles. 2018. Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940. *Quarterly Journal of Economics* 133 (3): 1459–501. https://doi.org/10.1093/qje/qjy003.

Beatty, Lauren G., and Tracy L. Snell. 2021. Profile of Prison Inmates, 2016. Bureau of Justice Statistics, NCJ 255037. https://bjs.ojp.gov/content/pub/pdf/ppi16.pdf.

Belton, Judy, and Dora Durante. 2021. 2020 Decennial Census Group Quarters Data Collection Operation. National Academies of Science. https://www.nationalacademies.org/documents/embed/link/LF2255DA3DD1C41C0A42D3BEF0989ACAECE3053A6A9B/file/D9C327E9F077C55EEBCB276E438BF7148649C4B78F4B.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review* 94 (4): 991–1013. https://doi.org/10.1257/0002828042002561.

Besemer, Sytske, David P. Farrington, and Catrien C. J. H. Bijleveld. 2017. Labeling and Intergenerational Transmission of Crime: The Interaction between Criminal Justice Intervention and a Convicted Parent. *PLOS ONE* 12 (3): e0172419. https://doi.org/10.1371/journal.pone.0172419.

Blair, Irene V., Charles M. Judd, Melody S. Sadler, and Christopher Jenkins. 2002. The Role of Afrocentric Features in Person Perception: Judging by Features and Categories. *Journal of Personality and Social Psychology* 83:5–25. https://doi.org/10.1037/0022-3514.83.1.5.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. Stereotypes. *Quarterly Journal of Economics* 131 (4): 1753–94. https://doi.org/10.1093/qje/qjw029.

Bronson, Jennifer, and E. Ann Carson. 2019. Prisoners in 2017. Bureau of Justice Statistics, NCJ 252156. https://bjs.ojp.gov/library/publications/prisoners-2017.

Bureau of Justice Statistics. 2019a. Survey of Inmates in State and Federal Correctional Facilities, 2004. Inter-university Consortium for Political and Social Research [distributor], 2019-12-12. https://doi.org/10.3886/ICPSR04572.v6.

———. 2019b. Survey of Inmates in State and Federal Correctional Facilities, 2004: Codebook Part 1—Federal Data [Public]. Inter-university Consortium for Political and Social Research [distributor], 2019-12-12. https://doi.org/10.3886/ICPSR04572.v6.

———. 2023. 2016 SPI Data Analysis Tool (SPI DAT). Data generated on 17 June 2024. https://spi-data.bjs.ojp.gov.

Carson, E. Ann. 2016. Prisoners in 2015. Bureau of Justice Statistics, NCJ 250229. https://bjs.ojp.gov/library/publications/prisoners-2015.

———. 2018. Prisoners in 2016. Bureau of Justice Statistics, NCJ 251149. https://bjs.ojp.gov/library/publications/prisoners-2016.

Carson, E. Ann. 2020a. Prisoners in 2018. Bureau of Justice Statistics, NCJ 253516. `https://bjs.ojp.gov/library/publications/prisoners-2018`.

———. 2020b. Prisoners in 2019. Bureau of Justice Statistics, NCJ 255115. `https://bjs.ojp.gov/library/publications/prisoners-2019`.

———. 2021. Prisoners in 2020—Statistical Tables. Bureau of Justice Statistics, NCJ 302776. `https://bjs.ojp.gov/library/publications/prisoners-2020-statistical-tables`.

———. 2022. Prisoners in 2021. Bureau of Justice Statistics, NCJ 305125. `https://bjs.ojp.gov/library/publications/prisoners-2021-statistical-tables`.

Carson, E. Ann, and Rich Kluckow. 2023. Prisoners in 2022. Bureau of Justice Statistics, NCJ 307149. `https://bjs.ojp.gov/library/publications/prisoners-2022-statistical-tables`.

Charles, Kerwin Kofi, and Jonathan Guryan. 2011. Studying Discrimination: Fundamental Challenges and Recent Progress. *Annual Review of Economics* 3 (1): 479–511. `https://doi.org/10.1146/annurev.economics.102308.124448`.

Collister, Brian. 2016. Finding Errors in Texas Data. Investigative Reporters & Editors, Accessed 16 February 2024. `https://www.ire.org/finding-errors-in-texas-data`.

Dallas County District Attorney's Office. 2024. The Criminal Justice Process: What Happens after an Arrest? Accessed on 2/17/2024. `https://www.dallascounty.org/government/district-attorney/cj-process.php`.

Darity, William A., Patrick L. Mason, and James B. Stewart. 2006. The Economics of Identity: The Origin and Persistence of Racial Identity Norms. *Journal of Economic Behavior & Organization* 60 (3): 283–305. `https://doi.org/10.1016/j.jebo.2004.09.005`.

Davenport, Lauren. 2020. The Fluidity of Racial Classifications. *Annual Review of Political Science* 23 (1): 221–40. `https://doi.org/10.1146/annurev-polisci-060418-042801`.

Decker, Scott, Cassia Spohn, Natalie Ortiz, and Eric Hedberg. 2014. Criminal Stigma, Race, Gender and Employment: An Expanded Assessment of the Consequences of Imprisonment for Employment. National Institute of Justice, NCJ 244756. https://nij.ojp.gov/library/publications/criminal-stigma-race-gender-and-employment-expanded-assessment-consequences.

Dixon, Angela R., and Edward E. Telles. 2017. Skin Color and Colorism: Global Research, Concepts, and Measurement. *Annual Review of Sociology* 43 (1): 405–24. https://doi.org/10.1146/annurev-soc-060116-053315.

Doleac, Jennifer L., and Benjamin Hansen. 2020. The Unintended Consequences of "Ban the Box": Statistical Discrimination and Employment Outcomes when Criminal Histories Are Hidden. *Journal of Labor Economics* 38 (2): 321–74. https://doi.org/10.1086/705880.

Doyle, Jamie Mihoko, and Grace Kao. 2007. Are Racial Identities of Multiracials Stable? Changing Self-Identification among Single and Multiple Race Individuals. *Social Psychology Quarterly* 70 (4): 405–23. https://doi.org/10.1177/019027250707000409.

Eberhardt, Jennifer L., Phillip Atiba Goff, Valerie J. Purdie, and Paul G. Davies. 2004. Seeing Black: Race, Crime, and Visual Processing. *Journal of Personality and Social Psychology* 87 (6): 876–93. https://doi.org/10.1037/0022-3514.87.6.876.

Ennis, Sharon R., Sonya R. Porter, James M. Noon, and Ellen Zapata. 2018. When Race and Hispanic Origin Reporting are Discrepant across Administrative Records and Third Party Sources: Exploring Methods to Assign Responses. *Statistical Journal of the IAOS* 34 (2): 179–89. https://doi.org/10.3233/sji-170374.

Finkeldey, Jessica G., and Stephen Demuth. 2021. Race/Ethnicity, Perceived Skin Color, and the Likelihood of Adult Arrest. *Race and Justice* 11 (4): 567–91. https://doi.org/10.1177/2153368719826269.

Finlay, Keith, and Michael Mueller-Smith. 2022. Criminal Justice Administrative Records System (CJARS) [dataset]. Ann Arbor, MI: University of Michigan. https://cjars.org.

Finlay, Keith, Michael Mueller-Smith, and Jordan Papp. 2022. The Criminal Justice Administrative Records System: A Next-Generation Research Data Platform. *Scientific Data* 9 (1): 562. https://doi.org/10.1038/s41597-022-01620-y.

Finlay, Keith, Michael Mueller-Smith, and Brittany Street. 2023. Children's Indirect Exposure to the U.S. Justice System: Evidence from Longitudinal Links between Survey and Administrative Data. *Quarterly Journal of Economics* 138 (4): 2181–224. https://doi.org/10.1093/qje/qjad021.

Freeman, Jonathan B., Andrew M. Penner, Aliya Saperstein, Matthias Scheutz, and Nalini Ambady. 2011. Looking the Part: Social Status Cues Shape Race Perception. *PLOS ONE* 6 (9): e25107. https://doi.org/10.1371/journal.pone.0025107.

Fryer, Roland G., and Steven D. Levitt. 2004. The Causes and Consequences of Distinctively Black Names. *Quarterly Journal of Economics* 119 (3): 767–805. https://doi.org/10.1162/0033553041502180.

Gaddis, S. Michael. 2017. How Black Are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies. *Sociological Science* 4:469–89. https://doi.org/10.15195/v4.a19.

Goel, Sharad, Justin M. Rao, and Ravi Shroff. 2016. Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy. *Annals of Applied Statistics* 10 (1): 365–94. https://doi.org/10.1214/15-aoas897.

Gyimah-Brempong, Kwabena, and Gregory N Price. 2006. Crime and Punishment: And Skin Hue Too? *American Economic Review* 96 (2): 246–50. https://doi.org/10.1257/000282806777212530.

Harrell, Erika. 2009. Asian, Native Hawaiian, and Pacific Islander Victims of Crime. Bureau of Justice Statistics, NCJ 225037. https://www.ojp.gov/library/publications/asian-native-hawaiian-and-pacific-islander-victims-crime.

Harris, David R. 2002. In the Eye of the Beholder: Observed Race and Observer Characteristics. PSC Research Report 02-522. https://web.archive.org/web/20220121011923/https://psc.isr.umich.edu/pubs/pdf/rr02-522.pdf.

Harris, David R., and Jeremiah Joseph Sim. 2002. Who is Multiracial? Assessing the Complexity of Lived Race. *American Sociological Review* 67 (4): 614–27. https://doi.org/10.2307/3088948.

Holland, Paul W. 2003. Causation and Race. *ETS Research Report Series* 2003 (1): i–21. https://doi.org/10.1002/j.2333-8504.2003.tb01895.x.

King, Ryan D., and Brian D. Johnson. 2016. A Punishing Look: Skin Tone and Afrocentric Features in the Halls of Justice. *American Journal of Sociology* 122 (1): 90–124. https://doi.org/10.1086/686941.

Latessa, Edward, Paula Smith, Richard Lemke, Matthew Makarios, and Christopher Lowenkamp. 2009. Creation and Validation of the Ohio Risk Assessment System: Final Report. Center for Criminal Justice Research, School of Criminal Justice, University of Cincinnati. https://www.uc.edu/content/dam/uc/ccjr/docs/reports/project_reports/ORAS_Final_Report.pdf.

Law Offices of the Public Defender, 11th Judicial Circuit of Florida. 2024. Adult Case Process: Frequently Asked Questions. Accessed 2/17/2024. https://www.pdmiami.com/public-defender/client-information/adult-case-process.page.

Liebler, Carolyn A., Sonya R. Porter, Leticia E. Fernandez, James M. Noon, and Sharon R. Ennis. 2017. America's Churning Races: Race and Ethnic Response Changes between Census 2000 and the 2010 Census. *Demography* 54 (1): 259–84. https://doi.org/10.1007/s13524-016-0544-0.

Limburg, Aubrey, Adam Kurczewski, and Victoria Udalova. 2023. Enhancing Race and Ethnicity Information in Medicaid Data: The Role of Census Bureau Data. Census Bureau SEHSD Working Paper 2023-05. https://www.census.gov/library/working-papers/2023/demo/SEHSD-WP2023-05.html.

Luh, Elizabeth. 2022. Not So Black and White: Uncovering Racial Bias from Systematically Misreported Trooper Reports. SSRN Working Paper 3357063. http://dx.doi.org/10.2139/ssrn.3357063.

Mason, Patrick. 2017. Not Black-Alone: The 2008 Presidential Election and Racial Self-Identification among African Americans. *Review of Black Political Economy* 44 (1–2): 55–76. https://doi.org/10.1007/s12114-017-9247-z.

Mays, Vickie M., Abigail Echo-Hawk, Susan D. Cochran, and Randall Akee. 2022. Data Equity in American Indian/Alaska Native Populations: Respecting Sovereign Nations' Right to Meaningful and Usable COVID-19 Data. *American Journal of Public Health* 112 (10): 1416–20.

Michigan Department of Corrections. 2022. Policy Directive: Pre-Sentence Investigation and Report. Accessed 2/17/2024. https://www.michigan.gov/corrections/-/media/Project/Websites/corrections/Files/Policy-Directives/PDs-06-Field-Operations/PD-06-01-General-Provisions/06-01-140-Pre-Sentence-Investigation-and-Report-effective-01-17-22.pdf?rev=32a8e45c295f405f9e92f0760c26b97b.

Michigan Legislature. 2023. The Code of Criminal Procedure (Excerpt): Act 175 Of 1927. Accessed 2/17/2024. http://www.legislature.mi.gov/(S(0iahpxqxedowy4kxh1hd5wyp))/mileg.aspx?page=GetObject&objectname=mcl-771-14.

Michigan State Police. 2024. Traffic Stop Data. Accessed 2/17/2024. https://www.michigan.gov/msp/public-information/transparency/accordion/reports/traffic-stop-data-main.

Monk, Ellis P. 2021. The Unceasing Significance of Colorism: Skin Tone Stratification in the United States. *Daedalus* 150 (2): 76–90. https://doi.org/10.1162/daed_a_01847.

Morning, Ann. 2007. "Everyone Knows It's a Social Construct": Contemporary Science and the Nature of Race. *Sociological Focus* 40 (4): 436–54. https://doi.org/10.1080/00380237.2007.10571319.

Mueller-Smith, Michael, and Kevin T. Schnepel. 2021. Diversion in the Criminal Justice System. *Review of Economic Studies* 88 (2): 883–936. https://doi.org/10.1093/restud/rdaa030.

Naik, Piyu Parth, and Syed Nadir Farrukh. 2022. Influence of Ethnicities and Skin Color Variations in Different Populations: A Review. *Skin Pharmacology and Physiology* 35 (2): 65–76. https://doi.org/10.1159/000518826.

National Academies of Sciences, Engineering, and Medicine. 2023. *Assessing the 2020 Census: Final Report.* Edited by Teresa A. Sullivan and Daniel L. Cork. Washington, DC: The National Academies Press. https://nap.nationalacademies.org/catalog/27150/assessing-the-2020-census-final-report.

Neal, Derek, and Armin Rick. 2014. *The Prison Boom and the Lack of Black Progress after Smith and Welch.* Working Paper, Working Paper Series 20283. National Bureau of Economic Research. http://www.nber.org/papers/w20283.

Nelson, Alondra. 2016. *The Social Life of DNA: Race, Reparations, and Reconciliation After the Genome.* Boston: Beacon Press. https://www.penguinrandomhouse.com/books/250539/the-social-life-of-dna-by-alondra-nelson.

Noe-Bustamante, Luis, Ana Gonzalez-Barrera, Khadijah Edwards, Lauren Mora, and Mark Hugo Lopez. 2021. Majority of Latinos Say Skin Color Impacts Opportunity in America and Shapes Daily Life. Pew Research Center Hispanic Trends Project. https://www.pewresearch.org/hispanic/2021/11/04/majority-of-latinos-say-skin-color-impacts-opportunity-in-america-and-shapes-daily-life.

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.

———. 2024. Revisions to OMB's Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity. *Federal Register* 89 (62): 22182–96. https://www.federalregister.gov/documents/2024/03/29/2024-06469/revisions-to-ombs-statistical-policy-directive-no-15-standards-for-maintaining-collecting-and.

Pager, Devah. 2003. The Mark of a Criminal Record. *American Journal of Sociology* 108 (5): 937–75. https://doi.org/doi/10.1086/374403.

Roberts, Dorothy. 2012. *Fatal Invention.* New York: New Press. https://thenewpress.com/books/fatal-invention.

Rodriguez, Nancy, and Rebecca Tublitz. 2023. Exploring Latino/a Representation in Local Criminal Justice Systems: A Review of Data Collection Practices and Systems-Involvement. UCI School of Social Ecology. https://socialecology.uci.edu/sites/default/files/users/mkcruz/sjc_latinos_in_cjs_march_2023.pdf.

Rose, Dina R., and Todd R. Clear. 1998. Incarceration, Social Capital, and Crime: Implications for Social Disorganization Theory. *Criminology* 36 (3): 441–80. https://doi.org/10.1111/j.1745-9125.1998.tb01255.x.

Rose, Evan K. 2023. A Constructivist Perspective on Empirical Discrimination Research. *Journal of Economic Literature* 61 (3): 906–23. https://doi.org/10.1257/jel.20221705.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rogers, and Megan Schouweiler. 2024. IPUMS USA: Version 15.0 [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/d010.v15.0.

Saperstein, Aliya, and Andrew M. Penner. 2012. Racial Fluidity and Inequality in the United States. *American Journal of Sociology* 118 (3): 676–727. https://doi.org/10.1086/667722.

Scott, Charles G. 1999. Identifying the Race and Ethnicity of SSI Recipients. *Social Security Bulletin* 62 (4): 9–20. https://www.ssa.gov/policy/docs/ssb/v62n4/v62n4p9.pdf.

Small-Rodriguez, Desi, and Randall Akee. 2021. Identifying Disparities in Health Outcomes and Mortality for American Indian and Alaska Native Populations Using Tribally Disaggregated Vital Statistics and Health Survey Data. *American Journal of Public Health* 111 (S2): S126–S132. https://doi.org/10.2105/AJPH.2021.306427.

Snipp, C. Matthew. 2003. Racial Measurement in the American Census: Past Practices and Implications for the Future. *Annual Review of Sociology* 29 (1): 563–88. https://doi.org/10.1146/annurev.soc.29.010202.100006.

Telles, Edward, and Tianna Paschel. 2014. Who Is Black, White, or Mixed Race? How Skin Color, Status, and Nation Shape Racial Classification in Latin America. *American Journal of Sociology* 120 (3): 864–907. https://doi.org/10.1086/679252.

Texas State Legislature. 2024. Code of Criminal Procedure. Accessed 2/17/2024. https://statutes.capitol.texas.gov/docs/sdocs/codeofcriminalprocedure.pdf.

Tuttle, Cody. 2019. Racial Disparities in Federal Sentencing: Evidence from Drug Mandatory Minimums. SSRN Working Paper 3080463. https://doi.org/10.2139/ssrn.3080463.

U.S. Census Bureau. 2022. American Community Survey and Puerto Rico Community Survey Design and Methodology: Version 3.0. Washington, DC: U.S. Census Bureau. https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/2022/acs_design_methodology_report_2022.pdf.

Wagner, Deborah, and Mary Lane. 2014. The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software. Census Bureau CARRA Working Paper 2014-01. https://www.census.gov/library/working-papers/2014/adrm/carra-wp-2014-01.html.

Webster, Richard A. 2021. "If Everybody's White, There Can't Be Any Racial Bias": The Disappearance of Hispanic Drivers From Traffic Records. *ProPublica* 22 November 2021. https://www.propublica.org/article/if-everybodys-white-there-cant-be-any-racial-bias-the-disappearance-of-hispanic-drivers-from-traffic-records.

Williams, Ben. 2023. Inmate Data Reallocation in the 2020 Redistricting Cycle. Washington, DC: National Conference of State Legislatures. https://www.ncsl.org/redistricting-and-census/inmate-data-reallocation-in-the-2020-redistricting-cycle.

York, Gary. 2021. Identifying Incarcerated Gang Members is Critical in Corrections. *Corrections1.com* 24 February 2021. https://www.corrections1.com/prison-gangs/articles/identifying-incarcerated-gang-members-is-critical-in-corrections-XyBFxjEhDs2lleFg.

6 Tables

Table 1: Race and ethnicity categories in raw CJARS data sourced from state departments of corrections

State	White	Black	Asian	Hispanic	AIAN	Other
Arizona	✓	✓	✓	✓	✓	✓
Arkansas	✓	✓	✓	✓	✓	✓
Florida	✓	✓	✓	✓	*	-
Georgia	✓	✓	✓	*	*	-
Illinois	✓	✓	✓	✓	✓	✓
Indiana	✓	✓	✓	✓	*	-
Iowa	✓	✓	✓	✓	✓	-
Kansas	✓	✓	✓	✓	✓	-
Michigan	✓	✓	✓	✓	✓	-
Minnesota	✓	✓	✓	-	✓	-
Mississippi	✓	✓	✓	✓	✓	-
Missouri	✓	✓	✓	-	✓	-
Montana	✓	✓	✓	-	✓	-
Nebraska	✓	✓	✓	✓	✓	✓
Nevada	✓	✓	✓	✓	✓	-
New Jersey	✓	✓	✓	✓	*	-
New York	✓	✓	✓	✓	✓	✓
Ohio	✓	✓	✓	-	✓	✓
Oklahoma	✓	✓	✓	✓	✓	✓
Oregon	✓	✓	✓	✓	✓	-
Pennsylvania	✓	✓	✓	✓	*	✓
Texas	✓	✓	✓	✓	✓	✓
Vermont	✓	✓	✓	-	✓	✓

Source: Authors' calculations from raw 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/ethnicity category. "-" indicates that the state DOC does not record that race/ethnicity category. "*" indicates that fewer than 0.5% of records from this state are associated with this race/ethnicity. Only statewide DOCs 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 National Prisoners Statistics program.

Table 2: Convenience sample used to study the life cycle of race and ethnicity data in criminal justice records management systems

Source	State	Stage of the justice system	Citation
Person A	Arizona	Department of corrections	
Person B	Illinois, Michigan	State prosecutor's office, local police	
Person C	Michigan	Department of corrections, local police	
Person D	Ohio	Department of corrections	
Person E	New Jersey	Community supervision, criminal courts	
Person F	New York	Criminal courts	
Person G	New York	Criminal courts	
Guidelines A	Florida	Public defender's office	Law Offices of the Public Defender, 11th Judicial Circuit of Florida (2024)
Guidelines B	Michigan	Department of corrections	Michigan Department of Corrections (2022),Michigan Legislature (2023)
Guidelines C	Michigan	State police	Michigan State Police (2024)
Guidelines D	Texas	All stages	Texas State Legislature (2024)
Guidelines E	Texas	Criminal courts	Dallas County District Attorney's Office (2024)

Source: Authors' summary of sources consulted to study life cycle of race and ethnicity data in criminal justice records management systems.

Table 3: Recoding Census Bureau race and ethnicity to BJS/CJARS race and ethnicity

BJS/CJARS race/ethnicity	Census Bureau race	Census Bureau ethnicity
White, non Hispanic	White	Not Hispanic
Black, non Hispanic	Black or African American	Not Hispanic
AIAN, non-Hispanic	American Indian, Alaska Native	Not Hispanic
Asian/Pacific Islander, non-Hispanic	Chinese, Indian, Filipino, Vietnamese, Korean, Japanese, Other Asian, Native Hawaiian, Guamanian or Chamorro, Samoan, Other Pacific Islander	Not Hispanic
Other, non-Hispanic	Some Other Race, Two or More Races/Multiracial	Not Hispanic
Hispanic	Any race	Hispanic

Source: Authors' mapping based on Finlay and Mueller-Smith (2022) and, for example, Carson (2022, p. 42).
Note: BJS uses the coding above in collections based on administrative records. In demographic surveys, BJS uses the separate race and ethnicity codes required by OMB (1997).

Table 4: Agreement between Census Bureau race/ethnicity composites and component data sources

Census Bureau race/ethnicity composite	Linkage rates to data sources			Among records linked to 1+ census				Among records linked to 1+ census and the Census Numident
	1–2 censuses	1 census	Census Numident	2000 Census agrees	2010 Census agrees	Either census agrees	Either census disagrees	Census Numident agrees
Felony and misdemeanor defendants, 2000–2019:								
White, not Hispanic	89%	31%	88%	98%	97%	100%	4%	99%
Black, not Hispanic	80%	39%	91%	96%	96%	99%	6%	98%
Asian/PI, not Hispanic	86%	44%	83%	78%	84%	89%	28%	80%
Hispanic	86%	41%	89%	91%	93%	97%	11%	80%
AIAN, not Hispanic	88%	49%	87%	83%	85%	93%	22%	60%
Other, not Hispanic	79%	49%	75%	76%	43%	82%	49%	29%
Ever incarcerated population, 2000–2019:								
White, not Hispanic	75%	40%	95%	98%	97%	100%	4%	99%
Black, not Hispanic	64%	41%	96%	96%	96%	98%	6%	98%
Asian/PI, not Hispanic	72%	51%	92%	71%	74%	77%	35%	70%
Hispanic	75%	46%	94%	93%	93%	96%	10%	83%
AIAN, not Hispanic	79%	54%	94%	86%	87%	92%	17%	61%
Other, not Hispanic	58%	44%	88%	81%	50%	83%	36%	23%

Source: Authors' calculations from the 2000 and 2010 Censuses, the Census Numident, and CJARS criminal histories (2022Q4 vintage). U.S. Census Bureau, project number P-7500378, approval numbers CBDRB-FY24-0101, CBDRB-FY24-0277.

Note: Each panel reports the share of observations which meet the listed criteria, conditional on being in each row. CJARS court data come from Arizona, Florida, Maryland, Michigan, North Carolina, Oregon, Pennsylvania, Texas, Virginia, Wisconsin. CJARS prison data come from Arizona, Florida, Illinois, Indiana, Kansas, Michigan, Minnesota, Montana, Nebraska, New York, North Carolina, Oklahoma, Pennsylvania, Texas.

Table 5: Concordance of agency-record race/ethnicity versus Census Bureau race/ethnicity composite, individuals charged with a felony or misdemeanor, 2000–2019, total table percents

| Census Bureau race/ethnicity composite | Race/ethnicity recorded by justice system | | | | | | % of agency-recorded equal to composite | Caseload share | Row total |
	White (1)	Black (2)	AIAN (3)	Asian/PI (4)	Other (5)	Hispanic (6)	(7)	(8)	(9)
White	45.70%	0.73%	0.05%	0.11%	1.56%	0.08%	94.76%	48.23%	737,000
Black	0.93%	33.70%	0.01%	0.03%	0.16%	0.02%	96.68%	34.85%	533,000
AIAN	0.87%	0.28%	0.76%	0.01%	0.04%	0.01%	38.39%	1.97%	30,000
Asian/PI	0.62%	0.21%	0.02%	0.38%	0.11%	0.01%	28.06%	1.35%	21,000
Other	0.16%	0.14%	0.01%	0.02%	0.02%	0.01%	4.62%	0.36%	5,000
Hispanic	9.16%	1.30%	0.09%	0.06%	0.50%	2.13%	16.10%	13.24%	202,000
% of composite equal to agency label	79.58%	92.68%	80.90%	61.15%	0.67%	94.15%	Concordance rate 82.69%		
Caseload share	57.44%	36.36%	0.94%	0.61%	2.39%	2.26%			
Column total	878,000	556,000	14,000	10,000	36,000	35,000			1,529,000

Source: Authors' calculations from the 2000 and 2010 Censuses, the Census Numident, and CJARS criminal histories (2022Q4 vintage). U.S. Census Bureau, project number P-7500378, approval numbers CBDRB-FY24-0101, CBDRB-FY24-0277.

Note: CJARS court data come from Arizona, Florida, Maryland, Michigan, North Carolina, Oregon, Pennsylvania, Texas, Virginia, Wisconsin. This table presents the percent of individuals with particular combinations of agency-recorded race/ethnicity and Census Bureau race/ethnicity composites out of the total sample. For example, *White, White* indicates that 45.7% of the 1,529,000 in the sample are labeled by justice agencies as White are also identified as White in the race/ethnicity composite. Race groups exclude people with Hispanic ethnicity; Hispanic includes people of all races. Estimates and sample sizes have been rounded according to Census Bureau DRB rules.

Table 6: Concordance of agency-recorded race/ethnicity versus Census Bureau race/ethnicity composite, population of people in prison, 2000–2019, total table percents

| Census Bureau race/ethnicity composite | Race/ethnicity recorded by justice system | | | | | | % of agency-recorded equal to composite | Caseload share | Row total |
	White (1)	Black (2)	AIAN (3)	Asian/PI (4)	Other (5)	Hispanic (6)	(7)	(8)	(9)
White	37.21%	0.51%	0.15%	0.01%	0.06%	0.51%	96.78%	38.45%	190,000
Black	0.38%	39.56%	0.02%	0.00%	0.11%	0.08%	98.49%	40.15%	198,000
AIAN	0.70%	0.25%	1.04%	0.00%	0.01%	0.06%	50.46%	2.06%	10,000
Asian/PI	0.34%	0.19%	0.01%	0.20%	0.05%	0.03%	24.49%	0.82%	4,000
Other	0.10%	0.12%	0.04%	0.02%	0.02%	0.09%	5.11%	0.39%	2,000
Hispanic	4.08%	1.20%	0.12%	0.02%	0.53%	12.16%	67.19%	18.11%	89,000
% of composite equal to agency label	86.91%	94.59%	75.72%	78.37%	2.54%	93.99%	Concordance rate 90.2%		
Caseload share	42.81%	41.83%	1.38%	0.25%	0.78%	12.93%			
Column total	211,000	206,000	7,000	1,000	4,000	64,000			493,000

Source: Authors' calculations from the 2000 and 2010 Censuses, the Census Numident, and CJARS criminal histories (2022Q4 vintage). U.S. Census Bureau, project number P-7500378, approval numbers CBDRB-FY24-0101, CBDRB-FY24-0277.

Note: CJARS prison data come from Arizona, Florida, Illinois, Indiana, Kansas, Michigan, Minnesota, Montana, Nebraska, New York, North Carolina, Oklahoma, Pennsylvania, Texas. This table presents the percent of individuals with particular combinations of agency-recorded race/ethnicity and Census Bureau race/ethnicity composites out of the total sample. For example, *White, White* indicates that 37.21% of the 493,000 in the sample are labeled by justice agencies as White are also identified as White in the race/ethnicity composite. Race groups exclude people with Hispanic ethnicity; Hispanic includes people of all races. Estimates and sample sizes have been rounded according to Census Bureau DRB rules.

Table 7: Self-reported race and ethnicity in the 2004 Survey of Inmates in State and Federal Correctional Facilities (SISFCF) and 2016 Survey of Prison Inmates (SPI)

Race/ethnicity	2004 SISFCF		2016 SPI	
	State	Federal	State	Federal
White, non-Hispanic	36.5%	29.2%	31.9%	21.0%
	(0.7%)	(1.3%)	(1.0%)	(2.1%)
Black, non-Hispanic	40.0%	39.0%	33.8%	32.2%
	(0.7%)	(1.4%)	(1.2%)	(3.0%)
Hispanic	17.5%	24.7%	20.7%	37.1%
	(0.6%)	(1.2%)	(0.9%)	(4.2%)
American Indian/Alaska Native, non-Hispanic	1.7%	2.8%	1.4%	1.7%
	(0.2%)	(0.5%)	(0.2%)	(0.5%)
Asian/Pacific Islander, non-Hispanic	1.0%	1.7%	0.9%	1.5%
	(0.1%)	(0.4%)	(0.1%)	(0.3%)
Two or more races, non-Hispanic	3.3%	2.6%	11.3%	6.5%
	(0.3%)	(0.4%)	(0.4%)	(0.7%)
N	14,477	3,680	20,064	4,784

Source: 2004 SISFCF estimates and standard errors from authors' calculations of person-weighted proportions from publicly available survey microdata and methodology (Bureau of Justice Statistics 2019a, 2019b). 2016 SPI estimates and standard errors from Beatty and Snell (2021, p. 14) and Bureau of Justice Statistics (2023).

Table 8: Socioeconomic characteristics of adults aged 18–65 years, American Community Survey, 2005–2019, by race and ethnicity

Characteristic	White	Black	AIAN	Asian/PI	Other	Hispanic
Educational attainment						
High school graduate	0.94	0.88	0.84	0.92	0.91	0.72
	(0.0001)	(0.0002)	(0.0010)	(0.0003)	(0.0005)	(0.0003)
Some college	0.58	0.46	0.40	0.69	0.56	0.34
	(0.0001)	(0.0004)	(0.0014)	(0.0005)	(0.0009)	(0.0003)
College graduate	0.32	0.18	0.13	0.49	0.27	0.13
	(0.0001)	(0.0003)	(0.0010)	(0.0005)	(0.0008)	(0.0002)
Employment/earnings						
Unemployed	0.04	0.09	0.09	0.04	0.07	0.06
	(0.0001)	(0.0002)	(0.0008)	(0.0002)	(0.0005)	(0.0002)
Wage income	$37,857	$23,693	$20,576	$41,099	$28,994	$23,340
	(15.0)	(24.6)	(91.2)	(59.1)	(77.7)	(20.3)
Total family income	$96,807	$59,874	$59,659	$110,250	$82,007	$67,506
	(26.5)	(45.5)	(176.6)	(106.6)	(151.3)	(40.2)
Family income below	0.08	0.18	0.21	0.10	0.13	0.16
federal poverty line	(0.0001)	(0.0003)	(0.0012)	(0.0003)	(0.0006)	(0.0002)
Household structure						
Single-parent household	0.09	0.23	0.16	0.08	0.15	0.15
	(0.0001)	(0.0003)	(0.0011)	(0.0003)	(0.0007)	(0.0002)
N	19,714,757	2,925,138	249,137	1,527,804	506,110	3,837,687

Source: Authors' calculations from publicly available 2005–2019 5-year American Community Survey (ACS) IPUMS files (Ruggles et al. 2024).

Note: ACS sample is restricted to respondents aged between 18 and 65 years and stratified by the mutually exclusive race/ethnicity categories used throughout this paper. Race groups exclude people with Hispanic ethnicity; Hispanic includes people of all races. For each group, we calculate person-weighted means for each characteristic.

7 Figures

Figure 1: Flow of information during criminal case adjudication
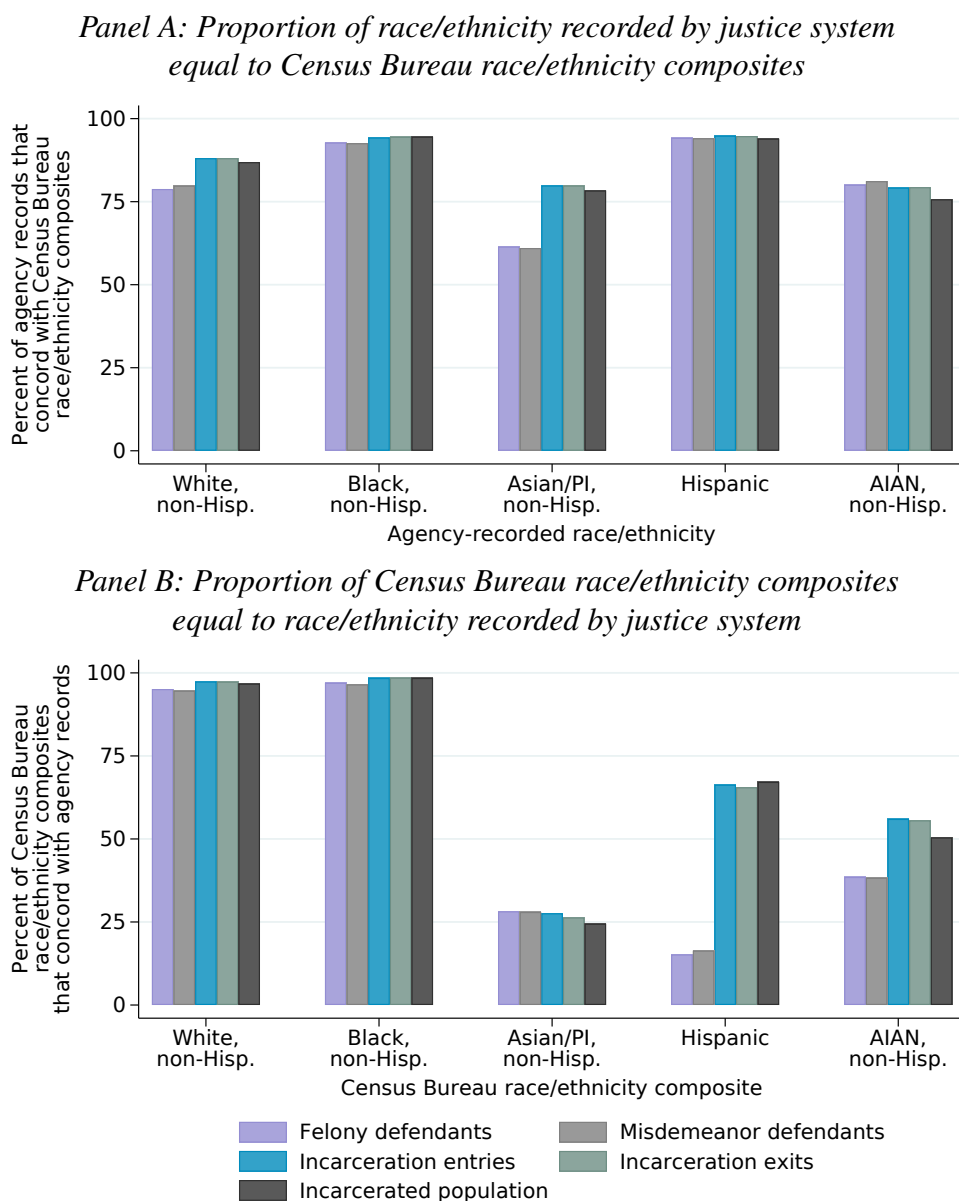**(a)** Case evolution prior to disposition
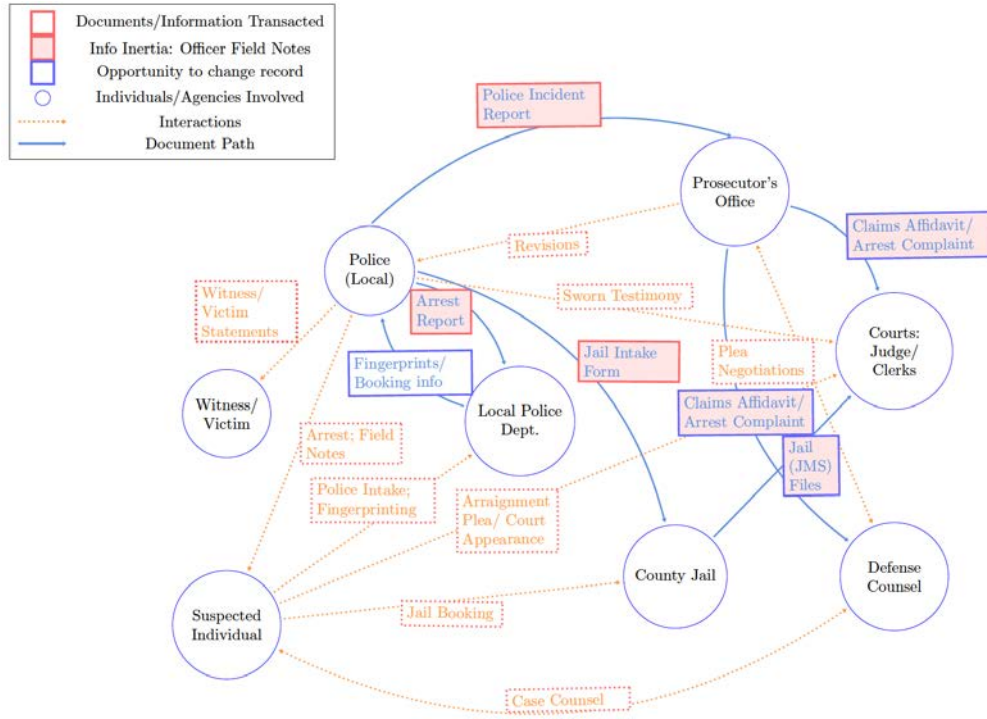
(b) Case evolution following disposition
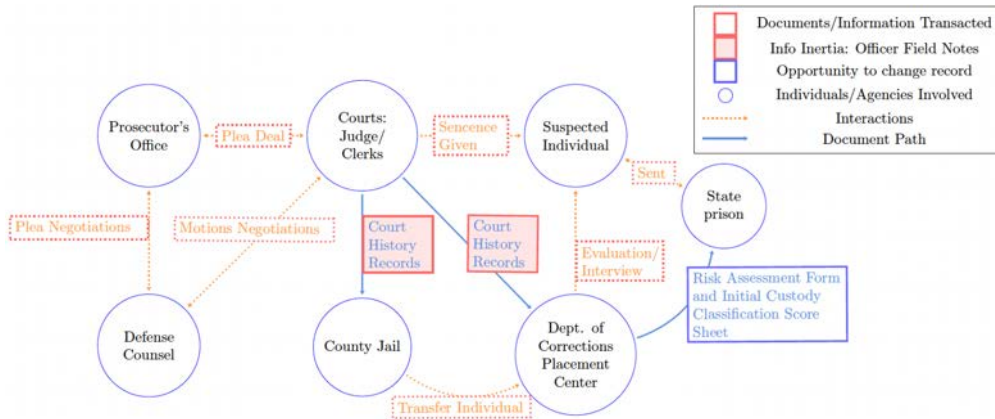
Source: Authors' summary of findings based on qualitative research. Sources consulted are enumerated in Table 2.
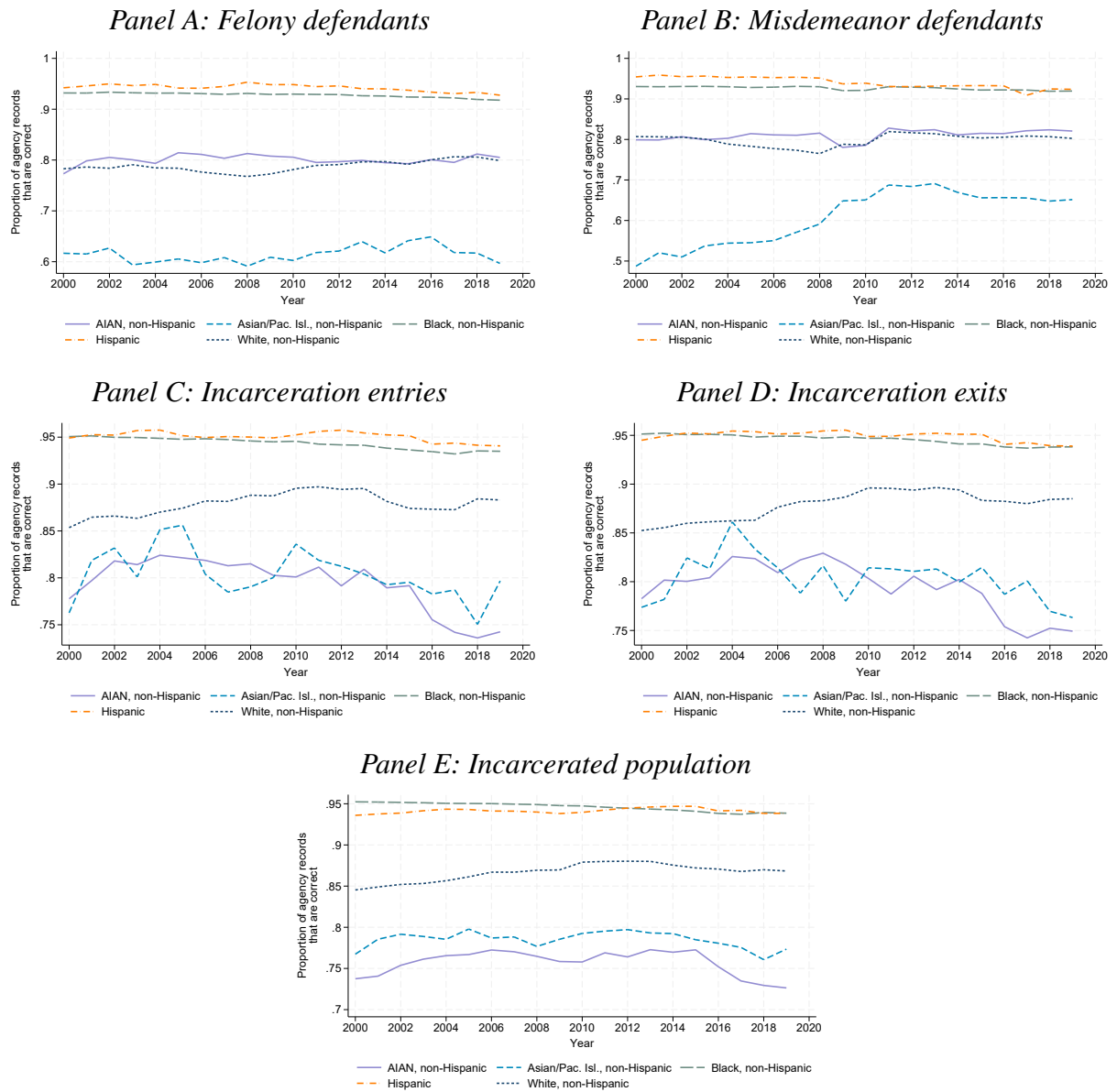
Figure 2: Concordance of agency-recorded race/ethnicity versus Census Bureau race/ethnicity composite, 2000–2019, by criminal justice caseload

Panel A: Proportion of race/ethnicity recorded by justice system equal to Census Bureau race/ethnicity composites

Panel B: Proportion of Census Bureau race/ethnicity composites equal to race/ethnicity recorded by justice system

Source: Authors' calculations from the 2000 and 2010 Censuses, the Census Numident, and CJARS criminal histories (2022Q4 vintage). U.S. Census Bureau, project number P-7500378, approval numbers CBDRB-FY24-0101, CBDRB-FY24-0277.

Note: CJARS court data come from Arizona, Florida, Maryland, Michigan, North Carolina, Oregon, Pennsylvania, Texas, Virginia, Wisconsin. CJARS prison data come from Arizona, Florida, Illinois, Indiana, Kansas, Michigan, Minnesota, Montana, Nebraska, New York, North Carolina, Oklahoma, Pennsylvania, Texas. This figure presents the concordance of agency-recorded race/ethnicity and Census Bureau race/ethnicity composites. The two panels alter the base measure/denominator. The bars represent different criminal justice caseloads. Estimates and sample sizes have been rounded according to Census Bureau DRB rules.
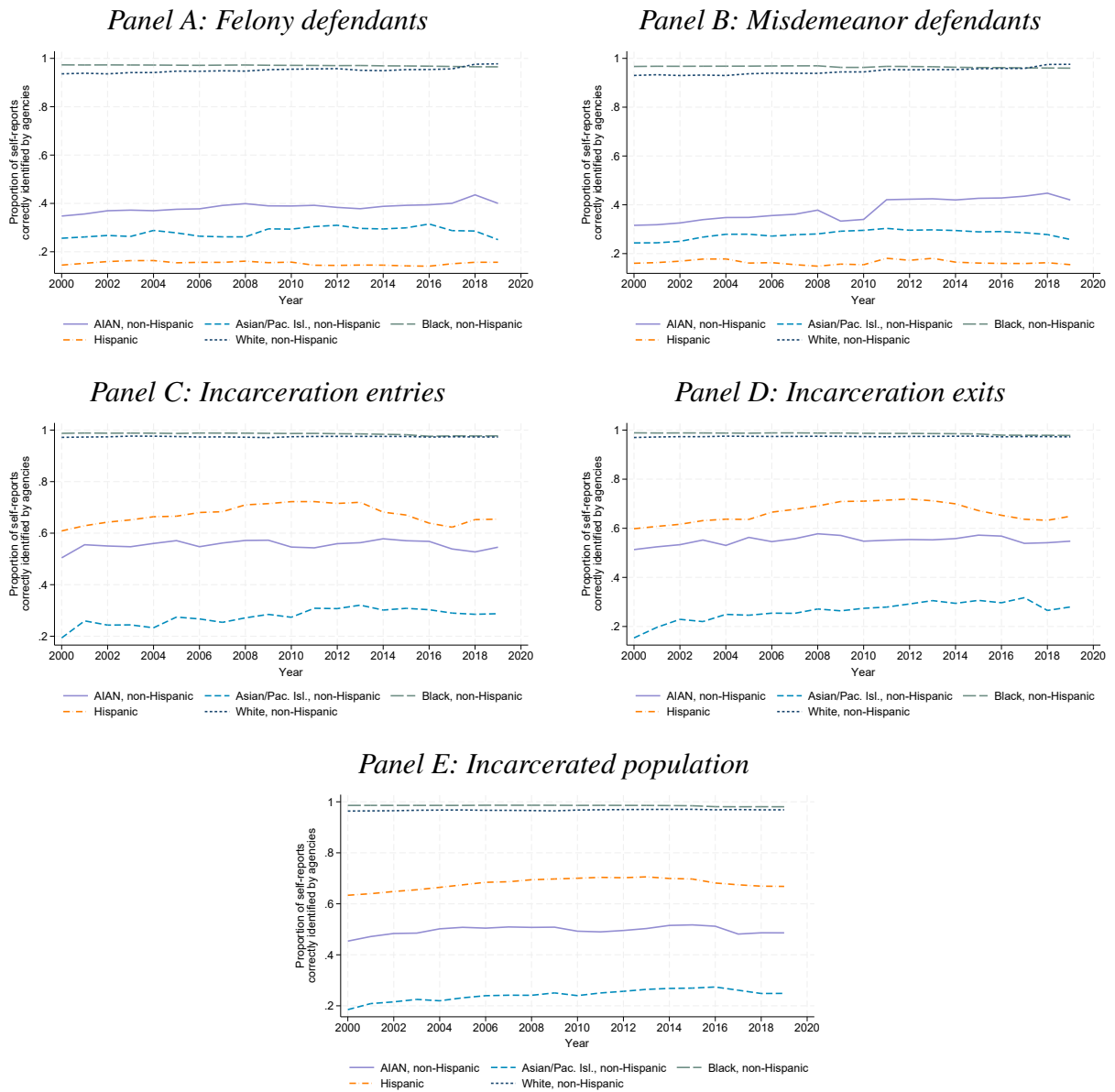
Figure 3: Proportion of race/ethnicity recorded by justice system equal to Census Bureau race/ethnicity composites, 2000–2019, by criminal justice caseload and year

Panel A: Felony defendants

Panel B: Misdemeanor defendants

Panel C: Incarceration entries

Panel D: Incarceration exits

Panel E: Incarcerated population

Note: CJARS court data come from Arizona, Florida, Maryland, Michigan, North Carolina, Oregon, Pennsylvania, Texas, Virginia, Wisconsin. CJARS prison data come from Arizona, Florida, Illinois, Indiana, Kansas, Michigan, Minnesota, Montana, Nebraska, New York, North Carolina, Oklahoma, Pennsylvania, Texas. These figures present the concordance of agency-recorded race/ethnicity and Census Bureau race/ethnicity composites over time. Each panel represents a different criminal justice caseload. Estimates and sample sizes have been rounded according to Census Bureau DRB rules.
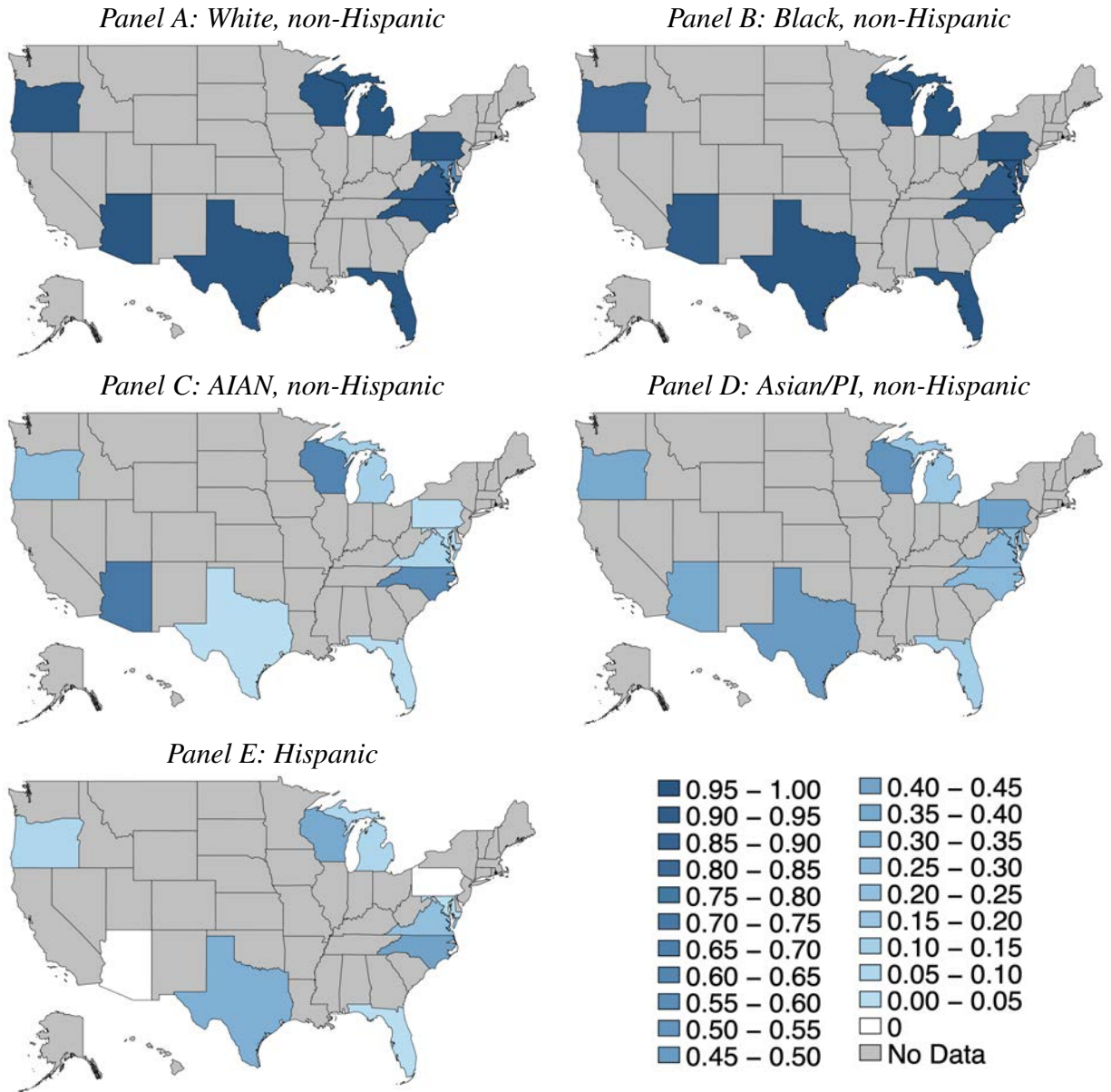
Figure 4: Proportion of Census Bureau race/ethnicity composites equal to race/ethnicity recorded by justice system, 2000–2019, by criminal justice caseload and year

Panel A: Felony defendants

Panel B: Misdemeanor defendants

Panel C: Incarceration entries

Panel D: Incarceration exits

Panel E: Incarcerated population

Source: Authors' calculations from the 2000 and 2010 Censuses, the Census Numident, and CJARS criminal histories (2022Q4 vintage). U.S. Census Bureau, project number P-7500378, approval numbers CBDRB-FY24-0101, CBDRB-FY24-0277.

Note: CJARS court data come from Arizona, Florida, Maryland, Michigan, North Carolina, Oregon, Pennsylvania, Texas, Virginia, Wisconsin. CJARS prison data come from Arizona, Florida, Illinois, Indiana, Kansas, Michigan, Minnesota, Montana, Nebraska, New York, North Carolina, Oklahoma, Pennsylvania, Texas. These figures present the concordance of agency-recorded race/ethnicity and Census Bureau race/ethnicity composites over time. Each panel represents a different criminal justice caseload. Estimates and sample sizes have been rounded according to Census Bureau DRB rules.
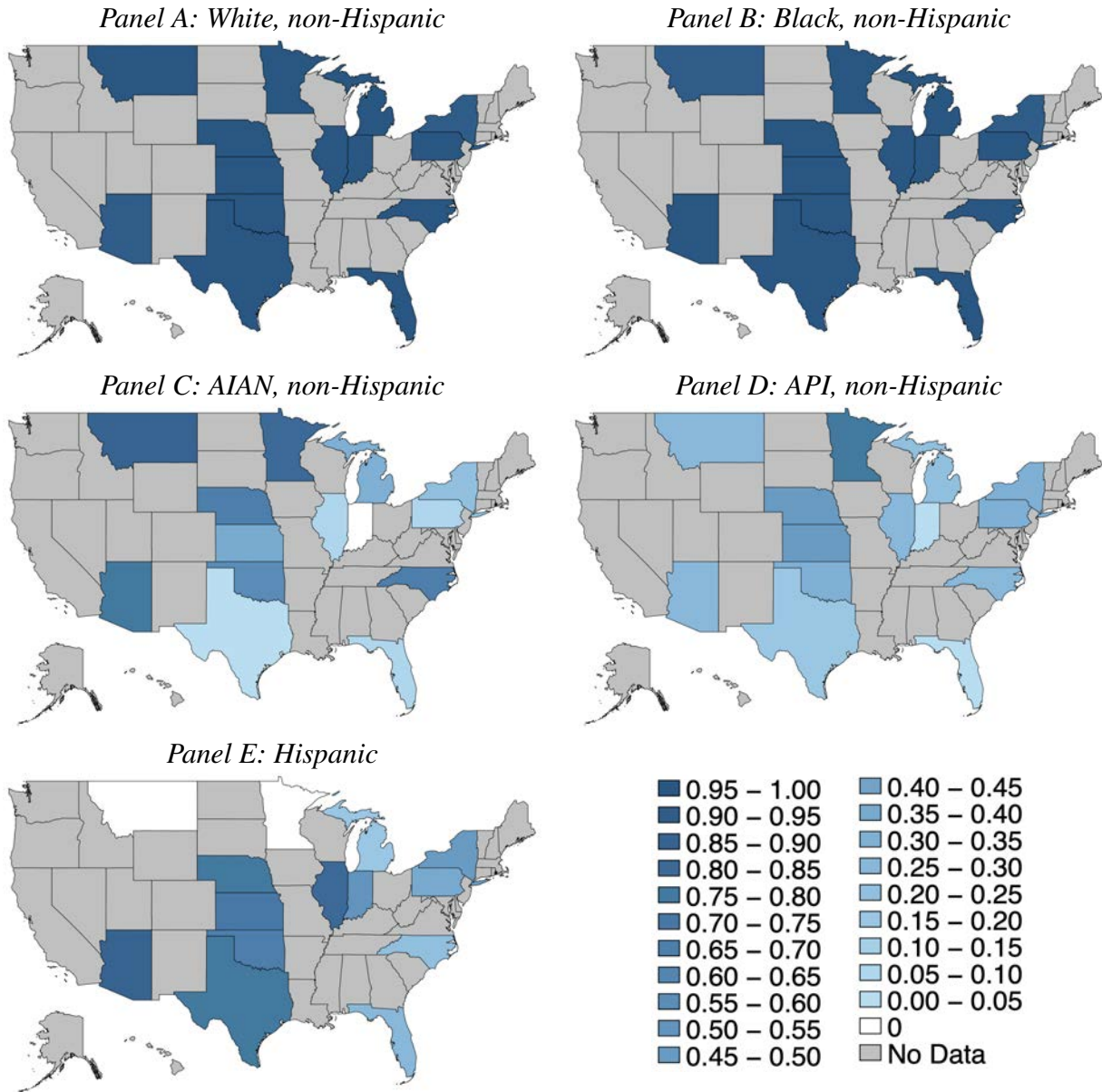
Figure 5: Geographic variation in the proportion of race/ethnicity recorded by justice system equal to Census Bureau race/ethnicity composites, misdemeanor and felony defendants, 2000–2019, by race/ethnicity and state

Panel A: White, non-Hispanic

Panel B: Black, non-Hispanic

Panel C: AIAN, non-Hispanic

Panel D: Asian/PI, non-Hispanic

Panel E: Hispanic

■ 0.95 – 1.00	■ 0.40 – 0.45
■ 0.90 – 0.95	■ 0.35 – 0.40
■ 0.85 – 0.90	■ 0.30 – 0.35
■ 0.80 – 0.85	■ 0.25 – 0.30
■ 0.75 – 0.80	■ 0.20 – 0.25
■ 0.70 – 0.75	■ 0.15 – 0.20
■ 0.65 – 0.70	■ 0.10 – 0.15
■ 0.60 – 0.65	■ 0.05 – 0.10
■ 0.55 – 0.60	□ 0.00 – 0.05
■ 0.50 – 0.55	□ 0
■ 0.45 – 0.50	■ No Data

Note: "0"/white means that no race/ethnicity records concord across the two data sources. CJARS court data come from Arizona, Florida, Maryland, Michigan, North Carolina, Oregon, Pennsylvania, Texas, Virginia, Wisconsin. These figures present the proportion of misdemeanor and felony defendants from different subgroups accurately labeled by justice agencies, by state of criminal charge. Estimates and sample sizes have been rounded according to Census Bureau DRB rules.

Figure 6: Geographic variation in the proportion of race/ethnicity recorded by justice system equal to Census Bureau race/ethnicity composites, population of prison inmates, 2000–2019, by race/ethnicity and state

Panel A: White, non-Hispanic *Panel B: Black, non-Hispanic*

Panel C: AIAN, non-Hispanic *Panel D: API, non-Hispanic*

Panel E: Hispanic

Source: Authors' calculations from the 2000 and 2010 Censuses, the Census Numident, and CJARS criminal histories (2022Q4 vintage). U.S. Census Bureau, project number P-7500378, approval numbers CBDRB-FY24-0101, CBDRB-FY24-0277.
Note: "0"/white means that no race/ethnicity records concord across the two data sources. CJARS prison data come from Arizona, Florida, Illinois, Indiana, Kansas, Michigan, Minnesota, Montana, Nebraska, New York, North Carolina, Oklahoma, Pennsylvania, Texas. These figures present the proportion of prison inmates from different subgroups accurately labeled by justice agencies, by state of incarceration. Estimates and sample sizes have been rounded according to Census Bureau DRB rules.

Figure 7: Percent of Census Bureau sex composites that concord with agency-recorded sex, 2000–2019, by criminal justice caseload
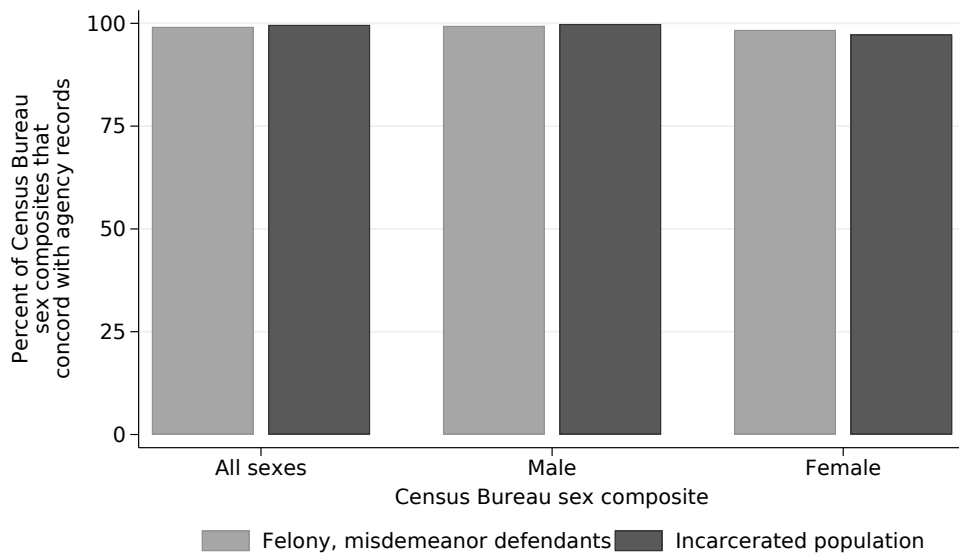
Source: Authors' calculations from the 2000 and 2010 Censuses, the Census Numident, and CJARS criminal histories (2022Q4 vintage). U.S. Census Bureau, project number P-7500378, approval numbers CBDRB-FY24-0101, CBDRB-FY24-0277.

Note: CJARS court data come from Arizona, Florida, Maryland, Michigan, North Carolina, Oregon, Pennsylvania, Texas, Virginia, Wisconsin. CJARS prison data come from Arizona, Florida, Illinois, Indiana, Kansas, Michigan, Minnesota, Montana, Nebraska, New York, North Carolina, Oklahoma, Pennsylvania, Texas. This figure presents the concordance of agency-recorded sex and Census Bureau sex composites by criminal justice caseload. Estimates and sample sizes have been rounded according to Census Bureau DRB rules.
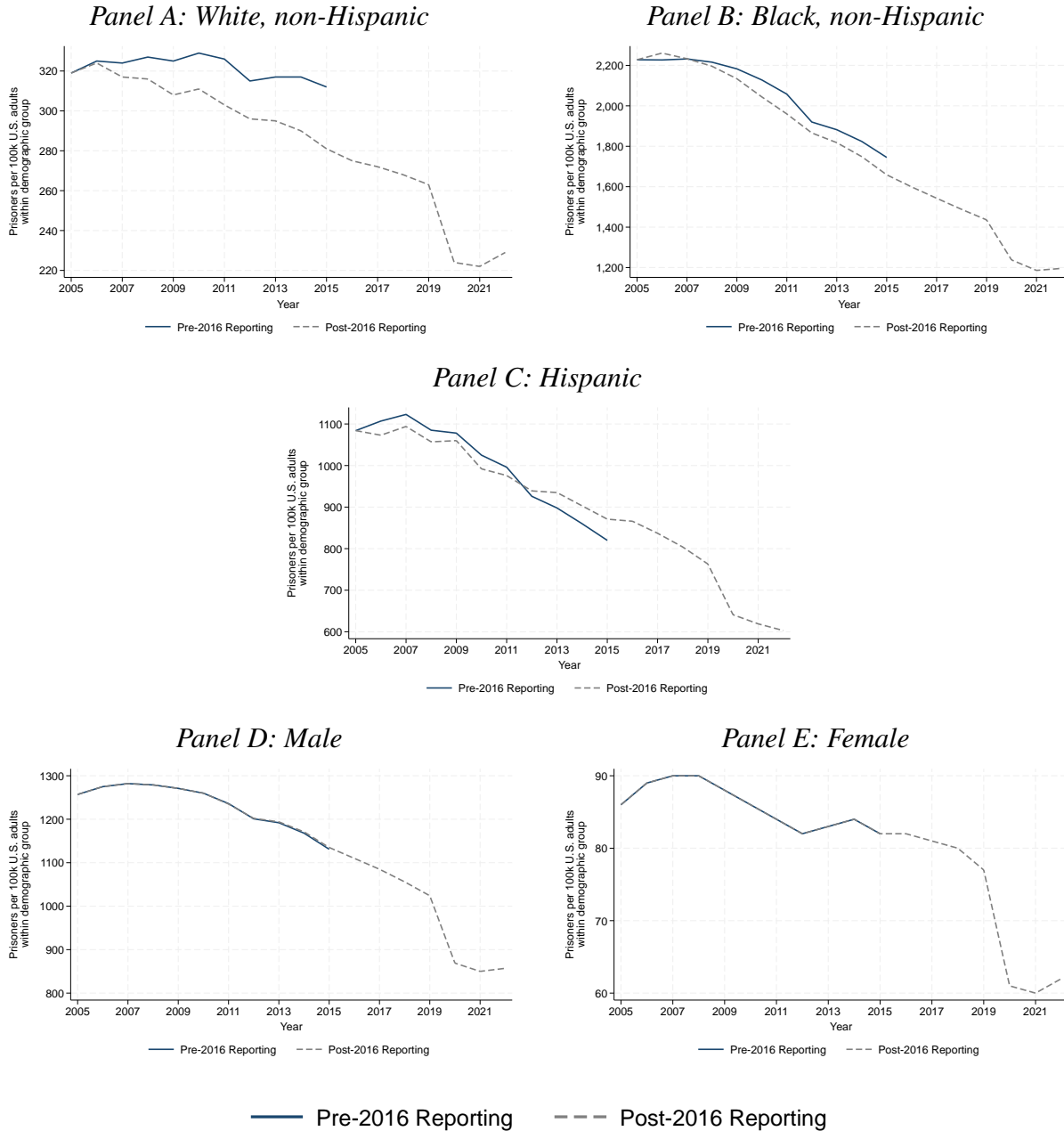
Figure 8: Percent of Census Bureau race/ethnicity composites that concord with justice or Medicaid agency records, 2000–2019, by criminal justice caseload
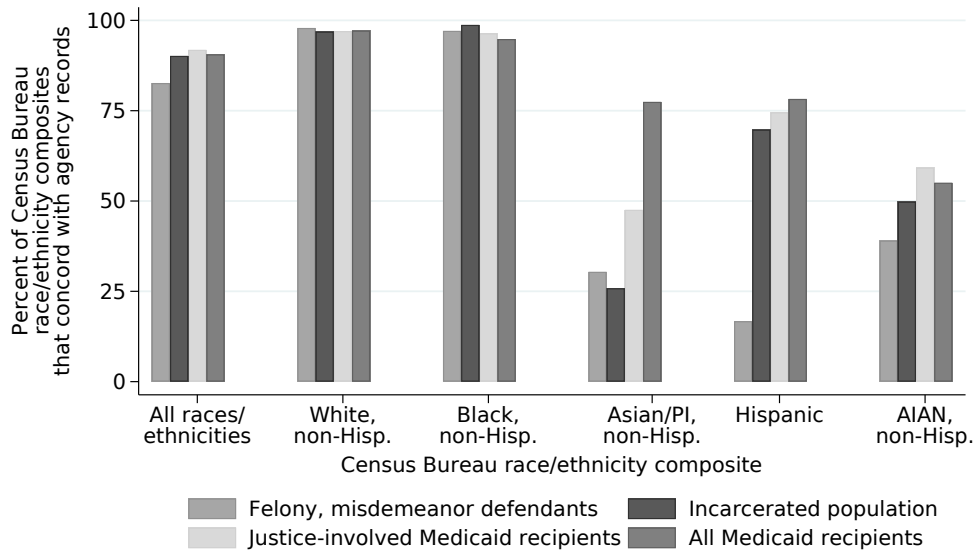
Source: Authors' calculations from the 2000 and 2010 Censuses, the Census Numident, CJARS criminal histories (2022Q4 vintage), and CMS Medicaid Statistical Information System (MSIS)/Transformed MSIS data. U.S. Census Bureau, project number P-7500378, approval numbers CBDRB-FY24-0101, CBDRB-FY24-0277.

Note: CJARS court data come from Arizona, Florida, Maryland, Michigan, North Carolina, Oregon, Pennsylvania, Texas, Virginia, Wisconsin. CJARS prison data come from Arizona, Florida, Illinois, Indiana, Kansas, Michigan, Minnesota, Montana, Nebraska, New York, North Carolina, Oklahoma, Pennsylvania, Texas. These figures present the proportion of individuals from different subgroups accurately labeled by justice agencies (in the case of the two justice populations) or state Medicaid agencies (in the case of the two Medicaid populations). Estimates and sample sizes have been rounded according to Census Bureau DRB rules.
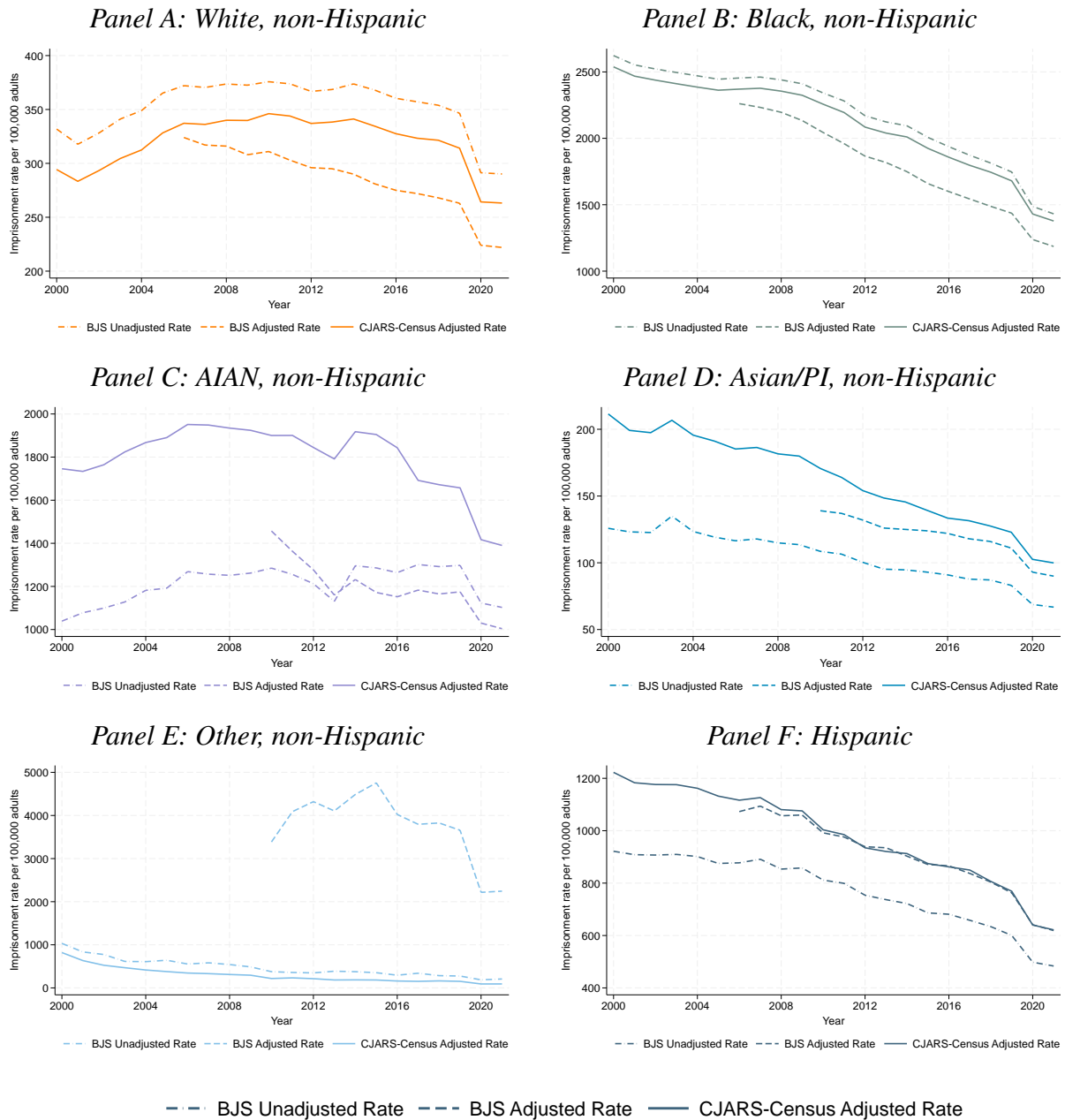
Figure 9: Imprisonment rates by race/ethnicity and sex adjusted using the 2004 Survey of Inmates in State and Federal Correctional Facilities (SISFCF) and 2016 Survey of Prison Inmates (SPI)

Panel A: White, non-Hispanic

Panel B: Black, non-Hispanic

Panel C: Hispanic

Panel D: Male

Panel E: Female

Pre-2016 Reporting --- Post-2016 Reporting

Source: These figures report BJS published imprisonment rates by reporting method (Carson 2016, 2018; Bronson and Carson 2019; Carson 2020a, 2020b, 2021, 2022; Carson and Kluckow 2023).

Note: Pre-2016 reporting exclusively uses the 2015 BJS report and post-2016 reporting uses the latest available report as the value for each year.

Figure 10: Imprisonment rates by race/ethnicity, 2000–2021: unadjusted NPS rates, rates adjusted using BJS surveys, rates adjusted using CJARS-Census Bureau data

Panel A: White, non-Hispanic

Panel B: Black, non-Hispanic

Panel C: AIAN, non-Hispanic

Panel D: Asian/PI, non-Hispanic

Panel E: Other, non-Hispanic

Panel F: Hispanic

— · — · BJS Unadjusted Rate   – – – BJS Adjusted Rate   —— CJARS-Census Adjusted Rate

Source: Authors' calculations from the 2000 and 2010 Censuses, the Census Numident, CJARS criminal histories (2022Q4 vintage), and National Prisoner Statistics Program data. U.S. Census Bureau, project number P-7500378, approval numbers CBDRB-FY24-0101, CBDRB-FY24-0277.

Note: Figures show estimated imprisonment rates for state and federal prisoners using three methods. The BJS-unadjusted rate shows the original racial and ethnic distribution reported by agencies to BJS in the NPS. The BJS-adjusted rate shows their published per capita rates which rely on imputing racial and ethnic population shares based on inmate survey data. The CJARS-Census Bureau-adjusted rate rescales the original unadjusted data reported by agencies in the NPS using year-specific versions of Table 6. Estimates and sample sizes have been rounded according to Census Bureau DRB rules.