Race and Ethnicity (Mis)measurement in the U.S. Criminal Justice System

Keith Finlay
U.S. Census Bureau

Elizabeth Luh
Securities & Exchange Commission

Michael Mueller-Smith
University of Michigan & NBER

July 24, 2024

Race research in the justice system

Significant concern among researchers, policy-makers, and the general public about racial disparities and discrimination in the U.S. justice system

Underappreciated fact about criminal justice data:

– Race/ethnicity data often populated by law enforcement

– Used for serving arrests warrants, tracking population, and allocating institutionalized population

– Operational purposes take priority over downstream statistical use

What do we do in this project?

1. Qualitative interviews & document reviews to understand how policies and practices may affect quality of race/ethnicity fields in criminal justice administrative records
   - Interviews with 7 criminal justice personnel; document review of 6 practice guidelines
   - Together cover: AZ, FL, IL, MI, NJ, NY, OH, and TX

2. Document the degree of "mislabeling" in criminal justice records
   - Possible through individual-level linkage of CJARS to Census Bureau survey/admin data
   - Disaggregate to consider differences by procedural stage, across geography, and over time

3. Reassess federal imprisonment estimates by racial/ethnic subgroups
   - Apply estimated correspondence between agency-recorded and self/family-reported identity to counts from the National Prisoner Statistics program
   - Compare with raw NPS data and current BJS imputation practices

What do we do in this project?

1. Qualitative interviews & document reviews to understand how policies and practices may affect quality of race/ethnicity fields in criminal justice administrative records

   - Interviews with 7 criminal justice personnel; document review of 6 practice guidelines
   - Together cover: AZ, FL, IL, MI, NJ, NY, OH, and TX

2. Document the degree of "mislabeling" in criminal justice records

   - Possible through individual-level linkage of CJARS to Census Bureau survey/admin data
   - Disaggregate to consider differences by procedural stage, across geography, and over time

3. Reassess federal imprisonment estimates by racial/ethnic subgroups

   - Apply estimated correspondence between agency-recorded and self/family-reported identity to counts from the National Prisoner Statistics program
   - Compare with raw NPS data and current BJS imputation practices

What do we do in this project?

1. Qualitative interviews & document reviews to understand how policies and practices may affect quality of race/ethnicity fields in criminal justice administrative records
   - Interviews with 7 criminal justice personnel; document review of 6 practice guidelines
   - Together cover: AZ, FL, IL, MI, NJ, NY, OH, and TX

2. Document the degree of "mislabeling" in criminal justice records
   - Possible through individual-level linkage of CJARS to Census Bureau survey/admin data
   - Disaggregate to consider differences by procedural stage, across geography, and over time

3. Reassess federal imprisonment estimates by racial/ethnic subgroups
   - Apply estimated correspondence between agency-recorded and self/family-reported identity to counts from the National Prisoner Statistics program
   - Compare with raw NPS data and current BJS imputation practices

What do we find?

→ Race and ethnicity data is viewed skeptically by those involved in the justice system
  – Information frictions permit erroneous information to propagate across agencies
  – Disincentives exist to discourage amending the record

→ Approximately 17% of criminal court records and 10% of prison inmates have agency-recorded information that does not align with Census-recorded identification
  – Substantial geographic variation, less temporal variation

→ Current BJS reported incarceration rates by race and ethnicity may be flawed
  – White and Black inmate populations may be underestimated by 20% and 17% respectively
  – American Indian/Alaskan Native incarceration rate since 2010 underestimated by 46%

What do we find?

→ Race and ethnicity data is viewed skeptically by those involved in the justice system
  – Information frictions permit erroneous information to propagate across agencies
  – Disincentives exist to discourage amending the record

→ Approximately 17% of criminal court records and 10% of prison inmates have agency-recorded information that does not align with Census-recorded identification
  – Substantial geographic variation, less temporal variation

→ Current BJS reported incarceration rates by race and ethnicity may be flawed
  – White and Black inmate populations may be underestimated by 20% and 17% respectively
  – American Indian/Alaskan Native incarceration rate since 2010 underestimated by 46%

What do we find?

→ Race and ethnicity data is viewed skeptically by those involved in the justice system
  - Information frictions permit erroneous information to propagate across agencies
  - Disincentives exist to discourage amending the record

→ Approximately 17% of criminal court records and 10% of prison inmates have agency-recorded information that does not align with Census-recorded identification
  - Substantial geographic variation, less temporal variation

→ Current BJS reported incarceration rates by race and ethnicity may be flawed
  - White and Black inmate populations may be underestimated by 20% and 17% respectively
  - American Indian/Alaskan Native incarceration rate since 2010 underestimated by 46%

Outline

1. **Collecting race and ethnicity data in the criminal justice context**

2. Concordance rates of agency-recorded and Census-measured race and ethnicity

3. Reassessing federal statistics on imprisonment rates by race and ethnicity

4. Conclusion

Interviews and document review

Qualitative research to better understand:

1. When race and ethnicity information is collected,

2. How it is transmitted across actors and offices,

3. Who uses this information, and

4. Why errors may not be corrected

Convenience sample of jurisdictions/personnel where we have built relationships with individuals experienced in criminal justice records management
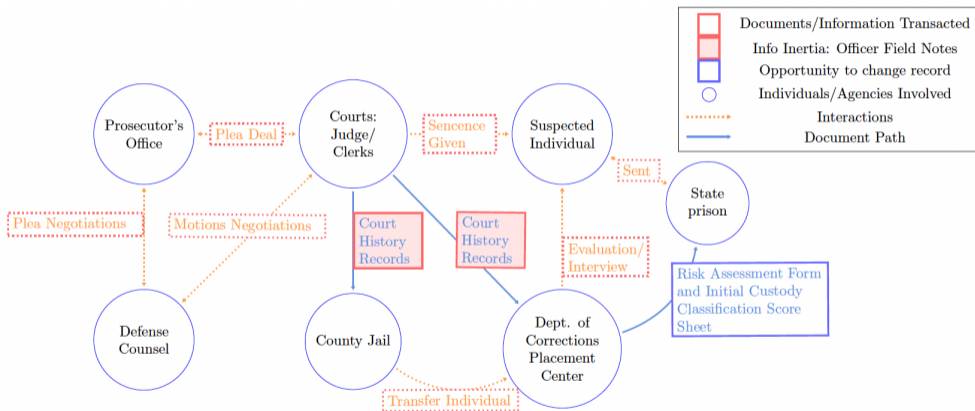
Sample characteristics

Source	State	Relevant procedural stage of the justice system
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
Guidelines B	Michigan	Department of corrections
Guidelines C	Michigan	State police
Guidelines D	Texas	All stages
Guidelines E	Texas	Criminal courts

Case evolution prior to disposition
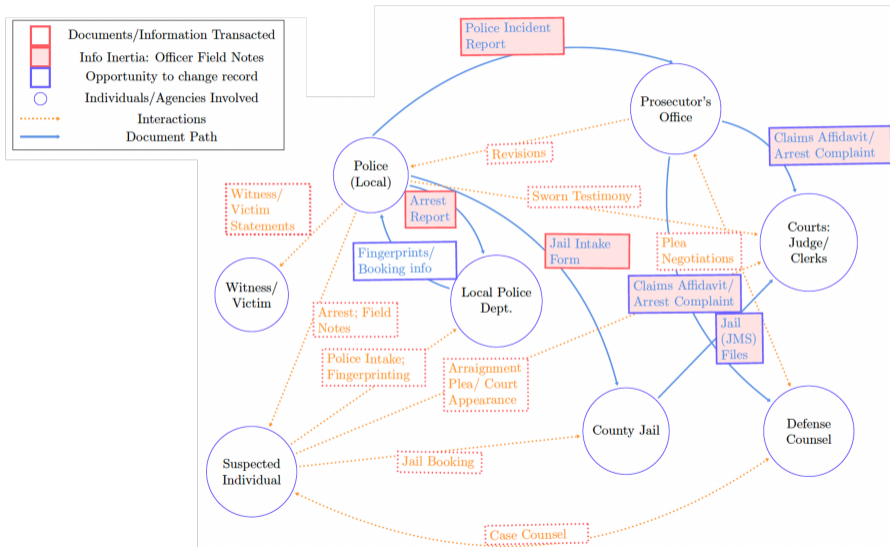
Case evolution following disposition

Broad takeaways

1. Information frictions arise due to the data decentralization
   - Information flows tend to be unidirectional, propagating errors and limiting recourse to amend the record

2. Internal views on the reliability of race and ethnicity data varies tremendously, depending on whether agencies view this information as operationally relevant

3. Correcting errors can be costly from a variety of perspectives, discouraging accurate record keeping

Broad takeaways

1. Information frictions arise due to the data decentralization
   – Information flows tend to be unidirectional, propagating errors and limiting recourse to amend the record

2. Internal views on the reliability of race and ethnicity data varies tremendously, depending on whether agencies view this information as operationally relevant

3. Correcting errors can be costly from a variety of perspectives, discouraging accurate record keeping

Broad takeaways

1. Information frictions arise due to the data decentralization
   - Information flows tend to be unidirectional, propagating errors and limiting recourse to amend the record

2. Internal views on the reliability of race and ethnicity data varies tremendously, depending on whether agencies view this information as operationally relevant

3. Correcting errors can be costly from a variety of perspectives, discouraging accurate record keeping

Outline

1. Collecting race and ethnicity data in the criminal justice context

2. Concordance rates of agency-recorded and Census-measured race and ethnicity

3. Reassessing federal statistics on imprisonment rates by race and ethnicity

4. Conclusion

Empirical exercise

Integrate CJARS (agency-recorded) race/ethnicity information with Census Bureau (self and family-reported) race/ethnicity microdata

To minimize measurement error, we create composite using modal identification across three sources:

1. Decennial Census 2000
2. Decennial Census 2010
3. Social Security Administration's Numident file

Responses may reflect individual or family reports of racial/enthic identification

Group Quarters responses and records with imputed race/ethnicity values are excluded since these may actually be agency-recorded values
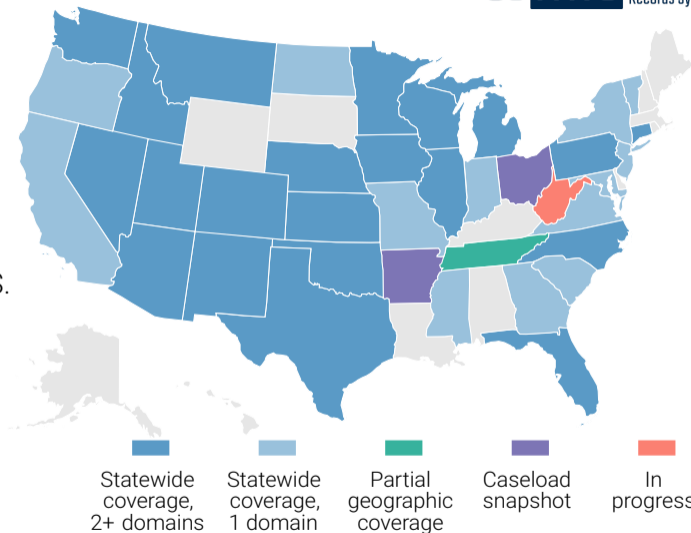
What is CJARS?

CJARS, founded in 2016, is a joint data infrastructure project between University of Michigan and the U.S. Census Bureau

Building a novel data platform to modernize research and statistical reporting on the U.S. criminal justice system, including:
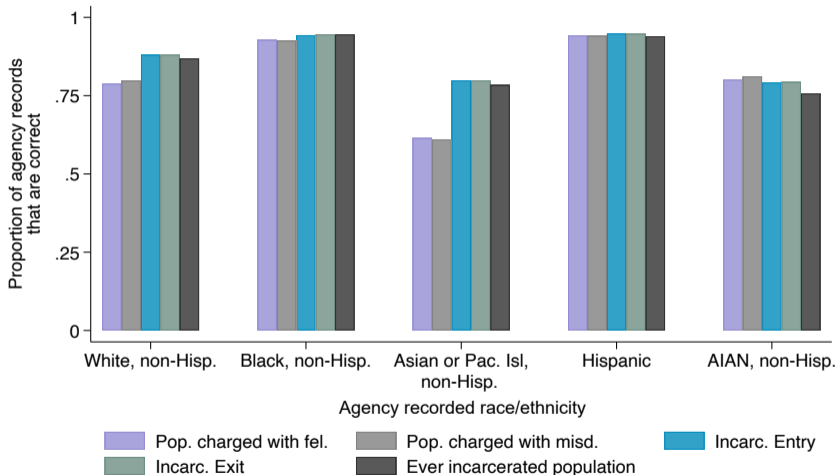
- Event-level criminal justice data with nationwide scope
- Tracking across key milestones in the justice system
- Capacity to link with individual-level survey and administrative data at the U.S. Census Bureau

CJARS data coverage

- Data from 2,635 counties, 38 states

- States with coverage of state court, state DOC, and/or state repository represent ~84% of U.S. population

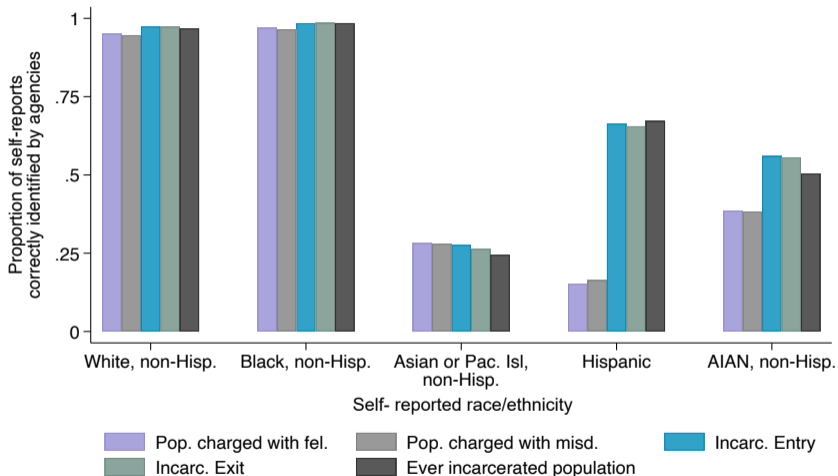- 3.5b records

- 215m CJ events

- 44m unique individuals

Statewide coverage, 2+ domains
Statewide coverage, 1 domain
Partial geographic coverage
Caseload snapshot
In progress

Share of agency-records correctly labeled

Share of composite-ID correctly labeled

Bar chart titled with y-axis "Proportion of self-reports correctly identified by agencies" ranging from 0 to 1, and x-axis "Self-reported race/ethnicity" with categories: White, non-Hisp.; Black, non-Hisp.; Asian or Pac. Isl, non-Hisp.; Hispanic; AIAN, non-Hisp.

Legend: Pop. charged with fel.; Pop. charged with misd.; Incarc. Entry; Incarc. Exit; Ever incarcerated population.

All results were approved for release by the Census Bureau, authorization numbers #CBDRB-FY24-0101 and #CBDRB-FY24-0277.

Fel/misd defendants correspondence matrix

Census Bureau race/ethnicity composite	Race/ethnicity recorded by justice system						% of agency-recorded equal to composite (7)	Caseload share (8)	Row total (9)
	White (1)	Black (2)	AIAN (3)	Asian/PI (4)	Other (5)	Hispanic (6)			
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

All results were approved for release by the Census Bureau, authorization numbers #CBDRB-FY24-0101 and #CBDRB-FY24-0277.

Prison inmate correspondence matrix

| Census Bureau race/ethnicity composite | Race/ethnicity recorded by justice system | | | | | | % of agency-recorded equal to composite (7) | Caseload share (8) | Row total (9) |
	White (1)	Black (2)	AIAN (3)	Asian/PI (4)	Other (5)	Hispanic (6)			
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

All results were approved for release by the Census Bureau, authorization numbers #CBDRB-FY24-0101 and #CBDRB-FY24-0277.

Additional exercises

1. Temporal variation, 2000-2019
   - Some instances of improving data quality, but largely unchanging over time
     smallskip

2. Geographic variation
   - Coverage of Hispanic and AI/AN populations tends to be stronger in areas with high
     prevailing rates in the resident populations

3. Concordance rates of sex fields

4. Comparison with racial/ethnic concordance in Medicaid records

5. Consistency of Census composite over source data files

Outline

1. Collecting race and ethnicity data in the criminal justice context

2. Concordance rates of agency-recorded and Census-measured race and ethnicity

3. Reassessing federal statistics on imprisonment rates by race and ethnicity

4. Conclusion

BJS "Prisoners" statistical series

Regular statistical reports on the size and composition of the carceral population since 1980 based on data reported through the National Prisoner Statistics program

OMB guidelines from 1997 required the collection of information on Hispanic origin in addition to data on race

– Preferred measure is self-reported identity although exceptions are allowed

Data submitted through NPS does not meet OMB guidelines

– Many DOCs have no measure of Hispanic ethnicity

– BJS imputes race/ethnicity based on available inmate surveys

BJS surveys for race/ethnicity imputation

Race/ethnicity	2004 SISFCF		2016 SPI	
	State	Federal	State	Federal
White, non-Hispanic	36.5%	29.2%	31.9%	21.0%
Black, non-Hispanic	40.0%	39.0%	33.8%	32.2%
Hispanic	17.5%	24.7%	20.7%	37.1%
American Indian/Alaskan Native, non-Hispanic	1.7%	2.8%	1.4%	1.7%
Asian/Pacific Islander, non-Hispanic	1.0%	1.7%	0.9%	1.5%
Multiple races reported, non-Hispanic	3.3%	2.6%	11.3%	6.5%
Sample size	14,477	3,680	20,064	4,784

Impact of the 2016 SPI

– Clear and growing
impact on racial and
ethnic measurement

White incarceration rate
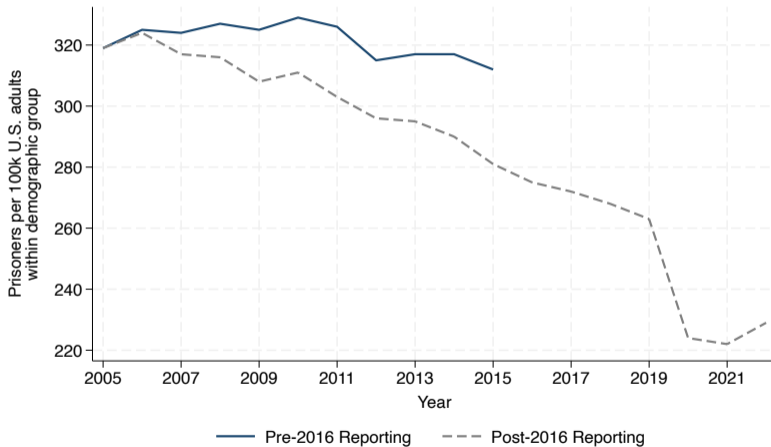
Impact of the 2016 SPI

– Clear and growing impact on racial and ethnic measurement

Black incarceration rate
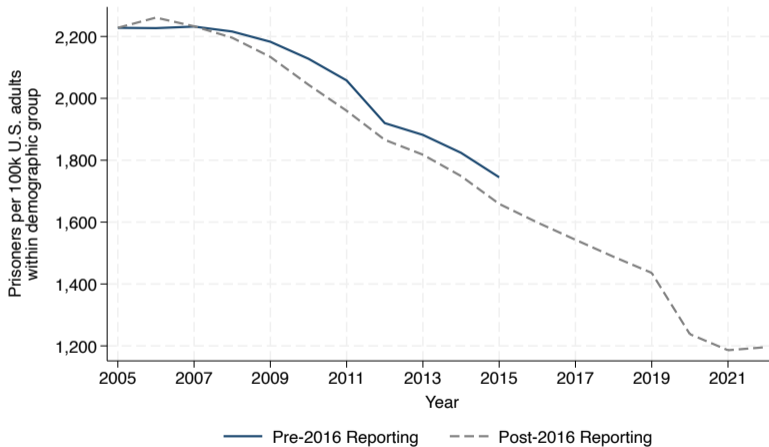
Impact of the 2016 SPI

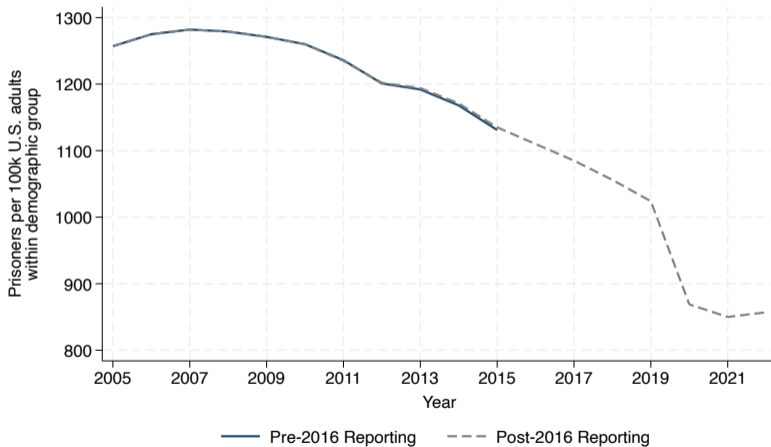– Clear and growing impact on racial and ethnic measurement

Hispanic incarceration rate
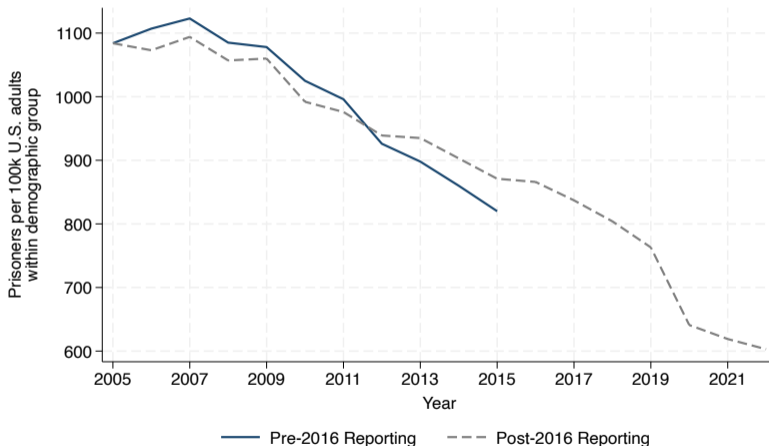
Impact of the 2016 SPI
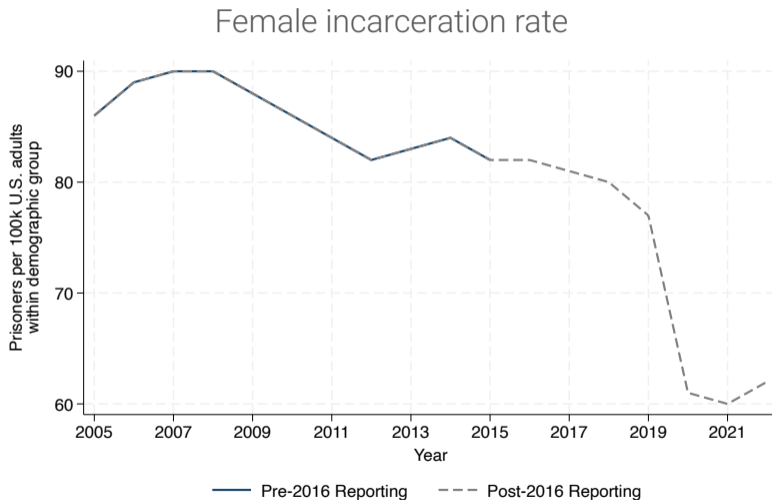
– No impact on other
  demographic
  breakouts like sex

Male incarceration rate

Impact of the 2016 SPI

– No impact on other
demographic
breakouts like sex

Female incarceration rate

Alternative strategy

Develop alternative strategy that leverages the original NPS raw race and ethnicity counts

Combine with annual estimated correspondence matrices based on linked CJARS records

Assumption:

- Error structure observed in CJARS-covered jurisdictions is externally valid to non-covered jurisdictions

Original NPS, BJS-imputed, CJARS-adjusted

White incarceration rate

- CJARS-adjustment tracks trend of original NPS reports, but at a lower level

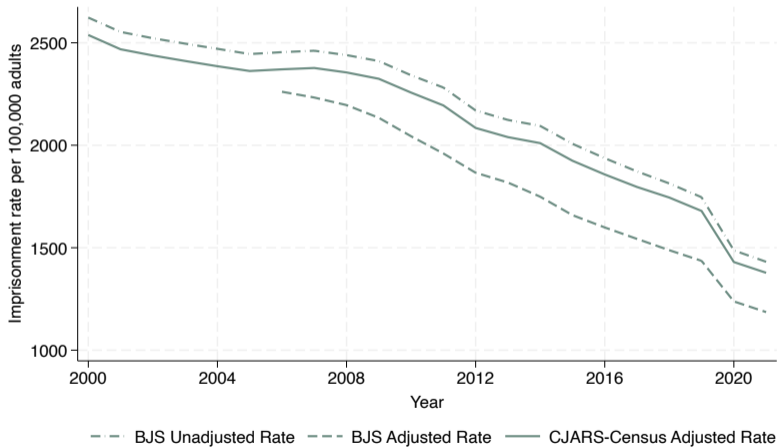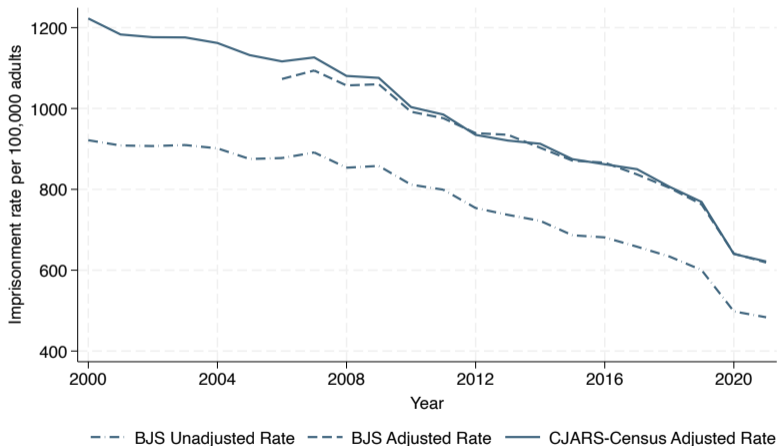- BJS imputation substantially lower with widening gap over time
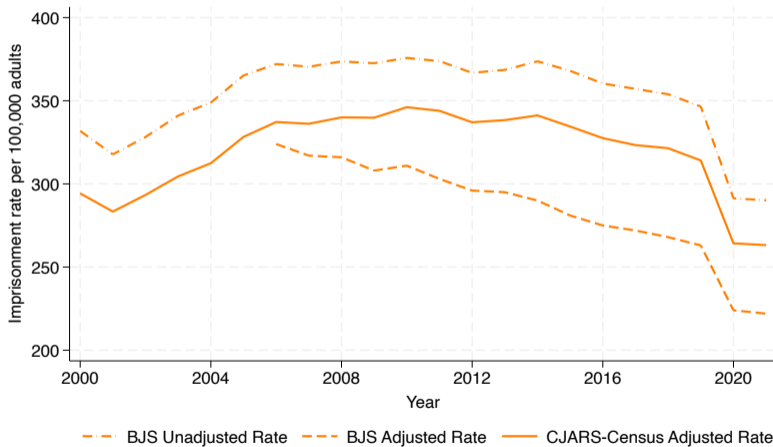
All results were approved for release by the Census Bureau, authorization numbers #CBDRB-FY24-0101 and #CBDRB-FY24-0277.

Original NPS, BJS-imputed, CJARS-adjusted

Black incarceration rate

- CJARS-adjustment tracks trend of original NPS reports, but at a lower level

- BJS imputation substantially lower with widening gap over time

All results were approved for release by the Census Bureau, authorization numbers #CBDRB-FY24-0101 and #CBDRB-FY24-0277.

Original NPS, BJS-imputed, CJARS-adjusted

Hispanic incarceration rate

– CJARS-adjustment
and BJS imputation
similarly resolve issue
of missing Hispanic
information in original
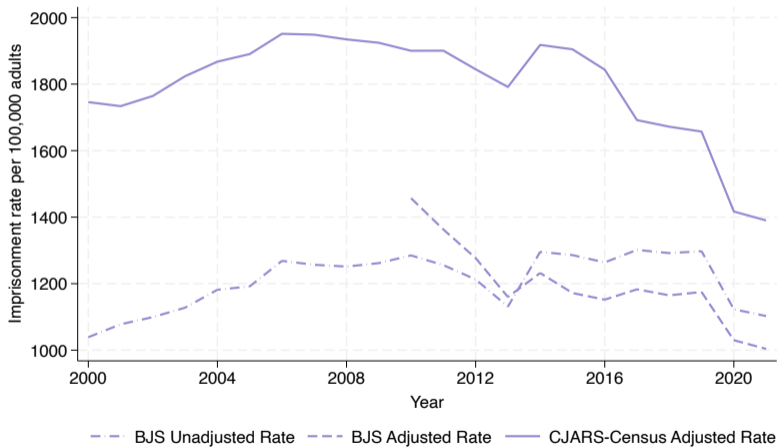reported NPS data

All results were approved for release by the Census Bureau, authorization numbers #CBDRB-FY24-0101 and #CBDRB-FY24-0277.

Original NPS, BJS-imputed, CJARS-adjusted

– CJARS-adjustment shows substantial underreporting of incarceration within the American Indian and Alaskan Native communities
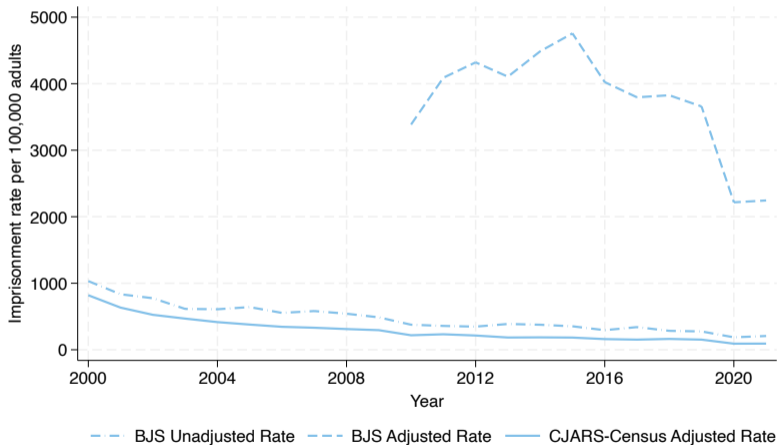

American Indian/Alaskan Native incarceration rate

All results were approved for release by the Census Bureau, authorization numbers #CBDRB-FY24-0101 and #CBDRB-FY24-0277.

Original NPS, BJS-imputed, CJARS-adjusted

- BJS imputation strategy implies high incarceration for non-Hispanic, multi-racial and other race group

- Rate twice as high as non-Hispanic Black adults

Implied "Other" incarceration rate

All results were approved for release by the Census Bureau, authorization numbers #CBDRB-FY24-0101 and #CBDRB-FY24-0277.

Other SES measures from ACS, 2005-2019

	White*	Black*	AIAN*	Asian or Pac. Isl.*	Other*	Hispanic
Educational attainment:						
High school graduate	0.94	0.88	0.84	0.92	0.91	0.72
Some college	0.58	0.46	0.4	0.69	0.56	0.34
College graduate	0.32	0.18	0.13	0.49	0.27	0.13
Employment/earnings:						
Unemployed	0.04	0.09	0.09	0.04	0.07	0.06
Earned income	$37,857	$23,693	$20,576	$41,099	$28,994	$23,340
Total family income	$96,807	$59,874	$59,659	$110,250	$82,007	$67,506
Family income below federal poverty line	0.08	0.18	0.21	0.10	0.13	0.16
Household structure:						
Single-parent household	0.09	0.23	0.16	0.08	0.15	0.15

Outline

Conclusion

1. Highlight structural factors that contribute to low quality race/ethnicity information in criminal justice administrative data
   - Major implications for researchers studying racial disparities and discrimination

2. Measure the degree of concordance/discordance between agency-recorded and self/family-identified race/ethnicity
   - Identify several important dimensions of heterogeneity: procedural stage and geography

3. Apply novel estimates to reevaluate imprisonment rates over the last two decades
   - Identify potential federal underestimates, which have become more pronounced over time
   - Develop new strategy leveraging data advances to provide actionable solution