



Justice Outcomes Explorer (JOE) Technical Documentation

The CJARS Team

University of Michigan

Vintage: 2023Q3, Version: v1

Abstract

The Justice Outcomes Explorer (JOE) is a new dataset and data visualization interface for understanding the scope and impact of the criminal justice system in the United States. Data from the Criminal Justice Administrative Records System (CJARS) have been linked with extensive socioeconomic data at the U.S. Census Bureau to provide an unprecedented window into the outcomes of people charged with misdemeanors or felonies, released from prison, or beginning probation or parole sentences. From recidivism to employment rates, public benefit take-up to mortality risk, and caseload size to sentencing, JOE aims to capture holistically how the justice system shapes lives and communities across the United States. Users can leverage the Explorer to compare outcomes across jurisdictions, view changes over time, or follow the evolution of outcomes for specific cohorts. Statistical series can be broken down by demographic group to consider differences by race, ethnicity, sex, or age. Criminal history information from CJARS allows users to examine the distinct trajectories of people with first versus repeat offenses. Outcomes can be contextualized relative to local crime rates, poverty levels, or public health environments. This document provides an overview of the project and details on the methodology used to create the statistics available on the website and for download.

The project website is <https://joe.cjars.org>. Data users who have questions about CJARS can contact cjars-joe@umich.edu.

Any conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. *All statistics in this document are based on CJARS data at the University of Michigan except for those explicitly labeled as sourced from Census Bureau data protected by 13 USC §9a.*



Contents

1	Project overview	5
1.1	Project description and objectives	5
1.2	Data sources	5
1.2.1	CJARS database	5
1.2.2	Census Bureau’s Federal Statistical Research Data Center (FSRDC) network	5
1.2.3	Context data	6
1.3	Coverage	6
1.4	Principal investigators	6
1.5	Project funding	6
1.6	Data citation	6
2	JOE Methodology	10
2.1	Overview	10
2.2	Creating cohorts and outcomes	10
2.2.1	CJARS Microdata	10
2.2.2	Census Microdata	10
2.2.3	Subpopulations	11
2.2.4	Imputation	11
2.3	Re-weighting outcome statistics by repeat contact	11
2.4	Geographies	11
2.4.1	County groups	12
2.4.2	Estimation groups	12
2.4.3	Geographical aggregations	13
2.5	Disclosure avoidance methodology	13
2.6	Technical notes	13
3	Context data	15
	Appendices	16
A	Socioeconomic Outcome Methodology	16
A.1	Introduction	16
A.2	Overview	16
A.3	Data sources and record linkage	17
A.3.1	Criminal Justice Administrative Records System (CJARS)	17
A.3.2	IRS W-2 information returns	17
A.3.3	Social Security Administration (SSA) Supplemental Security Record (SSR) data	17
A.3.4	Department of Housing and Urban Development Longitudinal Public and Indian Housing Information Center (PIC)/Tenant Rental Assistance Certification System (TRACS) data	17
A.3.5	Centers for Medicare and Medicaid Services (CMS) Medicaid Statistical Information System (MSIS) and Transformed Medicaid Statistical Information System (T-MSIS) data	18
A.3.6	CMS Electronic Database (EDB) data	18
A.3.7	Census Numident	18
A.3.8	Census Bureau Title 13 BestRace files	18
A.3.9	Record linkage	18
A.4	Concepts and methodology	19
A.4.1	Geographies	20
A.4.2	Using longer windows for cohorts with small populations	20
A.4.3	Removing cohorts with small populations	20
A.4.4	Reweighting to account for record linkage selection	21
A.4.5	Choropleth-like geographic aggregation of counties	22
A.4.6	Re-estimation within choropleth groupings	22

A.4.7 Rounding	23
A.5 Disclosure avoidance methodology	23
A.6 Reliability of the data	23
A.7 Glossary	23
A.8 References	24
B Code schemes	26
B.1 Geographic and demographic codes	26
B.1.1 Sex codes	26
B.1.2 Race and ethnicity codes	26
B.1.3 Age group codes	26
B.1.4 Offense type codes	26
B.1.5 Repeat contact codes	27
C Data notes by state	28
C.1 Idaho data notes	28
C.2 Minnesota data notes	28
C.3 Nevada data notes	28
C.4 Washington data notes	28
Bibliography	29
Index	30

List of Tables

1	List of statistics created from CJARS data along with the CJARS variable used to define cohort year.	11
2	Description of subpopulation breakdowns available in the data.	12
3	List of context statistics which are presented as time averages in the Data in Context section on the JOE portal.	15
6	Sex codes	26
7	Race and ethnicity codes	26
8	Age group codes	26
9	Offense type codes	27
10	Repeat contact codes	27

1 Project overview

1.1 Project description and objectives

The Justice Outcomes Explorer (JOE) is a new dataset and data visualization interface for understanding the scope and impact of the criminal justice system in the United States. Data from the Criminal Justice Administrative Records System (CJARS) at the University of Michigan have been linked with extensive socioeconomic data at the U.S. Census Bureau to provide an unprecedented window into the outcomes of people charged with misdemeanors or felonies, released from prison, or beginning probation or parole sentences. From recidivism to employment rates, public benefit take-up to mortality risk, and caseload size to sentencing, JOE aims to capture holistically how the justice system shapes lives and communities across the United States.

Users can leverage the Explorer to compare outcomes across jurisdictions, view changes over time, or follow the evolution of outcomes for specific cohorts. Statistical series can be broken down by demographic group to consider differences by race, ethnicity, sex, or age. Criminal history information from CJARS allows users to examine the distinct trajectories of people with first versus repeat offenses. Outcomes can be contextualized relative to local crime rates, poverty levels, or public health environments.

1.2 Data sources

There are three primary sources of data that are used to generate the statistical information available through JOE. They include the CJARS database, the Census Bureau's Federal Statistical Research Data Center (FSRDC) network and a variety of publicly available data resources.

1.2.1 CJARS database

The CJARS database serves as the basis for producing the statistics that are available through JOE. The 2023Q3 vintage of CJARS was used to identify the cohorts based on dates of misdemeanor charges and dispositions, felony charges and dispositions, prison entries and exits, probation starts and ends, and parole starts and ends. CJARS data do not cover all states and counties; this data reflects geographies covered by the 2023Q3 vintage. Similarly, CJARS longitudinal coverage varies; cohorts for JOE data are identified starting in 2000 or as early as are available after that.

The CJARS data documentation outlines the processes and data quality checks that are used to harmonize raw criminal justice administrative data into a high-quality dataset (Finlay and Mueller-Smith 2022). Every CJARS vintage is used to produce aggregate statistics that can be benchmarked against those published by the Bureau of Justice Statistics. These comparisons show that CJARS data are high quality and can reproduce aggregate statistics on the caseloads of criminal justice agencies (Papp and Mueller-Smith 2022).

1.2.2 Census Bureau's Federal Statistical Research Data Center (FSRDC) network

CJARS records are linked to other data resources available within the Census Bureau's Federal Statistical Research Data Center (FSRDC) network. Data linkage occurs by using an anonymous identification key called a Protected Identification Key (PIK) which is used to identify unique individuals in the data available within the FSRDC network. These datasets include Internal Revenue Service (IRS) W-2 information returns; Social Security Administration (SSA) Supplemental Security Record (SSR) data; Department of Housing and Urban Development Longitudinal Public and Indian Housing Information Center (PIC) and Tenant Rental Assistance Certification System (TRACS) data; Centers for Medicare and Medicaid Services (CMS) Medicaid Statistical Information System (MSIS) and Transformed Medicaid Statistical Information System (T-MSIS) data; CMS Electronic Database (EDB) data; Census Numident data and Census Bureau Title 13 BestRace files. All statistics that use these Census Bureau derived datasets are reviewed by the Census Bureau Disclosure Review Board in accordance with U.S. Code, Title 13, Section 9.

1.2.3 Context data

Publicly available data resources are leveraged to provide supplemental information against which criminal justice statistics and associated outcomes can be contextualized. The data resources used to generate the contextual data available include: the U.S. Census Bureau, the Federal Bureau of Investigation (FBI) Uniform Crime Reporting Program (UCR), the Federal Reserve Economic Data (FRED) from the Federal Reserve Bank of St. Louis and the County Health Rankings and Roadmaps (CHRR).

1.3 Coverage

JOE statistics are limited to the coverage of CJARS data and available Census data. JOE statistics also include the following further coverage restrictions:

- Cohort years before 2000 are not included.
- Partially covered years are not included (i.e., all 12 months of a year must be covered in the CJARS database).
- State-level statistics are only included when the geographic coverage of the CJARS database contains 80% or more of the state population in a given year.
- For recidivism statistics, the cohort coverage is described above. The recidivating event is considered covered if the state in which the initial contact with the criminal justice system has coverage at the time of followup.

For an overview of JOE coverage by state, year and type of criminal justice event, see Figure 1.

All statistical information that is generated for JOE undergoes extensive data quality review to ensure high fidelity. This is accomplished by implementing both automated and manual review systems. We accomplish this by deploying algorithms that analyze statistics and flag those that meet criteria that we have pre-defined as potentially problematic. Examples include outliers, significant changes in a caseload over time, large variation across geographic regions, etc. Statistics that are flagged by our automated systems are then reviewed manually by CJARS staff. Statistics confirmed to be problematic are redacted. Additionally, statistics generated that rely on the value of the redacted statistic are removed to avoid issues from propagating through the system.

1.4 Principal investigators

Michael Mueller-Smith. Assistant Professor, Department of Economics, University of Michigan; Faculty Associate, Population Studies Center, University of Michigan.

Keith Finlay. Research Economist, U.S. Census Bureau.

1.5 Project funding

The Justice Outcomes Explorer has been supported by the [National Science Foundation](#), [Arnold Ventures](#), the [Bill & Melinda Gates Foundation](#), the [Robert Wood Johnson Foundation](#), [The Annie E. Casey Foundation](#), and the [U.S. Census Bureau](#).

1.6 Data citation

Publications and research reports based on the JOE database should cite it appropriately. The citation should include the following:

Keith Finlay, Michael Mueller-Smith, and the CJARS team. 2024. Justice Outcomes Explorer (JOE) [dataset]. Ann Arbor, MI: University of Michigan. <https://joe.cjars.org>.

2 JOE Methodology

2.1 Overview

The JOE statistics are produced by aggregating administrative criminal justice microdata collected and processed by the CJARS team (Finlay and Mueller-Smith 2023). JOE data is currently based on the third quarter 2023 vintage (“2023Q3”) of CJARS data. The data are aggregated into cohorts based on five event types: felony case filings, misdemeanor case filings, incarceration, parole and probation. For each event type and year starting in 2000, we produce a per capita statistic based on the geographic population and size of the cohort and report relevant outcomes for the cohort. With the ability to track individuals through time, we report various outcomes for followup periods of one through five years after the initial interaction with the criminal justice system. Statistics are available for a variety of subpopulations and geographic breakdowns.

Section 2 starts by detailing the definition of cohorts for each event type from the CJARS microdata. It continues on to define each outcome reported on JOE. Next it discusses the variety of ways the data is grouped, including by demographics, previous interaction with the criminal justice system, and geography. Finally, it touches on the methods used to avoid disclosure of individual’s protected information. Technical notes are provided for users who want extra details on the production of statistics.

2.2 Creating cohorts and outcomes

2.2.1 CJARS Microdata

The CJARS database is used to create cohorts for each covered year and geographic region by totaling the individual cases for each subpopulation breakdown in that year. Please refer to technical note 1 for how CJARS data is turned into cohorts. See Table 1 for details on how the years are defined for cohorts. We provide per capita statistics, outcomes and recidivism rates for these cohorts.

Per capita statistics (felony charge rate, misdemeanor charge rate, incarceration population rate, parole entry rate, and probation entry rate) are calculated as size of caseload cohorts, broken down by subpopulation, divided by the geographical resident subpopulation, multiplied by 100,000.

Spell lengths for incarceration, parole and probation are the length of time between the entry and exit date of supervision. Processing times for felony and misdemeanor cases are the length of time between the filing date (if missing, offense date) and disposition date (if missing, sentence date) of a charge. Felony or misdemeanor charges that resulted in conviction are reported as the proportion of felony or misdemeanor charges with a guilty sentence.

Recidivism rates are calculated as the proportion of the cohort that has individuals who have re-entered the criminal justice system via the specified recidivating event within a specified followup period designated as years since their event date. Note that due to the structure of CJARS microdata, recidivism can only be observed if the next interaction with the criminal justice system happens in the same state as the first interaction.

2.2.2 Census Microdata

CJARS microdata are linked with Census microdata via a Protected Identification Key (PIK), an anonymous identifier that is assigned to data resources that have been incorporated into the FSRDC network. Socioeconomic outcomes for individuals are then identified from population-level administrative data, while demographic data come from population-level survey and administrative data.

Socioeconomic outcomes are processed in the Federal Statistical Research Data Center (FSRDC) and exported for further processing by CJARS staff at the University of Michigan. This export, which includes county and state aggregated statistics for the same subpopulations produced for CJARS-based statistics, can be found as a Census Bureau [Experimental Data Product](#). Details on the production of this exported socioeconomic data can be found in Appendix A. Please refer to the technical note 2 for details on cohort year definition and cohort size production for these socioeconomic outcome statistics.

Table 1: List of statistics created from CJARS data along with the CJARS variable used to define cohort year.

Statistic	Cohort Year
Incarceration population rate	between entry and exit date
Incarceration spell length	exit date
Recidivism from incarceration	exit date
Parole entry rate	entry date
Parole spell length	exit date
Recidivism from parole	entry date
Probation entry rate	entry date
Probation spell length	exit date
Recidivism from probation	entry date
Felony charge rate	charge filing date (if missing, disposition then sentencing then offense date)
Felony processing time	charge disposition date (if missing, sentence date)
Felony share conviction	charge disposition date (if missing, sentence date)
Recidivism from felony charge	charge disposition date (if missing, sentence date)
Misdemeanor charge rate	charge filing date (if missing, disposition then sentencing then offense date)
Misdemeanor processing time	charge disposition date (if missing, sentence date)
Misdemeanor share conviction	charge disposition date (if missing, sentence date)
Recidivism from misdemeanor charge	charge disposition date (if missing, sentence date)

Please note that currently all socioeconomic data on JOE are derived exclusively from the 2022Q4 CJARS vintage, regardless of the selected vintage, due to the time delay caused by the thorough disclosure review process prior to exporting data from the FSRDC network.

2.2.3 Subpopulations

Statistics are provided for subpopulations with differing levels of experience with the justice system and for different offense types. We sort experience with the criminal justice system into two categories: first contact and repeat contact. Statistics are further calculated for three types of demographic subpopulations including age, sex and race/ethnicity. See Table 2 for details on these subpopulations and technical note 3 for further discussion.

2.2.4 Imputation

For some geographies, incarceration entry and exit date are imputed in order to calculate more accurate population rates. Entry dates are estimated using the exit date and subtracting the median incarceration spell length for a given state and offense type. Life sentences are carefully excluded from being assigned exit dates. Similarly, missing exit dates are estimated using the entry date and adding the 50th percentile incarceration spell length for a given state and offense type.

2.3 Re-weighting outcome statistics by repeat contact

The calculation of some outcome statistics can be overly influenced by individuals that have high levels of contact with the criminal justice system. For this reason, we calculate re-weighted statistics that take into consideration the occurrence of individuals that have frequent contact with the system. Our nominal outcome statistic (repeat_contact=0) is simply the average statistic for the entire cohort. In addition to this, we provide another value for each total outcome statistic (repeat_contact=3) which de-weights cases that are repeat contacts with the criminal justice system. See technical note 4 for details on this reweighting.

2.4 Geographies

Event geographies are determined by the state and county of sentencing. Note that these may not be the states and counties of residence for people in a cohort.

Table 2: Description of subpopulation breakdowns available in the data.

Description	Variable Name	Category	Variable Value
Contact with criminal justice system	repeat_contact	All	0
		First contact	1
		Repeat contact	2
		Re-weighted all	3
Disposition (if missing, charge) offense type	off_type	All	0
		Violent	1
		Property	2
		Drug	3
Age	age_group	All	0
		15-24	1
		25-39	2
		40+	3
Sex	sex	All	0
		Male	1
		Female	2
Race/Ethnicity	race	All	0
		White	1
		Black	2
		Asian or Pacific Islander	3
		Hispanic	4
		American Indian or Alaska Native (AIAN)	5

2.4.1 County groups

Many counties in the United States have very few people entering the criminal justice system each year. In order to retain statistics while maintaining privacy, small counties are pooled into “county groups” (“county aggregations” in Appendix A) to create aggregate statistics. Each county within a county group is assigned the county group statistical value. Note that the cohort size associated with the county is still reported at the individual county level. The downloadable supplemental file contains a variable (variable name “f_CG”) which flags county statistics that belongs to a multi-county county group. To match counties with their county group for any statistic, please refer to the downloadable “county_to_CG_xwalk” crosswalk file. County group numerical IDs are six digits, ending with a zero if the county group contains multiple counties and ending with a one if it contains a single county.

The following is a description of the algorithm used to create county groups. Counties with total populations greater than 80,000 people are kept as singular entities (i.e. are in a county group consisting of only one county). The remaining counties are grouped into their census designated commuting zones (CZs). If the CZ crosses state lines, it is split into separate CZs for each state. If this (state delineated) CZ has a population greater than 80,000 people, then it is considered a county group. Finally, the remaining CZs in each state are all grouped together as the final county group of that state. Statistics for each county group are calculated via weighted averages of the counties within it. The weights are the county population for per capita statistics and the cohort size for outcome statistics. Finally, the averaged value for the county group is assigned to all of its contained counties.

2.4.2 Estimation groups

Estimation groups (“chloropleth groupings” in Appendix A) are aggregations of county groups with similar statistical values. Unlike counties within county groups, estimation groups do not necessarily contain geographically clustered county groups. For a full description of estimation groups, please refer to the section A.4.5.

Estimation grouping was only performed on socioeconomic outcome statistics due to their origination from Census data. Note that each county group (and therefore subsequent county) within an estimation group is assigned the estimation group statistical value, but the cohort size reported for the county is at the individual county level.

The downloadable supplemental file contains a variable (variable name “f_EG_statistic variable name”) which flags county statistics that belong to a multi-county-group estimation group. The file also contains a variable (variable name “EG_statistic variable name”) which provides the estimation group bin number. Each relevant statistic has its own flag and estimation group bin number variable, as they are not uniform across statistics.

2.4.3 Geographical aggregations

Beyond county and state statistics which are calculated directly from CJARS microdata, commuting zone (CZ) and nationwide aggregations are also provided.

Census designated CZs split at state lines are used. State-delineated CZs are necessary due to the state-delineation of the CJARS microdata; see technical note 1. The FIPS codes assigned to CZs are the two digit state FIPS followed by the three digit CZ FIPS. Statistics are calculated via weighted averages of the covered counties within state delineated CZs. The weights are the county population for per capita statistics and the cohort size (see Table 1) for outcome statistics. To match counties with their commuting zone, please refer to the downloadable “county_to_CZ_xwalk” crosswalk file.

Nationwide statistics are compiled via weighted averages of county data for adjudication based statistics and state data for incarceration, parole and probation based statistics. Note that these nationwide statistics are only based on counties or states that we have coverage over.

2.5 Disclosure avoidance methodology

Disclosure avoidance is the process used to protect the confidentiality of data and ensure individual privacy. A number of disclosure avoidance methods were used to protect the identity and characteristics of individuals in the input data:

- County groups: Counties with total populations less than 80,000 people are grouped.
- Rounding: At the end of data processing, a variety of rounding rules are applied to each statistic’s value and the associated population or cohort size.
- Removal of small populations (cohorts): Any observations with a population (cohort size) under 20 when generating rates (outcome statistics) are entirely omitted from data processing. This is stipulated based on the observation’s geographical location and demographic specifics.
- Limits to the interactions of demographic subpopulations: Only two-way interactions of the three demographic categorical variables (age groups, race/ethnicity, and sex) are calculated. Three-way interactions are not provided.

2.6 Technical notes

The following are technical notes on the processing of JOE statistics:

1. When this document refers to "individuals" or "cohorts", it is technically referring to "cjars_ids". From CJARS documentation, cjars_ids are assigned via "an algorithm that probabilistically matches records to individuals when no unique identifier is available by using names and dates of birth to identify individuals in the data. Once an individual has been identified, they are assigned an anonymized individual identifier (cjars_id)." One notable feature is that cjars_ids are defined within each state; a person who has contact with the criminal justice system in two different states would be assigned two different cjars_ids.
2. The nature of secure data processing and Census Bureau disclosure avoidance policies do not allow for release of cohort sizes. Cohort sizes for socioeconomic outcome related statistics are therefore estimates. For example, Census based statistics used Census-derived data for determining individual demographic information, but the reported cohort sizes use CJARS-derived data for determining demographic information. In addition, for felony and misdemeanor cohorts, outcome statistic cohort years are based on charge disposition date, but the reported cohort sizes are based on charge filing date. Incarceration socioeconomic outcome statistics as well as their cohort sizes are based on incarceration exit. Parole and probation socioeconomic outcome statistics as well as their cohort sizes are based on parole and probation entry dates.
3. Within CJARS microdata, if the year of date of birth is known but the month and day are unknown, the month is set to July and the date is set to the first. If multiple sexes, races/ethnicities, or dates of birth are

used for the same individual for different cases, the most frequent listing is used. If there is a tie for the most frequent listing, the minimum value (see variable values in Table 2) of the most frequent listings is used.

4. Repeat contact weights are created using an algorithm that counts the number of cases in each cohort that are repeat contacts within a five year window centered on the cohort year. Note that the year associated with a cohort used to calculate the new weights is always based on the entry date or filing date (if missing, disposition, sentence or offense date) for all statistics including outcomes regardless of the cohort listed in Table 1.

Socioeconomic outcome re-weighting by repeat_contact at the county level uses the weighted average of repeat_contact=1 and repeat_contact=2 even when those two statistics were not in the same estimation group. Therefore, an estimation group is not reported for repeat_contact=3, but the estimation group flag is still set to 1 as the statistic was created from base statistics that were a part of a defined estimation group.

5. Section A.4.2 describes using two-year windows for cohorts with small populations. In the data downloads, this is represented with a flag. The socioeconomic outcome technical document (Appendix A) states that "All state-level statistics use one-year windows of justice events". However, some post-processing on the statistics was performed which calculated state-level statistics by aggregating the counties within that state. Because some of these counties were calculated using two-year windows, some state statistics do end up using two-year windows of justice events.

3 Context data

Additional publicly available data resources were collected and minimally processed to be included in the Data in Context section on the JOE portal for comparison and contextualization of criminal justice caseload and socioeconomic outcome statistics. One data point is provided for each county, commuting zone (as defined in this document), or state, associated with a temporal average over the data. See Table 3 for details.

Table 3: List of context statistics which are presented as time averages in the Data in Context section on the JOE portal.

Statistic	Averaged Years	Source	Definition
Poverty rate	2010 to 2021	US Census Bureau: SAIPE (United States Census Bureau Small Area Income and Poverty Estimates (SAIPE) Program 2021)	Rate of people below the poverty line, regardless of age
Property crime rate	2010 to 2016	FBI: UCR (United States Department of Justice. Federal Bureau of Investigation. 2019)	Burglary, larceny or motor vehicle theft
Violent crime rate	2010 to 2016	FBI: UCR (United States Department of Justice. Federal Bureau of Investigation. 2019)	Murder, rape, robbery or aggravated assault
Unemployment rate	2010 to 2022	FRED (U.S. Bureau of Labor Statistics 2010)	Rate of unemployed population that is a part of the labor force
Health index	2010 to 2022	CHRR (University of Wisconsin Population Health Institute 2022)	CHRR health index representing the level of health related to physical and mental well-being. A value greater than zero indicates that the region is less healthy than average

Appendices

A Socioeconomic Outcome Methodology

Appendix A is an insertion of the technical documentation written by Keith Finlay describing the production of the socioeconomic statistics in the Census Bureau’s Federal Statistical Research Data Center (FSRDC).

A.1 Introduction

The Justice Outcomes Explorer (JOE) provides data on the outcomes of people who have interacted with the criminal justice system. JOE statistics measure the socioeconomic outcomes of people charged with criminal offenses, released from prison, or who began probation or parole. JOE is a product of the U.S. Census Bureau and the University of Michigan, which uses Criminal Justice Administrative Records System (CJARS) data to better understand the employment, program participation, and health outcomes of justice-involved people.

JOE covers justice events starting in 2000 and outcomes between 2001 and the present. Actual coverage of justice events depends on the coverage of agencies in the CJARS data, while coverage of outcomes depends on the availability of source data at the Census Bureau.

These statistics are part of a broader ensemble of statistics, which includes statistics about criminal justice caseloads and their justice outcomes that are produced by the University of Michigan without using Census Bureau data.

JOE is a Census Bureau [Experimental Data Product](#). Experimental Data Products are innovative statistical products created using new data sources or methodologies that benefit data users in the absence of other relevant products. We are seeking feedback from data users and stakeholders on the quality and usefulness of these new products. Census Bureau experimental data may not meet all of our quality standards. Because of this, we clearly identify experimental data products and include methodology and supporting research with their release.

A.2 Overview

Each statistic is calculated for a cohort associated with a justice event, a period of time since the justice event, and a socioeconomic outcome. Justice event cohorts are identified as the people associated with specific justice events (misdemeanor charge, felony charge, prison release, probation start, parole start) linked with a geography (county, county aggregation, or state) and a time period (1-year window of events for larger cohorts or 2-year window of events for smaller cohorts). Outcome statistics include:

- % employed in a year,
- average annual employment earnings,
- % with annual employment earnings above the poverty threshold for a single person under age 65,
- % receiving Supplemental Security Income,
- % receiving HUD rental housing assistance,
- % enrolled in Medicaid,
- % enrolled in Medicare, and
- % who have died since the event.

These outcomes are measured 1, 3, and 5 years after the justice event. Statistics are provided for subpopulations with differing levels of experience with the justice system (no previous justice events, any previous justice events) and for different offense categories (violent, property, and drug crimes, but only for charge events). Statistics are further calculated for three types of demographic subpopulations (3 age groups [15-24, 25-39, 40+], 2 sexes, and 5 racial and ethnic groups [NH White, NH Black, NH Asian, Hispanic, AIAN]); only one-way stratifications and two-way interactions of these demographic groupings are allowed. The following is an example of a statistic from this package: “The percent of White men with no previous justice involvement enrolled in Medicaid three years after a felony charge for a property offense in Harris County, Texas.”

A.3 Data sources and record linkage

Justice events and socioeconomic outcomes are identified from population-level administrative data, while demographic data come from population-level survey and administrative data. All records are linked by Protected Identification Key (PIK), an anonymous identifier that is assigned to demographic data upon receipt at the Census Bureau. This section describes the strengths and limitations of each source dataset, as well as the record linkage process.

A.3.1 Criminal Justice Administrative Records System (CJARS)

Criminal justice events are identified from Criminal Justice Administrative Records System (CJARS) data. The 2022 Q4 vintage of CJARS was used to identify the cohorts based on dates of misdemeanor charge, felony charge, prison release, probation start, and parole start. CJARS data do not cover all states and counties, and this package only covers geographies covered by the 2022 Q4 vintage. Similarly, CJARS longitudinal coverage varies and so cohorts are identified starting in 2000 or as early as are available after that.

The CJARS data documentation outlines the processes and data quality checks that are used to harmonize raw criminal justice administrative data into a high-quality dataset (Finlay and Mueller-Smith 2022). Every CJARS vintage is used to produce aggregate statistics that can be benchmarked against those published by the Bureau of Justice Statistics. These comparisons show that CJARS data are high quality and can reproduce aggregate statistics on the caseloads of criminal justice agencies (Papp and Mueller-Smith 2022).

A.3.2 IRS W-2 information returns

Labor market outcomes are derived from Internal Revenue Service (IRS) W-2 information returns. [Federal law requires](#) that employers file W-2 forms when an employee has earned more than \$600 in wages in a year. Employees who work for more than one employer in a tax year may receive more than one W-2 form.

A person is defined as employed in a year if at least one W-2 form was issued to them. For annual employment earnings, the sum of earnings is calculated using all of the W-2 forms associated with a person in a tax year. Earnings are indexed to 2017 using the Bureau of Labor Statistics Chained Consumer Price Index For All Urban Consumers (C-CPI-U). The third labor market outcome measures if the annual employment earnings is greater than the poverty threshold for a single person under age 65.

Labor market outcomes measured only from W-2 forms exclude earnings from other sources including contractor and self-employment income. W-2 information is only available starting in 2005.

Research has shown that IRS W-2 data are a high-quality alternative to survey data for measuring earnings from employers (O'Hara, Bee, and Mitchell 2017), and have been used to measure the employment and earnings of justice-involved individuals (Finlay and Mueller-Smith 2021).

A.3.3 Social Security Administration (SSA) Supplemental Security Record (SSR) data

The percent of justice-involved individuals receiving Supplemental Security Income (SSI) is identified from Social Security Administration (SSA) Supplemental Security Record (SSR) data. SSR data covering SSI benefits are only available during the period 2010-2015 and 2019-2020, so SSI take-up statistics are not available for follow-up periods outside of these years.

SSR data are used at the Census Bureau to impute SSI receipt in the Survey of Income and Program Participation (National Academies of Sciences, Engineering, and Medicine 2018).

A.3.4 Department of Housing and Urban Development Longitudinal Public and Indian Housing Information Center (PIC)/Tenant Rental Assistance Certification System (TRACS) data

The percent of justice-involved individuals receiving HUD rental housing assistance is identified from Department of Housing and Urban Development Longitudinal Public and Indian Housing Information Center (PIC)/Tenant Rental Assistance Certification System (TRACS) data. These data cover almost all

HUD-sponsored subsidized and public housing assistance programs (Meyer and Wu 2018; see list of programs in Supplemental online appendix).

The PIC/TRACS data exclude housing programs sponsored by the Department of Agriculture and rent caps for low-income tenants associated with the Low-Income Housing Tax Credit (Scally and Lipsetz 2017; Scally et al. 2018).

HUD's [Moving to Work](#) program, initiated in the late 1990s, allows participating local housing authorities to implement changes to policies and reporting standards relative to national HUD programs. As a result, JOE does not include statistics for geographic areas with [participating authorities](#) in the years of their participation.

A.3.5 Centers for Medicare and Medicaid Services (CMS) Medicaid Statistical Information System (MSIS) and Transformed Medicaid Statistical Information System (T-MSIS) data

The percent of people enrolled in Medicaid is identified from Centers for Medicare and Medicaid Services (CMS) Medicaid Statistical Information System (MSIS) and Transformed Medicaid Statistical Information System (T-MSIS) data. Medicaid data are available from 2005-2019. CMS transitioned from MSIS to T-MSIS between 2014 and 2016. States began reporting into T-MSIS at different times and it is noted below that record linkage rates for T-MSIS are 3-10% higher than for MSIS. Thus, there may be small measurement discontinuities during the transition period.

Research has shown that the CMS MSIS is a high-quality alternative to survey data for measuring Medicaid enrollment (Noon, Fernandez, and Porter 2016).

A.3.6 CMS Electronic Database (EDB) data

The percent of people enrolled in Medicare is identified from CMS Electronic Database (EDB) data. CMS Medicare data are available between 2000 and 2020. Research has shown that the CMS EDB is a high-quality alternative to survey data for measuring Medicare enrollment (Bhaskar et al. 2016).

A.3.7 Census Numident

The Census Numident is used to measure mortality, date of birth, and sex. The Census Numident is derived from the SSA Numident. The Census Numident is a high-quality source for identifying all-cause mortality for periods covered by the JOE statistics (Finlay and Genadek 2021).

A.3.8 Census Bureau Title 13 BestRace files

Race and ethnicity are identified from Census Bureau Title 13 BestRace files. These files use race and ethnicity data from a variety of survey and administrative records to identify high-quality race and ethnicity measures for each person with a PIK. Data sources include 2000 and 2010 decennial census responses, responses from all ACS years, the Census Numident, CMS MSIS, CMS T-MSIS, CMS EDB, HHS Temporary Assistance for Needy Families (TANF), HUD PIC, HUD TRACS, HUD Computerized Homes Underwriting Management System (CHUMS), and the Indian Health Service. When there are multiple discordant race or ethnicity measures from these sources for a person, responses associated with less populous racial groups are prioritized over those associated with more populous ones. This approach was developed based on research benchmarked against the 2010 Census population distribution (Ennis et al. 2018).

A.3.9 Record linkage

The datasets above were linked at the person level using the Census Bureau's Protected Identification Key (PIK), an anonymous person identifier that is assigned to demographic data upon receipt at the Census Bureau. PIKs can be assigned to records either probabilistically or deterministically. If records contain a mix of potentially imperfectly measured personally identifying information (PII), the Person Identification Validation System (PVS) is used to probabilistically match people to reference files at the Census Bureau that are primarily based on the Census Numident (Wagner and Layne 2014). When the Social Security Number (SSN) is the only or primary identifier on a file, the Quick PIK process is used to assign a PIK deterministically using a SSN-PIK lookup table.

The Census Numident covers all individuals with an SSN or Individual Taxpayer Identification Number (ITIN), and so excludes undocumented immigrants. In addition, probabilistic record linkage is imperfect and, therefore, not all individuals receive a PIK. Young people, immigrants, non-White people, and Hispanics are less likely to receive a PIK (Bond et al. 2014; Rastogi and O’Hara 2012).

The following table describes the record linkage methods used to assign a PIK to a person record for a particular data source. It also identifies the percent of records assigned a PIK for each source.

Dataset	Linkage method	% with PIKs	Citation/notes
CJARS	PVS	85%	Author’s calculation. Census Bureau DRB authorization number CBDRB-FY19-371.
IRS W-2	Quick PIK	~100%	There is no publicly available PIK assignment rate, but records are delivered with SSN.
SSA SSR	Quick PIK	100%	All records have an SSN.
HUD PIC/TRACS	Quick PIK, PVS	~100%	There is no publicly available PIK assignment rate, but records are delivered with SSN.
CMS MSIS	PVS	87.7-94.3%	Noon, Fernandez, and Porter (2016)
CMS T-MSIS	PVS	97.7%	Limburg, Kurczewski, and Udalova (2023)
CMS EDB	PVS	99.9%	Bhaskar et al. (2016)
Census Numident	n/a	100%	All records have an SSN.
Census BestRace	Quick PIK, PVS	100%	Only linkable records are included.

A.4 Concepts and methodology

Most JOE statistics measure the proportion of a justice-involved caseload who have a particular socioeconomic experience or outcome. For example, the proportion of justice-involved individuals i employed would be defined as:

$$y_{ejkgsrahc} = \frac{\sum_{iejkgsrahc} I_i EMP_i}{\sum_{iejkgsrahc} I_i}, \text{ where}$$

$$I_i = \begin{cases} 1 & \text{if } i \in ejkgsrahc \\ 0 & \text{otherwise} \end{cases},$$

$$EMP_i = \begin{cases} 1 & \text{if } i \text{ is employed} \\ 0 & \text{otherwise} \end{cases},$$

e = justice event \in {misdemeanor charge, felony charge, prison release, probation start, parole start},
 j = justice event year,
 k = years since justice event \in {1,3,5},
 g = geography: state, county, county aggregation,
 s = sex \in {male, female},
 r = race/ethnicity \in {NH White, NH Black, NH Asian/PI, Hispanic, AIAN},
 a = age group \in {15–24, 25–39, \geq 40},
 h = justice history \in {no previous justice involvement, previous justice, involvement}, and
 c = crime type \in {violent, property, drug}.

Average income is calculated as a mean in an analogous way.

A.4.1 Geographies

Cohort geographies are determined by the state and county of sentencing. These may not be the states and counties of residence for people in a cohort, but research has shown that most people convicted of crimes live close to where the crimes were committed (e.g., Bernasco, Block, and Ruiter 2012). Therefore, a county in this dataset is primarily identifying the outcomes of caseloads associated with a particular criminal justice agency, but can also serve as a proxy for the residence of the justice-involved individuals in a cohort.

In order to release statistics for less populous areas, small counties were aggregated. States and counties with average total populations between 2000 and 2020 of 80,000 or greater are used as-is. After removing those larger counties from 2000 Commuting Zones (CZs), any CZs or residual CZ county aggregations with total populations of 80,000 or greater are identified. After those larger CZs and CZ component county aggregations have been removed, all other counties in a state are combined into a single residual geographic aggregation. Under this scheme, the smallest rural counties in a state are aggregated together, reducing the risk of disclosure for their smaller justice-involved populations. From the beginning, commuting zones are cut off at state boundaries, so a CZ that has portions in two states is split into two separate CZ areas.

A.4.2 Using longer windows for cohorts with small populations

For county statistics, cohorts with any demographic disaggregation (using sex, race/ethnicity, or age group) are constructed from two-year windows of justice events instead of one-year windows. All state-level statistics use one-year windows of justice events.

A.4.3 Removing cohorts with small populations

After justice-involved cohorts have been identified from these initial geographies, cohorts that contain fewer than 20 people are removed. These cohorts are not included in any further steps and are not released. Because the statistics, which are all estimated proportions and averages, have no additive relationships, this step does not create any implicit samples that might require removal of additional cohorts.

A.4.4 Reweighting to account for record linkage selection

Not all CJARS records can be assigned a PIK. To account for potential record linkage selection, cohort statistics are reweighted based on the probability that a record was assigned a PIK, conditional on four-way interactions of the CJARS topical table from which the record is sourced, sex, race/ethnicity, and age group. Demographic characteristics for the model are defined as in the statistics. Some of the estimated record linkage probabilities may be disclosed in another information product. Because of this reweighting, final statistics are treated as estimated parameters.

The weights are calculated using data from all available states but separately for each topical table in the CJARS dataset (i.e., arrest, adjudication, probation, incarceration, and parole). Within each table, weights are calculated at the level of the justice event separately by demographic variables identified on the CJARS data including sex (male or female), race/ethnicity (non-Hispanic White, non-Hispanic Black, non-Hispanic Asian or Pacific Islander, Hispanic, and American Indian or Alaska Native), and age group on the date of the justice event (15-24, 25-39, 40+). These variables were used because they exist in common between the CJARS data and the other data sources, and they have been found to be correlated with the likelihood of a person record being assigned a PIK (Bond et al. 2014; Rastogi and O'Hara 2012). The CJARS table is included because we know that PII quality varies across the tables. A weight for a person i with CJARS topical table t , sex s , race/ethnicity r , and age group a is defined as:

$$w_{itsra} = \frac{\sum_{itsra} I_i}{\sum_{itsra} I_i PIK_i}, \text{ where}$$

$$I_i = \begin{cases} 1 & \text{if } i \in tsra \\ 0 & \text{otherwise} \end{cases},$$

$$PIK_i = \begin{cases} 1 & \text{if } i \text{ was assigned a PIK} \\ 0 & \text{otherwise} \end{cases},$$

$$t = \text{CJARS table} \in \{\text{arrest, adjudication, probation, incarceration, parole}\},$$

$$s = \text{sex} \in \{\text{male, female}\},$$

$$r = \text{race/ethnicity} \in \{\text{NH White, NH Black, NH Asian/PI, Hispanic, AIAN}\}, \text{ and}$$

$$a = \text{age group} \in \{15-24, 25-39, \geq 40\}.$$

The weight represents the inverse probability of a person being assigned a PIK conditional on the demographic categories listed above, as well as which CJARS topical table the person record comes from.

To apply the weights during estimation, the weighted proportion or average is calculated. For example, the reweighted proportion of justice-involved individuals i employed would be defined as:

$$y_{ejkgsrahc} = \frac{\sum_{iejkgsrahc} I_i EMP_i w_{itsra}}{\sum_{iejkgsrahc} I_i w_{itsra}}, \text{ where}$$

$$I_i = \begin{cases} 1 & \text{if } i \in ejkgsrahc \\ 0 & \text{otherwise} \end{cases},$$

$$EMP_i = \begin{cases} 1 & \text{if } i \text{ is employed} \\ 0 & \text{otherwise} \end{cases},$$

$t = \text{CJARS table} \in \{\text{arrest, adjudication, probation, incarceration, parole}\},$
 $e = \text{justice event} \in \{\text{misdemeanor charge, felony charge, prison release, probation start, parole start}\},$
 $j = \text{justice event year},$
 $k = \text{years since justice event} \in \{1, 3, 5\},$
 $g = \text{geography: state, county, county aggregation},$
 $s = \text{sex} \in \{\text{male, female}\},$
 $r = \text{race/ethnicity} \in \{\text{NH White, NH Black, NH Asian/PI, Hispanic, AIAN}\},$
 $a = \text{age group} \in \{15-24, 25-39, \geq 40\},$
 $h = \text{justice history} \in \{\text{no previous justice involvement, previous justice, involvement}\}, \text{ and}$
 $c = \text{crime type} \in \{\text{violent, property, drug}\}.$

The misdemeanor charge and felony charge events are sourced from the adjudication table; prison release is sourced from the incarceration table; probation start is sourced from the probation table; and parole start is sourced from the parole table.

It is important to note that inverse-probability weighting will not correct for any selection bias associated with PIK assignment if PIK assignment is correlated with the socioeconomic outcomes but is not fully accounted for by the demographic variables used to construct the weights.

A.4.5 Choropleth-like geographic aggregation of counties

At this point in estimation, many of the county or county aggregation geographies have total populations smaller than the threshold used to identify Geographic Areas of Small Population (GASP). The Census Bureau Disclosure Review Board (DRB) requires that statistics from geographies with total populations below the GASP threshold have noise added. To avoid this technique, counties are further aggregated into larger geographies. To further aggregate counties, consider the analogy of a choropleth map of one of the statistics with a fixed number of shades. In the shading of choropleth regions, geographies with statistics within an interval will receive the same color and be grouped together even if they are far away on the map. Counties and county aggregations are aggregated across the U.S. based on the closeness of their initial statistic after reweighting for record linkage bias. This aggregation is done separately for each cohort year, each post-event follow-up period, each outcome, each level of prior justice experience, and each demographic grouping. The algorithm used to create these aggregations can be provided upon request. After this aggregation is complete, every county aggregation in every year has a total population greater than the GASP threshold for that year.

A.4.6 Re-estimation within choropleth groupings

With counties further aggregated, statistics are re-estimated for the larger geographies that have total populations greater than the GASP threshold. These statistics are also re-weighted to account for record linkage selection, as described above.

A.4.7 Rounding

After the re-estimation within choropleth groupings, the estimated statistics are rounded. For proportions, estimates are rounded to two significant digits. Average annual income is rounded to three significant digits.

A.5 Disclosure avoidance methodology

In accordance with U.S. Code, Title 13, Section 9, no data are published that would disclose the identity or characteristics of any person.

Disclosure avoidance is the process used to protect the confidentiality of data provided by an individual. A number of disclosure avoidance methods were used to protect the identity and characteristics of individuals in the input datasets:

- Geographic aggregation. At multiple steps, geographies have been aggregated, both to ensure that geographies have total populations above the GASP threshold and to reduce the total number of statistics being released.
- Temporal aggregation. For county statistics, cohorts with any demographic disaggregation (using sex, race/ethnicity, or age group) are constructed from two-year windows of justice events instead of one-year windows.
- For smaller cohorts, two-year windows are used instead of one-year windows to identify cohorts associated with specific justice events.
- Rounding. All proportion statistics are rounded to two significant digits and average annual income to three significant digits. In all cases, this is coarser rounding than is required by Census Bureau Disclosure Review Board rounding rules for estimated parameters.
- Removal of cohorts with small populations. Cohorts that contain fewer than 20 people are removed. The statistics have no additive relationships so this step does not create any implicit samples that might require removal of additional cohorts.
- Limits to the interactions of demographic subpopulations. Only two-way interactions of the three demographic categorical variables (age groups, race and ethnicity, and sex) are allowed. Three-way interactions are not allowed.
- Re-weighting with record linkage selection weights. Estimates have been re-weighted using the probability of being assigned a PIK, conditional on all the demographic features described above. This reweighting further ambiguates the person-level data used to create the estimates.

A.6 Reliability of the data

The data sources used to produce these estimates are of the highest quality, and well-grounded in a body of proven administrative records research showing that the quality and suitability of those data sources to directly replace demographic information in household surveys (Rastogi and O'Hara 2012; Bhaskar et al. 2014; Mulry et al. 2021).

Because the JOE statistics are constructed from a combination of population-level administrative data, rather than a probability sample, sampling error does not apply. Non-sampling error, however, still exists. For instance, administrative records data may contain measurement error because of issues such as coverage problems (e.g., the data source may not cover certain populations as well as others); linking or matching issues which could affect accuracy or precision; conceptual and timing misalignments; reporting errors; definition and classification difficulties; errors in recording or coding the data obtained; and other errors of coverage, processing, and estimation for missing or misreported data.

It may be possible to provide measures of uncertainty about the statistics in this product, and such measures will be considered for future versions of JOE.

A.7 Glossary

Acronym	Definition
CJARS	Criminal Justice Administrative Records System
CMS	Centers for Medicare and Medicaid Services
DRB	Census Bureau Disclosure Review Board
EDB	Medicare Electronic Database
GASP	Geographic Areas of Small Population
HUD	Department of Housing and Urban Development
IRS	Internal Revenue Service
ITIN	Individual Taxpayer Identification Number
JOE	Justice Outcomes Explorer
MSIS	Medicaid Statistical Information System
PIC	Public and Indian Housing Information Center
PII	Personally Identifiable Information
PIK	Protected Identification Key
PVS	Person Identification Validation System
SSA	Social Security Administration
SSI	Supplemental Security Income
SSN	Social Security Number
SSR	Supplemental Security Record
T-MSIS	Transformed Medicaid Statistical Information System
TRACS	Tenant Rental Assistance Certification System

A.8 References

- Bernasco, Wim, Richard Block, and Stijn Ruiter. 2012. Go Where the Money Is: Modeling Street Robbers' Location Choices. *Journal of Economic Geography* 13(1): 119-43. doi:[10.1093/jeg/lbs005](https://doi.org/10.1093/jeg/lbs005).
- Bhaskar, Renuka, Adela Luque, Sonya Rastogi, and James Noon. 2014. Coverage and Agreement of Administrative Records and 2010 American Community Survey Demographic Data. Census Bureau CARRA Working Paper 2014-14. <https://www.census.gov/library/working-papers/2014/adrm/carra-wp-2014-14.html>.
- Bhaskar, Renuka, James Noon, Brett O'Hara, and Victoria Velkoff. 2016. Medicare Coverage and Reporting: A Comparison of the Current Population Survey and Administrative Records. Census Bureau CARRA Working Paper 2016-12. <https://www.census.gov/library/working-papers/2016/adrm/carra-wp-2016-12.html>.
- Bond, Brittany, J. David Brown, Adela Luque, and Amy O'Hara. 2014. The Nature of the Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey. Census Bureau CARRA Working Paper 2014-08. <https://www.census.gov/library/working-papers/2014/adrm/carra-wp-2014-08.html>.
- 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. doi:[10.3233/SJI-170374](https://doi.org/10.3233/SJI-170374). Also released as [Census Bureau CARRA Working Paper 2015-08](#).
- Finlay, Keith, and Katie Genadek. 2021. Measuring All-Cause Mortality with the Census Numident File. *American Journal of Public Health* 111(S2): S141-S148. doi:[10.2105/ajph.2021.306217](https://doi.org/10.2105/ajph.2021.306217). Also released as [Census Bureau ADEP Working Paper 2021-03](#).
- Finlay, Keith, and Michael Mueller-Smith. 2021. Justice-Involved Individuals in the Labor Market since the Great Recession. 2021. *ANNALS of the American Academy of Political and Social Science* 695(1): 107-22. doi:[10.1177/00027162211024532](https://doi.org/10.1177/00027162211024532). Also released as [Census Bureau ADEP Working Paper 2021-04](#).
- Finlay, Keith, and Michael Mueller-Smith. 2022. Criminal Justice Administrative Records System (CJARS). https://cjars.org/wp-content/uploads/CJARS_data_docs_2022q4.pdf.
- 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/sehsd/working-papers/2023-05>.

[//www.census.gov/content/dam/Census/library/working-papers/2023/demo/sehsd-wp2023-05.pdf](https://www.census.gov/content/dam/Census/library/working-papers/2023/demo/sehsd-wp2023-05.pdf).

Meyer, Bruce D., and Derek Wu. 2018. The Poverty Reduction of Social Security and Means-Tested Transfers. *ILR Review* 71(5): 1106-53. doi:10.1177/0019793918790220 Supplemental online appendix: https://journals.sagepub.com/doi/suppl/10.1177/0019793918790220/suppl_file/DS_10.1177_0019793918790220.pdf

Mulry, Mary H., Tom Mule, Andrew Keller, Scott Konicki. 2021. Administrative Record Modeling In the 2020 Census. Census Bureau Working Paper. <https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/administrative-record-modeling-in-the-2020-census.pdf>.

Noon, James M., Leticia E. Fernandez, and Sonya R. Porter. 2016. Response Error and the Medicaid Undercount in the Current Population Survey. Census Bureau CARRA Working Paper 2016-11. <https://www.census.gov/library/working-papers/2016/adrm/carra-wp-2016-11.html>.

O'Hara, Amy, Adam Bee, and Joshua Mitchell. 2017. Preliminary Research for Replacing or Supplementing the Income Question on the American Community Survey with Administrative Records. American Community Survey Research and Evaluation Report Memorandum Series ACS16-RER-6; Center for Administrative Records Research and Applications Memorandum Series 16-7. https://www.census.gov/content/dam/Census/library/working-papers/2016/acs/2016_Ohara_01.pdf

Papp, Jordan, and Michael Mueller-Smith. 2022. Benchmarking the 2022 Q4 Vintage of the Criminal Justice Records System's Data Infrastructure. https://cjars.org/wp-content/uploads/CJARS_Benchmarking_Report_2022q4.pdf

Rastogi, Sonya and Amy O'Hara. 2012. 2010 Census Match Study Report. Census Bureau Planning Memo 247. https://www.census.gov/library/publications/2012/dec/2010_cpex_247.html.

Scally, Corianne P., Amanda Gold, Carl Hedman, Matt Gerken, and Nicole DuBois. 2018. The Low-Income Housing Tax Credit: Past Achievements, Future Challenges. Washington, D.C.: Urban Institute. https://www.urban.org/sites/default/files/publication/98761/lithc_past_achievements_future_challenges_final_0.pdf

Scally, Corianne P., and David Lipsetz. 2017. New Public Data Available on USDA Rural Housing Service's Single-Family and Multifamily Programs. *Cityscape*, 19(1): 295-304. <https://www.huduser.gov/portal/periodicals/cityscpe/vol19num1/ch17.pdf>

Wagner, Deborah, and Mary Layne. 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>.

National Academies of Sciences, Engineering, and Medicine. 2018. The 2014 Redesign of the Survey of Income and Program Participation: An Assessment. Washington, DC: National Academies Press. doi:10.17226/24864

B Code schemes

B.1 Geographic and demographic codes

B.1.1 Sex codes

This scheme is used for the following variables:

- [Primary data: sex](#)
- [Supplementary: sex](#)

Table 6: Sex codes

Value	Label
0	All
1	Male
2	Female

B.1.2 Race and ethnicity codes

This scheme is used for the following variables:

- [Primary data: race](#)
- [Supplementary: race](#)

Table 7: Race and ethnicity codes

Value	Label
0	All
1	White, non-Hispanic
2	Black, non-Hispanic
3	Asian or Pacific Islander, non-Hispanic
4	Hispanic
5	American Indian or Alaska Native

B.1.3 Age group codes

This scheme is used for the following variables:

- [Primary data: age_group](#)
- [Supplementary: age_group](#)

Table 8: Age group codes

Value	Label
0	All
1	15-24
2	25-40
3	40+

B.1.4 Offense type codes

This scheme is used for the following variables:

- [Primary data: off_type](#)
- [Supplementary: off_type](#)

Table 9: Offense type codes

Value	Label
0	All
1	Violent
2	Property
3	Drug

B.1.5 Repeat contact codes

This scheme is used for the following variables:

- [Primary data: repeat_contact](#)
- [Supplementary: repeat_contact](#)

Table 10: Repeat contact codes

Value	Label
0	All
1	First-time contact only
2	Repeat contact only
3	All calculated with reweighted caseloads

C Data notes by state

The following are notes on the JOE statistics that originate from relevant CJARS data notes. These notes in full can be found in CJARS technical documentation (Finlay and Mueller-Smith 2022).

C.1 Idaho data notes

CJARS Note ID	Note
ID001	Idaho probation supervision is split between the Department of Corrections and courts, but CJARS data only covers probationers supervised by the Department of Corrections. Due to this partial coverage, Idaho is excluded from the national covered average for probation per capita.

C.2 Minnesota data notes

CJARS Note ID	Note
MN001	Adjudication data in Minnesota only covers convictions, so charge rate per capita is actually conviction rate per capita. Minnesota adjudication statistics are not included in nationwide aggregation.
MN002	Minnesota court cases with a criminal traffic offense as the controlling charge not covered.

C.3 Nevada data notes

CJARS Note ID	Note
NV001	Incarceration data in Nevada only covers the initial prison entry and exit for a given sentence; subsequent re-entries (e.g. parole revocations) are not available. Due to this, recidivism to prison for Nevada is not included. Additionally, Nevada incarceration stock and spell length statistics are not included in nationwide aggregation

C.4 Washington data notes

CJARS Note ID	Note
WA001	Washington probation supervision is split between the Department of Corrections and courts, but CJARS data only covers probationers supervised by the Department of Corrections. Due to this partial coverage, Washington is excluded from the national covered average for probation per capita.

Bibliography

- Finlay, Keith, and Michael Mueller-Smith. 2022. *Criminal Justice Administrative Records System (CJARS)*. Technical report. U.S. Census Bureau and University of Michigan. cjars.org/wp-content/uploads/CJARS_data_docs_2022q4.pdf.
- . 2023. Criminal Justice Administrative Records System (CJARS) [dataset]. Ann Arbor, MI: University of Michigan. <https://cjars.org>.
- Finlay, Keith, Michael Mueller-Smith, and the CJARS team. 2024. Justice Outcomes Explorer (JOE) [dataset]. Ann Arbor, MI: University of Michigan. <https://joe.cjars.org>.
- Papp, Jordan, and Michael Mueller-Smith. 2022. Benchmarking the 2022 Q4 Vintage of the Criminal Justice Records System's Data Infrastructure. https://cjars.org/wp-content/uploads/CJARS_Benchmarking_Report_2022q4.pdf.
- U.S. Bureau of Labor Statistics. 2010. *Unemployment Rate [UNRATE]*, retrieved from FRED, Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/UNRATE>.
- United States Census Bureau Small Area Income and Poverty Estimates (SAIPE) Program. 2021. *SAIPE State and County Estimates for 2010-2021*. Each year downloaded separately. <https://www.census.gov/data/datasets/2021/demo/saipe/2021-state-and-county.html>.
- United States Department of Justice. Federal Bureau of Investigation. 2019. *Uniform Crime Reporting Program Data: County-Level Detailed Arrest and Offense Data, United States, 2010-2016*. Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/ICPSR37059.v3>.
- University of Wisconsin Population Health Institute. 2022. *County Health Rankings & Roadmaps 2022*. www.countyhealthrankings.org.

Index

- age code, [26](#)
- aggregations, [13](#)
- Census microdata, [10](#)
- citation, [6](#)
- CJARS microdata, [10](#)
- code schemes, [26](#)
- cohorts, [10](#)
- context data, [15](#)
- county groups, [12](#)
- coverage, [6](#)
- data notes, [28](#)
- data sources, [5](#)
- disclosure avoidance, [13](#)
- estimation groups, [12](#)
- ethnicity code, [26](#)
- funders, [6](#)
- geographies, [11](#)
- Idaho
 - data notes, [28](#)
- imputation, [11](#)
- methodology overview, [10](#)
- Minnesota
 - data notes, [28](#)
- Nevada
 - data notes, [28](#)
- off type code, [26](#)
- principal investigators, [6](#)
- project description and objectives, [5](#)
- race code, [26](#)
- repeat contact, [11](#)
- repeat contact code, [27](#)
- reweighting, [11](#)
- sex code, [26](#)
- subpopulations, [11](#)
- technical notes, [13](#)
- Washington
 - data notes, [28](#)