September 2020

Incarceration Trends Project

Data and Methods for Historical Jail Populations in U.S. Counties, 1970-2018

Jacob Kang-Brown, Oliver Hinds, Eital Schattner-Elmaleh, and James Wallace-Lee

Vera Center on Sentencing + Corrections

Additional copies can be obtained from the communications department of the Vera Institute of Justice, 34 35th street, suite 4-2A, Brooklyn, NY, 11232, (212) 334-1300. An electronic version of this report is available for download on Vera's web site, www.vera.org.

Requests for additional information about the research described in this report should be directed to Jacob Kang-Brown at the above address or to jkangbrown@vera.org

Introduction

Incarceration Trends, a project of the Vera Institute of Justice, aims to inform the public debate on mass incarceration and help guide change by providing easily accessible information on the number of people in jails and prisons for every county in the United States.

One part of the project is a data tool—available at trends.vera.org—that displays publicly available but disparately located data about incarceration in counties and states. This tool can be used for reference and measurement by justice system stakeholders and others looking to understand how their jail and prison is being used and how it compares with other places, and to spot problem areas—such as excessive growth or racial or ethnic disparities. Currently, the *Incarceration Trends* tool displays jail data for every county with a jail, and prison data up to 2016 for all counties in 33 states: Alabama, Arizona, California, Colorado, Florida, Georgia, Indiana, Iowa, Kentucky, Maine, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Nebraska, Nevada, New Hampshire, New Jersey, New York, North Carolina, Ohio, Oklahoma, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Washington, Wisconsin, and Wyoming.[1]

To support further research on incarceration, Vera has released a complete dataset that backs the Incarceration Trends data tool. The data set is available as a county based file as well as a more detailed jail jurisdiction based file from Vera's github.[2]

The purpose of this document is to provide detail on the sources and methods used to compile the jail dataset. The first analysis using this dataset can be found in *In Our Own Backyard: Confronting Growth and Disparities in American Jails* (Subramanian, Henrichson, and Kang-Brown, 2015).

Background

[1] Six states (Alaska, Connecticut, Delaware, Hawaii, Rhode Island, and Vermont) do not participate in the Bureau of Justice Statistics jails data collections because they do not have local jails. In these places, the state corrections department operates both jails and prisons. Recent estimates of pretrial detention and short (under one year) sentences in these states would sum to an additional 2 percent of U.S. jail population on a single day (15,806 people).

[2] See the Incarceration Trends dataset and documentation at https://github.com/vera-institute/incarceration_trends

Research on mass incarceration usually centers on state-level data: state prison populations or the statewide combined prison and jail population. Using the state as the unit of analysis is sufficient for understanding the broad contours of incarceration in the United States, but it does not provide the level of detail necessary to help policymakers reduce the use of incarceration. This is because although *state* law dictates criminal penalties, these laws are interpreted and deployed in practice by police, prosecutors, judges, and others at the *local level* (counties and cities). Therefore, county-level data, specifically data on how counties use the local jail and how many people the local courts send to state prison, are fundamental to a thorough understanding of the justice-system. Yet comprehensive and comparable historical county-level information is not available. Vera researchers sought to address this gap by developing a dataset using the available federally-collected data.

This project uses data collected by the U.S. Department of Justice Bureau of Justice Statistics (BJS): The BJS Census of Jails, which covers *all* jails and is conducted every 5 to 8 years; the Deaths in Custody Reporting Program (DCRP), which collects facility-level admissions, daily count, and average daily population from about 2,800 local jails every year since 2000; and the Annual Survey of Jails which covers about one-third of jails—and includes nearly all of the largest jails—and has been conducted in years where the Census of Jails was not conducted, since 1985.[3] Vera researchers merge this data at the county-level, from every year for

[3] Data collections used: Annual Survey of Jails, 2018 (ICPSR 37392); Annual Survey of Jails, 2017 (ICPSR 37373); Annual Survey of Jails, 2016 (ICPSR 37135); Annual Survey of Jails, 2015 (ICPSR 36760); Annual Survey of Jails, 2014 (ICPSR 36274); Census of Jails, 2013 (ICPSR 36128); Annual Survey of Jails, 2013 (ICPSR 35517); Deaths in Custody Reporting Program: Jail Populations, 2000-2013 (ICPSR 36560); Annual Survey of Jails: Jail-Level Data, 2012 (ICPSR 34884); Annual Survey of Jails: Jail-Level Data, 2011 (ICPSR 33722); Annual Survey of Jails: Jail-Level Data, 2010 (ICPSR 31261); Annual Survey of Jails: Jail-Level Data, 2009 (ICPSR 29081); Annual Survey of Jails: Jail-Level Data, 2008 (ICPSR 28281); Annual Survey of Jails: Individual Reporting-Level Data, 2007 (ICPSR 24641); Annual Survey of Jails: Jurisdiction-Level Data, 2006 (ICPSR 20368); Census of Jail people: Individual-Level Data, 2005 (ICPSR 20367); Annual Survey of Jails: Jurisdiction-Level Data, 2004 (ICPSR 20200); Annual Survey of Jails: Jurisdiction-Level Data, 2003 (ICPSR 4635); Annual Survey of Jails: Jurisdiction-Level Data, 2002 (ICPSR 4428); Annual Survey of Jails: Jurisdiction-Level Data, 2001 (ICPSR 3883); Annual Survey of Jails: Jurisdiction-Level Data, 2000 (ICPSR 3882); National Jail Census, 1999 (ICPSR 3318); Annual Survey of Jails: Jurisdiction-Level Data, 1998 (ICPSR 2682); Annual Survey of Jails: Jurisdiction-Level Data, 1997 (ICPSR 2313); Annual Survey of Jails: Jurisdiction-Level Data, 1996 (ICPSR 6856); Annual Survey of Jails: Jurisdiction-Level Data, 1995 (ICPSR 6784); Annual Survey of Jails: Jurisdiction-Level Data, 1994 (ICPSR 6538); National Jail Census, 1993 (ICPSR 6648); Annual Survey of Jails: Jurisdiction-Level and Jail-Level Data, 1992 (ICPSR 6395); Annual Survey of Jails: Jurisdiction-Level and Jail-Level Data, 1991 (ICPSR 6511); Annual Survey of Jails: Jurisdiction-Level Data, 1990 (ICPSR 9569); Annual Survey of Jails: Jurisdiction-Level Data, 1989 (ICPSR 9373); National Jail Census, 1988 (ICPSR 9256); Annual Survey of Jails: Jurisdiction-Level Data, 1987 (ICPSR 9074); National Survey of Jails: Jurisdiction-Level and Jail-Level Data, 1986 (ICPSR 8871); Annual Survey of Jails: Jurisdiction-level and Jail-level Data, 1985 (ICPSR 8687); National Jail Census, 1983 (ICPSR 8203); National Jail Census, 1978 (ICPSR 7737); National Jail Census, 1972 (ICPSR 7638); and, finally, National Jail Census, 1970 (ICPSR

which it is collected, enabling an analysis of local-level change over time and exploration into the origins and growth of mass incarceration in the United States.

This report details the methodology used to create the *Incarceration Trends* dataset, and reviews the sources and data processing techniques. It concludes with a discussion of limitations of this data and the opportunities for future research this new resource provides.

Data Sources and Methods

The information Vera researchers collected centers on jail population data from the BJS Census of Jails, DCRP, and Annual Survey of Jails.[4] Annual demographic data for county resident population from the U.S Census Bureau are used in combination with the jail population to calculate incarceration and admission rates. This section provides further detail on these data and the methods used to organize jail data at the county-level.

Jails data

Why do we have national jail data since 1970? One reason is that in 1965, the President's Commission on Law Enforcement and Administration of Justice requested that the National Council of Crime and Delinquency conduct an analysis of correctional agencies and institutions.[5] This study worked with the U.S. Census to make a representative sample of 250 counties in 1965, but found that more data were needed.[6] A few years later, in 1970, the U.S. Census and the Department of Justice's Bureau of Justice Statistics conducted the first jail census.[7]

7641). The Survey of Jails was also conducted in 1982 and 1984, but these are not included in the Vera dataset because the results are not publically available in digital format.

[4] Although the BJS jail definition covers many local places where one is held in custody, it does not include police and sheriff lock-ups, halfway houses, or prison pre-release centers. See also Richard S. Frase, "Jails," in *The Handbook of Crime and Punishment*, ed. Michael Tonry, (New York: Oxford University Press, 1998), 476. ("The federal definition of a jail has changed over the years and has always been underinclusive").

[5] For summary of this request and research, see Frederick Ward, Jr. "Introduction," *Crime & Delinquency*, 13(1967): iii; and "6. Local Adult Correctional Institutions and Jails," *Crime & Delinquency*, 13(1967):136-157.

[6] See implication 9 of 10, "6. Local Adult Correctional Institutions and Jails," 156. The final report on this research, citing the Manhattan Bail Project, also concluded that "*(t)he number of persons held in jail awaiting trial can be sharply reduced.*" See ibid., 151-152. (Emphasis in original).

[7] National Jail Census, 1970 (ICPSR 7641).

For this project, Vera researchers use the information collected on people held in jail from the Census of Jails data collections in 1970, 1978, 1983, 1988, 1992, 1999, 2005, and 2013.[8] For the years 2000-2013, data from facilities where counties are included in the DCRP were used. The database also uses the Annual Surveys of Jails that were fielded in the intercensal years since 1985. The survey captures all of the nation's largest jails and a representative sample of hundreds others used to estimate the total U.S jail population.[9]

Data cleaning

Vera researchers reviewed COJ and ASJ data in order to identify data anomalies. Where erroneous data were observed, Vera researchers removed the data and linearly interpolated from the last data point available to the subsequent data point. If no subsequent data point existed Vera researchers used the last available value in the most recent year. A detailed description of this process is below.

The COJ/ASJ data are reported in some years by individual jail facility and in some years by jail jurisdiction.[10] Vera researchers first identified variables that, when combined, could identify unique jail facilities in order to assign facility identifiers for every jail in the dataset that can be compared over time. A subset of these variables were used to assign each facility a jurisdiction identifier, a unique number that tracks a legal jurisdiction within a county. In order to preserve the local justice focus, Vera researchers removed federal jail facilities, which were only collected in 1993, 1999, 2000, and 2005-2018. Then, Vera researchers identified variables of interest that were consistently reported over the years, such as jail population, and created a document identifying what each variable was named in each year's dataset. This document was used to combine individual year datasets into a single datafile.

In cases where data were collected at different scales over the years (variables such as admissions and discharge were sometimes collected over a week period, over a 24 hour period, or over a whole year) the data were normalized to reflect annual counts. In the years 2006 and 2013 both ASJ and COJ data were collected; Vera researchers used the COJ data except in specific cases where a respondent's data were either cleaner or more comprehensive in the ASJ dataset. In some years different wings of the same facility all reported the data separately; Vera researchers aggregated all data to the facility level.

To help identify data error, Vera researchers computed year over year change in every numerical variable for every jail and looked at the distribution of these values for every variable in the dataset. Based on the distributions, Vera researchers identified thresholds where the year over year change in values of variables was outside a normal range. Then, Vera researchers manually reviewed data for every variable and facility when the year over year change was more

[8] BJS also conducted a jail census in 1972, which is not included in the *Incarceration Trends* data collection because of its proximity to the 1970 census and lack of comparable information.

[9] For a discussion of sampling methodology see Todd D. Minton and Zhen Zeng, *Jail people at Midyear 2014,* Bulletin (Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics, 2015, NCJ 248629), 10. See also the codebook in Annual Survey of Jails, 2014 (ICPSR 36274), 63.

[10] See above, at footnote 3 for which years include individual jail facility data.

extreme than the threshold. Across all variables, the average threshold was a year over year change of 30 percent. In many variables, well over 1,000 possible errors were reviewed by hand. Possible errors identified this way were, more often than not, accurate and represented real changes in the jail populations.Vera researchers used a variety of methods to verify data, including calling facilities to determine if the reported data were correct. For some variables, such as the numbers of incarcerated men and women, the data were often reported incorrectly -- the men's number in the women's field and vice versa. This same problem would also often occur in the variables describing incarceration by racial category. Vera researchers would swap reported numbers when solution to the data error was obvious. Other erroneous data were removed from the dataset and replaced with linear interpolation.

The resulting dataset was aggregated at the jurisdiction level by summing each variable for each facility in a jurisdiction, within the specific year. This jurisdiction-level dataset was then interpolated to fill in values for years where the jurisdiction was not sampled in the ASJ, or not reported to COJ or ASJ. Variable values were extended from the last year that the jurisdiction reported to 2017, unless the most recent report was prior to the 2005 COJ.

The resulting jurisdiction-level jail dataset was used as the basis for a county-level dataset, which was obtained by summing variable values for each jurisdiction in a county, within the specific year. This ensures better, slightly higher county-level jail population estimates in states like Alabama, California, or Washington that frequently have both city jails and county jails in the same county, as the city jails tend not to be sampled in the ASJ.

Jurisdiction and County-level data

The *Incarceration Trends* project's emphasis on the local takes us to the level of the 3,143 U.S. counties or county equivalents.[11] (Herein, the term county is used to describe counties and county-equivalents.) This project aggregates individual jails to their jail jurisdiction, and then to their county. This approach is generally straight-forward since most counties have their own jail jurisdiction. But many (generally smaller) counties do not have a jail and thus rely on a multi-jurisdiction or regional jail. Vera is publicly releasing both jurisdiction-level and county-level data files.

Multi-jurisdictional jails

As of 2017, an estimated 192 jails hold people for multiple counties. These multi-jurisdictional jails serve 530 total counties, covering about 5 percent of the jail population,

[11] The U.S. Census uses county equivalent to describe divisions of a state comparable to counties for statistical purposes, such as parishes in Louisiana, boroughs in Alaska, the District of Columbia, and the independent cities of Virginia, Maryland, Missouri, and Nevada. In the data tables, the District of Columbia is considered both a county, and where appropriate, a state. New York City contains five counties but is reported as one for this analysis.

counties that appear to have no jail apart from a temporary holding facility (e.g. police lockup), and must therefore rely on another jurisdiction for pretrial detention or non-prison sentences of incarceration. Multi-jurisdictional jails may be "regional jails," such as those in Virginia and West Virginia, that are generally flagged as such in the BJS data collection. Or they may not be identified in a formal way, but provide jail beds to an often small neighboring county. For instance, Vera researchers found that Boise County (population 7,000) in Idaho sends people to a jail run by adjacent Ada County (population 390,000). Consequently, in these cases the number of people in multi-jurisdiction jails exceeds the number of people sent to the jail by the county where the jail is physically located. Without any further adjustment, this would overstate the jail population for the county where the multi-jurisdiction jail is located.

Vera researchers addressed this problem by identifying the sending and receiving counties (meaning the counties that use, and the counties that operate regional jails, respectively), and the year that the sending county last reported data to BJS. Then, Vera apportioned the multi-jurisdictional jail population, distributing jail data to the sending counties based on their share of the combined resident population. For example, if a regional jail holds 100 people and serves two counties, one with 90,000 residents, and the other with 10,000 residents, this method assumes 10 people are from the smaller, sending county and 90 from the larger, receiving county.[12]

In order to identify these cases, Vera researchers used the 2005 census of jail facilities and reviewed all of the counties that *did not* have a respondent to the jail survey. Vera researchers identified many of the nonrespondent counties as sending counties, through internet research, and paired them with their receiving counties that ran the multi-jurisdiction jail. For counties that researchers could not pair with a multi-jurisdiction jail through internet research, Vera placed telephone calls to sheriffs' departments in those counties that do not operate a jail, and asked whether they had any arrangements in place to send people to other jails. Most complied with the request, and shared this information. Through follow up research (completed using information from the 2015 ASJ, and updated based on the 2017 ASJ) Vera determined that 192 counties contain a multi-jurisdictional (or regional) jail. Of the counties with no respondents,

[12] While this approach does not account for jail-relevant factors like crime or police activity, it allows us to apply a fix nationwide, especially in cases that would otherwise grossly over-state incarceration rates in counties with multi-jurisdictional jails. These fixes are time sensitive. For instance, if a county reported jail numbers to BJS in the 1970s and 1980s, and stopped in the 1990s, the database would start the apportionment in the 1990s only.

530 utilize a single multi-jurisdictional jail.[13] These counties are weighted by population and jail data were apportioned from the receiving county as described above.

U.S. Census data

Vera researchers used the U.S. Census Bureau's decennial census as well as the American Community Survey (ACS) to source variables for community characteristics. The project uses intercensal population estimates for individual years between 1970 and 2010, and 2018 postcensal estimates, all available from U.S. Census.[14]

Resident Population, by Age. To get a more accurate picture of incarceration rates, people under the age of 15 and over 64 were removed from the general population for the purposes of calculating incarceration and admission rates, since these groups are at very low risk of jail incarceration. Because the proportion of these groups varies greatly by county, keeping them in would skew rates and make comparisons between counties difficult. Note that this method differs from most other calculations of statewide and national incarceration rates, which use either the total resident population or the population aged 18 and older.

Resident Population, by Race and by Gender. Using the same age limits as above (15-64), Vera researchers collected resident population by racial and gender categories available as single year estimates at the county level from the U.S. Census.[15] The federal Office of Management and Budget standards for the collection of data on race and ethnicity changed in 1977 and again in 1997. In order to preserve comparability to the jails data, which have used a bridged race and ethnicity question since 1985 (see "Confined Population by Race and by

---

[13] A smaller number of cases, 41 counties, use various jails in more than one county. Of the 41 counties, over half were from 5 states: Nebraska (6), Minnesota (5), Kansas (4), Montana (4), and South Dakota (4). Some of these counties have data for earlier years, but in later years that they did not respond to BJS data collections, these split counties were not covered in this database because Vera researchers could not determine from which county to apportion jail information. Vera was unable to identify the current jail arrangements in place for about 100 small counties.

[14] Vera used the following files: Intercensal County Estimates by Age, Sex, Race: 1970-1979; 1980-1989 Intercensal County Estimates by Age, Sex, Race; Intercensal Estimates (1990-2000) Age by Sex by Race by Hispanic Origin; 2000-2010 Intercensal Estimates - April 1, 2000 to July 1, 2010 by age, sex, race and Hispanic origin; and vintage 2018 Population by age, sex, race, and Hispanic origin. All current postcensal and historical inter-censal population estimates are available online at https://www.census.gov/programs-surveys/popest.html and https://www.cdc.gov/nchs/nvss/bridged_race.htm

[15] The following categories are available at the county level for ethnicity ("Hispanic or Latino" or "not Hispanic or Latino"), for race ("American Indian or Alaska Native (AIAN)," "Asian," "Black or African American (Black)," "Native Hawaiian or Other Pacific Islander (NHOPI)," and "White") and for gender ("female" and "male").

Gender"), Vera researchers only use the Census race data from 1990, when it is first available using bridged ethnicity and race categories.

Jail Variable Description

As discussed above, the jail admissions and population data are extracted from the BJS data collections. Vera researchers then make a few calculations to improve the comparability of these data across counties and to identify and remove data errors. More complete information on available variables is in the associated codebook, available on github.

Jail population. BJS collects two measures of the confined population: (1) the average daily jail population (often called "ADP") and (2) the confined population on a given day, usually the last weekday in June. The *Incarceration Trends* project's measure of the jail population is the ADP. The confined population is only used when ADP was not collected (as in some of the older surveys) or is missing.[16] This jail population figure is also the basis for the project's incarceration rates.

Confined Population by Race and by Gender. BJS has collected information on single day confined female and male population since 1970,[17] and confinement by a set of race and ethnicity categories since 1985.[18] Vera researchers use these data points to calculate group specific jail incarceration rates for county residents with the available census data described above. Because these questions on race and gender are not asked by custody status, the totals may sum to higher than the above local jail total, since the total subtracts the number of people held for other, non-local authorities.[19] Due to limits in the U.S. Census data, single year estimates at the county level using the current standard federal race and ethnicity categories are only available from 1990 to present.

---

[16] There appears to be data reporting or data entry error for a small number of cases. When that occurs, the database uses the confined population.

[17] BJS does not currently collect information on transgender people in jail through the jail survey or census.

[18] The current categories are substantially similar to those first used in 1985. They include "Hispanic or Latino," and the following, all specified as "not of Hispanic origin": "American Indian / Alaska Native," "Asian," "Black or African American," "Native Hawaiian or Other Pacific Islander, "Two or more races," and "White."

[19] This is primarily a concern in counties with large facilities that hold people as people for a local jurisdiction as well as for other local, state, and federal authorities. Because the numbers held for other authorities are often higher, this could distort various indicators, from jail admissions to gender and race demographic characteristics.

Jail admissions. Jail admissions are an important indicator of a jail's reach, since many people are detained or incarcerated for only short periods of time. The admissions number does not indicate the number of unique individuals, as some people may be sent to jail multiple times in a given year. Also note that this measure does not count movements to court or medical services, only new admissions to jail. BJS has asked this question at different time scales, collecting new admissions in the COJ and ASJ for a:

- Typical Week: 1978
- Year (July to June): 1983 -1991
- 24-hour Period (June 30): 1992 -1993
- Last Week in June: 1998-2014
- Year (Calendar year): 2015-2018

Vera adjusts these numbers so that they all reflect *annual* admissions. Because jail admissions vary greatly on a daily basis, the database does not use BJS admissions and discharge data for 1992-93 (when it was collected as a 24-hour count). Additionally, these data were not collected from 1994-97. Therefore, these data for 1992-1997 are estimated using linear interpolation. Jail admissions by gender are only reported through the DCRP, so are only available for the years 2000-2013 for counties with included facilities.

Conclusion

This 2020 version of the dataset includes the expanded local level jail statistics around specific federal agencies involved in contracting with local jails, jail rated capacity, and jail discharge information. It also includes a file that allows researchers to explore jurisdiction-specific data within a county, as well as more precise data related to race and gender. We encourage other researchers to explore questions of the causes and consequences of mass incarceration using this data on local jails.