Linking Administrative BJS Data: Understanding Jail Mortality through the Communities They Serve

Margaret Noonan
Bureau of Justice Statistics
810 7th Street, NW
Washington, D.C. 20531

Susan Brumbaugh
David Chrest
Christopher Ellis
RTI International
3040 East Cornwallis Road
Post Office Box 12194
Research Triangle Park, NC 27709-2194

Proceedings of the 2015 Federal Committee on Statistical Methodology (FCSM) Research Conference

Local jails are the entry point to the American correctional system. Jails process 12 million admissions annually, representing an estimated 9 million unique individuals. This is more than all other segments of the correctional system in the United States combined—prisons probation, and parole (Carson, 2014; Minton & Zeng, 2015). Jails are primarily local, county, and community-based institutions that confine persons before trial and/or after adjudication. In practice, jails are a custodial catch-all, holding an array of inmates with different correctional statuses, including but not limited to persons who are awaiting trial (pending either arraignment or bond) and persons who are convicted of a misdemeanor and serving a sentence of typically less than 1 year. Forty percent of the confined population are sentenced offenders or convicted offenders awaiting sentencing, and 60% are unconvicted inmates being held for a variety of reasons (e.g., inability to meet bail, awaiting trial, mental health holds, drug or alcohol detoxification, temporary holds for federal authorities).

Every year, the Bureau of Justice Statistics (BJS) collects information on jail populations through two collections: the Deaths in Custody Reporting Program (DCRP) and the Annual Survey of Jails. Twice each decade, BJS conducts its Census of Jail Facilities (Jail Census). Unlike BJS's other jail collections, the Jail Census collects data at the facility level, data about inmates both in custody and under the jurisdiction of jails, and information about the jail facility's management. The most recent Jail Census was conducted in 2013 through the DCRP because like the census, the DCRP is a full enumeration of jails in the United States. However, unlike the census, it is conducted annually, which meant that tasks and costs typically associated with the census (e.g., updating the jail frame to confirm jail eligibility) and respondent burden were minimized. As with earlier iterations, the 2013 Jail Census contains information on jail facilities in the United States, including an aggregate count of deaths during the year, jail population snapshots and averages, numbers of admissions and releases, jail capacity, counts of inmates being held for other jurisdictions, counts of inmates in programs that offer alternatives to incarceration, prison costs, and staffing.

BJS is able to supplement data on deaths in jails by linking the DCRP population files (i.e., the data collected via its annual summary forms) and Jail Census data to the DCRP death files, but the resulting analysis of mortality correlates is restricted to the study of the jail population and facility characteristics. Although this type of analysis is useful, it is limited in its ability to consider the impact of external, or community, factors on mortality. Thus, one of BJS's goals is to link criminal justice datasets to other datasets covering topics that BJS does not collect but that can better inform how and why persons enter the criminal justice system. In the case of jails, indicators of economics and poverty, health, crime, mental health, and substance abuse in the communities where jails are located are of particular interest. In an effort to better inform how and why persons enter the criminal justice system at the local jail level, BJS began exploring options to identify community-level files that could be linked to jail datasets.

To study jails in the context of the communities they serve, BJS tasked its data collection agent, RTI International, to undertake a data-linking project using the 2013 Jail Census and external community-level files. Because the 2013 Jail Census collected data at the community level, with the jail file aggregated to the jail jurisdiction level, any datasets identified for potential matching would also have to contain community-level information that could inform jail inmate mortality rates. This paper outlines the process that BJS and RTI used to identify nearly two dozen potential community files and selected datasets eligible for linking, and the process used to link the selected datasets with 2013 Jail Census data. The paper also describes BJS's plans for working with the linked data.

Results of Dataset Review

RTI reviewed 22 potential datasets to determine whether it would be feasible to link them with data from the 2013 Jail Census. Datasets were evaluated for their capacity to provide indicators and measures on demographics (age, sex, race, educational attainment), economics and poverty (median income, percent below poverty level), crime and social disorder (arrest rate in the community, presence of certain types of businesses such as gun shops or bars), mental health, substance abuse, and physical health (proximity to hospitals or mental health/substance abuse treatment centers, health care access and use, and health insurance). Considerations included availability/cost, level of data (i.e., whether data were available at the county level), and unique contributions to relevant domains. When possible, RTI downloaded the actual datasets to better inform the review.

Thirteen datasets were deemed inappropriate for linking because of one or more of the following reasons: (1) they were not comprehensive, (2) no county-level data were available, (3) permission for use could not be obtained, (4) better data were available from other sources, and (5) the dataset no longer existed.

Ultimately, nine community-level datasets were linked to the combined DCRP/Jail Census dataset, which created a community-level file that allowed BJS to group jails on community and jail characteristics, with the goal of understanding how these factors affect jail inmate mortality.

Datasets Selected for Linking

Exhibit 1 summarizes the nine datasets linked with Jail Census data. Following is a description of the datasets.

2009–2013 American Community Survey, U.S. Census Bureau. Data from the American Community Survey are available from the U.S. Census Bureau at no cost. Data are available at the county level and contain demographic, socioeconomic, and household characteristics indicators. This dataset includes many variables with information on population, age groups, sex, race, income, employment, family types, housing, occupation, marital status and household relationship, veteran status, disability status, foreign or native born, and transient status. These variables can be used to develop aggregate county-level indicators of demographics so that similar counties can be grouped and compared with regard to mortality rates.

American Hospital Association Annual Survey Database, American Hospital Association. The American Hospital Association dataset is available for purchase (about $8,800 for fiscal year 2013). Data exist at the hospital/facility level and contain indicators for a facility's services, size, and features. This includes hospital location, type(s), bed and operating room count, staff characteristics, and the availability of psychiatric care. The existence of hospitals in a particular county could be associated with mortality rates.

Area Health Resource File, U.S. Department of Health and Human Services, Health Resources and Services Administration. The Area Health Resource File is available from the Health Resources and Services Administration at no cost. Data are available at the county level. This dataset contains a wide variety of relevant variables, such as number of psychologists, use of food stamps, education information, supplemental income figures, poverty levels, labor force, health facilities, and health-related professionals. These economic indicators may be related to mortality rates and could allow similar communities to be grouped for comparison purposes.

Behavioral Health and Treatment Services Locator, U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration. The Behavioral Health and Treatment Services Locator is available from the Substance Abuse and Mental Health Services Administration at no cost. Variables

include locations of physical health, mental health, and substance abuse facilities, including county facilities. The availability of mental health or substance abuse treatment facilities per capita can be used to examine mortality rates.

Unemployment Rates, Bureau of Labor Statistics. Information on unemployment rates is contained in a dataset that is available from the Bureau of Labor Statistics at no cost. Data are available at the county level. The dataset includes unemployment rates by year (for 2013) and by month (for August 2013 through September 2014), making it possible to spot trends.

Dun and Bradstreet Business Listing File, Esri. The Dun and Bradstreet Business Listing File from Esri normally is available for purchase (about $12,000). However, RTI had already purchased the 2013 listing (the same time period as the reference year for the Jail Census), so there was no additional cost to BJS. This is a Geographic Information Systems (GIS) database of all U.S. business locations, totaling more than 18 million geographic points. It provides descriptive variables that include North American Industry Classification System and Standard Industrial Classification codes. Any type of business, including mental and physical health facilities, can be identified by county. Other business types such as bars, convenience stores, and alcohol outlets could be used as "negative" economic indicators. Calculating various types of business per capita may help inform county-level mortality rates.

Small Area Health Insurance Estimates, U.S. Census Bureau. Small Area Health Insurance Estimates data are available from the U.S. Census Bureau at no cost and are listed at the county level. Variables in the dataset include income information (e.g., percent of persons in the county at or below 200% of poverty) and percent uninsured overall and for various income levels by age, sex, and race categories. Insurance coverage, combined with the availability of health care from other datasets, could be useful information.

Uniform Crime Reports, FBI. Uniform Crime Reports data are available from the FBI at no cost. Data are available at the county level. Variables in the dataset include crime statistics for adults and juveniles. Crime rates per capita can be compared with mortality rates in each county to look for relationships with regard to overall rates or types of crimes.

WONDER (Wide-ranging Online Data for Epidemiologic Research), Centers for Disease Control and Prevention. The WONDER tool provides access to information on injuries and deaths from the Centers for Disease Control and Prevention at no cost. Data are available at the county level. Variables include cause of death, injury, and injury intent statistics. Mortality information in the general population can be compared with facility-specific mortality rates.

Exhibit 1. Available datasets that can be linked at the county level

Dataset	Agency Source	Cost	Latest Available Year	Level of Data	Domains					
					Demo-graphics	Poverty/ Economic	Crime/ Social Disorder	Mental Health	Physical Health	Substance Abuse
ACS Census Information	U.S. Census Bureau	No cost	2009–2013	County	X	X				
AHA Hospital Information	American Hospital Association	About $8,800	2013	Facility				X	X	
AHRF Health Indicators	HRSA	No cost	2013–2014	County	X	X		X	X	
Behavioral Health Treatment Services Locator	SAMHSA	No cost	2014	County/city				X	X	X
Unemployment Rates	BLS	No cost	2013–2014	County	X	X				
D&B Business Information	Esri	Already purchased (about $12,000)	2013	Geographic point		X	X	X	X	
SAHIE Insurance Information	U.S. Census Bureau	No cost	2012	County	X	X			X	
UCR Crime Information	FBI	No cost	2012	County			X			
WONDER Tool	CDC	No cost	2013	County	X				X	

Note: ACS = American Community Survey. AHA = American Hospital Association. AHRF = Area Health Resource File. BLS = Bureau of Labor Statistics. CDC = Centers for Disease Control and Prevention. D&B = Dun and Bradstreet. HRSA = Health Resources and Services Administration. SAHIE = Small Area Health Insurance Estimates. SAMHSA = Substance Abuse and Mental Health Services Administration. UCR = Uniform Crime Reports. WONDER = Wide-ranging Online Data for Epidemiologic Research.

Process for Merging Datasets

The dataset for the 2013 Jail Census includes data from 3,170 U.S. jails that RTI identified through the DCRP. For each jail, the dataset lists the county, latitude and longitude coordinates, ANSI codes, and FIPS codes. Nine datasets containing items in the domains of interest (demographic, sociological, economic, household, physical and behavioral health treatment, medical facility, business, crime, death, and insured information) were linked by county to the 3,170 records contained in the DCRP/Jail Census dataset.

Item selection. RTI and BJS determined the specific items from each dataset that should be included in the linking activity. A list of specific items from each dataset was provided to BJS, along with RTI's preliminary recommendations for the items that should be linked. BJS approved the final list.

Dataset preparation. Variables in the nine datasets were available in formats such as CSV, Excel, ASCII, and .txt files. RTI carried out preliminary editing and conversion to organize and prepare the datasets for linking to the counties where the jails were located. RTI then imported the datasets into a geodatabase to be edited further. RTI determined whether and how each variable of interest should be modified to create a county-level indicator. For example, the WONDER database contained records for each death in the county, including the sex and race of the decedent and the cause of death. RTI created aggregate county-level indicators for each variable of interest, such as percentage of deaths that involved females, along with the absolute number of females, males, and total deaths.

Dataset linking. The process of creating the final linked data table(s) was primarily carried out using Esri's GIS software product, ArcGIS 10.3. A GIS polygon data layer representing all 3,145 U.S. counties was used as the primary source for identifying and matching data to the counties either by geographic location or using any unique FIPS codes that existed in the data to be linked. Matching the nine datasets (each containing one record per county) to the Jail Census dataset (containing records at the facility level) was a complicated process because although nationally there is approximately a 1:1 match for jails to counties in the United States, the match varies at the state and community levels. For example, Alaska, Connecticut, Delaware, Hawaii, Rhode Island, and Vermont are all combined correctional systems, which means jails are under the authority of the state department of corrections and cannot be matched at the community level. Furthermore, some counties do not have a jail, other (typically more densely populated counties) have multiple jails serving a single county, and other counties have one jail serving multiple counties (regional jails).[1] Therefore, counties without a jail were excluded from the analysis because they could not be linked to the jail census file, counties with multiple jails were linked more than once, and regional jails were linked based on the facility's physical location.

Linked file delivery. The linked data tables delivered to BJS were designed so that each record represented a jail in the set of 3,170 facilities, which is the total number of facilities in the 2013 Jail Census. Facility-related variables (such as facility name, address, latitude/longitude coordinates, and ANSI/FIPS codes) and DCRP/Jail Census data were included, followed by county-level variables linked from the nine datasets. Because of the number of variables and the resulting file size, RTI delivered one file with variables from all datasets. This file included a complete version with all nine datasets plus the Jail Census data, and a public-use file that excluded data from the American Hospital Association and Dun & Bradstreet because the data-use agreements disallowed any public dissemination of data. Individual files of paired datasets were also provided. These files contained data on each of the nine datasets linked with the DCRP/Jail Census data so the end user could easily access a subset of the data, all linkable by the FIPS code. Data were delivered to BJS in .csv and .xlsx formats, and data dictionaries were included.

Next Steps for Using the Data

In addition to making the linked data available for ad hoc analyses and formal reporting, BJS will also fund several applications of the linked data.

Product 1: Thematic Map Series
RTI has worked with BJS to create two series of thematic maps, which are designed to communicate topical information using symbols and shading. One series includes a number of U.S. maps depicting county-level

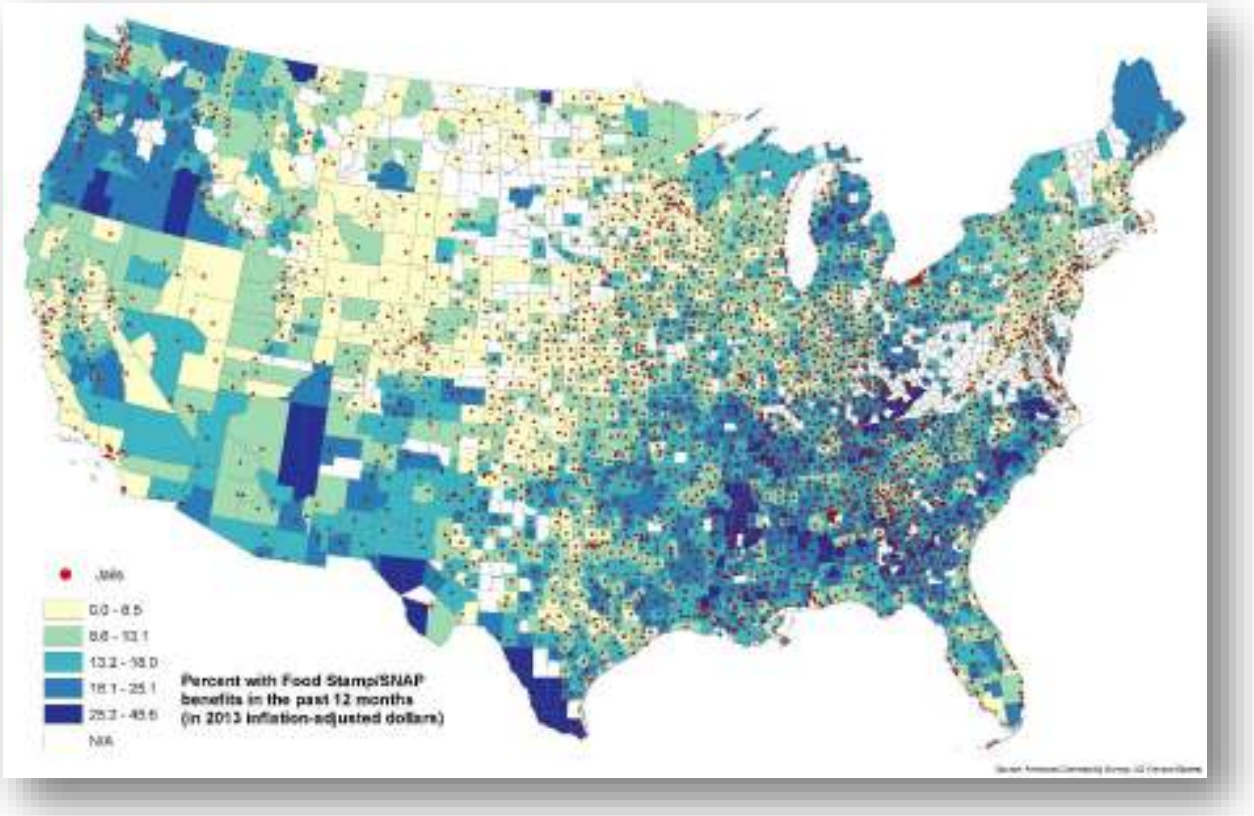
[1] In addition to the 64 counties in the 6 states with combined correctional systems, 440 counties do not have a jail that serves their populations.

information, and the other series includes city-specific or local-area maps with information about the areas immediately surrounding the jail. Following is a description of each map series.

Nationwide county-level maps. These maps show the county-level indicators from each of the linked datasets. The viewer can then quickly examine shared characteristics in counties (e.g., higher unemployment rates, crime rates, mortality). Demographic characteristics of each county can also be mapped. Jail locations are then overlaid on the map and their point symbols are sized proportionally to deaths in custody rates, which allows the user to view jail information against the backdrop of county-level information.

Exhibit 2 shows county-level data from one of the variables linked to Jail Census data, with jail location points overlaid. Additional maps can be made that use varying sizes for the dots representing jails, depending on the number of deaths reported by the jail.

Exhibit 2. County-level data showing jail locations and percent of individuals reporting that they received food stamp benefits in the past 12 months



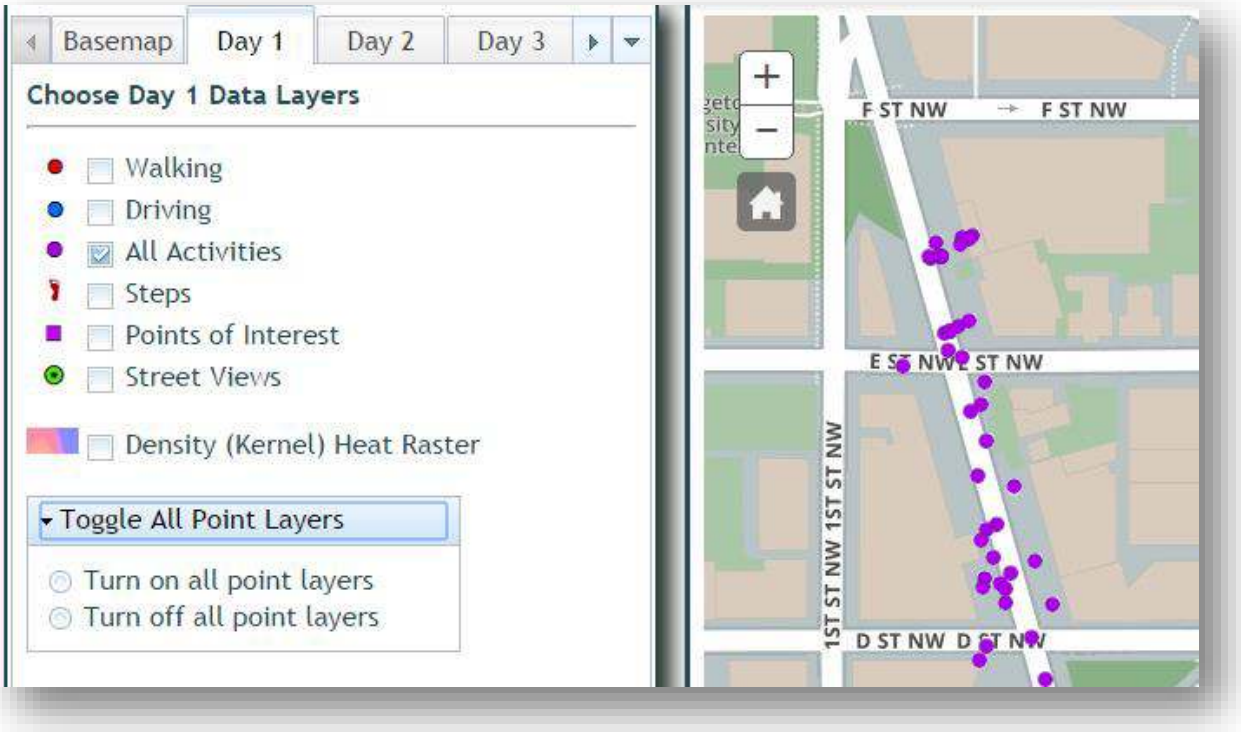
City maps. Local or multicounty urban area maps can be developed to show jail and social indicator/demographic data. These detailed local area maps can depict individual jails and their proximity to features such as health facilities or alcohol outlets and bars (see Exhibit 3 for an example from Cook County, IL). In addition, downloadable tables can be created that show distances in miles, kilometers, feet, etc. from each jail to the closest chosen feature such as hospitals, certain kinds of health facilities, treatment centers, and business types.

Exhibit 3. Map showing individual jails, health facilities, and business types in Chicago near the Cook County Department of Corrections facility

Product 2: Interactive Web Mapping Application
Interactive web mapping is a simplified custom GIS application that is accessed via a website. This type of tool allows the user to interact with geographic data using a map that can be zoomed in or out and panned to different geographic areas.

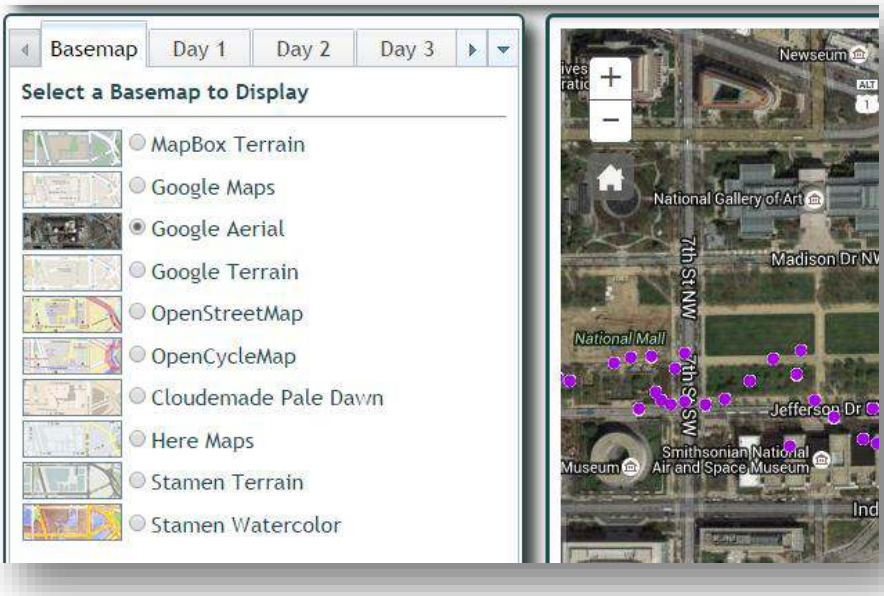
When a point is clicked in the map, jail and associated information is displayed in a popup window that allows the user to select and display data elements of interest. Information about other features (e.g., hospitals, businesses) appears when the user hovers over that feature in the map. Additionally, geographic data layers such as jails, facilities, businesses, and counties—each with their own set of indicators—can each be turned on or off in the map (see an example in Exhibit 4).

Exhibit 4. Ability to toggle geographic data layers



Different styles of base maps can be toggled in the map area. Exhibit 5 is an example of a Google aerial view.

Exhibit 5. Basemap display option: Google aerial



The tool will be fully searchable by jail name or identification number. A scale bar will change dynamically in the lower-left corner of each resulting map, with latitude and longitude continuously reflecting the location of the mouse pointer on the map. This gives the user a point of geographic reference and scale.

Features such as jails, businesses, and health facilities can be selected by drawing a polygon on the map. This can be useful if a user wants to see all indicators in a certain area, such as west of Detroit. Similarly, a buffer around a particular location (e.g., a jail) can be drawn for any distance in several units of measure. This can be useful for selecting features, such as health facilities, located within a certain distance of a jail. The features in the selected area (e.g., the hand-drawn area or the buffer around a jail) can be displayed in a downloadable table format.

Expected Reports

BJS expects to publish a series of reports using the linked data. Following the release of a data quality report, BJS will develop single-topic reports on domains such as demographics, mental health, physical health, and economic indicators. Reports that link mortality in jails to community indicators will provide context to patterns of jail mortality throughout the United States and allow for a better understanding of how the community surrounding a jail might influence jail inmate mortality.

References

Carson, E. (2014). *Prisoners in 2013,* NCJ 247282. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.

Minton, T. D. & Zeng, Z. (2015). *Jail Inmates at Midyear 2014,* NCJ 248629. Washington, D.C.: U.S. Department of Justice, Bureau of Justice Statistics.