Collecting and Processing Multistate Criminal-History Data for Statistical Analysis

Matthew R. Durose and Alexia D. Cooper, Ph.D., *BJS Statisticians*
Howard N. Snyder, Ph.D., *former BJS Statistician*

Introduction

The Bureau of Justice Statistics (BJS) aims to produce timely and accurate statistics on the criminal activities and recidivism patterns of offenders. To help achieve this goal, BJS has developed methods by which multistate criminal-history records are collected and processed for statistical and research purposes. This report discusses (1) BJS's methods to collect and process criminal-history data and (2) the substantive, technical, and methodological challenges faced in creating a new system to conduct this work.

Each state and the Federal Bureau of Investigation (FBI) operates a repository that maintains and disseminates official criminal-history records. The FBI's Criminal Justice Information Services (CJIS) Division maintains the repository of criminal-history records from the federal justice system and U.S. territories. It also maintains and disseminates criminal-history record information received from state and local agencies. The FBI's Interstate Identification Index (III) is an automated pointer system that allows criminal justice agencies to determine which repositories across the country have criminal-history records on an individual.

BJS's prior methods to collect and process criminal-history data for statistical analysis

Historically, conducting multistate criminal-history research has been a time-consuming process. Since the early 1980s, BJS has periodically used criminal-history data from the FBI and state repositories to study the recidivism patterns of various offenders. For example, in 1998, BJS began collecting information on the offending patterns of a sample of prisoners released in 15 states in 1994 that covered the 3 years following their release.[1] For this study, letters requesting the prisoners' criminal-history records were sent to

criminal-history repositories in states where sampled prisoners were released. Months of work followed to establish separate data-security agreements and file-transfer procedures with each repository. When all these agreements were in place, each of the 15 state criminal-history repositories provided BJS with an extract of the criminal-history data on the sampled prisoners released in that state.

For information on state and local arrests and prosecutions of the sampled prisoners that occurred outside of the state of release, the study of prisoners who were released in 1994 relied on state records submitted to the FBI's national criminal-history repository.[2] In addition to the out-of-state information, the FBI repository was also the source of criminal-history information on federal arrests and prosecutions of the sampled prisoners.

The content, format, and location of the fields in the extracts of criminal-history data varied widely by repository. For the study of prisoners who were released in 1994, which involved data extracts generated by each state repository and the FBI repository, extensive work was required to transform the jurisdiction-specific files and data fields into a single dataset that could support multistate analyses. This data-conversion process involved complex computer programs. BJS completed the final report for this study in 2002.

The transition to a new approach for data collection and processing

In 2008, BJS entered into a Memorandum of Understanding (MOU) with the FBI and Nlets (formerly the National Law Enforcement Telecommunications System), which for the first time allowed BJS to access the nation's criminal-history records for criminal justice research and evaluation

---

[1] See *Recidivism of Prisoners Released in 1994* (NCJ 193427, BJS web, June 2002).

[2] See *Recidivism of Prisoners Released in 1994* (NCJ 193427, BJS web, June 2002).

through the FBI's III system. Nlets is a central component of the nation's criminal-history record system. This non-profit organization operates a computer-based message-switching system that facilities the exchange of information among criminal justice agencies, including criminal-history record inquiries made through III to all the state repositories and the FBI repository. The FBI and state repositories automatically respond to these requests on the Nlets network, typically within seconds. In response to each III query, the requesting agency receives one or more messages from the repositories containing the subject's state and federal criminal-history records.

This new approach was aimed at reducing the reporting burden on the repositories and the time required to obtain criminal-history data for statistical purposes. From 2008 to 2012, BJS worked collaboratively with the FBI, state criminal-history repositories, and Nlets to develop an automated system to (1) collect multistate criminal-history records and (2) consolidate and transform these records into analytical datasets through the process outlined in the next section.

BJS's criminal-history data collection and processing system

The Bureau of Justice Statistics relies on Nlets for access to criminal–history data in the FBI repositories and in all state repositories through queries to FBI's III system. The data collection and processing system has five stages:

1. Accessing criminal-history data for research purposes

With the MOU in place, BJS partnered with the FBI's CJIS Division and Nlets to develop a technical solution that decreased the time and resources needed to acquire criminal-history data from all state criminal-history repositories through III. Nlets developed a secure, automated process to collect computerized criminal-history record information on samples of study subjects for statistical purposes from the FBI and all state repositories through III. As a result, BJS is able to obtain the most recent out-of-state criminal-history information available on samples of prisoners directly from the states. BJS only needs to access the FBI's national criminal-history repository for records that the FBI is solely responsible for disseminating (e.g., federal arrest and prosecution information and certain state records). This new data-retrieval process helps to reduce the time for data collection.

In addition to developing new automated systems to minimize the time and resources required to collect national criminal-history data, the need for BJS to change its data-collection methodology was also precipitated by the FBI's implementation of the National Fingerprint File (NFF) program. This program reduced the duplicate maintenance of state criminal-history records by both the states and the FBI. Under this program, the FBI maintains only an individual's initial identification record from states that participate in the NFF Program. NFF states are the sole disseminators of their criminal-history records and no longer submit duplicate records to the FBI. When an III request is made on an offender with criminal-history information in an NFF state, the FBI forwards the request to that state's repository.

2. Parsing criminal-history text into a set of fields within a relational dataset

Because the format of individual criminal-history records varies across states, BJS funded Nlets to develop a system to convert the numeric and text fields on the electronic criminal-history records into a relational dataset structure. This structure displays the state- and federal-specific fields in a uniform record layout and supports statistical analysis. BJS and Nlets modeled the uniform record layout of the parsed relational dataset after data elements found in the Joint Task Force on Rap Sheet Standardization's Interstate Criminal History Transmission Specification.[3]

The National Task Force on Increasing the Utility of the Criminal History Record developed a set of recommendations in 1995 to improve the readability of criminal-history records across states, including the implementation of a standardized format to support interstate sharing of criminal-history information. In response, the Joint Task Force on Rap Sheet Standardization was created to produce an extensible markup language (XML) transmission format for criminal-history records. Since 2004, the task force has implemented several versions of the standardized transmission format based on the Global Justice XML Data Model and the National Information Exchange Model.

While the criminal-history information transmitted by the FBI and many state repositories is in a structured XML format, some states transmit their

---

[3] Interstate Criminal History Transmission Specification, XML Version 4.0, Joint Task Force on Rap Sheet Standardization, August 2009. Retrieved from http://wiki.nlets.org/images/8/86/Rap_V4.0_Full_Documentation_August_2009.pdf

criminal-history records in state-specific text-based formats. The automated parsing logic is more complex for states that transmit criminal-history records in non-standard text-based formats. To ensure the parsing is accurate, Nlets works with the state repositories to determine the intricacies in the structure of their criminal-history records.

BJS first used this new criminal-history collection and parsing system to assess the criminal offending and recidivism patterns of approximately 70,000 prisoners who were randomly sampled to represent the more than 400,000 state prisoners released in 2005 in 30 states.[4] This sample of prisoners released in 2005 was developed using data reported by state departments of corrections to BJS's National Corrections Reporting Program (NCRP). In 2010, BJS and Nlets tested the criminal-history collection and parsing system to ensure the record-request process provided complete data and effectively handled the wide variations in the content and structure of the criminal-history records on the sample of prisoners released in 2005. To verify the parsing logic, BJS staff reviewed the criminal-history data to ensure the relevant fields were accurately captured in the parsed relational dataset. When BJS staff found an issue with the parsed data, Nlets worked to identify the source of the problem, revise the parsing logic, and generate a revised version of the parsed relational dataset.

After Nlets completed the development of this new data-collection process in 2011, BJS submitted the state and FBI identification numbers supplied by the state departments of corrections to the FBI's III, via Nlets, to collect the criminal-history data on the sampled prisoners. Within two weeks of initiating the record-request process, Nlets had collected and assembled the multistate criminal-history data from all repositories on the sample of approximately 70,000 prisoners released in 30 states in 2005. The criminal-history data included a combined total of more than 800,000 arrests and dispositions from all 50 states and the District of Columbia.

3. Converting the state- and federal-specific criminal-history fields in the parsed relational dataset into nationally standardized codes

Building on the criminal-history parsing programs developed by Nlets, BJS funded NORC at the University of Chicago to assist with developing computer programs that convert the contents of the parsed relational datasets into standardized research datasets. The standardized

research datasets provide aggregate statistical counts of the number of arrests, convictions, and incarcerations in a subject's criminal career before and after a particular event (e.g., a release from prison). The research dataset also summarizes the number of days between the date of this event and the first subsequent arrest, conviction, and incarceration.

The parsed relational datasets generated by Nlets contain the exact strings of text (numeric or alphanumeric) found on the criminal-history records. BJS worked with NORC to convert those state- and federal-specific text strings into a standardized national coding structure. The state offense fields (originally state statutes or free-text fields) are recoded into a standardized set of BJS offense categories. For example, the BJS numeric offense code for robbery replaced such text strings as "Armed Robbery 3rd degree" and "Robbery-Armed" in the standardized research dataset. Similarly, court disposition fields (often long text strings in the criminal-history data) are converted into a standardized set of BJS court disposition categories that represent whether the defendant was acquitted or convicted. If the defendant was convicted, sentencing categories capture whether the person received probation, jail, or prison. For example, text strings on the criminal-history records such as "Convicted" and "Felony CNV" receive the BJS numeric disposition code for court convictions in the standardized research dataset.

To produce consistently coded research datasets that are comparable over time, BJS designed the standardization system to be a repeatable process that ensures reliable recoding across datasets. This approach combines a large set of data-processing rules into a single system that may be modified to produce customized research files. A key component of the standardization system is a set of crosswalk tables that include BJS's national recode of state- and federal-specific data fields found in the parsed relational datasets. Many of the coding rules within the crosswalks rely on multiple fields in the parsed dataset to arrive at the appropriate BJS standardized code. The crosswalk tables in the standardization system greatly reduce the time and effort to process criminal-history datasets.

For the 2012 BJS study on the recidivism of federal prisoners who were placed on community supervision in 2005, the majority of the offenses in the criminal-history data were already in the crosswalks due to the large sample of criminal-history data already converted for the study on state prisoners released that same year. Crosswalk matches were obtained on more than 70% of the offenses in the criminal-history data for the federal study.

---

[4]*Recidivism of Prisoners Released in 30 States in 2005: Patterns from 2005 to 2010* (NCJ 244205, BJS web, April 2014).

Throughout the development of the standardization system, BJS contacted the criminal-history repositories and other criminal justice agencies to—

- obtain their translation tables for coding criminal justice information

- seek clarification when a criminal-history record or data field was difficult to interpret (e.g., an abbreviation, acronym, or code that is unique to a local jurisdiction)

- understand the criminal-history reporting practices in each state and develop uniformity in the state-specific content.

4. Generating standardized research datasets that account for variations in the criminal-history data

BJS worked with NORC to develop computer programs to generate standardized research datasets that can be analyzed with statistical software such as SPSS, STATA, or SAS. In developing rules to create the research datasets, BJS implemented procedures to identify and address anomalies within each state's data (e.g., arrest dates in a criminal history that were not in chronological order, or state arrest records received by the FBI that duplicated the same arrest record received by a state repository). BJS developed a set of data-processing rules to create research datasets that adjust for certain state or temporal variations in the criminal-history data. Some examples are—

- Traffic violations (with the exception of vehicular manslaughter, DWI, and hit-and-run) are excluded from the research datasets because the coverage of these events in the criminal-history data varies widely by state.

- Administrative and procedural records embedded in the criminal-history data that do not refer to an actual arrest are also identified and excluded from the research datasets. These records include actions indicating a criminal registration, a custody transfer or intake, or the issuance of a warrant that can duplicate a record reported by the agency responsible for charging or prosecuting the subject.

- While criminal-history records offer comprehensive details on arrests for new crimes, the completeness of the data on technical violations (such as failing a drug test or missing an appointment with a parole officer) varies widely by state. Given the inconsistent reporting of these actions in the criminal-history data, BJS uses the National Corrections Reporting Program (NCRP) data to supplement the criminal-history

data by indicating additional technical violations that led to state prisoners being returned to prison following release.

To verify the criminal-history parsing, data standardization, and generation of the research datasets, quality-control checks are implemented at each stage, including an analysis that compares the pre-processed fields to the output in the standardized variables. This analysis also verifies that the appropriate selection, filter, and coding was utilized for the data extracted into the research datasets.

5. Assessing the quality, completeness, and variability of the criminal-history data

Criminal-history records are designed to provide detailed information on the adjudicated guilt or innocence and, if convicted, the sentence imposed (e.g., prison, jail, probation) for each arrest. However, criminal-history records vary in content due to state-specific policies and practices. When using administrative data for statistical purposes, BJS aims to develop and maintain an understanding of the contents of these information systems.

BJS implements a set of assessments to identify variations in the criminal-history reporting practices of federal and state criminal justice agencies. BJS also assesses the quality and completeness of the criminal-history data by comparing the consistency of key fields reported through different sources.

Because demographic information reported in administrative data may occasionally be inconsistent, BJS validates the criminal-history data by comparing available information on each offender from independent data sources. For the recidivism study of state prisoners who were released in 2005, BJS used NCRP information from the state departments of corrections to develop a sample of all persons released from prison in 2005 in the study's 30 states. BJS compared individual identifiers in the NCRP data (e.g., sex, race, and date of birth) to those reported in the criminal-history data to ensure that the demographic information was accurate and complete. For the 2005 study, a released prisoner's date of birth in the NCRP data matched his or her birthdate in the criminal-history data 98% of the time, and nearly 100% of the NCRP and criminal-history data matched on sex and race.

Another technique to assess the completeness of criminal-history data focuses on unrecognized identification numbers. At times during the collection of criminal-history information for a recidivism study, a

fingerprint-based identification number is sent through III to request a criminal-history record, and none is returned. BJS examines the III response messages to determine why the request was rejected.

To understand the reasons behind non-responses from III, BJS submits a secondary set of record requests directly to the state repositories via Nlets. BJS works with the FBI's CJIS Division to understand why certain identification records are valid at the state level but not indexed in III. BJS has found that such response problems occur because (1) the state has not forwarded the information to the FBI, or (2) the fingerprint image quality is poor and has not been accepted by the FBI. These understandings caused BJS to incorporate into its system an automatic second set of requests to state repositories when the initial request to III fails.

BJS also examines the completeness of disposition reporting in the state and federal criminal-history data and the sources of missing disposition information. The proportion of arrests that provide a case outcome varies across states. These variations may be caused by several factors, including the reporting practices of local law enforcement agencies, prosecutor offices, and courts; variations in the nature of criminal justice system processing within the state; and the inability of the state's central repository to connect a reported disposition to a specific arrest. Low disposition reporting rates are not necessarily an indication of missing or incomplete court data. Some examples are—

- Two separate arrest events may be consolidated into one court case, and the disposition may be attached to only one of the arrests.

- An arrestee could be transferred from one agency to another, and only the receiving agency may report the disposition on its arrest.

- Subsequent arrests for contempt of court, failure to appear, or a parole or probation violation, which result in sentence modifications, may only include disposition information back on the arrest for the underlying conviction offense.

BJS uses the NCRP data to assess the completeness of the disposition and incarceration information reported in the criminal-history data collected for recidivism studies on state prisoners. For the recidivism study of state prisoners who were released in 2005, 93% of the released prisoners had their incarceration sentence that ended in 2005 reported in their criminal-history record.

Summary

Administrative records can provide government agencies with a valuable source of data to support their statistical programs. However, each administrative dataset presents its own unique set of challenges, strengths, and weaknesses. This report illustrates how BJS addresses the challenges involved with one form of administrative data. For many years, BJS has used criminal-history data to study the recidivism patterns of offenders.

Historically, criminal-history research has been a time-consuming and labor-intensive process. One of the highest priorities of BJS has been to develop the capacity to conduct criminal-history research on a more frequent basis and address the growing demand for information on reentry and recidivism. To support this effort, BJS has partnered with Nlets to develop an automated data-collection process that helps to reduce the time and resources needed to collect criminal-history data from repositories nationwide. This system enables BJS (1) to request criminal-history data on samples of study subjects through the FBI's III system for statistical purposes, (2) to reduce the burden placed on the FBI and state repositories to supply these data, (3) to receive the criminal-history records in an automated way, and (4) to parse their contents into a relational dataset.

Building on the criminal-history collection and parsing system, BJS has developed a separate system that converts the state- and federal-specific fields extracted from the criminal-history records into research datasets with standardized national coding structures. Each time a new set of criminal-history data is passed through the standardization system, the conversion process uses the knowledge developed from the prior processing of criminal-history data (with coding rules from previous studies included in the crosswalk tables). As the crosswalk tables expand, the time and resources needed to process a new set of parsed criminal-history datasets is reduced. These systems have expanded both the ability and the potential of the administrative data housed in the U.S. criminal-history repositories to yield informative statistics on recidivism.

NCJ250750

Office of Justice Programs
Building Solutions • Supporting Communities • Advancing Justice
www.ojp.gov