

## **NIBRS Data Available for Secondary Analysis**

**Christopher S. Dunn<sup>1</sup> and Thomas J. Zelenock<sup>2</sup>**

---

The NIBRS data program currently being implemented by the FBI and local law enforcement agencies has by now produced sufficient data for archiving and distribution. Although not representative of crime in the United States, existing NIBRS data can be used to investigate the nature of crimes known to the police compared to the traditional UCR data. The Bureau of Justice Statistics has requested the National Archive of Criminal Justice Data to store and make NIBRS data available to interested users. The data from 1996 will shortly be available from the NACJD web site. The 1996 data contain almost 6.5 million records and the FBI's full file includes about 361 Mbytes of data. The data have been disaggregated from the FBI's complex single file into 11 segment levels or record types. This makes the individual record types easier and faster to analyze than using the full file, which more closely resembles a relational database than a hierarchical file. However, splitting apart the record types requires that special procedures be used to merge files of different record types, which would be necessary if a user were interested in analyzing variables appearing in more than one record type (e.g., comparing offender and victim ages). These procedures are described, and a test comparing the time to run a simple frequency count using the full file against the merged files shows that using the merged files is considerably more efficient. Also discussed are some future developments to facilitate the analysis of NIBRS data.

---

**KEY WORDS:** NIBRS; UCR; crime statistics; secondary analysis; archive.

### **1. INTRODUCTION**

The National Incident-Based Reporting System (NIBRS) has developed from an idea to a plan to a program undergoing implementation to production and submission of data to the FBI from some states in place of summary data from the Uniform Crime Reporting (UCR) program. The time has finally arrived when it is possible for NIBRS data to be made available for research and analysis. The Bureau of Justice Statistics (BJS)

<sup>1</sup>National Archive of Criminal Justice Data, Inter-university Consortium for Political and Social Research, University of Michigan, Ann Arbor, Michigan.

<sup>2</sup>Inter-university Consortium for Political and Social Research, University of Michigan, Ann Arbor, Michigan.

has asked the National Archive of Criminal Justice Data (NACJD) at the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan to archive NIBRS data and make them available to the research community. This article describes how that is being done and where and how the data can be obtained. The data described in this article are the 1996 NIBRS data acquired from the FBI. Any differences in future releases of data will be documented in a revised edition of the data collection codebook.

## 2. FBI DATA AVAILABLE AT ICPSR

Data from the UCR program that are currently archived and available at the NACJD include agency-level monthly data about index crimes reported to police and crimes cleared by arrest, data about property stolen and recovered, incident-level data about homicide from the Supplementary Homicide Reports (SHR), data about law enforcement personnel, and county-level data about offenses known to police and persons arrested. NIBRS data are being archived and made available in the same way; all of these data are available free of charge to users at the NCJD World Wide Web site: <http://www.icpsr.umich.edu/NACJD>.

The simplest way to find the data from the FBI archived at the NACJD is to use the Site Guide selection on the NACJD home page and choose "Data from the FBI" under "Major Data Collections from Other Government Agencies" (about halfway down the *Site Guide* page). If you want go directly there at the outset, add `/otherdata.html#FBI` to the URL address above. From that point on, just follow the links to the NIBRS data.

## 3. TRANSFORMING NIBRS DATA INTO AN ICPSR COLLECTION

Each data collection is prepared according to longstanding ICPSR standards. Raw data are stored in fixed-field ASCII text files, which means that they can be imported into any computer programs that accept ASCII text input, including SPSS, SAS, spreadsheets, and database programs. A number of other files are also prepared to accompany each data file. Codebook files, now commonly in Adobe Acrobat (.pdf) format, describe the data collection and present the record layout and variable names and contents of the collection. The codebook also includes at least a brief description of the data collection procedures, sampling, and other methodological information necessary to use and interpret the data correctly. Usually SPSS and/or SAS data definition syntax files are also prepared. These syntax files

can be run to produce an SPSS system file or an SAS data set file subsequently used for data transformation or statistical analysis purposes.

When data are acquired by the NACJD, a number of checks are run on the data and possible corrections and/or enhancements considered. The data are checked to see if they can be read correctly using the record layout and formatting specifications supplied by the data producer. They are also checked against the coding information supplied with the documentation. If no variable or value labels exist in the data set, these are prepared and added to the syntax files. Frequently, "wild" codes are found. The data producer is always contacted to resolve these undocumented codes by providing the correct information. If the wild codes are meaningless to the data producer, then agreement is obtained to recode them to missing data codes.

The structure of the data file is also checked to see if it is the most efficient structure to release publicly. Often, data producers will output data to the NACJD according to standard output formats. These standard output formats often add unnecessary columns to single or double digit integer codes. If this is discovered in the data file, the wasted space is eliminated and the record layout and codebook are modified accordingly. The 1996 NIBRS data were subjected to these checks and enhancements as well as an overall review of the data structure.

### 3. "Full-File" vs Rectangular-File Formats

NIBRS data as compiled by the FBI are stored in one file. These data are organized by various segment levels; for a description, see the Appendix. Together these various segments total almost 6.5 million records and 361 Mbytes of data for the 1996 data.

Significant computing resources are necessary to work with the data in its full, single-file form. In addition, the user must be prepared to work with data in complex file types. Although it is possible to think of NIBRS as a hierarchical file, its real structure is more similar to a relational database in that there are "key" variables that link the different segment levels together.

In view of the large NIBRS single-file size, its complex relational structure, and the likely interest of many users in only one or two segment levels, the ICPSR (with BJS oversight) decided to make the data available as a set of segment-level files. Thus, each NIBRS segment level in the full file has been made into a separate rectangular raw data file. Linkage (key) variables are used to perform analyses that involve two or more segment levels. Table I shows summary counts of records in each segment-level file for 1996. Empirical differences in analyzing the two approaches (full-file analysis vs merged segment-file analysis) are reported below.

If the user is interested in variables contained on one segment level only (e.g., victim records or offender records), then the data are easy to work

**Table I.** File Types in the NACJD 1996 NIBRS Collection

File type <sup>a</sup>	Record counts	Variables	Total columns
Batch header	18,932	42	272
Administrative	1,064,767	24	91
Offense	1,138,781	20	62
Property	1,178,891	33	135
Victim	1,132,821	50	126
Offender	1,195,543	9	42
Arrestee	284,998	30	107
Group B arrest report	387,622	19	59
Window Clearance 1	371	24	62
Window Clearance 3	357	33	135
Window Clearance 6	17,925	30	107

<sup>a</sup>Segment types are described in the Appendix.

with since each segment-level file is simply a rectangular raw data file. SPSS setups are available to read each segment level. Also, with only one segment level, there is no choice regarding the unit of analysis. The unit of analysis is the segment type, for example, incident, victim, or offender.

However, if the user is interested in variables that are contained on two or more segment levels, then it is necessary to merge files, decide on the unit of analysis, and decide on the number of records per level to read. Decisions about unit of analysis and number of records per level to read are necessary regardless of whether the data are stored in one large file or in separate files.

The 1996 codebook contains a number of examples of SPSS code that can be used to create custom SPSS setups to merge segment level files. In addition, if the user needs assistance in creating a merge program, NACJD staff can give some pointers.

Consider an example involving variables from the victim (04) and offender (05) segment levels. Three units of analysis are available: the victim, the offender, and the incident itself. In addition, there may be up to 999 victim records per incident and up to 99 offender records per incident. The vast majority of incidents has only one or two victims and offenders. If the user does not limit the number of records read per incident, the analysis file will contain mostly missing data and will be excessively large. So even in this simple example, the user must choose the unit of analysis and the number of records per segment level to read and then merge the victim and offender files accordingly. SPSS code for this example is listed in the ICPSR codebook.

### 3.2. Linkage Variables Between the Segments

To work with variables from two or more segment levels, the ICPSR segment-level files must be merged. Variables used to merge the files are the

**Table II.** Linkage Variables by Segment Type

Segment type	Linkage variables
01—Administrative	ORI, incident number
02—Offense	ORI, incident number, UCR offense code
03—Property	ORI, incident number, property description (may not be unique)
04—Victim	ORI, incident number, victim sequence number
05—Offender	ORI, incident number, offender sequence number
06—Arrestee	ORI, incident number, arrestee sequence number
W1—Incident	ORI, incident number
W3—Property	ORI, incident number, property description
W6—Arrestee	ORI, incident number, arrestee sequence number

**Table III.** Structure of Segment Record Count Variable

Variable	<i>N</i> of records per ORI-incident
Victim record 1	3
Victim record 2	.
Victim record 3	.

originating agency identifier, or ORI, and the incident number, as shown in Table II. All segment levels except the batch header contain both the ORI and the incident number. The batch header contains only the ORI. Within each segment level, these variables taken together uniquely identify a record.

On each file a variable has been added to indicate the number of records that comprise an incident. This variable has a valid value for the first record in an incident, and it is missing for subsequent records. For example, if the victim segment file contains three victim records for a particular ORI and incident number, these records would be shown as in Table III. This record count variable is essential for creating files using multiple segment levels.

Each segment-level data file is sorted by ORI and incident number. The ICPSR has checked that records are uniquely identified as described above. All possible code values are documented for each variable. SPSS data definition statements have been prepared to read each segment level. Examples of merging multiple segment levels are available in the codebook and will be available at the NACJD web site.

**4. FULL-FILE vs MERGED SEGMENT-FILE COMPARISON**

To determine the comparative efficiency of processing NIBRS analytical computations using the full file vs merged segment files, we wrote SPSS file input program codes and frequency codes for six variables: age, sex and race of the victim, and age, sex, and race of the offender. We ran these two jobs on a Sun 1000 computer operating under Solaris 2.2 using SPSS for

**Table IV.** Comparison of Time Required to Produce Frequency Counts on Age, Sex, and Race of Victims and Offenders

File	Record count	Processing time (min)
Full file	6,458,872	52
Merged segments		38
Victim segment	1,132,821	
Offender segment	1,195,543	

Unix Release 5.0. The Sun 1000 is a typical medium-sized computer having significantly faster processing capabilities than desktop personal computers.

Table IV shows that using the full file to obtain frequencies of the age, sex, and race of both the offenders and the victims, the job had to read 6,458,872 records, execute an input program that separated victim and offender records from the total number of records, and then count the frequency distribution for the six specified variables. The table shows that this job used 52 min of CPU time to execute a job just to get six frequency distributions.

Using the merged segment files produces a substantial time savings. Table IV shows that the job read only 1,132,821 victim records in separately defining a victim segment system file in SPSS and 1,195,543 records in separately defining an offender segment system file. The job also merged the two files by ORI and incident number, selected incident as the unit of analysis, and calculated the same frequency distributions of the six variables. The CPU time for the merged segment files run was 38 min.

Further testing revealed that simply running the full file to define all variables and read all records took 15 min, almost exactly the time difference between the two jobs.

Thus, we conclude that using the merged segment files procedure represents a more efficient way to process the NIBRS data for analytic purposes when variables from multiple segments are the focus of the analysis. The efficiency would be even greater if the analysis were focused on a single segment-level file because fewer records would be read and fewer variables defined, compared to the full file.

## 5. FUTURE DEVELOPMENTS

Currently, the NIBRS file structure and size of the data set make analysis of the data very difficult for all but the most experienced data analysts supported by computer programmers. We have formulated and tested some input programs for using NIBRS data in a familiar program package, SPSS. We have organized the data for distribution in a format that makes the

analysis easier and more efficient. The documentation provides examples of how to construct SPSS code that facilitates the analysis of variables from multiple segments.

Nevertheless, more improvements are needed. One area involves developing custom NIBRS analysis jobs. A “smart” program is being written and tested that will allow users to respond to English language queries about the variables and unit of analysis desired and that will output the SPSS code based on the analyst’s selections and run the job. Another area involves finding software applications that will accept the front-end analysis definition program but run the job significantly faster than packages like SPSS and SAS. The range of duration for six simple frequency counts in the current test, from just under 40 min to nearly an hour, is unacceptably long.

These and other developments are being supported by the BJS to provide every opportunity to take advantage of the potential of NIBRS data. Interested persons and potential uses can follow these developments at both the BJS web set (<http://www.ojp.usdoj.gov/bjs/>) and the National Archive of Criminal Justice Data web site given above. NIBRS data and documentation are also freely available at the NACJD web site.

## **APPENDIX: NIBRS SEGMENT DESCRIPTIONS**

The NIBRS segment levels are described below. These segment levels are discussed in the order they would appear in the “full” file. Since the ICPSR has separated the full file, each segment level is an individual data file. Examples of SPSS code to conduct analyses on multiple segments using the linkage variables are provided in the ICPSR codebook.

### **Batch Header (BH) Segment**

This separates and identifies individual police agencies by ORI. An individual police agency will appear in the batch header segment once.

### **Group A Offenses**

Each Group “A” Crime Incident Report can contain up to six distinct records, each of which is referred to as a “segment.” In addition, most segments can occur multiple times for a given crime incident. An incident report may consist of many possible combinations of circumstances, ranging from a single administrative, one-offense, one-victim, and one-offender situation to a complex set of multiple offenses, property losses, victims, offenders, and arrestees. In addition, each of the victims may not be

involved in each of the offenses. In other words, one, some, or all of the victims may be connected to each applicable offense.

- *Administrative (01) Segment.* The administrative segment of the Group A Incident Report allows the national UCR program to identify uniquely each criminal incident reported under NIBRS, along with common characteristics of all offenses within each incident, e.g., the date and hour the incident occurred. Each crime incident has one administrative segment record. Thus the number of crime incidents included in NIBRS data is equal to the number of administrative records.
- *Offense (02) Segment.* All offenses, up to 10, associated with a crime incident are listed on separate records. In NIBRS, all Group A offenses occurring in an incident are to be reported. Care must be taken to identify all such offenses involved in an incident. For example, a rape case might also involve the offenses of motor vehicle theft and kidnapping. All three offenses in the incident should be reported. Care must also be taken to ensure that each offense which is reported is a separate, distinct crime, rather than just a part of another offense. For example, every robbery includes some type of assault. Because the assault is an element that makes up the crime of robbery, only robbery should be reported. If during a robbery, however, the victim is forced to engage in sexual relations, both robbery and forcible rape should be reported, because forced sexual intercourse is not an element of robbery. Information on at least one offense must be included in each Group A Incident Report.
- *Property (03) Segment.* Property data are collected to describe the type, value, and, for drugs and narcotics seized in drug cases, quantity of property involved in the incident. Property information is to be submitted separately for each type of property loss, i.e., burned, counterfeited, forged, destroyed, recovered, seized, etc. Consequently, there is no limit to the number of property records that can be recorded for an incident. The following offenses/offense categories can have property records associated with the incident.

Arson

Bribery

Burglary/breaking and entering

Counterfeiting/forgery

Destruction/damage/vandalism of property

Drug/narcotic offenses

Embezzlement

Extortion/blackmail



Fraud offenses  
Gambling offenses  
Kidnapping/abduction  
Larceny, Theft offenses  
Motor vehicle theft  
Stolen property offenses

- *Victim (04) Segment.* Victim data are collected to describe the victims involved in the incident. A separate set of victim data is recorded for each of the victims (up to 999) involved in the incident. There must be at least one set of victim data for each crime incident.
- *Offender (05) Segment.* Offender data include the characteristics (age, sex, and race) of each offender (up to 99) involved in a crime incident whether or not an arrest has been made. The objective is to capture any information known to law enforcement concerning the offenders even though they may not have been identified. There are, of course, instances where no information about perpetrators is known.
- *Arrestee (06) Segment.* Arrestee data are to be reported for all persons apprehended for the commission of Group A or Group B crimes, that is, all offenses except justifiable homicide (not a crime). It must be remembered that the objective here is to collect data on persons arrested, not on charges lodged. For example, a person may be arrested on several charges in connection with the same incident. In this situation, only one set of arrestee data would be reported. Likewise, one person may be arrested many times during a given time span for similar or different violations within a jurisdiction. Because of a separation of time between the arrests, a set of arrestee data is to be reported for each separate arrest since it involves separate crime incidents.

### Group B Offenses

- *Group B Arrest Report (07) Segment.* Because of the different natures of Group A and B offenses, not all details required for Group A Incident Reports are requested for Group B Arrest Reports. Only arrestee data are required for Group B crimes. No other information need be supplied for Group B offenses.

### Window Clearances

Window Clearance Segments reflect data in which the complete Group “A” Incident Report was not submitted to the FBI; rather, only a portion

of the incident was submitted. One reason that would generate a window submission to the FBI is if a summary-based agency converted over to incident-based reporting and, say, an arrest was subsequently made. Only the arrest would be submitted.

- *Window Clearance (W1)*. The same data elements as Segment Level 01 plus up to 10 Group “A” offenses associated with the incident.
- *Window Clearance (W3)*. The same data elements as Segment Level 03 plus up to 10 Group “A” property offenses associated with the incident.
- *Window Clearance (W6)*. The same data elements as Segment Level 06 plus up to 10 Group “A” offenses associated with the incident.