National Center for Juvenile Justice
710 Fifth Avenue, Suite 3000
Pittsburgh, PA 15219-3000

The Reliability of the FBI's NIBRS Data: Five Case Studies

by

Eileen Poe
and
Howard N. Snyder

October 1995

This report was prepared under Cooperative Agreement #92-MC-CX-K002 from the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice.

Introduction

Researchers who utilize the FBI's National Incident-Based Reporting System (NIBRS) data files must be assured of the reliability of these data — that is, the ability of this complex national information resource to replicate the information stored in the local law enforcement agencies that contribute to the NIBRS effort. The data received by the FBI have passed through many stages:

- data entry into the local incident-based reporting (IBR) system,
- transformation of the local data into a State-specified format,
- transmission to the State data collector and entry into the State IBR system,
- conversion of the State's data into the FBI-specified NIBRS format,
- transmission to the FBI, and
- updating the data at all levels when new information on an incident becomes available.

Inconsistencies can develop between any of these stages in the process. A study was designed to evaluate the consistency of the incident records stored in the FBI's NIBRS file with their antecedents in the local and State IBR systems. Consistency was evaluated by comparing the data on a sample of incidents that were likely to involve a non-family abduction of a child. In doing so, this work identified where changes in or loss of information occurred and speculates on possible reasons for such occurrences.

Methodology

Site Selection. In 1992, a telephone survey was conducted of 20 State UCR programs and selected local sites submitting test or production NIBRS data to the FBI. During this phase, basic information was gathered about program status and operation, as well as State/local specifications for data collection and file preparation. A total of five local law enforcement agencies from two states were determined to be the most appropriate study sites.

Site selection criteria included the availability of the site's IBR data, the caseload size for relevant incident types, and the site's ability and willingness to participate in the project's case study activities.

Data Collection. Prior to site visits the five local agencies were asked to prepare a list of the identification numbers of all incidents that had occurred in 1991 or 1992 that met the following criteria: an incident involving a victim under the age of 18, an offender who was a non-family member, and an offense of either murder, violent sexual assault, or kidnapping. These selection criteria paralleled the essential elements of the selection criteria used in the Police Record Study component of the National Incidence Studies of Missing, Abducted, Runaway, and Thrownaway Children (NISMART). From the five local law enforcement agencies, a total of 532 incidents met the selection criteria.

During the site visits to the five local agencies, in-depth interviews were conducted with IBR system personnel. Data development, coding, and transmission processes were discussed and documented. At the end of each interview, the research team asked the local staff to provide a data file containing their automated data on the incidents that met the selection criteria. Similar visits were made to the corresponding State data collectors, and their automated versions of the sample incidents were requested.

With the exception of the data file from one local law enforcement agency, data files provided by the local and State law enforcement agencies contained the IBR records that the agency maintained for its own use. One local agency could only extract their incident records in the State's reporting format.

Members of the research team also met with representatives of the FBI's NIBRS program. Their data collection and processing procedures were reviewed. The research team requested and received a copy of the FBI's 1991 and 1992 NIBRS data files. The use of these data was complicated by the fact that the FBI routinely encrypts incident numbers during the processing of State data to maintain confidentiality. The FBI agreed to pass the sample's local incident numbers through its encryption algorithm, so that the research team could extract the

sample from the FBI's master file. Authorization to provide the research team with this information was obtained through written consent from the chief of police or sheriff of each local agency.

Standardization of Data. The local, State, and Federal incident-based reporting systems maintain different record structures and coding categories. Therefore, it was necessary to standardize the data from each source to compare the contents of these files. Standardization was achieved by extracting a common set of variables with common coding structures from each of the eight files (five local, two State, and one Federal).

To ensure the accuracy of the extracted data, two individuals independently encoded the IBR records. Coders were first trained in the coding rules and were tested until they were yielding similar results on test data. The coders then were given copies of the eight raw data files from which they extracted incident-level data in the standardized format. The standardized data were recorded on data entry forms. The data entry forms from both coders were periodically "spot-checked" by a third individual to verify the coder's understanding of the file format and coding structure. Each coder's data forms were then entered into separate data files, which were compared for discrepancies. Differences were flagged, and the discrepancies were reviewed and corrected.

Analysis

The reliability of the FBI's NIBRS data is related to both (a) sample attrition [i.e., the proportion of local incidents available for analysis at the Federal level] and (b) the consistency of the incident characteristics [i.e., the proportion of local incident characteristics that are replicated in the FBI file]. To test the reliability of the NIBRS data, local incident records were compared to the State and Federal versions of these incidents.

Sample Attrition. Sample attrition was measured by the proportion of the local incident records that were not found in the FBI's NIBRS file. Analysis found that 12% of the sample incidents that were expected to be in the FBI data were in fact not available at the

Federal level (Table 1). Both States had a 12% attrition rate, although the point at which incidents were lost varied by State and local law enforcement agency.

Table 1

Proportion of Local Incidents Found in State and Federal Data Files

	Percent of Local Incidents		
	Not in State File	Not in FBI File	Available in FBI File
All Agencies in Sample	3%	12%	88%
State A Agencies	11	12	88
Agency 1	7	7	93
Agency 2	14	16	84
State B Agencies	1	12	88
Agency 3	0	6	94
Agency 4	0	10	90
Agency 5	1	15	85

Nearly all of the State A incidents that were not available at the Federal level were also unavailable at the State level. This loss may be due to local agencies not transmitting their data to the State. However, another possible reason for the loss at the State level may be related to data-processing problems at the State level. The basis for this speculation is found in the unique character of the data contributed by Agency 2. Agency 2 is the agency that could not provide this project with data in the locally-defined format, but only in the format it transmitted to the State (i.e., we knew what the State had received from this agency). It was expected that the incident records from Agency 2 would be identical in the local and State files. They were not. Therefore, it is possible that at least some of the incidents were lost after they were transmitted to the State.

In contrast to State A, only 1% of local incidents were not available in the State B file; however, a 11% of the incidents in the State B file were not found in the Federal data. Discussions with State personnel point to one possible reason for this loss of information. When the FBI receives records with coding or logical errors, the records are returned to the

State for revision and are not entered into the Federal file. The loss of State B incidents between the State and Federal level may indicate that the State did not revise and then return these records. Although the FBI has no record of rejected records from State B and believes the missing incident records were never received, a contact at the State office recalls there were many incident records rejected by the FBI during this time. However, the State could not confirm that it had resubmitted the ejected records.

Reliability of Incident Characteristics. The usefulness of information at the Federal level is also affected by the stability of the incident characteristics across the local, State, and Federal data files. In other words, do the incident records in the FBI file contain the same information that exists at the local level? The consistency of the data was analyzed variable by variable. Consistency was quantified by the proportion of available incidents at the Federal level with the same coding value for a specific variable in all three versions of the incident (local, State, and Federal).

Overall, the consistency of the data in the local and FBI files is quite high. In fact, with the exception of Agency 3 data, the average consistency rate for incident level data is well above 95% (Table 2). For example, all (100%) of State A incidents in the FBI file had the same *most serious offense* as did the local records. In comparison, the *most serious offense* consistency rate for State B was lower (86%). This lower joint rate was caused by the relatively low rate for Agency 3 (64%); data from both Agencies 4 and 5 displayed very high consistency rates (97% and 95%, respectively).

Inter-coder Reliability. A major reason for the low consistency rate of Agency 3's data is, unlike the other local agencies studied, Agency 3 sends the State copies of the paper incident reports, hand-written by officers, rather than automated data files. The local and State data processing staffs, therefore, independently interpret these paper reports and enter their decisions into their own IBR systems. Consequently, Agency 3 incidents are more likely to contain discrepancies between the local and State level than if automated records of these incidents had been sent to the State. Fortunately for this project, a closer look at Agency 3's

Table 2

Proportion of Cases with Consistent Codes Across Local, State and Federal Levels by Variable and Reporting Agency

Variable	All Agencies	State A			State B			
		Total	Agency 1	Agency 2	Total	Agency 3	Agency 4	Agency 5
Counts								
# of offenders	94%	100%	100%	100%	92%	73%	100%	99%
# of victims	96	100	100	100	95	85	99	99
# of offenses	95	100	100	100	93	82	100	96
# of arrestees	88	93	92	94	86	66	93	96
Incident Information								
Year	99	100	100	100	99	96	100	100
Month	99	100	100	100	99	96	100	100
Date	99	100	100	100	99	97	100	100
Hour	92	99	100	98	90	67	100	100
Excpt clearance code	93	96	94	98	92	75	99	99
Excpt clearance year	99	97	94	100	99	n/a	99	99
Excpt clearance month	99	97	94	100	99	n/a	99	99
Excpt clearance date	99	97	94	100	99	n/a	99	99
Offense Information[1]								
Offense	89	100	100	100	86	64	97	95
Offense location	79	100	100	100	73	32	90	91
Weapon	99	99	100	98	99	n/a	100	99
Victim Information[2]								
Age	95	100	100	100	94	78	100	100
Sex	98	100	100	100	98	92	100	100
Race	99	100	100	100	99	97	100	100
Ethnicity	85	100	100	100	80	32	100	100
Resident status	84	100	100	100	80	32	100	99
Relationship	100	100	100	100	99	n/a	100	99
Victim injury	100	99	98	100	100	n/a	100	99
Offense	87	100	100	100	83	55	96	95
Offender Information[2]								
Age	94	100	100	100	92	77	99	98
Sex	96	100	100	100	95	85	100	99
Race	97	100	100	100	96	86	100	99
Arrestee Information[2]								
Age	86	82	92	71	87	67	94	96
Sex	90	95	94	96	88	69	96	96
Race	90	95	94	96	88	69	96	96
Arrest year	89	95	94	96	87	66	96	96
Arrest month	89	95	94	96	87	64	96	96
Arrest date	88	95	94	96	86	60	96	96

[1] Offense information is provided for the most serious offense associated with the incident.

[2] Victim, offender, and arrestee information is provided for the youngest actor associated with the incident.

N/A indicates that the data element was not routinely available in the local data system.

data at the local and State levels provides a natural test of the effect of inter-coder reliability on the quality of the Federal NIBRS data.

The majority of inconsistencies found when comparing Agency 3 incident data at local, State, and Federal levels occur between the local and State levels, likely reflecting the independent coding decisions made by Agency 3 and State data-processing personnel. For instance, it is believed that inconsistent interpretations of the paper records by the local and State data-processing personnel accounted for slightly more than one-half of the differences between local and State data files for the variable *most serious offense*. The majority of incidents with differences in local and State coding revealed a code of *90Z* or *other type B offenses* on the local file. Data at the State level reflect a more specific interpretation of the same incident as a *sexual offense* (offense codes 11A through 11D).

Observed differences in the Agency 3 coding of both *victim age* and *offender age* may also be attributable to inconsistent coding at the local or State levels. In fact, examination of these variables revealed that a large portion of the differences among both variables is the result of *missing data* codes in the local file.

A study of the *offense location* variable in the Agency 3 data also indicates problems with inter-coder reliability. Comparison of the local, State, and Federal data files revealed differences in 67 of Agency 3's 99 incidents for this variable. (This total number of incidents reflect those with a record available at the local, State, and national levels). Although there is no definite explanation for these inconsistencies, it is curious that inconsistencies in the coding of this variable by Agency 3 and State personnel markedly decreased from 1991 to 1992.

Differences also existed among 33 of the 99 incidents in the variable *incident hour*. Examination of this variable revealed the majority of these differences was due to the State practice of simply dropping the minutes from incident hour (e.g., 11:45 AM became 11 AM), while Agency 3 coded both the hour and the minute of the incident. This project's recoding of the Agency 3 data (as per FBI coding guidelines) rounded the time values to the nearest whole hour for analysis, resulting in the inconsistencies in the two data sets.

Incompatible System Designs. There is an additional reasons why certain data elements from incidents originating in Agency 3 exhibited lower consistency rates than data from the other law enforcement agencies studied. For part of the study period Agency 3's IBR system did not capture the following variables: *victim resident status and ethnicity; number of arrestees; arrestee age, sex, and race; and year, month and date of arrest.* By 1992, however, Agency 3 began to code these variables. In contrast, the State had been distilling this information from Agency 3's paper files and entering the information into the State system during the entire study period. Consequently, discrepancies resulted between the local and State automated files and exceptionally low consistency rates among these variables.[1]

If the consistency analyses had been run differently (distinguishing incidents which had missing information at the local level by design for a limited time period), a much higher consistency rate would have resulted. For instance, the original consistency rate in Agency 3's data for *victim ethnicity* was 32% (Table 2). However, nearly half (48%) of Agency 3's incidents did not contain values for this variable in large part because the data element was excluded by design from the local system for part of the study period (Table 3). If these incidents were classified as not having inconsistent data, the consistency rate for Agency 3 for this variable increases to 81%. However, this new consistency rate assumes no discrepancies in Agency 3's incidents at the local, State, and national levels had complete information been available. Therefore, this new rate may be somewhat higher than what would actually be found.

Data Revision Problems. Data revisions problems were discussed earlier in regard to sample attrition in the State B data. A possible explanation for this attrition is that, once the FBI rejects a submission for coding irregularities, State data processing does not resubmit the

[1]The following additional variables were not collected by Agency 3 during the entire study period: victim injury; the relationship of victim to offender; year, month, and date of exceptional clearance; and weapon connected to offense. This project's calculation of consistency rates for these variables were not influenced by this total lack of reporting at the local level because variables never collected at the local level was excluded from consistency analysis and not considered inconsistent.

corrected record. Another possible data revision problem includes information about an incident that is developed after the transmission of the data to the next level has occurred. For example, inconsistencies in arrest-related and exceptional clearance-related variables may be due to the addition of new information or changing of information to the incident records at the local level that is not communicated to the State. However, the consistency rates for arrestee and clearance variables are high for incidents originating from Agencies 1, 2, 4, and 5 (Table 2). The low rates in Agency 3 can largely be attributed to the aforementioned problem of coding reliability. Therefore, the submission of revisions after an incident has been accepted at the next level does not appear to be a significant problem for NIBRS.

Table 3

Proportion of Agency 3's Incidents with Consistent Coding Given Different Interpretation of "Missing Data"

	Victim Ethnicity
Total	100%
Incidents with Consistent Codes	81
Incidents with Consistent Nonmissing Codes	32
Incidents with Local Data Missing by Design	48
Incidents with Inconsistent Codes	19

Data Conversion Problems. The reliability of Federal data may also be influenced by reliability of automated transformations that occur between the local and State and the State and Federal levels. For example, Table 2 reveals a consistency rate of 71% among incidents originating in Agency 2 for the variable *arrestee age*. Examination of this variable revealed 14 differences between the local and State files. Twelve of these differences originate from incidents in which the arrestee is reported as a 24-year-old in the State file, while the ages in the local Agency 2 data varied. As previously stated, the records the study received from Agency 2 were identical to the records sent for State processing (i.e., there were no coding or

formatting differences between the two files). Consequently, the problem is likely to have developed through a conversion problem at the State level.

Another example of data conversion problems occurs when local systems use coding structures that are inconsistent with State and Federal formats. For example, Agencies 4 and 5 use an offense location code that does not exist at the Federal level (*condominium*). For these agencies, most of the differences in the coding of the variable *offense location* can be attributed to the use of an offense location code of *unknown* for these incidents at the Federal level. Apparently, either the process by which local data are transformed into the State format or State data into the FBI format is not designed to handle these *out-of-bounds* codes.

Conclusions

To assess the potential of using NIBRS to study non-family abductions and related child victimizations, the Federal data were evaluated with respect to the consistency of information stored at the local and Federal levels. The following findings reflect the results of this analysis:

- The completeness of the Federal data is reflected by the proportion of local incidents available for analysis at the Federal level. Overall, 88% of all incidents reported to local law enforcement agencies were available for analysis at the Federal level. The point at which incidents were lost varied by State and local law enforcement agency.

- In 4 of the 5 agencies studied, analysis revealed that to a very high degree the incident characteristics found in the local IBR systems were also found in the Federal data.

- The independent IBR coding by local and State personnel of the paper files from one local agency provided an opportunity to assess the impact of inter-coder reliability on the quality of the NIBRS data. The low consistency rate of the data from this one agency at the local and State level raises serious concerns about the effect of inter-coder reliability of the Federal NIBRS data.

- Although there is evidence from a study of sample attrition that about 12% of incidents never find their way into the FBI data file, updated incident

information developed after the initial transfer of data files appears to have consistency rates that are relatively high.

- Incompatibly designed system coding structures at the local or State levels (i.e., coding structures that are incompatible with the NIBRS format) may result in inconsistent reporting of information.

- Finally, data transformation procedures at the local or State level may also effect the reliability of data at the Federal level.

Recommendations

NIBRS data maintained by the FBI appear to reasonably reflect the information stored in local and State information systems. However, some problems were discovered. To increase the ability of the Federal NIBRS data to support research on the non-family abductions of children, the following recommendations are offered:

- FBI and State data collectors should develop procedures to ensure that all local incidents are reported to them. They should also develop a process to ensure that incidents returned for corrections are, in fact, corrected and returned.

- Data transformation programs should be routinely reviewed and updated when the feeder system changes. Problems associated with improperly functioning data transformation programs should be resolved.

- Local and State information system designers should be strongly encouraged to develop and enhance their systems to capture information at as detailed a level as possible. However, these coding structures must be compatible with the NIBRS format. Recoding rules for transformations from the local to the State IBR system, and from the State to the Federal NIBRS system, should be a required and reviewed component of all system designs.

- It is likely that more training is needed to improve the inter-coder reliability of the NIBRS data. Systematic auditing of local IBR records is also encouraged.