This report was prepared by RTI International using federal funding provided by the Bureau of Justice Statistics.

Document title: NIBRS Estimation and Extract File Creation Methodology Report

Authors: Marcus Berzofsky, DrPH, RTI International
Dan Liao, PhD, RTI International
G. Lance Couzens, RTI International
Ian Thomas, RTI International
Caroline Kery, RTI International
Kim Janda, RTI International
Mark Pope, MS, RTI International

BJS Project Managers: Lizabeth Remrey, PhD, Statistician, Law Enforcement Incident-Based Statistics Unit

Erica L. Smith, Unit Chief, Law Enforcement Incident-Based Statistics Unit

Document No.: NCJ 310516

Publication Date: December 2025

Award No.: This project was supported by award number 15PBJS-22-GK-00713-BJSBS

Abstract:

The National Incident-Based Reporting System (NIBRS) offers the federal government and researchers a much richer set of information about crimes recorded by the police than the summary format previously used by law enforcement agencies. While the summary format simply provided a count of offenses recorded in a given month, NIBRS provides detailed information at the incident level about characteristics related to the incident, offense, victim, offender, and arrestee. However, because not all law enforcement agencies have transitioned to submitting information to the Federal Bureau of Investigation (FBI) using NIBRS and not all NIBRS reporting agencies provide a full 12 months of data, a robust and complex statistical estimation process is needed to produce representative and unbiased estimates of crimes recorded by police.

Beginning with the 2021 data year, RTI International has produced representative estimates of crimes recorded by the police for the use as official statistics by BJS and FBI. Estimates are produced across several levels of geography, including the national, regional, state, metropolitan statistical area, judicial district, and FBI Field Office levels. Estimates cover topics related to the characteristics of incidents, offenses, victims, and arrestees. Additional estimates are produced specifically on the topics of assaults to law enforcement officers, gun violence, and details of drug offenses. Estimates of victims and

arrestees are produced among all persons as well as by population subsets by age, sex, and race (e.g., Females, aged 18–24).

In this report, the methodology implemented to produce the annual estimates of crimes recorded by police is described in detail. The report covers the following topic areas: (1) the procedures for ingesting the FBI Criminal Justice Information Services (CJIS) NIBRS database and implement the transformations needed for the estimation process; (2) the structure of the indicators for which estimates are produced; (3) the procedures for accounting for agencies that do not participate in NIBRS through weighting; (4) the procedures for handling missing or unknown values through imputation; (5) the procedures for developing population estimates; (6) the procedures for measuring statistical uncertainty; (7) the process for constructing confidence intervals; (8) the publication rules for suppressing estimates; (9) the structure and format of the output; (10) how to interpret the estimates; (11) how to conduct statistical tests using the estimates; and (12) the procedures for creating the NIBRS Extract Files.

This page intentionally left blank.

NIBRS Estimation and Extract File Creation

Methodology Report

Prepared for

U.S. Department of Justice
Bureau of Justice Statistics

999 North Capitol Street, NE
Washington, DC 20531

Prepared by

Marcus Berzofsky, DrPH; Dan Liao, PhD; G. Lance Couzens; Ian Thomas; Caroline Kery; Kim Janda; and Mark Pope, MS

RTI International
3040 East Cornwallis Road
Research Triangle Park, NC 27709

Federal Award Number: 15PBJS-22-GK-00713-BJSBS
RTI Project Number: 0218750
December 2025

RTI INTERNATIONAL

Contents

1. Introduction

The National Incident-Based Reporting System (NIBRS) is part of the Federal Bureau of Investigation's (FBI) Uniform Crime Reporting (UCR) Program. NIBRS was developed in the 1980s through a collaboration between the Bureau of Justice Statistics (BJS) and the FBI that resulted in the ***Blueprint for the Future of the Uniform Crime Reporting Program – Final Report of the UCR Study***. The NIBRS data collection, maintained and managed by the FBI's Criminal Justice Information Services (CJIS) Division, officially began receiving data from states and local law enforcement agencies in 1989. Prior to launching NIBRS, crime data had been collected as monthly summary tallies of a limited number of offenses through the FBI's Summary Reporting System (SRS). If multiple offense types were involved, SRS guidelines indicated that reporting agencies should employ the *hierarchy rule* and provide a count of the most serious offense in the incident.

NIBRS was designed differently from the SRS in that it was *incident-based* rather than *offense-based*. This means that NIBRS does not apply the hierarchy rule whereby the most serious offense was the only offense reported; instead, NIBRS counts each offense that is part of a crime incident. NIBRS also collects more details on criminal incidents such as victim and offender demographics, the date and time the crime incident occurred, the relationship(s) between victim(s) and offender(s), property involved, location types, injury types, and weapon involvement. Additionally, NIBRS-reportable offenses are broader in scope than the traditional SRS offenses, covering crimes such as human trafficking, simple assault, drug offenses, and fraud, among others. NIBRS captures detailed incident and arrest characteristics for 52 different offenses that can occur within a crime incident and collects arrest-only information for an additional 10 offenses.[1]

In 2015, recognizing the lack of detailed national information about crime, BJS and the FBI began major efforts to transition all states and law enforcement agencies to NIBRS reporting. They established the National Crime Statistics Exchange (NCS-X), a BJS and FBI joint effort that provided grant funding to states and agencies to transition to NIBRS.[2] NIBRS became the primary national law enforcement crime reporting standard as of January 1, 2021. Based on data from 2023, about 83% of the nation's population is covered by agencies certified to report data to NIBRS; these agencies comprise about 73% of all crime-reporting law enforcement agencies in the United States.[3]

With the expansion of NIBRS, BJS identified two key priorities: (1) use NIBRS data to comprehensively characterize crime nationally and at other levels of geography, including state and region; and (2) make NIBRS data more accessible to a range of constituents, including other federal agencies, researchers, state and local governments, and communities. To meet these priorities, BJS produces two primary data products based on annual NIBRS data:

1. NIBRS Estimation Program (EP), which generates national and subnational estimates of crime that are imputed and weighted to account for missing data, and

[1] Federal and tribal agencies can report an additional 19 offense types and 3 arrest-only offense types. For more information, see *Indicators for Crime Estimates Using NIBRS Data* at ***https://bjs.ojp.gov/content/pub/pdf/iceunibrsd.pdf***.

[2] For more information, see *Better Data for Evolving Crime Trends 2012–2022: NIBRS Transition* at ***https://www.theiacp.org/sites/default/files/2022-08/August%202022_Better%20Data.pdf***.

[3] This includes agencies that reported at least 1 month of NIBRS data during 2023. See page 4 of *UCR Summary of Crime in the Nation: 2023* at ***https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/explorer/crime/special-reports*** .

2. NIBRS Data Extract Files, which aggregate NIBRS relational data files into a series of flat files that are more conducive to research and analysis using statistical software and business intelligence tools.

The purpose of this document is to describe the methodology and procedures used by BJS to create the annual NIBRS National Estimates and the NIBRS Data Extract Files. ***Chapter 2*** describes the receipt and processing of NIBRS data from the FBI and the data quality checks performed on these data. ***Chapter 3*** describes the procedures used for the NIBRS EP, including the indicators to be estimated, imputation, and weighting. Finally, ***Chapter 4*** describes how the NIBRS Data Extract Files, which facilitate analysis by researchers and others, are created and validated, as well as use cases for when to use specific files and variables based on the research question or topic being considered.

2. Receipt and Processing of Base NIBRS Data

2.1. Introduction

The initial step to both the production of National Incident-Based Reporting System (NIBRS) estimates and the creation of the NIBRS Extract Files is to obtain, process, and transform the FBI Uniform Crime Reporting (UCR) database. A multistep process is required to ensure data received by the FBI is properly processed and of the quality necessary to produce estimates and extract files that meet the Bureau of Justice Statistics' (BJS) standards for official release. This chapter details the procedures used to prepare data for the creation of estimates and extract files.

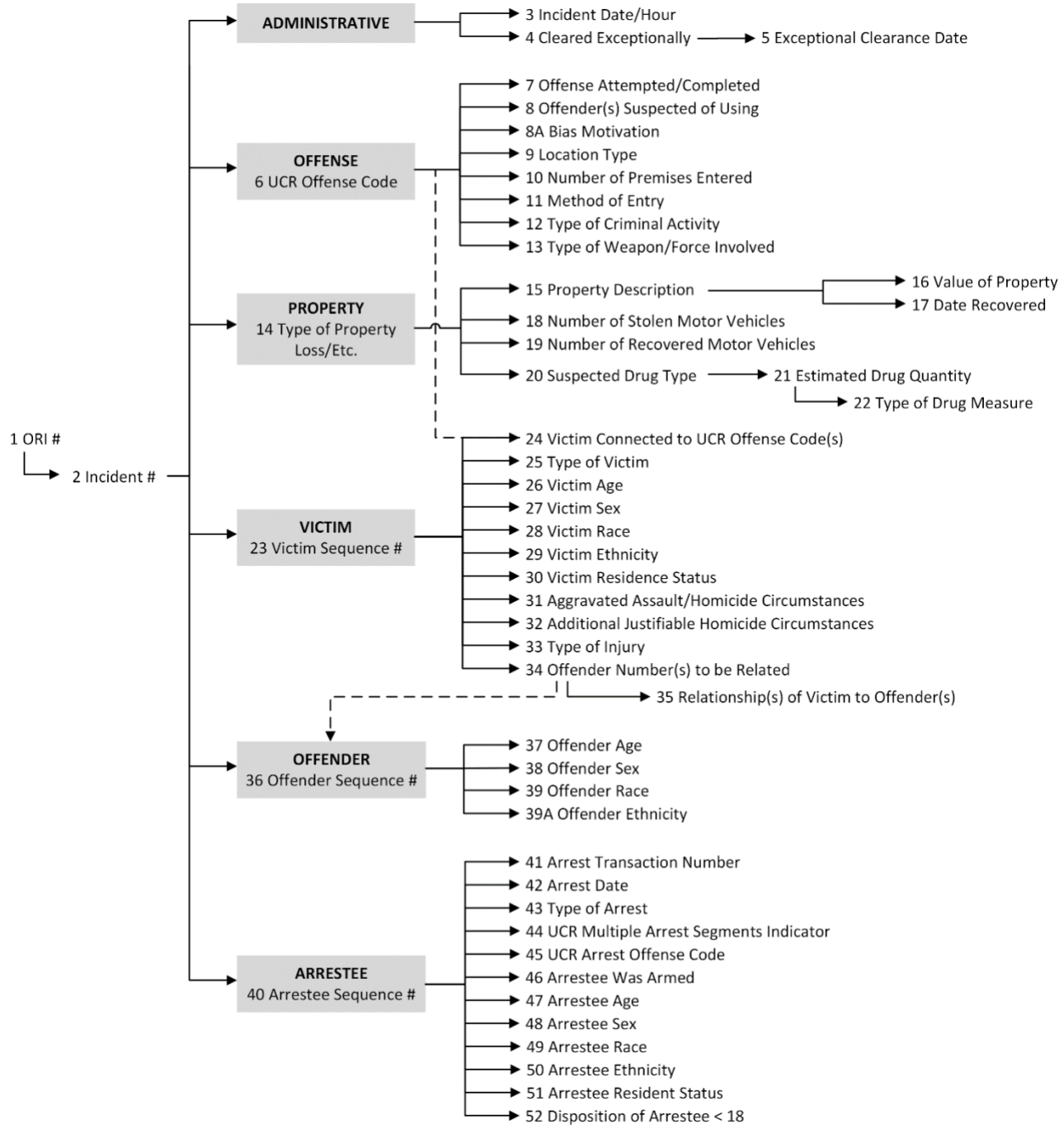
2.2. Background on Data Received from the FBI CJIS

2.2.1. Data received from the FBI and its structure

To produce the yearly NIBRS estimates using the NIBRS Estimation System (NES), RTI receives a copy of the production UCR database. This database is provided twice annually, once in February for testing and once post-reporting cutoff in mid-April for final estimates. The uncompressed size of the database is substantial and increases annually, currently at 438 GB for the April 2024 version. This includes the full database, operational tables, and NIBRS data.

NIBRS follows a hierarchical structure and contains a vast amount of data. ***Exhibit 2-1*** illustrates the six main file-type levels in this hierarchical system: administrative, offense, property, victim, offender, and arrestee. These levels correspond to different segments in the NIBRS data. Each agency maintains a record for every reported incident, which includes details on offenses, offenders, and victims. When an incident involves multiple offenses, victims, offenders, or arrestees, a separate record is created for each. The hierarchical structure is preserved through unique identifiers (IDs), enabling users to link offenses, victims, offenders, and arrestees within a given incident and agency.

Exhibit 2-1. Hierarchical structure of NIBRS data



Source: FBI's *Crime Data Explorer.*

2.2.2. Details on loading the CJIS database into RTI's database

Currently, the estimation process queries a copy of the FBI's operational UCR production database, hosted on a PostgreSQL server that was exported and delivered after the April cutoff for each data year.

The files are generated by CJIS with the following commands (environment variables are denoted with a $ followed by capital letters):

- pg_dump -h $PGHOST -p $PGPORT -U $PGUSER -d $DBNAME --create --no-owner --no-acl --no-privileges --format=plain --section=pre-data --file=$DBNAME-$EXPORT_YEAR-$EXPORT_MONTH-pre.sql $DBNAME
- pg_dump -h $PGHOST -p $PGPORT -U $PGUSER -d $DBNAME --no-owner --no-acl --no-privileges --format=custom --section=data --compress=9 --file=$DBNAME-$EXPORT_YEAR-$EXPORT_MONTH-data.dmp.gz $DBNAME
- pg_dump -h $PGHOST -p $PGPORT -U $PGUSER -d $DBNAME --no-owner --no-acl --no-privileges --format=plain --section=post-data --file=$DBNAME-$EXPORT_YEAR-$EXPORT_MONTH-post.sql $DBNAME

CJIS uploads these files via Teleporter, and RTI receives the export of the database. This export is provided in three key files (using the May 2022 export of the ucr_prd database as an example):

1. ucr_prd_2022_05-pre.sql: Contains the structured query language (SQL) for setting up the database, including table creation, views, and table functions.
2. ucr_prd_2022_05-data.dmp.gz (27GB): A compressed file containing the main data.
3. ucr_prd_2022_05-post.sql: Includes final SQL setup logic, such as creating indexes and constraints.

These files are loaded into the RTI database using the following commands:

- $ psql -h $PGHOST -p $PGPORT -U $PGUSER --password < ./ $DBNAME-$EXPORT_YEAR-$EXPORT_MONTH-pre.sql
- $ pg_restore -h $PGHOST -p $PGPORT -U $PGUSER -W --clean --no-acl --no-owner -d $DBNAME-$EXPORT_YEAR-$EXPORT_MONTH-data.dmp.gz
- $ psql -h $PGHOST -p $PGPORT -U $PGUSER --password < ./$DBNAME-$EXPORT_YEAR-$EXPORT_MONTH-post.sql

The database is modified to add a backward-compatible layer to the new database in the following steps:

1. A schema named public is created.
2. Reference tables are created using data provided by CJIS.
3. Backward-compatible views are created to behave like the tables from the original database schema.
4. Some of the views are materialized to improve read access speed.

In general database terminology, the term "schema," as referenced in step 1, is used to describe the structure of the database. In PostgreSQL, the same term is used to describe an independent area of the database.

A "view," as referenced in step 3, is effectively a saved query that is run each time the view is accessed. When a view is "materialized," the queried data from the view are copied and saved, which enables the creation of indexes on the copied data. Data in the materialized view are still dependent on the referenced tables and can be manually refreshed if the underlying data have changed.

Once this is complete, queries to the tables in ucr_prd 'schema' (e.g., ucr_prd.nibrs_incidents) will return results in the new schema definition, and queries to the views in public 'schema' (e.g., public.nibrs_incidents) will return results as if querying the original database.

Impact of database restructuring by the FBI

The original queries used for the NES were based on the schema definition (database structure) of the NIBRS data for the 2021 data year. The entire process for extracting and transforming the data and running the imputation and weighting procedures was based on these queries. Then, in 2022, the FBI CJIS completed a restructure of the main UCR database. This resulted in a database structured with a different schema definition than the version of the database against which the NES was developed.

The new schema definition fixed challenges related to standardization (ensuring all field values have the same type and structure) and normalization (the process of structuring a database to reduce data redundancy) that were present in the original NIBRS database structure. An example of standardization is that all yes and no responses are now returned as only 'Y', 'N', or (blank). In the original schema, the yes/no responses could have been returned as 'yes', 'no', NULL, 'true', 'false', 1, or 0, which are all commonly used ways to convey the same concept. The variation in response options reduced the ability to determine the quality and completeness of responses across reporting entities; the new schema improved the analyzability of the data.

Database normalization is a way of organizing data to reduce duplication and ensure consistency. It involves structuring data into related tables and defining clear relationships between them, making the information easier to manage and update. This process enhances data accuracy, efficiency, and flexibility. The normalization activity is simple in principle but can be difficult in implementation, as structuring a large, complex database requires making decisions informed by what the data will be and how the data will be used. For NIBRS data, the challenges related to normalization were exacerbated by changes made to the data collection over the years, including adding new data elements and augmenting existing data elements with new data values. The result was that the original schema included structural choices that did not work as well once the database had been in use for decades. Some examples include age being underspecified and the inability to connect an agency to more than one county or metropolitan statistical area (MSA).

Exhibit 2-2 presents the *nibrs_age* reference table from the original schema. As shown in the table, *nibrs_victim* had an *age_id* and *age_num* column. The *age_id* referenced what *age_num* meant. If *age_id* was 5, then *age_num* was the age of the victim in years. If it was 3, then it referenced the number of days between 7 and 364 that a baby had been alive. ***Exhibit 2-3*** presents the age lookup under the new (current) database schema. In the current schema, *nibrs_victim* references only one age field, *age_code*; this code is either the age in years or a string representing the special age cases (NN, NB, BB, 0, or NS). This can be joined to *age_name* to get the exact meaning written out (Under 24 Hours, Unknown, or 30 years old, etc.). Therefore, in the new schema, no join is needed to know the victim age via *age_code*, and one simple join will retrieve the unambiguous text-based age. By comparison, in the original schema, extra logic had to be built into the query to retrieve the same information.

Exhibit 2-2. Reference table for original NIBRS age schema

age_id	age_num	age_name
1	NN	Under 24 Hours
2	NB	1–6 Days Old
3	BB	7–364 Days Old
4	0	Unknown
5	AG	Age in Years
6	99	Over 98 Years Old

Exhibit 2-3. Reference table for new NIBRS age schema

age_code	age_name	sort_order	is_juvenile	asr_age_range_code
NN	Under 24 Hours	1	TRUE	
NB	1–6 Days Old	2	TRUE	
BB	7–364 Days Old	3	TRUE	
0	Unknown	103	FALSE	
NS	Not Specified	104	TRUE	
1	1 Years Old	4	TRUE	0_9
2	2 Years Old	5	TRUE	0_9
3	3 Years Old	6	TRUE	0_9
4	4 Years Old	7	TRUE	0_9
5	5 Years Old	8	TRUE	0_9
6	6 Years Old	9	TRUE	0_9
7	7 Years Old	10	TRUE	0_9
8	8 Years Old	11	TRUE	0_9
9	9 Years Old	12	TRUE	0_9
10	10 Years Old	13	TRUE	10_12
11	11 Years Old	14	TRUE	10_12
12	12 Years Old	15	TRUE	10_12
13	13 Years Old	16	TRUE	13_14
14	14 Years Old	17	TRUE	13_14
15	15 Years Old	18	TRUE	15
16	16 Years Old	19	TRUE	16
17	17 Years Old	20	TRUE	17
18	18 Years Old	21	FALSE	18
19	19 Years Old	22	FALSE	19
20	20 Years Old	23	FALSE	20
21	21 Years Old	24	FALSE	21

age_code	age_name	sort_order	is_juvenile	asr_age_range_code
22	22 Years Old	25	FALSE	22
23	23 Years Old	26	FALSE	23
24	24 Years Old	27	FALSE	24
25	25 Years Old	28	FALSE	25_29
26	26 Years Old	29	FALSE	25_29
27	27 Years Old	30	FALSE	25_29
28	28 Years Old	31	FALSE	25_29
29	29 Years Old	32	FALSE	25_29
30	30 Years Old	33	FALSE	30_34
31	31 Years Old	34	FALSE	30_34
32	32 Years Old	35	FALSE	30_34
33	33 Years Old	36	FALSE	30_34
34	34 Years Old	37	FALSE	30_34
35	35 Years Old	38	FALSE	35_39
36	36 Years Old	39	FALSE	35_39
37	37 Years Old	40	FALSE	35_39
38	38 Years Old	41	FALSE	35_39
39	39 Years Old	42	FALSE	35_39
40	40 Years Old	43	FALSE	40_44
41	41 Years Old	44	FALSE	40_44
42	42 Years Old	45	FALSE	40_44
43	43 Years Old	46	FALSE	40_44
44	44 Years Old	47	FALSE	40_44
45	45 Years Old	48	FALSE	45_49
46	46 Years Old	49	FALSE	45_49
47	47 Years Old	50	FALSE	45_49
48	48 Years Old	51	FALSE	45_49
49	49 Years Old	52	FALSE	45_49
50	50 Years Old	53	FALSE	50_54
51	51 Years Old	54	FALSE	50_54
52	52 Years Old	55	FALSE	50_54
53	53 Years Old	56	FALSE	50_54
54	54 Years Old	57	FALSE	50_54
55	55 Years Old	58	FALSE	55_59
56	56 Years Old	59	FALSE	55_59
57	57 Years Old	60	FALSE	55_59
58	58 Years Old	61	FALSE	55_59

age_code	age_name	sort_order	is_juvenile	asr_age_range_code
59	59 Years Old	62	FALSE	55_59
60	60 Years Old	63	FALSE	60_64
61	61 Years Old	64	FALSE	60_64
62	62 Years Old	65	FALSE	60_64
63	63 Years Old	66	FALSE	60_64
64	64 Years Old	67	FALSE	60_64
65	65 Years Old	68	FALSE	65P
66	66 Years Old	69	FALSE	65P
67	67 Years Old	70	FALSE	65P
68	68 Years Old	71	FALSE	65P
69	69 Years Old	72	FALSE	65P
70	70 Years Old	73	FALSE	65P
71	71 Years Old	74	FALSE	65P
72	72 Years Old	75	FALSE	65P
73	73 Years Old	76	FALSE	65P
74	74 Years Old	77	FALSE	65P
75	75 Years Old	78	FALSE	65P
76	76 Years Old	79	FALSE	65P
77	77 Years Old	80	FALSE	65P
78	78 Years Old	81	FALSE	65P
79	79 Years Old	82	FALSE	65P
80	80 Years Old	83	FALSE	65P
81	81 Years Old	84	FALSE	65P
82	82 Years Old	85	FALSE	65P
83	83 Years Old	86	FALSE	65P
84	84 Years Old	87	FALSE	65P
85	85 Years Old	88	FALSE	65P
86	86 Years Old	89	FALSE	65P
87	87 Years Old	90	FALSE	65P
88	88 Years Old	91	FALSE	65P
89	89 Years Old	92	FALSE	65P
90	90 Years Old	93	FALSE	65P
91	91 Years Old	94	FALSE	65P
92	92 Years Old	95	FALSE	65P
93	93 Years Old	96	FALSE	65P
94	94 Years Old	97	FALSE	65P
95	95 Years Old	98	FALSE	65P

age_code	age_name	sort_order	is_juvenile	asr_age_range_code
96	96 Years Old	99	FALSE	65P
97	97 Years Old	100	FALSE	65P
98	98 Years Old	101	FALSE	65P
99	Over 98 Years Old	102	FALSE	65P

In the original schema definition, the reference for county and for MSA was stored in the agencies table as *county_name* and *msa_name*. Agencies belonging to more than one county or MSA would require a comma-separated list stored in their *county_name* or *msa_name* fields. The new schema now stores that relationship in the *ref_agency_county* table that matches *agency_id* to *county_id* and *metro_div_id*. If an agency is in only one county, then its *agency_id* is stored only once in that table. However, if an agency has multiple associations (i.e., the agency jurisdiction crosses county or MSA boundary lines), then the *agency_id* can be entered multiple times into the table to associate the agency with each *county_id* and *metro_div_id* with which it has a relationship.

CJIS made a number of these types of schema changes to the database. While the underlying information held in the database was the same, changes to the structure of the database required significant modification to the queries the NES uses.

RTI's solution to this was to create views that translate the new schema into the original one. For example, the public.agencies view is created in the public "schema" of the new database by querying the ucr_prd.agencies table and joining it with other tables to pull together all data presented in the original table. This allows the NES pipeline to continue using the queries as developed until the project has sufficient time and resources to convert the existing queries into the new schema and test them.

2.2.3. Transformations made prior to starting the estimation process and extract file creation

Before any additional processing began, RTI engaged in a database schema transformation process. First, the 20 reference tables with predefined mappings of IDs to codes and names were created. They serve as a "dictionary" or "lookup" to translate original identifiers into the new schema's format. There are 27 views, which are virtual tables created by executing SQL queries to join tables from the new schema and make use of the reference tables. The views combine data from multiple tables to create a unified structure. Of the 27 views, 13 are materialized views; these are database objects that store the results of the queries physically, rather than generating them dynamically when queried. Materialized views are used here to precompute and store results, which can include complex joins or aggregations. In addition to improving query speed, this process helps ensure the new schema works efficiently while maintaining compatibility with the original data structure. The indices generated from this transformation process are listed in ***Exhibit 2-4***.

Exhibit 2-4. Reference tables, views, and materialized views in NIBRS backward-compatible database*

Reference tables	Views	Materialized views
nibrs_age	agencies_yearly	agencies
nibrs_bias_list	nibrs_activity_type	agency_county
nibrs_circumstances	nibrs_arrest_type	county
nibrs_cleared_except	nibrs_arrestee_weapon	nibrs_arrestee
nibrs_criminal_act_type	nibrs_assignment_type	nibrs_bias_motivation
nibrs_drug_measure_type	nibrs_criminal_act	nibrs_incident
nibrs_ethnicity	nibrs_month	nibrs_offender

Reference tables	Views	Materialized views
nibrs_injury	nibrs_relationship	nibrs_offense
nibrs_justifiable_force	nibrs_suspect_using	nibrs_property
nibrs_location_type	nibrs_victim_circumstances	nibrs_property_desc
nibrs_offense_type	nibrs_victim_injury	nibrs_suspected_drug
nibrs_prop_desc_type	nibrs_victim_offender_rel	nibrs_victim
nibrs_prop_loss_type	nibrs_weapon	nibrs_victim_offense
nibrs_relationship_type	ref_agency_county	
nibrs_suspected_drug_type		
nibrs_using_list		
nibrs_victim_type		
nibrs_weapon_type		
ref_agency_type		
ref_race		

*Once created, these can all be referenced in a new database using the prefix "public." (e.g., public.nibrs_age or public.agencies).

Since the original schema definition (database structure) was in the "public" schema (the PostgreSQL organizational term) and the new database schema was placed under "ucr," RTI was able to create and use "public" as the location for this translation layer of reference tables, views, and materialized views. References to tables and views are in the backward-compatible layer using public.<table or view name>, and tables in the new database are named as ucr_prd.<table name>. Tables in the original database are referenced without a prefix. See ***Exhibit 2-5*** for a crosswalk between reference table in the new database and the backward-compatible view.

As an example of the transformations required to convert the new FBI schema to the original FBI schema to recreate the original **NIBRS Agencies** table, a public.agencies view is created by joining 15 tables together and recoding 6 variables (see ***Exhibit 2-5***). For the recoding, four variables are transformed into Y/N from T/F, meaning that the field in the original database was either 'Y' or 'N', but the same information in the new database is conveyed as 'T' or 'F'. Additionally, two fields are recreated since the information has been moved. Specifically:

- **Dormant Flag:** sets 'Y' if *agency_status* is 'D', or 'N' otherwise
    - This creates a field in the view called dormant_flag that returns 'Y' or 'N'. In the original schema definition, this information was kept in agency_status with the value 'D' for dormant.
- **Dormant Year:** sets *data_year* to the associated in *ucr_prd.ref_agency_yearly.data_year* if *agency_status* is 'D', or NULL otherwise
    - This creates a field in the view that returns the data_year of the incident if the agency_status is 'D'.

Additionally, the public.agencies view aggregates multiple values from the *ref_metro_division* and *ref_agency_county* tables into County Name, Metro Division Name, and MSA Name fields.

The full and exact details can be found in the SQL source code at the following GitHub repository, *https://github.com/RTIInternational/NIBRS_Load_UCR_DB*. Some examples from the public.nibrs_incident and public.nibrs_arrestee (materialized) views include:

- public.nibrs_incident: Joins on two tables and includes:
    - **Incident Date Transformation**: Converts *incident_date* to a timestamp and aliases it.
    - **Data Year Extraction**: Extracts and aliases the year from *incident_date*.
    - **Cargo Theft Flag**: Sets 'Y' if true, 'N' if false, or NULL otherwise.
    - **Report Date Flag**: Sets 'R' if true, '' if false, or NULL otherwise.
    - **NIBRS Month ID Alias**: Aliases *form_month_id* as *nibrs_month_id*.
- public.nibrs_arrestee: Joins on nine tables and includes:
    - **Incident Data Year**: Aliases *data_year* from *nibrs_incident*.
    - **Arrestee Data Year**: Extracts and aliases the year from *arrest_date*.
    - **Age ID Logic**: Sets *age_id* based on specific *age_code* values.
    - **Age Number Logic**: Sets *age_num* to NULL for specific values.
    - **Resident Code Logic**: Sets to a space '' if NULL, otherwise retains value.
    - **Under 18 Disposition Code Logic**: Sets to a space '' if NULL, otherwise retains value.
    - **Aliases and Duplication**: Duplicates and aliases certain fields for clarity.

Exhibit 2-5. Tables used to create backward-compatible views (listed in order of creation)

View name	Reference to tables in new database	References to lookup tables or views in the backward-compatible public schema
public.nibrs_incident	ucr_prd.nibrs_incident	public.nibrs_cleared_except
public.agencies_yearly	ucr_prd.ref_agency_yearly, ucr_prd.ref_agency, ucr_prd.ref_state, ucr_prd.ref_division, ucr_prd.ref_region, ucr_prd.ref_agency_status, ucr_prd.ref_submitting_agency, ucr_prd.ref_agency_type, ucr_prd.ref_agency_covered_by	
public.nibrs_activity_type	ucr_prd.lkup_nibrs_activity	
public.nibrs_arrest_type	ucr_prd.lkup_nibrs_arrest_type	
public.nibrs_arrestee_weapon	ucr_prd.nibrs_arrestee_weapon, ucr_prd.nibrs_arrestee	public.nibrs_weapon_type
public.nibrs_assignment_type	ucr_prd.lkup_assignment	
public.nibrs_month	ucr_prd.form_month	
public.nibrs_relationship	ucr_prd.lkup_nibrs_relationship	
public.nibrs_victim_circumstances	ucr_prd.nibrs_victim_circumstance, ucr_prd.nibrs_victim, ucr_prd.nibrs_incident	public.nibrs_circumstances, public.nibrs_justifiable_force

View name	Reference to tables in new database	References to lookup tables or views in the backward-compatible public schema
public.nibrs_victim_injury	ucr_prd.nibrs_victim_injury, ucr_prd.nibrs_victim, ucr_prd.nibrs_incident	public.nibrs_injury
public.nibrs_victim_offender_rel	ucr_prd.nibrs_victim_offender_relationship, ucr_prd.lkup_nibrs_relationship	
public.nibrs_offense	ucr_prd.nibrs_offense	public.nibrs_offense_type, public.nibrs_incident, public.nibrs_location_type
public.nibrs_suspect_using	ucr_prd.nibrs_suspect_using, ucr_prd.lkup_nibrs_using_list	public.nibrs_offense
public.nibrs_weapon	ucr_prd.nibrs_offense_weapon	public.nibrs_weapon_type, public.nibrs_offense
public.nibrs_criminal_act	ucr_prd.nibrs_criminal_activity	public.nibrs_criminal_act_type, public.nibrs_offense
public.agencies	ucr_prd.ref_agency_yearly, ucr_prd.ref_agency, ucr_prd.ref_state, ucr_prd.ref_division, ucr_prd.ref_region, ucr_prd.ref_agency_status, ucr_prd.ref_submitting_agency, ucr_prd.ref_agency_type, ucr_prd.ref_agency_covered_by, ucr_prd.ref_population_group, ucr_prd.ref_parent_population_group, ucr_prd.ref_agency_county, ucr_prd.ref_county, ucr_prd.ref_metro_division, ucr_prd.ref_msa	
public.agency_county	ucr_prd.ref_agency_county, ucr_prd.ref_county, ucr_prd.ref_agency_yearly, ucr_prd.ref_state	
public.county	ucr_prd.ref_county, ucr_prd.ref_state	
public.nibrs_arrestee	ucr_prd.nibrs_arrestee, ucr_prd.lkup_sex, ucr_prd.lkup_nibrs_resident_status	public.nibrs_incident, public.nibrs_arrest_type, public.nibrs_offense_type, public.nibrs_age,

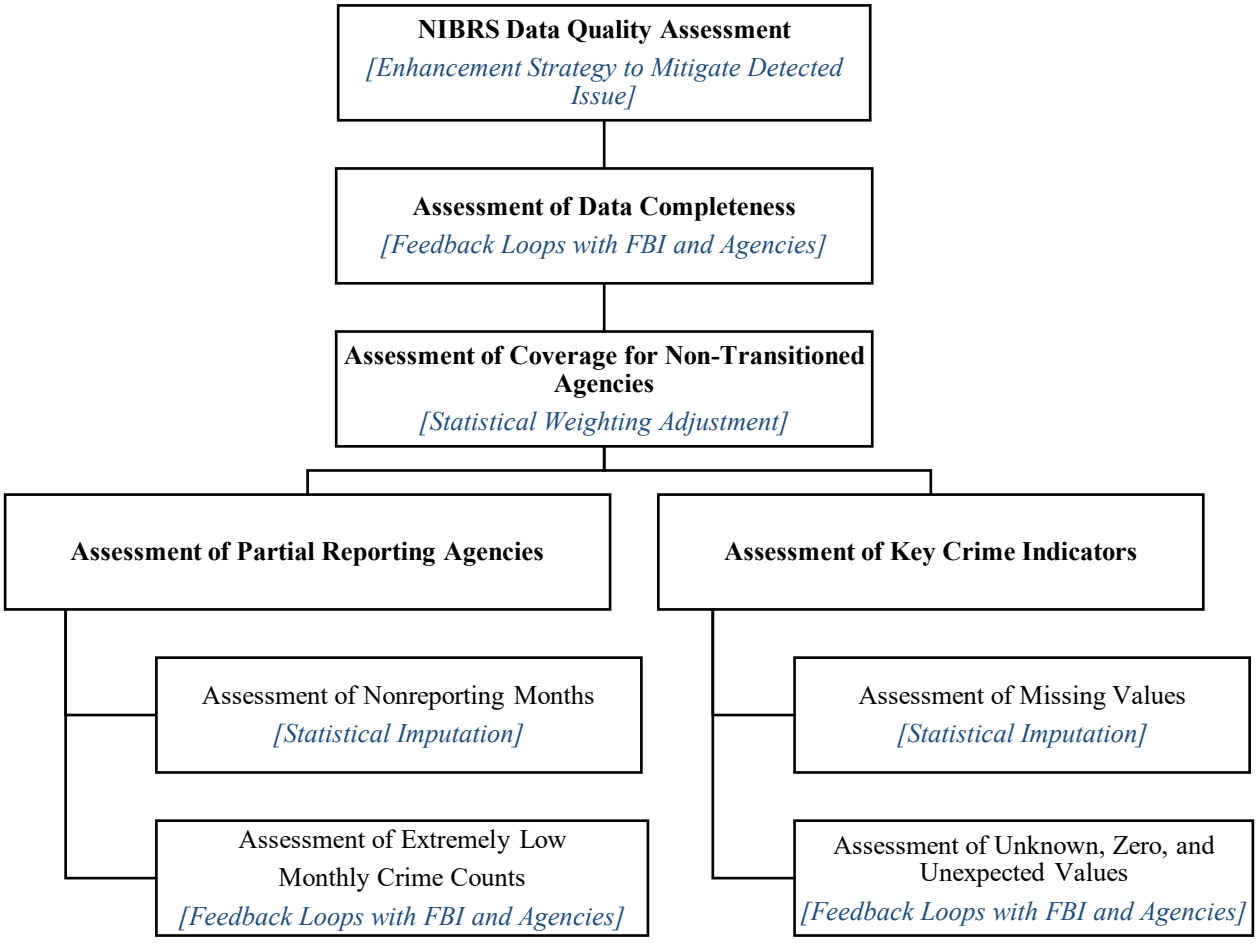
View name	Reference to tables in new database	References to lookup tables or views in the backward-compatible public schema
		public.nibrs_ethnicity, public.ref_race
public.nibrs_bias_motivation	ucr_prd.nibrs_bias_motivation	public.nibrs_bias_list, public.nibrs_offense
public.nibrs_offender	ucr_prd.nibrs_offender	public.nibrs_age, public.nibrs_incident, public.ref_race, public.nibrs_ethnicity
public.nibrs_property	ucr_prd.nibrs_property	public.nibrs_incident
public.nibrs_property_desc	ucr_prd.nibrs_property_description, ucr_prd.nibrs_property,	public.nibrs_prop_desc_type, public.nibrs_incident
public.nibrs_suspected_drug	ucr_prd.nibrs_suspected_drug, ucr_prd.nibrs_property, ucr_prd.nibrs_incident	public.nibrs_suspected_drug _type, public.nibrs_drug_measure_t ype
public.nibrs_victim	ucr_prd.nibrs_victim, ucr_prd.lkup_age, ucr_prd.lkup_assignment, ucr_prd.lkup_nibrs_activity,	public.nibrs_victim_type, public.nibrs_age, public.nibrs_incident, public.ref_race, public.nibrs_ethnicity,
public.nibrs_victim_offense	ucr_prd.nibrs_victim_offense	public.nibrs_offense

2.3. Overview of Quality Procedures of NIBRS Data

Before generating final estimates, NIBRS data undergo a thorough assessment for quality and completeness, and error mitigation strategies are implemented. Like other administrative data, NIBRS data may encounter quality issues at various stages, including coverage, reporting, submission, and processing. These efforts aim to enhance the reliability of the estimates and help prevent potentially biased or inaccurate conclusions about crime levels and characteristics.

As demonstrated in ***Exhibit 2-6***, several key aspects of NIBRS data are assessed and improved during the NIBRS estimation process using statistical methods and information feedback loops with data providers, including the FBI and law enforcement agencies.

Exhibit 2-6. NIBRS data quality assessment and enhancement flowchart



These aspects include:

- **Assessment of Data Completeness.** Data completeness measures whether all NIBRS data submitted by law enforcement agencies are included in the NIBRS database and available to use as input data for the estimation process. NIBRS has a large volume of data, with each reporting agency submitting records for incidents, detailing offenses, offenders, and victims. If there are multiple offenses, offenders, or victims for an incident, a specific record is associated with each. The record structure is maintained through a series of unique identifiers (IDs) that allow users to link victims and offenders within an incident and agency. The initial step of the data quality assessment is ensuring all submitted data are included, with no missing segments (e.g., victim and offender segments) and with proper links between them. If issues arise, the FBI examines and addresses the detected problems. When necessary, the FBI's quality assurance personnel work directly with reporting agencies to resolve data entry and submission challenges.
- **Assessment of Coverage for Non-Transitioned Agencies.** Coverage refers to the extent to which all agencies (and the population they serve) in the United States report their crime incidents to NIBRS. A non-transitioned agency is an agency that had not transitioned to being NIBRS-compliant during the reporting year. According to data available on the FBI's CDE website, as of 2023, about 13,730 law enforcement agencies, covering about 82% of the U.S. population, had

transitioned to NIBRS and submitted data for 3 months or more.[4] This rate was only 33% in 2017 and has been increasing rapidly in recent years as the 2021 sunset of the UCR Summary Reporting System (SRS) approached. The coverage rate also varies across states and by agency type. Some states had nearly 100% population coverage while other states still had relatively low coverage (less than 50%) in 2023. FBI personnel are actively working with states with low coverage to help law enforcement agencies complete their transition. To address gaps, statistical weighting processes at the agency level have been developed to compensate for nonreporting agencies, ensuring that NIBRS data represent all crimes reported to law enforcement in the United States. Separate sets of weights were created for producing estimates at the national, regional, and state levels. See ***Section 3.3*** for details.

- **Assessment of Partial Reporting Agencies.** Partial reporting agencies are those that only report a subset of their total crime incidents during the year. In NIBRS, partial reporting for an agency can take two different forms:
    - Partial Reporters—Agencies that report no incident for certain month(s).
    - Outliers—Agencies with monthly crime counts significantly lower than expected based on historical reporting data.

  Partial reporters are identified using NIBRS monthly tracking data maintained by the FBI's UCR Program. For each state, the UCR Program tracks whether each agency reported data in each month. When these tracking data indicate that an agency did not submit any incident or arrest records in a month and NIBRS did not include any incidents from this agency, the crime count for that month is coded as missing.

  A method that combines hierarchical clustering and median absolute deviation techniques is used to detect outliers among the monthly total crime counts reported by each agency in NIBRS. This approach helps identify whether any monthly crime counts are significantly larger or smaller than the others within the same agency. A monthly crime count that is significantly larger than the rest is classified as a "large outlier," while one that is significantly smaller is classified as a "small outlier." Because NIBRS includes automated checks for duplicate records, large outliers are rare and typically arise from seasonality or extreme events affecting the agency during that month. In contrast, small outliers are more common and may indicate that an agency did not report all its crime incidents for that month.

  For partial reporting agencies (including both partial reporters and outliers), a statistical imputation procedure was developed to impute missing data for nonreporting months or months detected as extremely small outliers. (For additional information, see ***Section 3.4***.) Statistical imputation is the general term for a procedure that fills in missing information based on the known information in the current database. In the FBI's former estimation process to generate crime statistics with SRS crime counts, a simple imputation procedure was implemented to compensate for nonreporting months among reporting agencies. Because NIBRS captures many more details about crime incidents than SRS, a different imputation approach was needed to address missing data collected from this system.

[4] See *https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/downloads*. These numbers differ from the numbers reported in the *Introduction*. The numbers here reference agencies that reported 3 or more months of NIBRS data, while the numbers in the *Introduction* reference agencies that reported 1 or more months of NIBRS data.

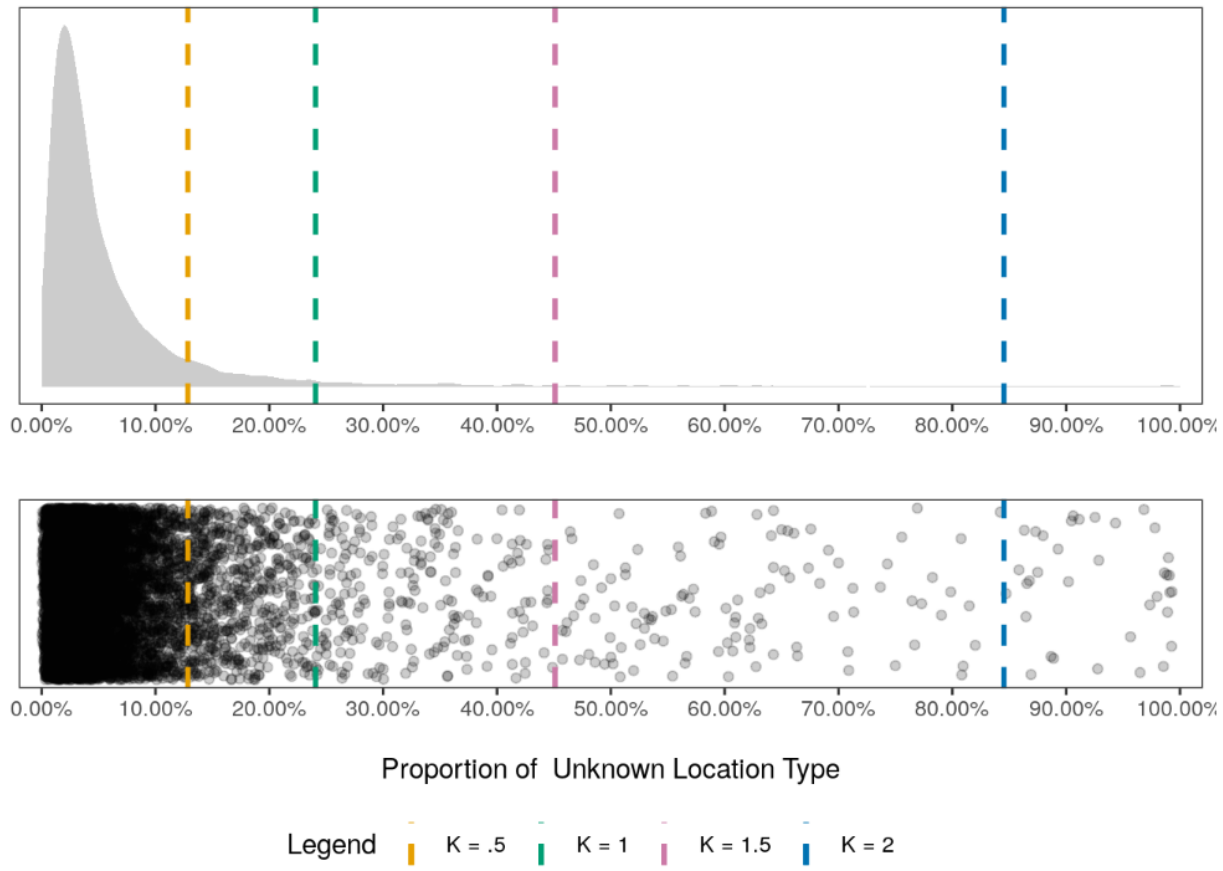
- **Assessment of Key Crime Indicators**. The NIBRS Estimation Program (EP) identified a set of key crime indicators to guide the estimation process (see ***Section 3.2*** for more details). These key indicators guided the development of the estimation procedures, starting with the data quality assessment. The assessment of key indicators comprises two components:
    1. Assessing the prevalence of missing values.
    2. Examining the distributions of unknown, zero, and unexpected values.

  For the first component, findings indicated that the missing rates for most of the key data elements was low (less than 6%) in the 2023 NIBRS data. The only exception was the victim-to-offender relationship variable, which was missing (3.7%) or unknown (14.5%) in 18.2% of incidents for which that variable was required.

  For the second component, the distributions of unknown, zero, and reported values were examined for key data elements within each agency. The checks determined if a high number of incidents (as a percentage of total incidents) had an unexpected value. By reviewing each agency at the aggregate level, rather than the incident level, the estimation system was able to identify agencies that overuse a technically valid value for a data element an unexpected amount of the time. The NIBRS estimation system performs two types of value checks on data elements:

    1. **100% Unexpected Value Check:** This check identifies agencies where all values for a data element are assigned to the unexpected value. It is conducted only for agencies with 15 or more incidents for each applicable data element, except for the following: (1) bias motivation, (2) additional justifiable homicide circumstances, (3) age of arrestee, and (4) race of arrestee. These four elements must never have an unexpected value.
    2. **Interquartile Range (IQR) Check:** This check identifies agencies with an unusually high percentage of unexpected values using the IQR outlier detection method. The IQR is defined as the distance between the 25th and 75th percentiles of a distribution—in this case, the distribution of percentages of unexpected values across all agencies. The IQR check identifies agencies whose percentage of unexpected values is k times larger than the 75th percentile in this distribution. The IQR check is performed on all agencies except for agencies with less than 15 incidents and agencies already identified by the 100% unexpected value check. For example, ***Exhibit 2-7*** shows the distribution of agencies by the percentage of incidents labeled “other/unknown” for the “location type” element among agencies with 15 or more incidents that do not exclusively use “other/unknown” using 2023 NIBRS data. The exhibit illustrates both the distribution and density of agencies, with varying values of k indicating those with extreme use of “other/unknown.” For instance, a total of 28 agencies assigned “other/unknown” location types more than 2 * IQR units beyond the 75th percentile. Location type is an example of a data element with a relatively narrow IQR range, as shown in the exhibit.

Exhibit 2-7. Distribution of agencies' percentage of incidents with a location type of "other/unknown," 2023



A statistical imputation procedure, as described in ***Section 3.4***, was developed to impute missing values in key data elements, including victim-and-offender relationship. Detected issues with unknown, zero, and other unexpected values were addressed by the FBI via direct feedback from the reporting agencies.

3. Methodology for Estimation Procedures

3.1. Introduction

The National Incident-Based Reporting System (NIBRS) estimation process provides representative estimates of crimes recorded by law enforcement and their corresponding characteristics. The NIBRS estimation process contains several key components to achieve this purpose—each enhancing the relevance, accuracy, and interpretability of final estimates. The process involves:

- Developing an indicator structure to capture key offense types and characteristics (***Section 3.2***).
- Using statistical weighting to address coverage errors from nonreporting agencies (***Section 3.3***).
- Developing a block imputation method to fill in missing incidents for partial reporting agencies and using logic and hot deck imputation methods to address item-level missingness within incident reports (***Section 3.4***).
- Estimating the population served by each agency to normalize crime data and enable meaningful cross-jurisdictional comparisons (***Section 3.5***).
- Addressing the statistical uncertainty in the estimates due to NIBRS being a large nonprobability sample of the United States (***Section 3.6***).
- Incorporating confidence intervals to show the range in which true values likely fall (***Section 3.7***).
- Applying statistical suppression rules to omit unreliable estimates in publications (***Section 3.8***).

The output datasets (***Section 3.9***), proper interpretation of the final estimates (***Section 3.10***), and statistical testing (***Section 3.11***) can be used to produce reliable crime statistics while ensuring transparency. Each component of this process is described in detail in the following sections.

3.2. Indicator Structure

3.2.1. Data segments in NIBRS

NIBRS estimates are produced for four of the five segments in the NIBRS database. Specifically, NIBRS estimates are produced for the following:

- Incident segment
- Offense segment
- Victim segment
- Arrestee segment.

Estimates are not currently produced for the offender segment because of the higher level of missing data elements.

3.2.2. Offense types collected in NIBRS

Individual offense types

Estimates are produced for most offenses for which NIBRS collects information. ***Exhibit 3-1*** presents the individual offense types for which estimates are produced by their offense type: crimes against persons, crimes against property, and crimes against society.

Exhibit 3-1. Individual offenses estimated by offense type

Crimes against persons	Crimes against property	Crimes against society	Group B offenses (arrestees only)
Aggravated assault	Arson	Animal cruelty	Import violations
Simple assault	Bribery	Drug/narcotic offenses	Export violations
Intimidation	Burglary	Gambling offenses	Federal liquor offenses
Murder and nonnegligent manslaughter	Counterfeiting/forgery	Prostitution/obscene material	Federal tobacco offenses
Negligent manslaughter	Destruction, damage, and vandalism	Prostitution offenses	Wildlife trafficking
Kidnapping and abduction	Embezzlement	Weapon law violations	Espionage
Sex trafficking	Extortion/blackmail		Money laundering
Labor trafficking	Fraud offenses		Harboring escapee/concealing from arrest
Rape	Larceny/theft offenses		Flight to avoid prosecution
Sodomy	Motor vehicle theft		Flight to avoid deportation
Sexual assault with an object	Robbery		Illegal entry into the United States
Fondling	Carjacking		False citizenship
Nonforcible sex offenses	Stolen property offenses		Smuggling aliens
	Hacking/computer invasion		Reentry after deportation
			Failure to register as a sex offender
			Treason
			Violation of National Firearms Act of 1934
			Weapons of mass destruction
			Explosives violation
			Family offenses, nonviolent
			Trespass of real property
			Curfew/loitering/vagrancy violations
			Liquor law violations
			Disorderly conduct
			Failure to appear
			Federal resource violations
			Perjury
			Driving under the influence
			All other offenses

Aggregate offense types

In addition to the individual offense types, estimates for aggregate offense types are produced. ***Exhibit 3-2*** lists each aggregate offense type and details the individual offense types included. An aggregate offense type is counted when one or more of the underlying individual offense types have occurred.

Exhibit 3-2. Definition of aggregate offenses

Aggregate measure	Individual offense components
Violent crime	Murder and nonnegligent manslaughter, legacy rape, sodomy, sexual assault with an object, aggravated assault, robbery
Rape*	Legacy rape, sodomy, sexual assault with an object
Crimes against persons	Aggravated assault, simple assault, intimidation, murder and nonnegligent manslaughter, negligent manslaughter, kidnapping and abduction, sex trafficking, labor trafficking, legacy rape, incest, fondling, statutory rape, sodomy, sexual assault with an object
Fraud offenses	False pretenses/swindle/confidence game, welfare fraud, wire fraud, credit card/automated teller machine fraud, impersonation, identity theft, hacking/computer invasion
Larceny/theft offenses	Theft of motor vehicle parts or accessories, pocket-picking, theft from motor vehicle, purse-snatching, shoplifting, all other larceny, theft from building, theft from coin-operated machine or device
Property crime	Burglary/breaking and entering, larceny/theft offenses, motor vehicle theft
Crimes against property	Arson, bribery, burglary/breaking and entering, counterfeiting/forgery, destruction/damage/ vandalism, embezzlement, extortion/blackmail, fraud offenses, larceny/theft offenses, motor vehicle theft, robbery, stolen property offenses
Corruption	Bribery, embezzlement, extortion/blackmail
Other acts of corruption	Embezzlement, extortion/blackmail
Gambling offenses	Sports tampering, operating/promoting/assisting gambling, betting/wagering, gambling equipment violation
Prostitution offenses	Assisting or promoting prostitution, prostitution, purchasing prostitution
Drug/narcotic offenses	Drug/narcotic violations, drug equipment violations
Other offenses	Money laundering, treason, espionage, illegal entry into the United States, false citizenship, smuggling aliens, reentry after deportation, failure to register as a sex offender, harboring, escapee/concealing from arrest, flight to avoid prosecution, flight to avoid deportation, violation of National Firearms Act of 1934, weapons of mass destruction, explosives violation, import

Aggregate measure	Individual offense components
	violations, export violations, federal liquor offenses, federal tobacco offenses, wildlife trafficking
Crimes against society	Animal cruelty, drug/narcotic offenses, prostitution/obscene material, prostitution offenses, weapon law violation, gambling offenses, and other offenses
Total gun violence offenses	Murder and nonnegligent manslaughter, negligent manslaughter, legacy rape, sodomy, sexual assault with an object, robbery, aggravated assault, kidnapping/abduction, human trafficking (sex and labor)
Fatal gun violence offenses	Murder and nonnegligent manslaughter
Nonfatal gun violence offenses	Legacy rape, sodomy, sexual assault with an object, robbery, aggravated assault, kidnapping/abduction, human trafficking (sex and labor)
Nonfatal gun violence offenses (violent crime)	Legacy rape, sodomy, sexual assault with an object, robbery, aggravated assault

*Previously referenced as "revised rape."

Note:

- Robbery is counted as a person crime for purposes of estimating violent crime.
    - For the violent crime indicator in the context of counting offenses, because robbery is a property crime, per the FBI's counting rule for offense, only one offense counts per incident.
    - In terms of counting the number of person victims of violent crimes, for the offense of robbery, each person victim that is associated with the offense of robbery is counted as a victim.
- For the property crime indicator in the context of counting offenses, the offense of motor vehicle theft has distinct counting rules for offenses.
    - If the motor vehicle theft offense is completed and the number of stolen vehicles is 0, then it counts as one offense.
    - If the motor vehicle theft offense is completed and the number of stolen vehicles is reported (one or more), then each stolen vehicle counts as an offense.
    - If the motor vehicle theft offense is attempted, then it counts as one offense.

Special case offenses

In addition to the standard offenses in NIBRS, estimates are produced for special case offenses, which are subsets of standard individual offenses.

Carjacking. Carjacking is a subset of robbery. ***Exhibit 3-3*** illustrates how it is constructed. A robbery is considered a carjacking when all the following circumstances are met:

- victim type is individual
- offense is recorded as completed
- property type is marked as automobiles, buses, RVs, trucks, or other motor vehicles
- property activity is marked as either stolen or recovered.

Exhibit 3-3. Definition of carjacking (bolded)

UCR offense	Victim type	Complete status	Property type	Property activity
1. Robbery	1. All ***2. Individual***	1. All ***2. Completed***	1. Aircraft ***2. Automobiles*** 3. Bicycles ***4. Buses*** ***5. Other MVs*** ***6. RVs*** ***7. Trucks*** 8. Vehicle parts 9. Watercraft 10. Aircraft parts 11. Fuel 12. Trailers 13. Watercraft parts	1. All ***2. Stolen*** ***3. Recovered***

Drug offense categories. Individual drug offenses are estimated using NIBRS Data Element 20 (drug type/category) and associated with an offense code of 35A (i.e., Drug/Narcotic Violations) with a completion status of completed. Data element 20 is aggregated into the following drug types (element value in parentheses)

- Cocaine/crack cocaine (A, B)
- Marijuana/hashish (C, E)
- Opiate/narcotic (D, F, G, H)
- Hallucinogen (I, J, K)
- Stimulant (L, M)
- Depressant (N, O)
- Other (P)
- Unknown (U)
- More than 3 types (X)

Gun violence. Gun violence offenses and victim characteristics are estimated for all person-level offenses. A gun violence offense is defined as a weapon type (Data element 13) of (element value in parentheses):

- Firearm (11)
- Handgun (12)
- Rifle (13)
- Shotgun (14)
- Other firearm (15)

Note: Any weapon type that is a gun can be further specified as an automatic (e.g., automatic firearm) if it is designed to shoot more than one shot at a time by a single pull of the trigger without a manual reloading.[5]

[5] See the *NIBRS User Manual*.

3.2.3. Characteristics of all four segments

General characteristics

Exhibit 3-4 presents the general characteristics (indicator and categories) that are estimated at the incident, offense, victim, or arrestee level. ***Exhibit 3-5*** presents additional characteristics that are only estimated at the victim or arrestee level.

Exhibit 3-4. General characteristics by reporting unit of analysis

Indicator	Categories	Reporting unit of analysis			
		Incident	Offense	Victim	Arrestee
Weapon involved[a,b]	Personal, firearms, knives/cutting instruments, blunt instruments, other non-personal, unknown	X	X	X	X
Weapon involved - Yes 2[a,b]	Firearms or explosives, firearms, another weapon other than firearms or explosives, knives and other cutting instruments, unknown			X	
Weapon involved - Yes 3[a,b]	Personal weapons, firearms, other non-personal, unknown			X	
Weapon involved hierarchy[a]	Handgun, firearm, rifle, shotgun, other firearm, knife/cutting instrument, blunt object, motor vehicle, personal weapons (hands, feet, teeth, etc.), asphyxiation, drugs/narcotics/sleeping pills, poison (include gas), explosives, fire/incendiary device, other, no weapon, unknown, not applicable		X	X	
Weapon involved hierarchy within offense[a]	Handgun, firearm, rifle, shotgun, other firearm, knife/cutting instrument, blunt object, motor vehicle, personal weapons (hands, feet, teeth, etc.), asphyxiation, drugs/narcotics/sleeping pills, poison (include gas), explosives, fire/incendiary device, other, no weapon, unknown, not applicable			X	
Injury[a,b]	Yes, no	X	X	X	
Multiple victims[a]	1 victim, 2+ victims	X	X		
Number of victims summarized at	1, 2, 3, 4+			X	

Indicator	Categories	Reporting unit of analysis			
		Incident	Offense	Victim	Arrestee
incident level within offense[a]					
Multiple offenders[a]	1 offender, 2+ offenders, unknown offenders	X	X		
Multiple offense incident[a]	1 offense, 2 offenses, 3+ offenses	X			
Victim-offender relationship[a,b]	Intimate partner, other family, outside family but known to victim, stranger, victim was offender, unknown relationship	X	X	X	
Victim-offender relationship 2[a,b]	Intimate partner plus family, outside family but known to victim, stranger, victim was offender, unknown relationship	X	X	X	
Victim-offender relationship hierarchy[a,b]	Intimate partner, other family, outside family but known to victim, stranger, victim was offender, unknown relationship, unknown offender incidents, missing from uncleared incidents			X	
Location type[a,b]	Residence/hotel, transportation hub/outdoor public locations, schools/daycares/universities, retail/financial/other commercial establishment, restaurant/bar/sports or entertainment venue, religious buildings, government/public buildings, jail/prison, shelter-mission/homeless, other/unknown location	X	X		
Location type 2[a,b]	Residence/hotel, transportation hub/outdoor public locations, schools, daycares, and universities, retail/financial/other commercial establishment, restaurant/bar/sports or entertainment venue, religious buildings, government/public buildings, jail/prison, shelter-mission/homeless, drug store/doctor's office/hospital, other/unknown location	X	X		

Indicator	Categories	Reporting unit of analysis			
		Incident	Offense	Victim	Arrestee
Location type 3[a,b]	Residence, not residence	X	X		
Location type hierarchy within offense[a,b]	Residence, not residence			X	
Location cyberspace[a,b]	Cyberspace		X		
Time of day: Incident time[a,b]	Midnight–4am, 4–8am, 8am–noon, noon–4pm, 4–8pm, 8pm–midnight, unknown	X	X		
Time of day: Report time[a,b]	Midnight–4am, 4–8am, 8am–noon, noon–4pm, 4–8pm, 8pm–midnight, unknown	X	X		
Population group[a,b]	Cities and counties 100,000 and over, cities and counties 25,000–99,999, cities and counties 10,000–24,999, cities and counties under 10,000, state police	X	X		
Agency indicator[a,b]	City, county, university or college, state police, other state agencies, tribal agencies, federal agencies	X	X		
Clearance[a]	Not cleared, cleared through arrest, exceptional clearance	X		X	
Clearance 2[a]	Cleared incident, not cleared incident		X		
Cleared through arrest[a]	Not cleared through arrest, cleared through arrest		X		
Exceptional clearance[a]	Death of offender, prosecution declined, in custody of other jurisdiction, victim refused to cooperate, juvenile/no custody	X		X	
Property loss[a]	None, burned, counterfeited/forged, destroyed/damaged/vandalized, recovered, seized, stolen, unknown	X			
MSA[a]	MSA counties, outside MSA, non-MSA counties, missing	X	X		
Gang involvement[a,b]	None/unknown, juvenile or other gang involvement		X	X	

[a] Counts and percentages will be calculated.

[b] These categories have been aggregated from the full list of responses collected in NIBRS. *Appendix A* includes a crosswalk of the original and aggregated categories.

Exhibit 3-5. General characteristics of victims and arrestees

Indicator	Categories	Reporting unit of analysis	
		Victim	Arrestee
Age[b]	Under 5, 5–14, 15, 16, 17, 18–24, 25–34, 35–64, 65+, Under 12, 12–17, 12–14, 15–17, 18+	X	X
Sex[b]	Male, female	X	X
Race[b]	White, Black, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander	X	X
Sex and race[b]	Same as above	X	X
Victim age category by offender age category[a,c]	Victim juvenile X offender juvenile, victim juvenile X offender adult, victim adult X offender adult, victim adult X offender juvenile, unknown victim age or unknown offender age	X	
Victim sex by offender sex[a,c]	Victim male X offender male, victim male X offender female, Victim female X offender female, victim female X offender male, unknown victim sex or unknown offender sex	X	
Victim race by offender race[a,c]	Victim White X Offender White, Victim White X Offender All Other Races Except White, Victim Black X Offender Black, Victim Black X Offender All Other Races Except Black, Victim AIAN X Offender AIAN, Victim AIAN X Offender All Other Races Except AIAN, Victim Asian X Offender Asian, Victim Asian X Offender All Other Races Except Asian, Victim NHOPI X Offender NHOPI, Victim NHOPI X Offender All Other Races Except NHOPI, Unknown victim race or unknown offender race	X	
Juvenile disposition[a]	Handled within department, referred to other authorities, not applicable, unknown		X
Multiple arrest indicator[a]	Multiple, count, not applicable		X
Arrestee armed (yes/no)[a]	Firearm, lethal cutting instrument, club/blackjack/brass knuckles		X
Arrest type[b]	On-view arrest, summoned/cited, taken into custody		X

[a] Counts and percentages will be calculated.
[b] Counts, rates, and percentages will be calculated.
[c] Offender characteristics will only be imputed in cleared cases (see DM0026).

Drug module characteristics

For the detailed drug module estimates, characteristics of the offense and victim or arrestee are detailed in ***Exhibit 3-6***.

Exhibit 3-6. Characteristics of detailed drug offenses

Indicator	Categories	Notes
Criminal activity	Buying/receiving, cultivating/manufacturing/publishing, distributing/selling, exploiting children, operating/promoting/assisting, possessing/concealing, transporting/transmitting/importing, using/consuming	
Property type	Drugs/narcotics, drug/narcotic equipment, chemicals, firearms, money	
Location type	Residence/hotel, transportation hub/outdoor public locations, schools/daycares/universities, retail/financial/other commercial establishment, restaurant/bar/sports or entertainment venue, religious buildings, government/public buildings, jail/prison, shelter-mission/homeless, other/unknown location	
Location type 2	Residence/hotel, transportation hub/outdoor public locations, schools, daycares, and universities, retail/financial/other commercial establishment, restaurant/bar/sports or entertainment venue, religious buildings, government/public buildings, jail/prison, shelter-mission/homeless, drug store/doctor's office/hospital, other/unknown location	
Time of day: Incident time	Midnight–4am, 4–8am, 8am–noon, noon–4pm, 4–8pm, 8pm–midnight, unknown	
Time of day: Report time	Midnight–4am, 4–8am, 8am–noon, noon–4pm, 4–8pm, 8pm–midnight, unknown	
Population group	Cities and counties 100,000 and over, cities and counties 25,000–99,999, cities and counties 10,000–24,999, cities and counties under 10,000, state police	
Agency indicator	City, county, university or college, state police, other state agencies, tribal agencies	
MSA	MSA counties, outside MSA, non-MSA counties, missing	
Arrestee age	Under 5, 5–14, 15–17, 18–24, 25–34, 35–64, 65+, Under 12, 12-17, 12–14, 15–17, 18+	
Arrestee sex	Male, female	

Indicator	Categories	Notes
Arrestee race	White, Black, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander	
Arrestee sex by race	Defined as above	
Juvenile disposition	Handled within department, referred to authorities, not applicable, unknown	
Multiple arrests	Multiple count, not applicable	This is data element 44 – Multiple Arrestee Segment Indicator.
Arrestee armed	Yes, no	

Gun violence characteristics

For the detailed estimates on gun violence, additional victim-level characteristics are estimated, as shown in ***Exhibit 3-7***.

Exhibit 3-7. Victim-level characteristics for gun violence module

Indicator	Categories	Notes
Age[b]	Under 5, 5–14, 15–17, 18–24, 25–34, 35–64, 65+, Under 18, Under 12, 12–17, 18+	
Sex[b]	Male, female	
Race[b]	White, Black, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander	
Number of victims[a]	1, 2, 3, 4+	Count of person victims aggregated at the incident level that includes offenses that may or may not be associated with a firearm.
Number of victims summarized at incident level[a]	1, 2, 3, 4+	Count of the number of incidents for the specified offense and demographics.
Number of firearm victims[a]	1, 2, 3, 4+	Count of person victims aggregated at the incident level that includes offenses associated with a firearm.
Number of victims murdered[a]	Yes, No	

[a]Counts and percentages will be calculated.
[b]Counts, rates, and percentages will be calculated.

3.3. Weighting to Compensate for Nonparticipating Agencies

3.3.1. Overview of statistical weighting to compensate for coverage error in NIBRS

Crime estimates produced from NIBRS are based on data provided by a subset of eligible law enforcement agencies, as not all agencies submit data. For example, in 2023, about 13,730 agencies (representing 82% of the U.S. population) submitted at least 3 months of data through NIBRS. The remaining 18% of the U.S. population is covered by law enforcement agencies that had not transitioned to NIBRS by 2023, which leads to coverage error. To compensate for this coverage error, agency weights are developed and assigned to the around 13,730 reporting agencies at both national and subnational levels.

When complete coverage of the population is not possible, statistical weights enable a sample to represent the entire population. Weights indicate the influence each observed record has in estimating population totals. When estimating specific characteristics (e.g., all U.S. law enforcement agencies), weights are constructed to reflect the total size of the target population. Typically, weights are 1 or greater: a weight of 1 means the unit represents only itself, while a weight greater than 1 means it represents itself plus other similar, unobserved units. For example, a unit with a weight of 3 represents itself plus two additional unobserved units, giving it three times the influence of a unit with a weight of 1 during estimation. The weights can be used to estimate the number of incidents or incident characteristics in the population by applying the agency weight to each incident submitted by an agency. Therefore, if an agency has a weight of 3, then each incident record will count three times when computing a total.

The current NIBRS estimation process produces estimates at the following geographic levels:

- national (all agencies, tribal agencies, and university agencies)
- regional
- state
- judicial district
- field office
- metropolitan statistical areas (MSA).

Because NIBRS reporting agencies are not evenly distributed across the different levels (e.g., some states have nearly all agencies reporting to NIBRS while others have coverages rates that are 50% or lower), a separate set of weights is created for each geographic level to ensure the estimates are representative of all agencies in those areas.

3.3.2. Creation of the national and regional weights

National NIBRS weights are developed through a process known as calibration, which adjusts weights to align the weighted sums of certain calibration variables based on the NIBRS data to equal their known population total. Auxiliary data sources, such as the Summary Reporting System (SRS), play a key role in optimizing these weights to better estimate the influence of each agency in the final weighted estimates. SRS is a legacy crime reporting system that collects summary-level data from almost all law enforcement agencies across the United States, providing counts of major crime categories like violent and property crimes. Although less detailed than NIBRS, SRS offers reliable, long-term benchmarks for total crime counts.

For example, calibration ensures that NIBRS weighted estimates for violent crimes align with SRS totals, which serve as population-level crime benchmarks. The weighting process in the NIBRS estimation system incorporates key variables as calibration variables, such as total crime counts and individual

offense categories—including murder, rape, aggravated assault, burglary, robbery, larceny, and motor vehicle theft. To ensure that the nonreporting agencies in the population are represented by reporting agencies with similar characteristics, the calibration is conducted within groups where the types and number of reported crimes are likely to be similar (referred to as "post-strata"). ***Exhibit 3-8*** presents the 15 post-strata used to group law enforcement agencies for constructing the national and regional weights.

Exhibit 3-8. Weighting post-strata created for constructing the national-level and regional-level weights

Post-stratum number	Post-stratum description
1	Cities 250,000 persons or more
2	Cities from 100,000 to 249,999 persons
3	Cities from 50,000 to 99,999 persons
4	Cities from 25,000 to 49,999 persons
5	Cities from 10,000 to 24,999 persons
6	Cities under 10,000 persons
7	MSA counties 100,000 persons or more
8	MSA counties from 25,000 to 99,999 persons
9	MSA counties under 25,000 persons
10	Non-MSA counties 25,000 persons or more
11	Non-MSA counties from 10,000 to 24,999 persons
12	Non-MSA counties under 10,000 persons
13	City agencies where no population is attributed[a]
14	MSA counties and state agencies where no population is attributed
15	Non-MSA counties and state agencies where no population is attributed

[a]For more information on populations and their attribution to general and special purpose agencies, please see ***Section 3.5***.

Within each post-stratum, weights are created using a calibration process that forces the NIBRS weighted crime totals of violent and property crimes and their subtypes to equal their SRS counterparts in the same year. The SRS totals used in the calibration procedure include:

- all violent and property crime
- murder/nonnegligent manslaughter
- rape
- robbery
- aggravated assault
- burglary
- larceny-theft
- motor vehicle theft.

Although SRS was phased out in 2021, the FBI continued to aggregate incident-level data from NIBRS into summary crime counts to maintain consistency and comparability with the traditional SRS format. Starting in 2022, the FBI allowed SRS counts to be submitted from agencies that have not yet transitioned to NIBRS. According to data available on the FBI's CDE website,[6] the SRS dataset combines data from NIBRS reporting agencies and those still using SRS, covering 94% of the U.S. population in 2023. For nonreporting agencies to SRS, their SRS counts are imputed based on their historical SRS data or SRS data from similar agencies. The imputed SRS data are then used in the calibration procedure.

A key benefit of using post-strata to calibrate the weights is the ability to generate domain estimates—crime metrics for subsets of agencies based on characteristics like population group or agency type. For the national estimates, estimates are produced in these domains, which align with the post-strata shown in ***Exhibit 3-8***, enabling more detailed analysis of crime statistics.

Regional weights are constructed using the same approach as the national weights with some modifications to the calibration models and restricted to agencies in each of the four regions that subdivide the nation: 1) Northeast, 2) South, 3) Midwest, and 4) West. At the regional level, each post-stratum has fewer agencies than at the national level, limiting the number of calibration variables the model can handle. Therefore, in the 2023 weighting process, the calibration models in some post-strata do not have all the calibration variables as used when creating the national weights. If it is desirable to include more calibration variables in the weighting process, some post-strata or all the post-strata will need to be combined, which increases the sample size, allows for additional calibration variables, and improves alignment with SRS data benchmarks for multiple offense types. The post-strata that are combined will vary across state, but the algorithm applied is the same. The algorithm consists of the following steps:

1. Run the calibration model using all offense types and post-strata.
2. If the model converges, stop.
3. If the model does not converge, collapse adjacent post-strata beginning with the post-strata with the lowest population coverage.
4. Run the calibration model using all offense types with collapsed post-strata.
5. If the model converges, stop.
6. If the model does not converge, repeat steps (3) and (4) until model convergence is obtained.

3.3.3. Creation of state and substate weights

State-level weights are created for each of the 50 states and Washington, D.C. Within each state, a calibration method is applied using the SRS data, following the same principles as the national and regional weights but with fewer post-strata and fewer calibration variables. To ensure a sufficient sample size for reliable state-level calibration, the 15 post-strata used for national and regional weights are consolidated into four broader groups, as shown in ***Exhibit 3-9***. This merging, guided by regression tree analysis of crime counts across agencies, groups agencies with similar crime volumes and patterns to enhance consistency within each post-stratum. In states with fewer agencies, such as Hawaii, some post-strata may lack NIBRS reporting agencies. When this happens, the four post-strata are further combined to support the calibration process.

[6] This includes agencies that reported at least 1 month of NIBRS data during 2023. See page 4 of *UCR Summary of Crime in the Nation: 2023* at ***https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/explorer/crime/special-reports***.

Exhibit 3-9. Description of the weighting post-strata created for constructing the state-level weights

Post-stratum number	Post-stratum description
1	Large general-purpose agencies* including: • Cities 250,000 persons or more
2	Relatively large general-purpose agencies, including: • Cities from 100,000 to 249,999 persons • MSA counties 100,000 persons or over
3	Other general-purpose agencies, including: • Cities under 100,000 persons • MSA counties under 100,000 persons • MSA state police • Non-MSA counties • Non-MSA state police
4	Zero-population agencies, including: • City agencies where no population is attributed • MSA counties and state agencies where no population is attributed • Non-MSA counties and state agencies where no population is attributed

*General-purpose agencies are those that serve as the primary law enforcement entities for some portion of their jurisdictions. Includes local and regional police departments, most sheriff's offices, county police departments, and in some cases, state police agencies.

Substate-level weights are constructed separately at the judicial district level, the field office level, and MSA level. Weights at each geographical level (judicial district, field office, or MSA) are constructed using a similar calibration approach that is used for state weights, except that the four post-strata in ***Exhibit 3-9*** are further collapsed into two post-strata to ensure sufficient sample size in each post-stratum for calibration. These two post-strata are: (1) general-purpose agencies with a non-zero population size and (2) zero-population agencies. Like the state-level weighting, the two post-strata can be collapsed when one post-stratum has no or very few reporting agencies.

3.3.4. Validating the NIBRS weights for estimation

To ensure the NIBRS weights produce accurate estimates, they are validated using the SRS data. Weighted Part I crime totals from NIBRS reporters are compared to SRS estimates, which sum crime totals from all in-scope agencies based on the imputed SRS data. Since SRS estimates include additional non-NIBRS reporting agencies, they are used to benchmark the NIBRS-based estimates. The difference between the weighted NIBRS estimates and SRS estimates is known as statistical bias. If the bias is small, the NIBRS weights are considered accurate; if the bias is large, the weights are not considered accurate. Although the NIBRS weights are generated through a calibration process, the number of calibration variables that can be used depends on the sample size and the differences between reporting and nonreporting agencies within each weighting group (i.e., post-stratum). As a result, NIBRS weighted estimates may still differ from SRS estimates, particularly for non-national estimates when the sample size is relatively small in a post-stratum.

Validation analysis of the 2023 NIBRS weights found that, at the national level, most weighted Part I crime totals had relative bias below 10%, indicating low bias. However, some specialized agencies showed higher bias. At regional, state, and substate levels, bias rates were strongly tied to population coverage—areas with higher coverage had lower bias. For example, a judicial district with 91.9%

coverage had a bias of -0.65% for murder, while another with 60.3% coverage showed a bias of 47.88% for murder. Since NIBRS weights do not always produce reliable estimates, statistical suppression rules were developed to determine when estimates should be published. For example, estimates from a district with low coverage and high bias would likely be suppressed. More information on these rules can be found in ***Section 3.8***.

3.4. Imputation for Missing or Unknown Response

The NIBRS estimation process imputes two types of missing or unknown responses: (1) partial reporting agencies and (2) item-level missingness. Imputation is essential for NIBRS estimation for two key reasons:

1. **Reducing potential bias:** Missing data in NIBRS is nonrandom, meaning analyses that exclude missing data may suffer from nonresponse errors. When substantial, nonresponse errors can skew estimates, pushing them higher or lower than their true values, potentially leading to invalid results and erroneous conclusions.
2. **Enabling disaggregated estimates:** NIBRS enables the generation of rate estimates by demographics (such as age, race, and sex) for victims, offenders, and arrestees. These rates standardize crime totals relative to the population served by the area being analyzed. To calculate these rates accurately, all individuals must be assigned to a known demographic category. Therefore, missing or unknown demographic values require imputation.

Exhibit 3-10 presents the data elements on which imputation is conducted in the NIBRS estimation process. Item-level missingness in reported incidents is addressed first to ensure that any incidents used during the imputation of partial reporting agencies are complete. For partial reporting agencies, the missing months and low outlier months are treated as missing and are entirely imputed as full months.

3.4.1. Imputation for item-level missingness

Item-level missingness in NIBRS data occurs when a data element is either missing or contains a value that cannot be used for estimation (e.g., "unknown"). ***Exhibit 3-10*** lists the four data elements chosen for imputation in the NIBRS estimation process: (1) age, (2) race, (3) sex, and (4) victim-offender relationship.

Exhibit 3-10. Imputed data elements in the 2023 NIBRS estimation process

	Victim segment		**Offender segment**		**Persons arrested segment**	
Data Element	Cleared case	Uncleared case	Cleared case	Uncleared case	Cleared case	Uncleared case
Age	✓	✓	✓	×	✓	✓
Race	✓	✓	✓	×	✓	✓
Sex	✓	✓	✓	×	✓	✓
Victim-offender relationship	✓	✓	✓	×	✓	×

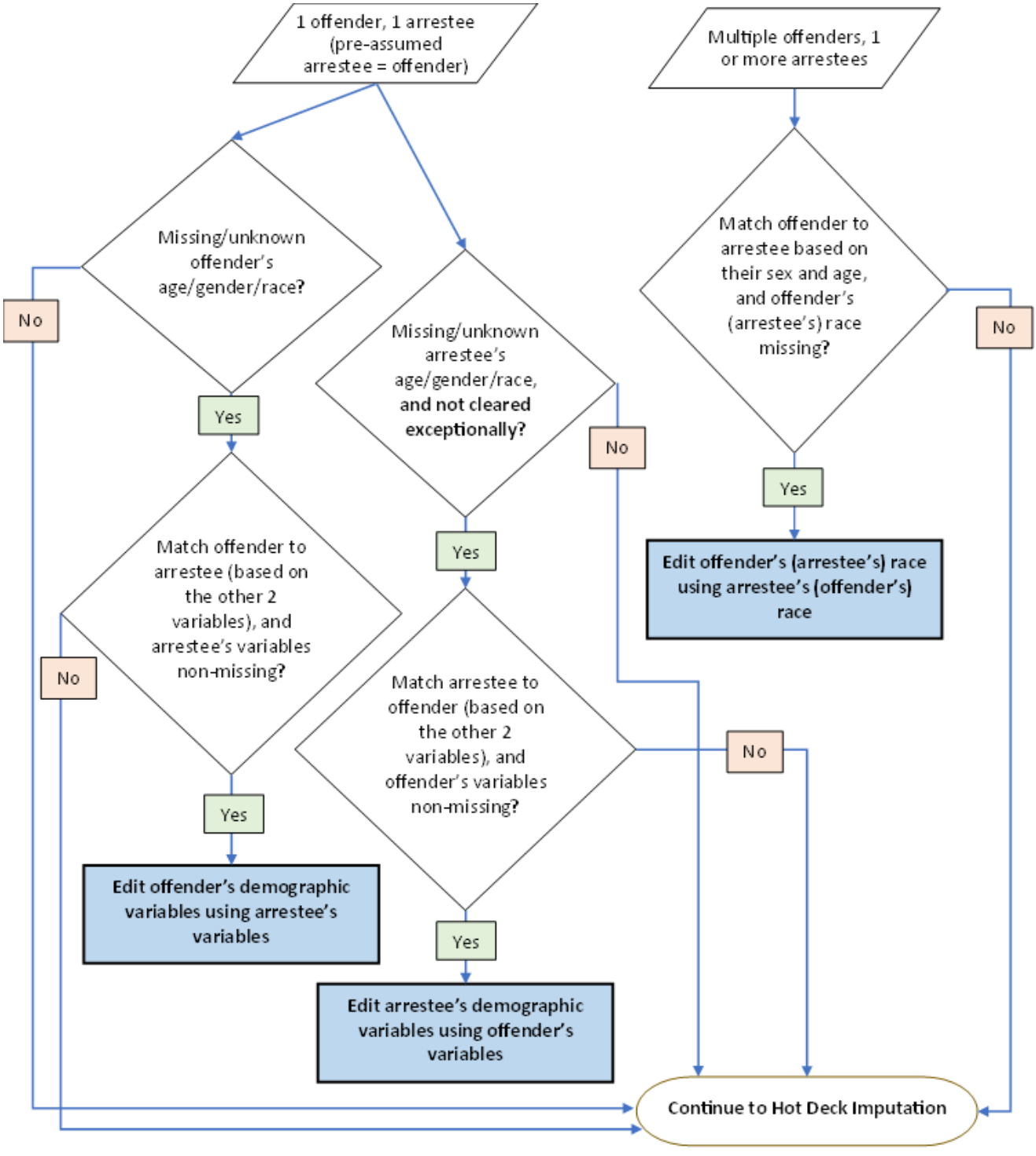
✓Data element is imputed.
×Data element is not imputed.

Each data segment (i.e., victim, offender, or persons arrested) is evaluated to determine if sufficient information is available in the NIBRS data to support imputation. For the victim segment, there is enough information to impute all four data elements, regardless of the incident's clearance status. Since cleared incidents often differ from uncleared ones in both characteristics and the degree of item missingness, they are treated separately in the imputation process. For the offender segment, only data elements in cleared cases are suitable for imputation. In the persons arrested segment, all data elements can be imputed, except for the victim-offender relationship in uncleared cases.

Item-level missingness is addressed using logical and hot deck imputation methods. In logical imputation, information within the incident record is used to determine the most likely value for the missing element. In NIBRS, missing offender data elements in cleared cases can be logically imputed if an arrestee is linked to an offender. ***Exhibit 3-11*** presents the detailed process for conducting logical imputation.

Hot deck imputation replaces a missing value in one incident with a value from a "donor" incident, identified using a predictive mean matching (PMM) model. This model selects donor incidents based on similar agency and offense characteristics (e.g., offense type). The missing value is then filled with the corresponding value from the donor incident.

Exhibit 3-11. Logical imputation for demographic NIBRS data



3.4.2. Imputation for partial reporting agencies

A partial reporting agency is an agency that submitted at least 3 but less than 12 months of incidents. An agency is considered to have not submitted their data in a given month if 1) the NIBRS tracking system indicates that it did not report in the corresponding month and no incidents are recorded in the NIBRS system for that month (referred to as a "missing month") or 2) if the outlier detection algorithm, developed as part of the estimation process, flags the agency's total crime count in that month as extremely low (referred to as "low outliers").

Outliers are identified using a set of mathematical algorithms applied to the monthly total crime counts reported by an agency in NIBRS. These algorithms help detect counts that deviate significantly—either higher or lower—compared to other counts within the same agency. A monthly crime count that is significantly higher is classified as a "high outlier," while one that is notably lower is classified as a "low outlier." High outliers are uncommon in NIBRS, as automated duplicate checks typically prevent them, and they often result from seasonality or extreme events. In contrast, low outliers are more frequent and may indicate that the agency did not report all crime incidents for that month. In the estimation process, data from months identified as low outliers is excluded from the NIBRS dataset and treated as missing.

For estimation purposes, agencies that report fewer than three months of incidents (i.e., with 10 or more months classified as either "missing months" or "low outliers") are categorized as non-reporting agencies. Their data will be excluded from the NIBRS dataset and compensated through statistical weighting, adjusting the weights of other reporting agencies to account for the missing data of these agencies. This rule is consistent with how partial reporting agencies were defined under the SRS estimation approach. For partial reporting agencies, the full set of incident records for each missing month or low outlier month is imputed. This imputation allows for estimation of not only the number of incidents and offenses but also for the characteristics of those incidents.

Partial reporting is addressed using a block imputation approach, which involves two main steps:

1. **Identify a Donor Agency:** A "donor" agency is selected based on similarity to the agency with partial reporting, specifically in terms of characteristics such as population size and agency type. This selection uses a predictive mean matching (PMM) model, which estimates the total number of incidents for each agency over the year while controlling for agency-level attributes. The donor agency is then randomly chosen from those with the closest number of predicted incidents.
2. **Impute Missing Months:** For the identified missing months, incident records from the donor agency are copied and assigned to the partial reporting agency. Only the missing months are imputed; months with available data remain unaltered during the block imputation process.

3.5. Estimating Population Served by Law Enforcement Agencies

3.5.1. The need for population estimates

Crime totals produced by the NIBRS estimation process are difficult to meaningfully compare across estimation domains (e.g., agency type or race and age groups), geographic groupings (e.g., states and regions), or years. A large, heavily populated state may account for many more crime incidents in a given year than a smaller, less populous state, but simply comparing two totals cannot convey whether crime was higher or lower in either of these states relative to the other.

Accounting for population size when examining criminal activity standardizes the measurement of crime, which is critical for making comparisons across place and over time. To do so requires accurate estimates of the population served for each of the domains to be estimated (e.g., geographic area or agencies of a certain population size).

3.5.2. Putting population estimates to use in crime rate estimation

Presenting totals as population-relative rates is a common practice—such rate estimates are often described as totals per population unit, such as per household or per person. For crime estimates, rates are presented as totals per 100,000 persons. These rates are calculated by dividing the total number of incidents by the population size and multiplying the result by the person group size (i.e., 100,000).

For example, if a state reported 30,800 aggravated assaults in 2021, and the state population that year was estimated at 11 million persons, the resulting rate of aggravated assault for that state in 2021 would be 280 aggravated assaults per 100,000 persons.

3.5.3. The types of population estimates created for NIBRS

NIBRS data are collected at the crime incident level. Each incident is reported by, and therefore associated with, a single law enforcement agency. The incident data includes information on the nature of the crime and on the characteristics of victims, offenders, and persons arrested. This level of detail enables the calculation of crime totals at multiple levels of aggregation—by geographic area (e.g., national, state, county, city), by agency type (e.g., city/municipal agency, county sheriff, tribal), and by individual characteristics (e.g., age, sex, race).

To place crime totals in context, population-relative rates should be used. To calculate a population-relative rate, population sizes are first estimated for each agency eligible to report NIBRS, regardless of its reporting status. Estimates are made for the size of the total population served by the agency and for the sizes of each of the demographic categories for which NIBRS-based victim, offender, and arrest estimates will be generated. Annually, the following population estimates are calculated for each law enforcement agency:

- Total population served
- Population of persons by age
    - 17 years or younger
        - 14 years or younger
            - 4 years or younger
            - 5 to 11 years
            - 5 to 14 years
            - 11 years or younger
            - 12-14 years
        - 15-17 years
            - 15 years
            - 16 years
            - 17 years
    - 18 years old and older
        - 18 to 24 years
        - 25 to 34 years
        - 35 to 64 years
        - 65 years and older

- Population of persons by sex
    - Male
    - Female
- Population of persons by race
    - White
    - Black or African American
    - Other Race
        - American Indian or Alaska Native
        - Asian
        - Native Hawaiian or Other Pacific Islander
- Population by the 2- and 3-way interactions of age group, race, and sex.

3.5.4. Estimating the size and characteristics of populations served by law enforcement agencies

Two broad categories of law enforcement agencies can be assigned estimates for population served in the NIBRS estimation process: *general-purpose agencies* and *special-purpose agencies*.

General-purpose agencies are those that serve as the primary law enforcement entities for some portion of their jurisdictions. General-purpose agencies include local and regional police departments, most sheriff's offices, county police departments, and in some cases, state police agencies. General-purpose agencies cover jurisdictions that align with standard U.S. statistical geographies, and population totals for these geographies are estimated annually by the U.S. Census Bureau. Every year, the Population Estimates Program[7] (PEP) disseminates official population estimates for the geographies that comprise general-purpose agency jurisdictions. In addition, estimates of population composition for these same geographies are released under the American Community Survey[8] (ACS) program. For each general-purpose agency, the PEP provides the source of data for the total population served and the ACS provides population distributions by person-level demographics.

Special-purpose agencies are agencies that are not the primary law enforcement entities for any portion of their jurisdictions. Special-purpose agencies include university police departments, tribal agencies, some sheriff's offices, and most state police agencies. As of the 2023 data year, the only special-purpose agencies with designated population estimates in the NIBRS estimation process are university and college agencies[9]. University and college agency jurisdictions do not align with standard geographies, such that the PEP and ACS population data cannot be directly associated with these agencies. The alternative source of data on university and college populations is student enrollment and staff size estimates produced by the National Center for Education Statistics through the Integrated Postsecondary Education Data System (IPEDS). The IPEDS can generate overall population sizes, as well as disaggregation by sex and race, for each university and college agency.

Because university and college agency jurisdictions overlap with those of general-purpose agencies, university and college agency estimates are displayed on their own and are not included in general crime rate estimates.

[7] See ***https://www.census.gov/programs-surveys/popest.html***.
[8] See ***https://www.census.gov/programs-surveys/acs***.
[9] In future years, the FBI and BJS will evaluate the ability to designate population estimates for other special-purpose agency types.

3.5.5. Ensuring the NIBRS estimation system accounts for overlapping agency jurisdictions

Across the U.S., every home, business, and other type of location where criminal activity can take place falls within the jurisdiction of at least one general-purpose agency[10], and in many cases, they fall within the overlapping jurisdictions of two general-purpose agencies. For example, a city with its own police department may reside within a county served by a sheriff's department. However, crimes occurring within this city—like all other crime reported to law enforcement—can be attributed to only one agency in the NIBRS database. It is therefore important for the population residing in that city to also only be attributed to one agency.

To ensure that population estimates are unique (i.e., mutually exclusive) within these overlapping jurisdictions, the population associated with each general-purpose agency represents only the portion of the agency's jurisdiction for which the agency is the primary law enforcement entity. In the city-within-a-county example, the county sheriff's office is not considered the most direct enforcement entity within the city limits, and therefore, the population estimate for the sheriff's office is reduced by the size of the city population. This process of deduction—termed "residualization" in the NIBRS estimation process—is used to reconcile the overlapping jurisdictions of city agencies (which are typically the primary law enforcement entities) with other general-purpose agencies. This ensures that state- and national-level crime rates are calculated accurately, with each incident's relevant characteristic (victim, offender, or arrestee count) contributing once to the numerator and each person in the population contributing once to the denominator.

3.6. Measuring Statistical Uncertainty

3.6.1. Statistical uncertainty

Estimates based on statistical samples depend on measurement, recording, and processing procedures that have inherent limitations. These limitations impact how effectively an estimate approximates the population value (i.e., the value of an outcome if all reporting agencies in the population had submitted 12 months of data). *Statistical uncertainty* is a way to measure how well the estimate approximates the population value.

Three common indicators of statistical uncertainty are sample variance, sample standard error, and mean squared error (MSE). These measures indicate the magnitude of potential error in an estimate attributable to the data collection and estimation process. In all cases, the smaller the value of uncertainty, the closer the estimate will be, on average, to the population value.

Federal statistical agencies establish thresholds for acceptable or unacceptable estimates based on the measure of uncertainty. The measures of uncertainly used to assess the quality of NIBRS estimates for publication are based on relative values to allow for a consistent interpretation across all measures.

The two main drivers of statistical uncertainty are *sampling error* and *non-sampling error*.

- **Sampling error** refers to the uncertainty introduced when an estimate is based on only part of the population of interest. When only part of the population is used, the value of an estimate is directly related to the subset of the population used in the estimation process. Ideally, the average of all population subsets should be the population value. Sampling error is a function of the total range of plausible alternatives for the population value. The goal is to achieve a range that is as small as resources and effort allow. Sampling error leads to statistical variance. In NIBRS, the variance is affected by the specific set of agencies that have transitioned to NIBRS.

[10] Excluding tribal areas, which may fall under the exclusive jurisdictions of non-general-purpose tribal agencies.

- **Non-sampling error** encompasses statistical uncertainty arising from limitations in measurement, recording, and processing procedures. Major aspects of non-sampling error in NIBRS statistical estimates are:
    - *Coverage error*. Coverage error occurs when not all law enforcement agencies are represented in NIBRS or able to submit data to it.
    - *Nonresponse error*. Nonresponse error occurs when a law enforcement agency can submit data but neglects to do so or does not provide all the data required.
    - *Measurement error*. Measurement error occurs when a law enforcement agency reports the wrong value or code for a NIBRS data element.
    - *Estimation error*. Estimation error occurs when there is uncertainty in the adjustments intended to compensate for limitations in the sample data.

Non-sampling error can lead to bias. Bias means estimates vary from the population value in a systematic and predictable way. Positive bias means estimates are consistently above the population value. Negative bias means estimates are consistently below the population value. In the absence of bias, the average of estimates across repeated samples converges very closely to the population value for the population of interest.

NIBRS measures statistical uncertainty by the MSE.[11] MSE accounts for the contributions to uncertainty of both sampling error and non-sampling error. For an estimate (θ), the MSE is a function of the sampling error (variance) and non-sampling error (bias) and defined as:

$$MSE = Var(\theta) + Bias^2$$

3.6.2. Measuring variance

In NIBRS, because the analysis weights for reporting agencies were generated through a calibration process, which aligned the weighted sum of reported crime totals by offense type with the population benchmarks from the SRS across different weighting groups (denoted as *h=1,….H*), we applied the variance estimation method appropriate for a calibrated estimator.[12] The base weight for each reporting agency *hi* is defined as $d_{hi}=N/n$, where *N* is the total number of agencies in the population and *n* is the total number of reporting agencies in the population.

Calibration weighting is applied by each weighting group using auxiliary variables $\boldsymbol{x}_{hi}$, which are SRS crime counts by offense type known for every agency in the entire population. The calibrated weights w_i are constructed to satisfy the calibration equation:

$$\sum_{i \in s_h} w_{hi}\boldsymbol{x}_{hi} = \sum_{i \in U_h} \boldsymbol{x}_{hi}$$

where s_h denotes the set of reporting agencies in weighting group *h* and U_h denotes the set of all agencies in weighting group *h.*

The calibration estimator for the total of the study variable *y* is then given by:

[11] In a sample survey, like the Bureau of Justice Statistics' (BJS) National Crime Victimization Survey (NCVS), the bias is assumed to be zero because a random sample of the population is drawn. Therefore, in the case of the NCVS, the measure of uncertainty is the sampling variance. However, for NIBRS estimates, the participating agencies are not random. Therefore, the measure of uncertainty needs to account for the bias component.

[12] Särndal, C.-E., Swensson, B., & Wretman, J. (1992). Model Assisted Survey Sampling. Springer-Verlag.

$$\hat{t}_y^{cal} = \sum_{h=1}^{H} \sum_{i \in s_h} w_{hi} y_{hi}$$

The design-based variance estimator for the calibration total $\hat{t}_y^{cal}$, incorporating the finite population correction (FPC) in each weight group, is then given by:

$$\hat{V}\left(\hat{t}_y^{cal}\right) = \sum_{h=1}^{H} \left(1 - \frac{n_h}{N_h}\right) \frac{n_h}{n_h - 1} \sum_{i \in s_h} \left(w_{hi} e_{hi} - \frac{1}{n_h} \sum_{j \in s_h} w_{hj} e_{hj} \right)^2$$

where:

N_h is the total number of agencies in weighting group *h*;

n_h is the total number of reporting agencies in weighting group *h*; and

$e_{hi} = y_{hi} - \boldsymbol{x}_{hi}^T \widehat{\boldsymbol{\beta}}_h$ is the residual for agency *hi* with $\widehat{\boldsymbol{\beta}}_h = \left(\sum_{i \in s_h} d_{hi} \boldsymbol{x}_{hi} \boldsymbol{x}_{hi}^T\right)^{-1} \sum_{i \in s_h} d_{hi} \boldsymbol{x}_{hi} y_{hi}$.

3.6.3. Measuring bias

Statistical bias is defined as the difference between the expected value (the average across all possible samples) for a parameter estimate ($E[\theta]$) and the true value for a parameter (Θ) in the population of interest. In other words,

$$Bias(\theta) = E[\theta] - \Theta$$

However, for crime estimates, like most estimates produced, the true value for a parameter and the expected value of the estimate are not known. Therefore, bias can only be estimated. For NIBRS estimates, bias is estimated by comparing two crime estimates via distinct methodologies (weighting vs. imputation) conditioned on different predictors (SRS crime totals vs. population served totals across characteristic groups based on race, age, gender, and poverty ratio). The official crime statistic is based on the weighted estimate and the estimated bias is used to gauge uncertainty in that estimate.

3.6.4. How statistical uncertainty is used for NIBRS estimates

Statistical uncertainty, as expressed through MSE, is used for two critical components of the NIBRS estimation process.

1. MSE is the building block for the confidence intervals associated with each of the estimates generated by the NIBRS estimation process (see ***Section 3.7***).
2. MSE is used to determine whether an estimate is of sufficient quality to publish, based on criteria established for NIBRS estimation (see ***Section 3.8***).

3.7. Construction of Confidence Intervals for Point Estimates

3.7.1. Constructing a confidence interval using MSE

A statistical confidence interval is the set of plausible values within which the population value for an indicator resides for a given confidence level. For example, using a 95 percent confidence level, in 95 out of 100 samples drawn, the confidence interval would be expected to contain the population value.

A NIBRS confidence interval requires:

- the value of the estimate (θ) calculated from the NIBRS data,
- the value of the root mean squared error (RMSE), which is the square root of the MSE, and
- a confidence level factor (*z*), a quantile from the standard normal distribution corresponding to the desired confidence level (i.e., for a 95% CI, z=1.96).[13]

Using these three values, the formula for a confidence interval is

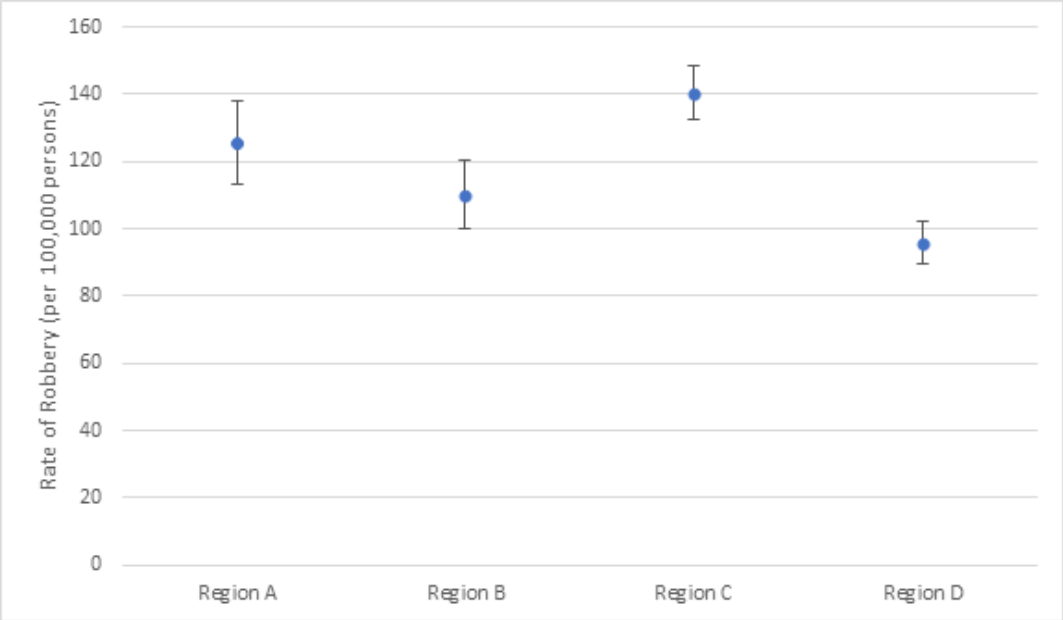
$$\theta \pm z \cdot RMSE$$

which can be explained as the point estimate plus or minus the value of the statistical uncertainty based on the desired level of confidence. Therefore, a confidence interval based on a 95% confidence level factor (*z*) represents the plausible range of values within which the population value will be found 95% of the time.

3.7.2. Using a confidence interval to understand estimate uncertainty

Often confidence intervals are shown pictorially to help visualize the potential range for the population value. ***Exhibit 3-12*** illustrates a hypothetical rate of robberies per 100,000 persons by region and the corresponding confidence intervals around them. For example, Region A has a rate of 125.5 per 100,000 persons with a confidence interval of ±12.5.

Exhibit 3-12. Example of confidence intervals around a hypothetical point estimate



The value of *z* is chosen to ensure that the confidence interval captures the true population parameter a specified percentage (or more) of the time in repeated sampling. When a confidence level of 95 percent is desired, the *z*-value is 1.96.

[13] While the t-value is more appropriate when the population standard deviation is unknown and estimated from the sample, it is typically used for smaller sample sizes. In contrast, the z-value is used when the sample size is large, as the t-distribution approaches the normal distribution.

The lower endpoint of a confidence interval is called the lower confidence limit. The upper endpoint is the upper confidence limit. The range spanned by the confidence interval represents the number of plausible alternatives for the population value. The closer the endpoints are to the point estimate (i.e., the smaller the range of plausible values), the higher the quality of the estimate.

For NIBRS estimates, the confidence interval should be used when discussing an estimate. For example, if the number of robberies in the year is 1,000 and its 95 percent confidence interval is 900 to 1,100, the appropriate interpretation of the data is: the estimated number of robberies in the year was 1,000 and, with 95 percent confidence, the population value is between 900 and 1,100.

3.8. Statistical Suppression Rules for Publishing NIBRS Estimates

3.8.1. What are statistical suppression rules

Statistical suppression is a decision not to publish an estimate because that estimate does not meet a minimum threshold for statistical quality. Ideally, all estimates produced through the NIBRS estimation process would be made available to the public. However, because NIBRS estimates each have varying levels of uncertainty, each estimate is reviewed to determine if it meets the FBI's and the Bureau of Justice Statistics' (BJS) publication standards.

3.8.2. NIBRS statistical suppression rules

In NIBRS, the decision to suppress an estimate (θ) is guided by the following rules:

- If $RMSE/\theta \geq 0.3$ and the NIBRS population coverage for the estimation domain (e.g., a region or state) is < 80%, then the estimate is suppressed.
- If $\theta = 0$ and the population coverage for the estimation domain is < 80%, then the estimate is suppressed.
- If population coverage for the estimation domain is < 80% and $RMSE/\theta \geq 0.3$ is true for 75% or more of key estimates, then all estimates in the domain are suppressed.
- If one or more must-have agencies[14] are missing from NIBRS for an estimation domain, the entire domain is suppressed.
- Otherwise, if population coverage for the estimation domain is > 80%, all estimates for that domain will be published regardless of the value of the ratio $RMSE/\theta$.

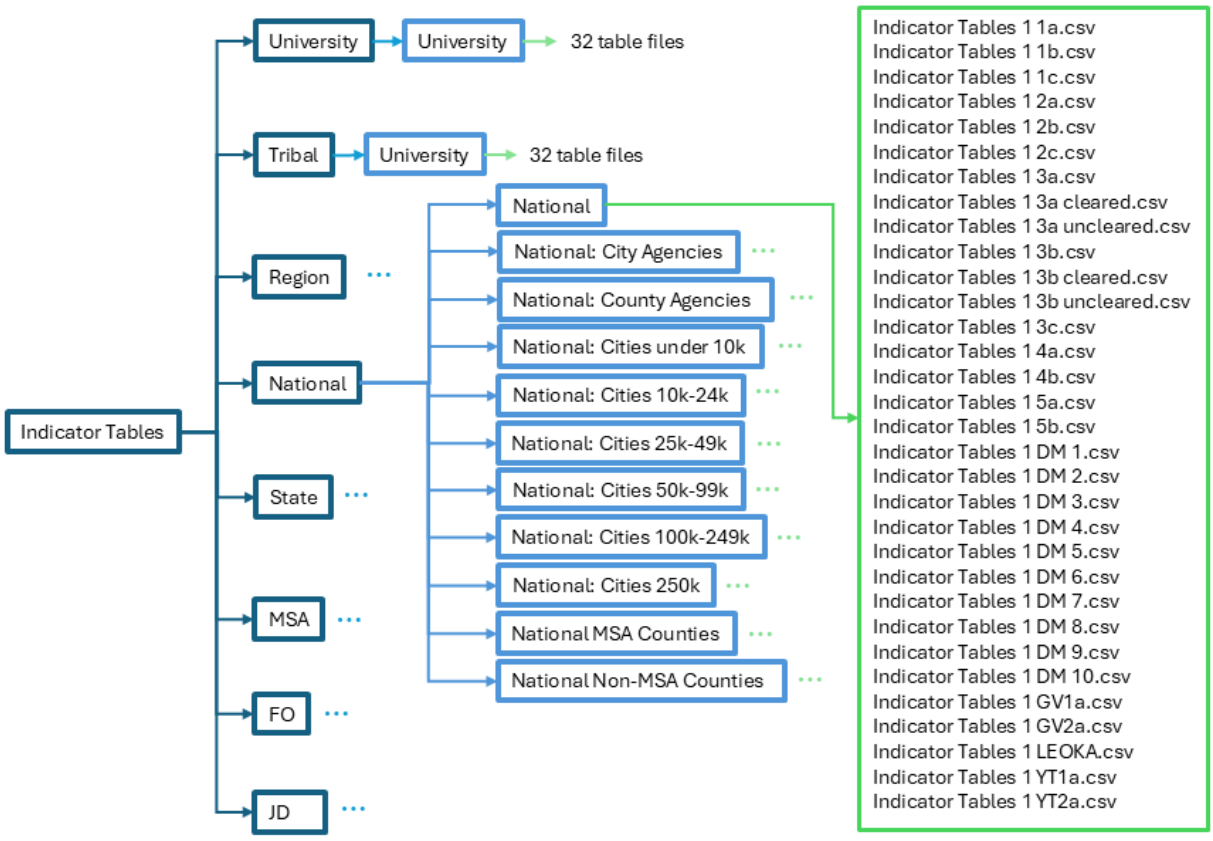
3.9. Output Datasets with Final Estimates

3.9.1. Output format

The estimates produced are output into a reporting database. Due to the number of estimates produced, the reporting database is a series of files divided by a domain of analytic interest and geographic area. To be useable by as many users as possible, the reporting database is output in CSV files. ***Exhibit 3-13*** details the schema used in the creation of the reporting database CSV files. ***Exhibit 3-14*** provides a crosswalk between the naming convention of the reporting database and the types of estimates included.

[14] For non-MSA geographic estimation domains, must-have agencies are those comprising 10% or more of the population for the entire domain. For MSAs, must-have agencies are those designated as principal cities.

Exhibit 3-13. Reporting database schema



The Estimate Folder Contains: One folder per broad geographic type, which each contain one folder per specific geographic subset. Each subset contains one file for each of the 32 tables.

Files are named by "Indicator Tables [GEO ID] [TABLE ID].csv"

Exhibit 3-14. Reporting database naming convention crosswalk

Table	Unit of analysis	Indicators included	Estimate type
1a	Incident	Crimes against persons	Count, percentage, rate
1b	Incident	Crimes against property	Count, percentage, rate
1c	Incident	Crimes against society	Count, percentage, rate
2a	Offense	Crimes against persons	Count, percentage, rate
2b	Offense	Crimes against property	Count, percentage, rate
2c	Offense	Crimes against society	Count, percentage, rate
3a	Victim	Person victimizations for crimes against persons and crimes against property	Count, percentage
3a-clear	Victim	Person victimizations for crimes against persons and crimes against property (cleared incidents only)	Count, percentage

Table	Unit of analysis	Indicators included	Estimate type
3a-unclear	Victim	Person victimizations for crimes against persons and crimes against property (uncleared incidents only)	Count, percentage
3b	Victim	Person victimizations for crimes against persons and crimes against property	Rate
3b-clear	Victim	Person victimizations for crimes against persons and crimes against property (cleared incidents only)	Rate
3b-unclear	Victim	Person victimizations for crimes against persons and crimes against property (uncleared incidents only)	Rate
3c	Victim	Non-person victimizations for crimes against property	Count, rate
4a	Arrestee	Group A offenses resulting in arrest	Count, percentage
4b	Arrestee	Group A offenses resulting in arrest	Rate
5a	Arrestee	Persons arrested for Group A and Group B offenses	Count, percentage
5b	Arrestee	Persons arrested for Group A and Group B offenses	Rate
DM1	Incident	Completed drug-related offenses	Count, percentage
DM2	Incident	Completed drug/narcotic violations by criminal activity	Count
DM3	Incident	Completed drug/narcotic violations by property type	Count, percentage
DM4	Incident	Completed drug/narcotic violations by drug type	Count
DM5	Incident	Completed drug/narcotic violations by drug type and criminal activity	Count
DM6	Incident	Drug type by criminal activity (incidents with one suspected drug type)	Count
DM7	Arrestee	Persons arrested for drug type by criminal activity (incidents with one suspected drug type)	Count
DM8	Incident	Drug type by criminal activity	Count
DM9	Arrestee	Persons arrested for drug type by criminal activity	Count
DM10	Arrestee	Type of criminal activity resulting in arrest	Count, percentage
LEOKA	Victim	Law enforcement officers killed and assaulted	Count, percentage, rate
GV1a	Offense	Crimes against persons involving guns	Count, rate
GV2a	Victim	Person victimizations for crimes against persons involving guns	Count, percentage, rate
YT1	Victim	Youth homicide victimizations	Count, percentage
YT2	Victim	Youth homicide offenders	Count, percentage

3.9.2. Structure of reporting database

The reporting database CSV files are all structured the same way. As such, a user can append files together as needed for their analytic needs. Each file contains the point estimates and a series of additional statistical attributes that provide additional information about the estimate or other data quality information. ***Exhibit 3-15*** lists the statistical attributes contained in the reporting database.

Exhibit 3-15. Statistical attributes contained in the NIBRS estimation reporting database

Statistical attribute	Reporting database field name	Attribute detail
Indicator name	indicator_name	This field contains the name of the indicator being estimated, for example, "Total violent crime" or "Property crime."
Estimate (weighted)	estimate	The weighted estimate for a particular indicator.
Estimate (unweighted)	estimate_unweighted	The unweighted estimate for a particular indicator.
Estimate geographic location	estimate_geographic_location	This field indicates if the estimate is at the national level, region, state, MSA, etc. Also, shows if subset by the various demographics for certain tables. Note the geographic permutations contain labels such as National, State, MSA, etc., whereas the demographic permutations contain labels such as Age: 12 or Older, Sex: Female, Race: White in addition to the geographic permutation. If the label is just the geographic permutation, then all persons within an incident were used for the estimate and, for the purpose of this document, it is called the main geographic permutation to calculate various columns in this database.
Estimate type	estimate_type	This field indicates if the estimate is of the type (1) count, (2) percentage, or (3) rate.
Estimate type (number)	estimate_type_num	This field contains a numeric value representing the estimate type: 1=count; 2=percentage; 3=rate.
Estimate type detail: percentage	estimate_type_detail_percentage	This field indicates the type of percentage being estimated. For example, incident level or victim level. If the estimate is not a percentage, this field will be blank.

Statistical attribute	Reporting database field name	Attribute detail
Estimate type detail: rate	estimate_type_detail_rate	This field indicates the type of rate being estimated. For example, crime rate per 100,000 persons or arrest rate per 100,000 persons. If the estimate is not a rate, this field will be blank.
Estimate domain Part 1	estimate_domain_1	This field indicates if the estimate is at the overall level or the subdomain level (e.g., by age category, race/ethnicity, or sex).
Estimate domain Part 2	estimate_domain_2	This field indicates if the estimate is at a more granular level (e.g., No, Yes, Male, Female, etc.).
Estimate standard error	estimate_standard_error	This field contains the measure of uncertainty for an estimate. The methodology for computing the standard error will be detailed in a future decision memo.
Estimate upper bound	estimate_upper_bound	This field represents the upper bound of the confidence interval for the estimate. Currently, this is a symmetrical confidence interval (e.g., a Wald confidence interval), but asymmetrical confidence intervals may need to be considered for percentages and rates since they have one or more boundary limit (i.e., 0 and 100 for a percentage and 0 for a rate).
Estimate lower bound	estimate_lower_bound	This field represents the lower bound of the confidence interval for the estimate. Presently, the expectation is to use a symmetrical confidence interval (e.g., a Wald confidence interval), but asymmetrical confidence intervals may need to be considered for percentages and rates since they have one or more boundary limit (i.e., 0 and 100 for a percentage and 0 for a rate).
Relative standard error (RSE)	relative_standard_error	This field contains the relative standard error (RSE) of the estimate. The RSE is the ratio of the standard error and the point estimate. RSEs are a good indicator of the quality of the estimate. For example, RSE values > 30% are typically considered poor.

Statistical attribute	Reporting database field name	Attribute detail
Analysis weight name	analysis_weight_name	This field contains the name of the weight variable used to produce the estimate. Because some years may have separate weights depending on the level of analysis (e.g., national or state) this field will document which weight was used.
Estimate of potential bias	estimate_prb	This field, where applicable, contains an estimate of the direction and magnitude of the bias associated with an estimate. The formula for this field is: 100 * ((weighted_estimate – imputed_estimate*) / imputed_estimate) Note: For percentages, the estimate_prb value for the corresponding count is used.
Estimate of bias	estimate_bias	The calculated bias for an estimate. Calculated for each estimate type as: TOTAL: Total * (estimate_prb / 100) RATE: Total * (1000/population_estimate) * estimate_prb PERCENTAGE: Percentage_NIBRS – (ALL_TOTAL_NUM) / (ALL_TOTAL_DEN) * 100
Root mean squared error (RMSE)	estimate_rmse	This field contains the relative mean squared error. Calculated as: RMSE=sqrt(SE^2 + bias^2)
Relative RMSE	relative_rmse	This field contains the relative root mean squared error. Calculated as: RRMSE: RMSE/estimate

Statistical attribute	Reporting database field name	Attribute detail
Suppression flag indicator	suppression_flag_indicator	This field recommends whether an estimate should be suppressed (i.e., not published at all). Depending on the geographic permutation in column estimate_geographic_location, it considers and applies the following rules, where appropriate: a. Using the main geographic permutation, it will consider if a key agency that accounts for more than 10% of the population served is a non-NIBRS reporter or for the MSA geographic permutations, a non-NIBRS reporter is in a principal city for the MSA. b. The unsuppression rule (i.e., der_perm_group_unsuppression_flag) c. The momentum rule (i.e., der_perm_group_suppression_flag) d. Use the column der_rrmse_gt_30_se_estimate_0_2_cond to determine if the estimate should be suppressed. This field has values of 1 for suppressed, 0 for not suppressed, and blanks if the estimate is the NA code (i.e., -9).
Estimate eligible for suppression	der_elig_suppression	This field contains the flag for an estimate eligible for suppression (i.e., when agency_counts is not NA and estimate is not the NA code (-9)).
Estimate geographic location coverage ratio	pop_cov	Grouping by estimate_geographic_location, calculate the population coverage for the geographic permutation and demographic permutation group.
Agency counts	agency_counts	The number of unweighted pseudo-agencies (i.e., agency crossed county, agencies count more than one if an agency served in more than one county) contributing to the estimate. If NA, then no agency report information.

Statistical attribute	Reporting database field name	Attribute detail
Estimate eligible for suppression RRMSE	der_rrmse_30	Flag when RRMSE is > 30%.
Estimate level suppression	der_rrmse_gt_30_se_estimate_0_2_cond	Estimate level suppression flag gets a value of 1 for the following conditions: 1. Any estimate with > 30% %RRMSE OR 2. The following agency-level criteria is used to suppress estimates or standard errors that are 0: a) The column estimate_geographic_location has a value of “National Agency Type Tribal” and the agency tribal coverage is < 80%. b) The column estimate_domain_1 has a value of “Agency indicator,” estimate_domain_2 has a value of “Tribal agencies,” and the geographic agency tribal coverage is < 80%. c) The column estimate_domain_1 has a value of “Agency indicator,” estimate_domain_2 has a value of “State police,” and the geographic state police coverage is < 80%. d) The column estimate_domain_1 has a value of “Agency indicator,” estimate_domain_2 has a value of “Other state agencies,” and the geographic other state agencies coverage is < 80%. e) For all other estimates or standard errors that are 0, suppress if geographic coverage is < 80%. Else this flag gets a value of 0 or is blank if the estimate is the NA code (-9).
Estimate geographic location percentage of suppressed estimates	der_rrmse_gt_30_se_estimate_0_2_cond_top	Grouping by the main geographic permutation (i.e., no demographic permutation) on column estimate_geographic_location, calculate the % of key estimates with a value of 1 in the flag from der_rrmse_gt_30_se_estimate_0_2_cond.

Statistical attribute	Reporting database field name	Attribute detail
Estimate geographic location unsuppression flag	der_perm_group_unsuppression_flag	Any main geographic permutation group (i.e., no demographic permutation) on column estimate_geographic_location level with a value from pop_cov > 80% gets a permutation group level force unsuppression flag of 1; else 0.
Estimate geographic location suppression flag	der_perm_group_suppression_flag	Any permutation group (estimate_geographic_location level) with a value from der_rrmse_gt_30_se_estimate _0_2_cond_top > 75% AND a value from the main geographic permutation pop_cov < 80% gets a permutation group level suppression flag of 1; else 0.
Population estimate	population_estimate	This field contains the population estimate associated with the estimate's geographic level and domain. For example, if the estimate is representing males (domain) in Texas (geographic level), then this field will contain the number of males in Texas. This field is the denominator for the calculation of any rate estimates.
Time series start year	time_series_start_year	This field indicates the first year for which estimates were produced—either original or revised.
Table from the Indicator Table Task	full_table	Table from the Indicator Table Task.
Variable name	der_variable_name	Code name for the cells from the Indicator Table Task. Need to cross the variables PERMUTATION_NUMBER, estimate_geographic_location, der_variable_name, and estimate_type to have unique identifier in database.

Statistical attribute	Reporting database field name	Attribute detail
Permutation number	PERMUTATION_NUMBER	This is the numeric version of the variable "Estimate geographic location." Note the main geographic permutations have values less than 1,000.
Actual bias	PRB_ACTUAL	This has the same value as estimate_prb after implementing the updated methodology of the bias calculation to use the Copula method.
Non-NIBRS agency accounting for over 10% of population served	POPTOTAL_ORIG_UNIV_ELIG_PERM_ AGENCY_MISSING_OVER_10_PERCENT	A Boolean variable that indicates for the main geographic permutation if a non-NIBRS agency accounting for over 10% of the population served is flagged. Has values of TRUE if there is such an agency and FALSE otherwise. This column is used in the indicator suppression_flag_indicator, for the geographic permutation of the following: a. National crossed agency type b. National crossed agency size c. Regional d. State e. Judicial district f. Field office
Non-NIBRS agency is a principal city in MSA	POPTOTAL_ORIG_ELIG_PERM_ AGENCY_ MISSING_PRINCIPAL_CITY	A Boolean variable that indicates for the main geographic permutation if a non-NIBRS agency is a principal city in MSA. Has values of TRUE if there is such an agency and FALSE otherwise. This column is used in the indicator suppression_flag_indicator, for the geographic permutation of the following: MSA
Proportion eligible pseudo-agency	PROP_ELIG_ORIS_NONZERO_COUNT	A proportion variable that is defined by the number of unweighted pseudo-agency (i.e., agency_counts) divided by the number of pseudo-agencies within a geographic permutation.

Statistical attribute	Reporting database field name	Attribute detail
Correlation with prior year	CORRELATION_WITH_PRIOR_YEAR	A pre-defined correlation variable that is to be used when comparing the current year estimates to the prior year estimates.

* The imputed estimate, which is not shown in the reporting database, is an alternative estimate of the population outcome of interest (i.e., counts, rates, or percentages of incidents/offenses/victims/arrestees). It uses a Gaussian copula model to impute for missing outcomes for nonparticipating NIBRS agencies. The primary NIBRS estimates, reported in the estimate column of the reporting database, are based on a weighted estimator, as described in ***Section 3.3***. The imputed estimates use an imputation estimator rather than a weighted estimator and are conditioned on a different set of covariates than the primary NIBRS estimates. Specifically, the weighted estimator covariates are crime counts from the SRS interacted with geography and post-stratum, whereas the imputation estimator covariates are the number of sworn officers and the size and demographics of the populations served by each contributing agency. The difference between the weighted estimate and the imputed estimate is considered a proxy for the unknown bias of the weighted estimator and contributes to an expansion of the confidence intervals.

3.10. Interpretation of NIBRS Estimates

3.10.1. Estimate domains

The NIBRS estimation process uses statistical weights to ensure the resulting estimate is representative of the population being analyzed. The NIBRS estimation process allows for inference to a large set of populations, including:

- The entire United States
- Population size groups
    - All cities 250,000 inhabitants or over
    - Cities 100,000 to 249,999 inhabitants
    - Cities 50,000 to 99,999 inhabitants
    - Cities 25,000 to 49,999 inhabitants
    - Cities 10,000 to 24,999 inhabitants
    - Metropolitan Statistical Area (MSA) counties
    - Non-MSA counties
- General-purpose agency types (i.e., city, county)
- Special agency types (i.e., university, tribal)
- Regions of the United States (Northeast, Midwest, South, West)
- Regions by population size group
- Regions by general-purpose agency type
- States (50 states plus Washington, D.C.)
- Metropolitan Statistical Areas (MSAs)
- Field offices
- Judicial districts.

Each area is considered an *estimate domain*. The estimate domain is the area for which inference is being made by an estimate. The NIBRS estimation process will facilitate inference in 709 estimate domains.

In addition to estimate domains, NIBRS can be generalized to the population by person characteristics such as age, race, and sex. The person domains for which estimates can be produced include:

- Detailed age categories (under 5, 5–11, 12–14, 15–17, 15, 16, 17, 18–24, 25–34, 35–64, 65 or older)
- Summary age categories (Under 12, Under 15, Under 18, 18 or older)
- Race categories (White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander)
- Aggregate "Other" race categories (combined American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander)
- Sex (male, female)
- Two-way or three-way combinations across the demographic domains
- Law enforcement officer assaults in the United States.

Estimates based on 2023 data are not provided by Hispanic origin due to potential nonresponse and measurement errors in this NIBRS data element for victims, offenders, and persons arrested. Algorithms are currently under development to address these errors and to enable estimates by Hispanic origin for future years of estimates.

3.10.2. Information needed to interpret estimates

The NIBRS estimation process provides four pieces of information that are needed to understand the estimate's meaning and give it context:

1. *Weighted point estimate (E)*. The weighted point estimate is a point estimate representing a population parameter (like the mean or proportion) under analysis.
2. *Confidence interval—lower bound (LB)*. The confidence interval—lower bound represents the smallest value, with 95% confidence, that the population value may be.
3. *Confidence interval—upper bound (UB)*. The confidence interval—upper bound represents the largest value, with 95% confidence, that the population value may be.
4. *Root mean squared error (RMSE)*. The mean squared error (MSE) is the measure of uncertainty used for NIBRS estimates (see ***Section 3.6***). It is utilized for calculating the RMSE confidence interval and for performing statistical comparisons between two estimates.

3.10.3. Types of estimates produced

The NIBRS estimation process produces three types of estimates:

1. *Totals*. Totals estimate the volume of crime or characteristics of a crime in a domain.
2. *Rates*. Rates represent the per capita level (e.g., per 100,000 persons) of crime in a domain. Rates are produced for an entire estimate domain or for an age, sex, or race category.
3. *Percentages*. Percentages are produced for characteristics of an incident (e.g., location).

NIBRS estimates can be produced for characteristics in the following NIBRS segments (see ***Section 3.2***):

- Incidents (person crimes, property crimes, crimes against society)
- Offenses (person crimes, property crimes, crimes against society)

- Crime victims (person victims, non-person victims)
- Persons arrested.

As of the 2023 data, estimates at the offender level were not produced due to higher rates of missing information in cases that are not cleared. Analysis is currently underway to determine how to account for the high rate of missing information and produce accurate offender-level estimates.

3.10.4. Interpreting NIBRS estimates

Point estimates derived by the NIBRS estimation process are representative of the estimate domain of interest. However, each point estimate has some amount of uncertainty associated with it, represented by the confidence intervals around the point estimate. Confidence intervals should be used in tandem with the point estimates when interpreting crime data findings.

For example, the estimation process may provide the following information for the number of violent crimes in the United States in 2023:

Violent crime estimate type	Weighted estimate	Confidence interval (lower bound)	Confidence interval (upper bound)	Root mean squared error (RMSE)
Total	1,298,843	1,265,917	1,331,769	16,799
Rate per 100,000 persons	387.8	378.0	397.7	5.0

The results can be interpreted as:

- The total volume of violent crime in the United States in 2023 is estimated to be 1,298,843, where the population value, with 95% confidence, is between 1,265,917 and 1,331,769.
- The rate of violent crime in the United States in 2023 is estimated to be 387.8 per 100,000 persons, where the actual rate in the population, with 95% confidence, is between 378.0 and 397.7.

3.10.5. Presenting NIBRS estimates

The NIBRS estimates are stored as their actual (unrounded) estimates. However, because of the uncertainty in the estimates, BJS uses the following rounding rules when publishing or citing an estimate:

- *Totals*. Round totals to the nearest 10. For example, 3,456 will be rounded to 3,460.
- *Rates*. Round rates to the first decimal place. For example, a rate of 34.23 persons per 100,000 should be rounded to 34.2 persons per 100,000.
- *Percentages*. Round percentages to the first decimal place. For example, a percentage of 32.782% should be rounded to 32.8%.

3.10.6. What it means when an estimate is not available

As a federal statistical agency, BJS needs to ensure that all official statistics that are released meet a set of quality standards. In general, if an estimate's level of uncertainty falls outside preestablished limits, the estimate may be withheld (i.e., suppressed) from publication and not available for release from the

NIBRS estimation process.[15] As the NIBRS estimation process produces over 829 million estimates for each data year, it is inevitable that some resulting estimates will not meet the established quality standards.

3.11. Performing Statistical Testing with NIBRS Estimates

3.11.1. Types of statistical tests

The information provided by the NIBRS estimation process allows for two different methods of statistical comparison, each with strengths and limitations.

Comparisons can be made by:

- *Examining overlapping confidence intervals*. The overlapping confidence interval method involves checking whether confidence intervals of the compared two estimates overlap. If there is an overlap, the two estimates are not statistically different; otherwise, they are considered statistically different. This check can be performed by comparing the upper bound of the smaller estimate to the lower bound of the larger estimate. A statistically significant difference is detected when the upper bound of the smaller estimate is less than the lower bound of the larger estimate.
    - *Strengths*: The overlapping confidence interval method is easy to implement.
    - *Limitations*: The overlapping confidence interval method is considered a conservative test (i.e., less likely to indicate a statistical difference even when one exists); unable to provide a statistical p-value, only a binary indication of a statistical difference; estimates are assumed to be independent.
- *Conducting a statistical test*. The statistical test method can take different forms, such as a t-test (used when comparing totals, rates, or percentage change between two groups) or a chi-squared test of proportions (used when comparing percentages across two or more groups). A difference is statistically significant when the resulting measure of association (i.e., p-value) corresponding to the resulting test statistic is less than 0.05.
    - *Strengths*: The statistical test method provides a p-value to assess strength of difference; a less conservative test than the overlapping confidence interval method; can incorporate correlation when comparing the same indicator across consecutive years.
    - *Limitations:* The statistical test method is more computationally challenging.

The NIBRS estimation process enables comparisons:

- Across estimate domains, such as comparing the same indicator between two domains (e.g., violent crime rate in Region A and Region B).
- Over time, such as comparing the same indicator across two different years (e.g., murder rate in 2022 and 2023).

3.11.2. Accounting for correlation in statistical tests

Many comparisons may be made between pairs of NIBRS estimates, including comparisons of a single indicator across 2 years (e.g., the difference in robbery rates between the current NIBRS year and the

[15] The reported NIBRS data are available to users from which estimates based on reporting agencies can be made for any outcome of interest.

previous year) and comparisons between two different domains within a single year (e.g., the difference in robbery rates between two regions for the current NIBRS year). Just as the individual estimates have statistical uncertainty (as discussed in ***Section 3.6***), there is statistical uncertainty in the difference between any pair of estimates, which, like the uncertainty in the individual estimates, can also be measured with an estimate of the MSE. That MSE depends on the variances and biases of the two individual estimators, as well as the correlation between the two estimators.

Due to the volume of NIBRS estimates produced each year and the number of potential comparisons that could be made between estimates, it would be computationally challenging to estimate correlations for every potential pair of estimators. Instead, two approaches are recommended for statistical tests of the difference between two NIBRS estimates:

- For the difference in estimates between two different domains within a single year: Assume the correlation is zero. That assumption will be true for domains that have no overlap in agencies (e.g., in a comparison of two different regions, no agency simultaneously appears in both regions). But the assumption will likely be false for domains where there is overlap in agencies (e.g., if both domains can appear in a single agency, such as in a comparison between males and females). In those latter cases, the correlation is likely not zero and is likely positive. If the unestimated correlation is positive, then assuming it is zero would increase the estimated MSE. In other words, assuming the two estimators are uncorrelated is likely a conservative approach that exaggerates the statistical uncertainty in the difference between the two.
- For the difference in estimates for a single indicator between 2 years: The NIBRS estimation output reports a generic estimate of the year-to-year correlation that may be used as an approximate correlation in comparisons of any indicator between the current year and prior year. The year-to-year correlation is associated with the proportion of eligible agencies (i.e., Originating Agency Identifiers, ORIs) that reported one or more incidents used to construct the corresponding estimate among all eligible agencies in the analytic domain (e.g., the entire nation or a region) for the current year. Since different estimates can have varying degrees of year-to-year correlation, the approximate correlation is estimated based on this proportion to provide a more general approach for accounting for year-to-year correlation. ***Exhibit 3-16*** shows the estimated correlations.

Exhibit 3-16. Estimated year-to-year correlation based on proportion of eligible ORIs reporting one or more incidents used to construct the corresponding estimate*

Proportion of eligible ORIs reporting 1+ incidents	Estimated year-to-year correlation
0–0.009	0
0.01–0.029	0.239
0.03–0.049	0.325
0.05–0.099	0.410
0.10–0.149	0.511
0.15–0.249	0.600
0.25–0.349	0.701
0.35–0.499	0.717
0.50–0.699	0.802
0.70–1.00	0.846

*Both values shown in Exhibit 3-16 are included on the most recent year's data files. The proportion of eligible ORIs reporting one or more incidents is labeled as "PROP_ELIG_ORIS_NONZERO_COUNT." The estimated correlation is labeled as "CORRELATION_WITH_PRIOR_YEAR."

3.11.3. Example of overlapping confidence intervals method

To determine if the robbery rate in Region A is different from the robbery rate in Region B or Region C, the following information needs to be generated from the NIBRS estimation process:[16]

Domain	Weighted estimate	Upper bound	Lower bound
Region A	54.3	60.1	48.6
Region B	61.4	93.6	29.1
Region C	60.2	64.6	55.8
Region D	82.6	95.1	70.1

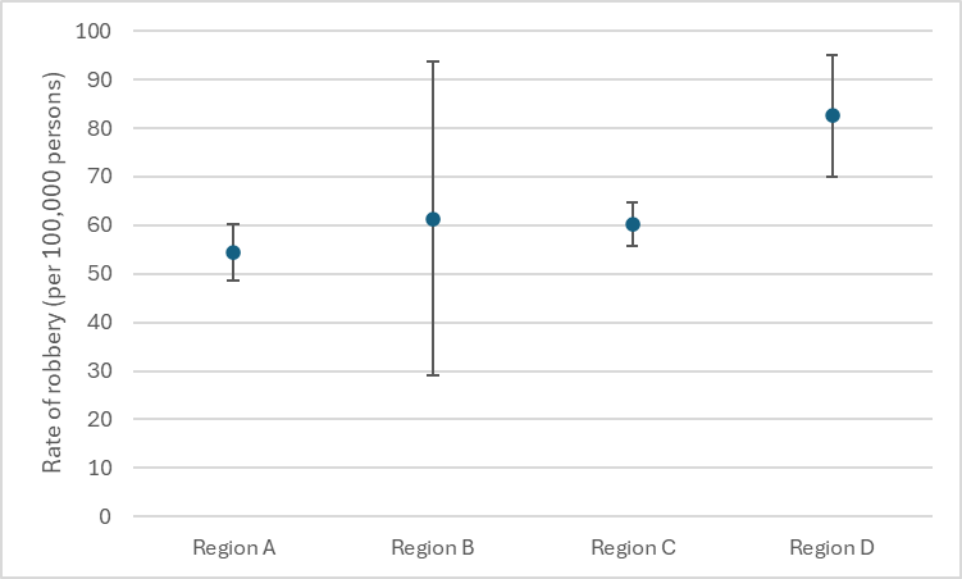
To determine if the robbery rate is statistically different in Region A and Region B, the upper bound from the smaller estimate (Region A) is compared to the lower bound of the larger estimate (Region B). Because 60.1 is not less than 29.1, the robbery rates in Region A and Region B are *not* statistically different from each other.

The confidence interval overlap method can also be conducted visually. ***Exhibit 3-17*** presents the estimates and confidence intervals for robbery in each region. From the exhibit, Region A has overlapping confidence intervals with Region B and Region C, but its confidence interval does not overlap with the confidence interval of Region D. Therefore, the robbery rate in Region A is statistically

[16] On the annual data files, the weighted estimate is labeled as "estimate." The confidence interval bounds are labeled as "estimate_upper_bound" and "estimate_lower_bound."

different from the violent crime rate in Region D, but the robbery rate of Region A is not statistically different from the rate in Region B or Region C.

Exhibit 3-17. Robbery offense rate (per 100,000 persons) by region



3.11.4. Example of statistical test method for estimates assumed to be uncorrelated

To determine if the robbery rate in Region A is different from the robbery rate in another region, the following information needs to be generated from the NIBRS estimation process:[17]

Domain	Weighted Estimate	Root Mean Square Error (RMSE)	Bias
Region A	54.3	2.9	-1.0
Region B	61.4	16.5	-6.4
Region C	60.2	2.2	-1.0
Region D	82.6	6.4	-1.6

The test statistic to compare the rates in Region A and Region B is calculated as:

$$t = \frac{E_1 - E_2}{\sqrt{RMSE_1^2 + RMSE_2^2 - 2 * B_1 * B_2}}$$

$$t = \frac{61.4 - 54.3}{\sqrt{16.5^2 + 2.9^2 - 2 * -6.4 * -1.0}} = 0.434$$

where E_1 (weighted estimate), $RMSE_1$ (root mean square error), B_1 (bias) are for the larger estimate and E_2, $RMSE_2$, B_2 are for the smaller estimate. Because the correlation for different indicators in the same year was not estimated, the estimates (E_1 and E_2) are assumed to be uncorrelated (and in this example, that is a good assumption because there is no overlap in agencies between Regions A and B).

[17] On the annual data files, the weighted estimate is labeled as "estimate." The RMSE is labeled as "estimate_rmse." The bias is labeled as "estimate_bias."

The corresponding p-value for the t-test statistic of 0.434 is 0.664 when the sample size in each domain is relatively large (larger than 30). Because 0.664 is greater than 0.05 the robbery rates in Region A and Region B are not statistically different from each other.

A second test can be conducted to compare the rates for Region A and Region D. The test statistic for this comparison is calculated as:

$$t = \frac{E_1 - E_2}{\sqrt{RMSE_1^2 + RMSE_2^2 - 2 * B_1 * B_2}}$$

$$t = \frac{82.6 - 54.3}{\sqrt{6.4^2 + 2.9^2 - 2 * -1.6 * -1.0}} = 4.165$$

The corresponding p-value for the t-test statistic of 4.165 is 0.000. Because 0.000 is less than 0.05, the robbery rate in Region A is statistically different (smaller) than the robbery rate in Region D.

3.11.5. Example of statistical test method for correlated estimates

The statistical test method can also be used to compare the same indicator across consecutive years. For example, to compare the national violent crime offense rate between the current year and the prior year, the following information is generated from the NIBRS estimation process:[18]

Domain	Weighted Estimate	Standard Error (SE)	Root Mean Square Error (RMSE)	Bias	Estimated Correlation with Prior Year
Current Year	387.8	3.2	5.0	-3.8	0.846
Prior Year	407.3	1.8	4.4	4.0	--

Because the same indicator is being compared across consecutive years, the correlation between estimates can be incorporated into the test statistic. The test statistic to compare the rates across years is calculated as:

$$t = \frac{E_1 - E_2}{\sqrt{RMSE_1^2 + RMSE_2^2 - 2 * B_1 * B_2 - 2 * CORR * SE_1 * SE_2}}$$

$$t = \frac{407.3 - 387.8}{\sqrt{4.4^2 + 5.0^2 - 2 * 4.0 * -3.8 - 2 * 0.846 * 1.8 * 3.2}} = 2.418$$

where E_1 (weighted estimate), $RMSE_1$ (root mean square error), B_1 (bias), and SE_1 (standard error) are for the larger estimate; E_2, $RMSE_2$, B_2, and SE_2 are for the smaller estimate; and CORR is the estimated correlation between E_1 and E_2.

The corresponding p-value for the t-test statistic of 2.418 is 0.016. Because 0.016 is less than 0.05, the violent crime offense rate in the current year is statistically different (smaller) than the violent crime offense rate in the prior year.

[18] On the annual data files, the weighted estimate is labeled as "estimate." The standard error is labeled as "estimate_standard_error." The RMSE is labeled as "estimate_rmse." The bias is labeled as "estimate_bias." The values in the last two columns are only available on the most recent year's data file. The proportion of eligible ORIs reporting one or more incidents is labeled as "PROP_ELIG_ORIS_NONZERO_COUNT." The estimated correlation is labeled as "CORRELATION_WITH_PRIOR_YEAR."

3.12. Estimation Procedures for Summary-Level Estimates

The NIBRS Estimation Program (EP) also produces summary-level estimates (i.e., NIBRS EP Summary Estimates) to examine trends in crime that include data before 2021. The methodology for creating NIBRS EP Summary Estimates is similar to the methodology for creating incident-level estimates (i.e., NIBRS Estimates) described earlier in this chapter, with additional steps to aggregate NIBRS incident data to the SRS format and to incorporate data from agencies that only report data to the Summary Reporting System (SRS-only). SRS-only agencies have not transitioned to NIBRS and submit their crime data only in summary format. The process for calculating the NIBRS EP Summary Estimates is as follows:

1. Area-specific weights (e.g., national, regional, state) are created to account for agencies that provide no data or 1–2 months of data, as described in ***Section 3.3***.
2. NIBRS reporting agencies that report 3–11 months of data have the non-reported months imputed using block imputation, as described in ***Section 3.4***.
3. The data for NIBRS reporting agencies are converted to SRS (i.e., SRS-conversion agencies) using the hierarchy rule.
4. SRS-only reporting agencies that report 3–11 months of data have the non-reported months imputed using block imputation, as described in ***Section 3.4***.
5. The population served by agencies (SRS-conversion plus SRS-only) is estimated, as described in ***Section 3.5***.
6. Uncertainty is estimated, as described in ***Section 3.6***.
7. Confidence intervals are constructed, as described in ***Section 3.7***.
8. Statistical suppression rules are applied, as described in ***Section 3.8***.
9. Output datasets are formatted, as described in ***Section 3.9***.
10. NIBRS EP Summary Estimates are interpreted, as described in ***Section 3.10***.
11. Statistical testing can be performed on NIBRS EP Summary Estimates, as described in ***Section 3.11***.

The main differences between the methodology for NIBRS Estimates and NIBRS EP Summary Estimates are the conversion of NIBRS data to SRS data (Step 3) and the block imputation of the SRS-only reporting agencies (Step 4). All other steps follow the same process as the NIBRS Estimates, with the inclusion of the SRS-only reporting agencies.

4. NIBRS Extract Files Creation Methodology

This chapter provides background and technical details for the annual National Incident-Based Reporting System (NIBRS) Data Extract Files produced by the Bureau of Justice Statistics (BJS). These files are made available to researchers and other stakeholders by BJS through the National Archive of Criminal Justice Data (NACJD), part of the Inter-university Consortium for Political and Social Research (ICPSR). The chapter gives background information on the:

1. NIBRS data structure and the need for analytic files
2. Purpose and structure of the NIBRS Master File and NIBRS Data Extract Files
3. Process for creating the NIBRS Master File
4. Process for creating the NIBRS Data Extract Files
5. Description of a unique, BJS-produced victim-level extract file that enables detailed analysis of victimizations involving a firearm
6. Use cases demonstrating which NIBRS data extract file could be used to answer different research questions.

The goal of providing this detail on the NIBRS Data Extract Files is to enable NIBRS data users and researchers to understand how these files can be used for analysis and to provide transparency for how these files are created.

4.1. Background and File Overview

BJS has produced annual NIBRS Data Extract Files since 1991 to facilitate wider use of the NIBRS data and to make them more accessible for analysis. Between 1991 and 2016, NACJD produced the files on behalf of BJS, and the files were made available through the ICPSR website.[19] RTI International, through the NCS-X Initiative (Cooperative Agreement 2017-BJ-CX-K054), recreated and validated the extract file structure, data elements, and data values used by NACJD and then produced the 2017–2022 NIBRS Data Extract Files. This approach was used to ensure continuity between the NIBRS Data Extract Files over time. Extract files for 2023 were produced by RTI through the NIBRS National Estimation Project (Cooperative Agreement 15PBJS-22-GK-00713-BJSB). The 2017–2023 extract files are also available on the ICPSR website through the National Incident-Based Reporting System Series at ***https://www.icpsr.umich.edu/web/ICPSR/series/128***. The code that RTI developed and used to create the 2017–2023 NIBRS Data Extract Files is available at the following GitHub repository, ***https://github.com/RTIInternational/extracts_icpsr_nibrs***; BJS is making this code available to promote transparency and future improvements.

The NIBRS data structure is complex given the substantial number of required and conditional NIBRS data elements, the multiple instances of the NIBRS segments depending on the crime incident, and how the data elements and segments are joined to form a NIBRS incident record. Properly conducting analyses using the raw NIBRS data requires a detailed understanding of how data are structured and linked if you were to create an analytic file. The NIBRS Data Extract Files are designed to enable various analytic levels and perspectives to capitalize on the rich information contained in the NIBRS data, provide standardized and validated NIBRS analytic files, and create efficiency by not requiring each user to create their own analytic files. ***Exhibit 4-1*** provides an overview of the NIBRS Data Extract Files.

[19] See ***https://www.icpsr.umich.edu/web/ICPSR/series/128*** and ***https://www.icpsr.umich.edu/web/pages/NACJD/NIBRS/***.

Exhibit 4-1. Overview of BJS NIBRS Data Extract Files

Extract file	Description
Incident	Each row represents a single crime incident with summarized information about any victims, offenders, offenses, and arrests associated with the incident.
Victim	Each row represents a single victimization with summarized information about the incident and any offenses, offenders, and arrests associated with the victimization.
Arrestee	Each row represents a single arrestee with summarized information about the incident and any offenses, offenders, and victims associated with the arrestee.
Offender	Each row represents a single offender with summarized information about the incident and any offenses, victims, and arrests associated with the offender.
Batch Header	Contains basic information about each NIBRS incident such as ORI, incident number, jurisdiction population, and county where the jurisdiction is located.
Admin	Contains administrative data about each NIBRS incident such as counts of the different data segments in the incident, incident clearance status, and cargo theft.

4.2. NIBRS Master File Methodology

The NIBRS Master File serves as the input file to create the NIBRS Data Extract Files and consists of all NIBRS segment files for the given data year appended together into a single text file. The FBI Criminal Justice Information Services (CJIS) Division releases a NIBRS Master File for each data year when they release the NIBRS data, which usually occurs in September for the previous year's data. ***Section 4.2.3*** describes how the NIBRS Master File can be downloaded from the FBI's Crime Data Explorer (CDE).

4.2.1. Steps for generating the NIBRS Master File

For the 2023 data year, RTI was able to produce the NIBRS Master File independently using code contained in the Uniform Crime Reporting (UCR) Production Database that is received from the FBI for the NIBRS National Estimation Project. By creating the NIBRS Master File using this code, it enabled the NIBRS Data Extract Files to be created earlier in the year, validated, and available for release through ICPSR in a timely manner after the FBI releases data in September. This ensures that the NIBRS Extract Files are available for analysis as soon as possible.

The steps for creating the NIBRS Master File are in ***Sections 4.2.2*** and ***4.2.3*** and use the UCR Production Database. Please note that there are two different nibrs_master functions in the UCR Production Database code. One function takes an array of agency ids (ORIs) to generate the master file for specific agencies. The other function includes optional state abbreviation to generate either the entire master file (RTI used 48 GB of RAM and 144 GB of disk space) or the master file for a specific state. Review the function parameters to determine what you want to run.

4.2.2. Running the NIBRS Master File extract process

1. Connect to the database with PGAdmin or another database query tool.
2. Modify logger__ins_log_atx in Schemas/ucr_prd/procedures to return without doing anything: as the database as shipped from the FBI does not have the linked database(s) required for this logging procedure to run.

3. Set your search path so the procedures and functions used may be called correctly:

```
set search_path = ucr_prd, public;
```

4. Set CSV quoting to None in your output options.
    a. Preferences/CSV/Text output.
5. Modify the 'replace null values' output setting with empty/nothing.
6. Run the desired nibrs_master function with the appropriate parameter settings for your run:

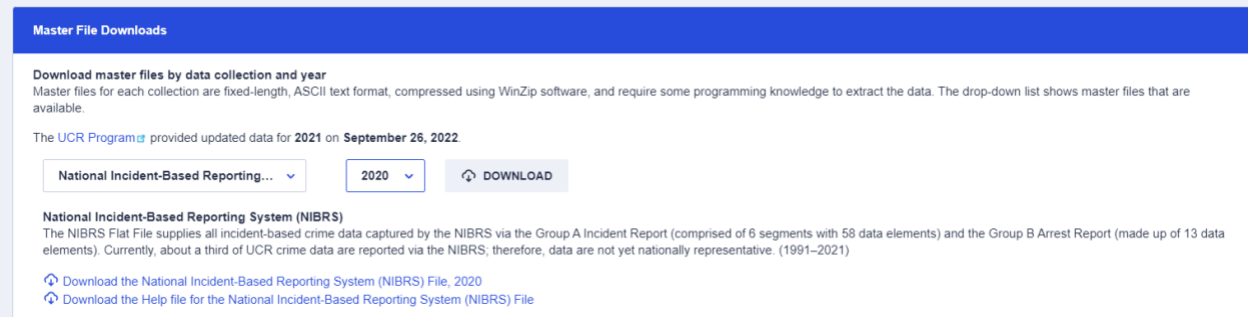
```
SELECT master_nibrs(
    2022,
    FALSE,
    'NC',
    ''
)
```

    a. Both nibrs_master functions are in: Schemas/ucr_prd/Functions.
    b. The function parameters should be reviewed to determine which function you want to run; the example above generates the master file for North Carolina. Replacing 'NC' with '' would generate the entire master file for 2022 and would NOT encode the incident ID.
7. Remove the first line / header line in the file manually or via a command line utility like sed:

```
(base) → projects sed -i '.bak' '1d' ~/Downloads/data-1711474527414.csv
```

4.2.3. Steps for downloading the NIBRS Master File

1. Download the target year NIBRS master files. The remainder of these steps will use 2020 as an example. The files are available here: ***https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/downloads*** (see screenshot). The downloaded file will be a zip file of several hundred megabytes in size and contains a text file consisting of all segments combined into a single file. The master file is a fixed-width file and documentation about the format of the file is contained in the docs folder of the source code repository. The format mostly follows the documented format but does differ slightly due to changes over the year and have not yet been updated in the documentation provided by the FBI.



2. Store the zip file in a local folder where you intend to point the extract code to.
    a. For example, the downloaded zip file is moved to a **data/raw_data/nibrs_from_fbi** folder under the **icpsr_nibrs_extracts** project directory.
    b. Unzip the compressed file using the unzip command and it results in exporting the **2020_NIBRS_NATIONAL_MASTER_FILE_ENC.txt** text file.

c. Change directories to the main project folder: **icpsr_nibrs_extracts.**

4.3. NIBRS Data Extract File Methodology

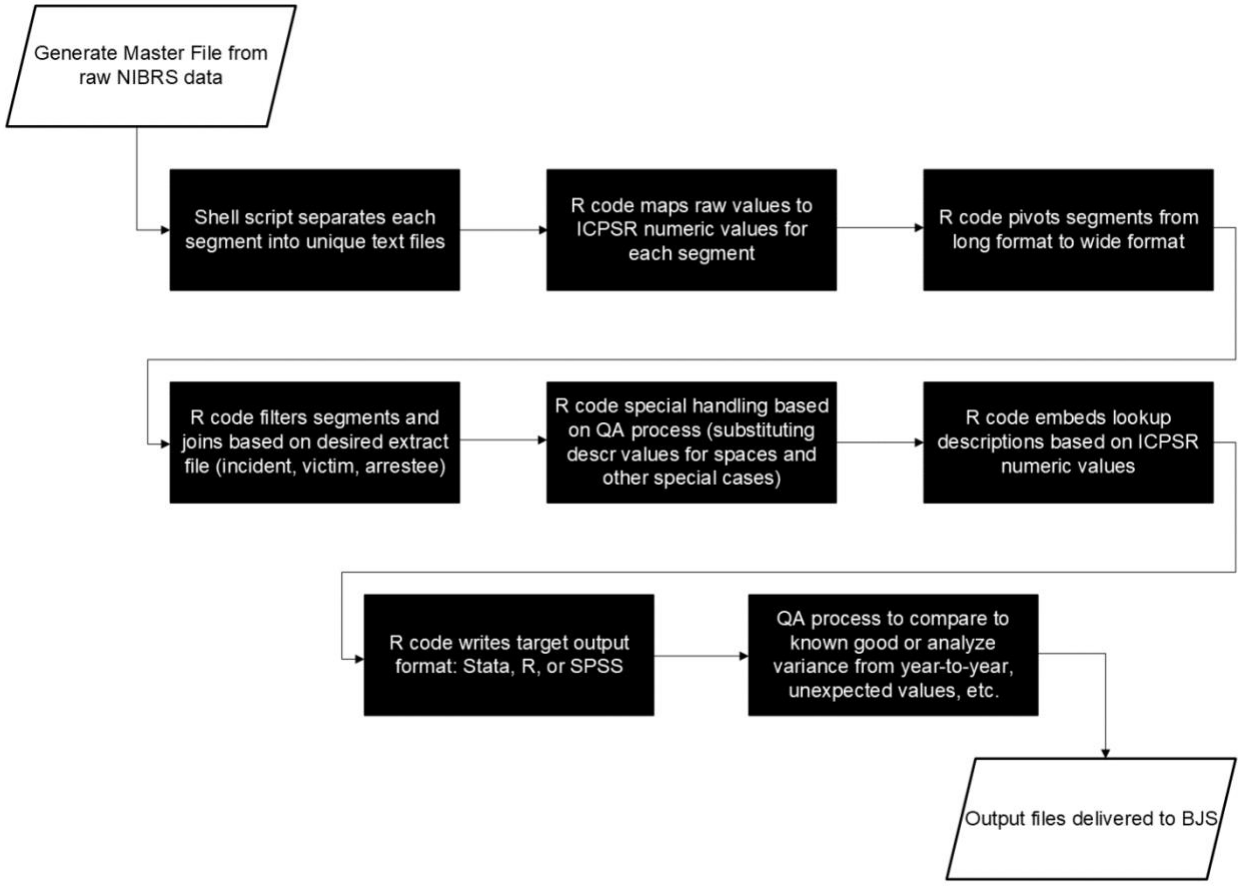
4.3.1. Process for creating the NIBRS Extract Files

In this section, the process for creating the NIBRS Data Extract Files from the NIBRS Master File for a given year is described in detail. This process uses the code developed by RTI, which is available in the above referenced GitHub repository, to create the NIBRS Data Extract Files. ***Exhibit 4-2*** provides a high-level overview of the steps in this process. The software and system requirements to create the NIBRS Data File Extracts are as follows:

- R 4.2.1 and greater
- R Studio (latest)
- A system that supports running bash shell scripts,[20] which is supported on:
    - Windows 11 +
    - Mac OS
    - Linux
- At least 32 GB of Ram
- 200 GB of available drive space.

[20] Bash scripts are text files containing a series of commands that can be executed by the Bash shell, used for automating tasks in Linux and macOS environments.

Exhibit 4-2. NIBRS extract file creation process



Step 1. Generate Raw Segment Data Files

1. Run the generate_raw_segments.sh script.

    a. From a bash prompt within the **icpsr_nibrs_extract** folder, run the **generate_raw_segments.sh** script, which provides the relative path to the data and a data year that will be used to generate another folder that stores the individual segments parsed from the master file.

```
./generate_raw_segments.sh data/raw_data/nibrs_from_fbi/2020_NIBRS_National_Master_File_ENC.text 2020
```

    b. Once the script finishes, it will have generated individual text files containing each segment that may be used by the R code. The 2020 example would generate the following segment files in the **data/raw_data/2020** folder:

        i. admin_segment.txt
        ii. arrestee_segment.txt
        iii. arrestee_segment_groupb.txt
        iv. arrestee_segment_window.txt
        v. batch_header_segment.txt
        vi. offender_segment.txt
        vii. offense_segment.txt
        viii. property_segment.txt

ix. victim_segment.txt

Step 2. Generate Extract Files

1. Launch R Studio to finish generating an extract.
2. Open the raw_to_extract.R file and specify your configuration parameters.
    a. Modify the year variable to your target year (e.g., 2020).
    b. Data_source should be “master.”
    c. Do_debug should remain false unless you need additional debug output.
    d. Make sure to modify the setwd path to match your local path to the icpsr_nibrs_extracts project director.
    e. Modify the extract variable to the options listed in the comments for your target extract: victim, incident, or arrestee.
3. You may load all the segments as a block of code by highlighting all the load statements and running them.

```
# loads the named segment and sorts as-needed
segment_victim <- get_segment("victim_segment", year) %>%
  arrange(V4003, V4004, V4006) # ORI, INCNUM, SEQUENCE #

segment_offense <- get_segment("offense_segment", year)

segment_batch_header <- get_segment("batch_header_segment", year)

segment_admin <- get_segment("admin_segment", year)

segment_property <- get_segment("property_segment", year)

segment_offender <- get_segment("offender_segment", year) %>%
  arrange(V5003, V5004, V5006)

segment_arrestee <- get_segment("arrestee_segment", year) %>%
  arrange(V6003, V6004, V6006)

segment_arrestee_window <- get_segment("arrestee_segment_window", year) %>%
  arrange(V6003, V6004, V6006)

segment_arrestee_groupb <- get_segment("arrestee_segment_groupb", year) %>%
  arrange(V6003, V6004, V6006)
```

```r
# ----------------------  CONVERT TO ICPSR values ----------------------

convert_to_icpsr_values <- function(segment) {
  # uses the RTI created cross reference data to map raw FBI values to the correct ICPSR numeric values
  # converts fields from segment to ICPSR values if we have a current cross reference from the raw data to
  segment_field_names <- names(segment)
  for (field_name in segment_field_names){
    data_file <- paste("./data/lookup_tables/", field_name, "_xref.Rda", sep = "")
    if (file.exists(data_file) == T && length(na.omit(segment[[field_name]])) > 0){
      field_xref <- readRDS(data_file)
      class(segment[[field_name]]) <- class(field_xref[[field_name]])
      segment <- left_join(segment, field_xref)
      segment[[field_name]] <- segment[[paste(field_name, ".icpsr", sep = "")]]
    }
  }
  #colnames(segment) <- lapply(names(segment), str_remove, pattern = ".icpsr")
  segment %>%
    subset(select = segment_field_names) %>%
    distinct()
}

# converts raw values for each segment to ICPSR values using the convert function defined above
segment_victim <- convert_to_icpsr_values(segment_victim)
segment_offense <- convert_to_icpsr_values(segment_offense)
segment_batch_header <- convert_to_icpsr_values(segment_batch_header)
segment_admin <- convert_to_icpsr_values(segment_admin)
segment_property <- convert_to_icpsr_values(segment_property)
segment_offender <- convert_to_icpsr_values(segment_offender)
segment_arrestee <- convert_to_icpsr_values(segment_arrestee)
segment_arrestee_groupb <- convert_to_icpsr_values(segment_arrestee_groupb)
segment_arrestee_window <- convert_to_icpsr_values(segment_arrestee_window)
```

4. After the segments are loaded and converted to extract File values (which are also the historic data values used in the NACJD-created extract files), you will want to interactively run each transformation section of code so you do not run out of memory. For example, run the segment_admin block of transformation code, then the victim_wide block of code and wait for each block of code to finish to avoid running out of memory.
5. The final blocks of code generate the final output as SPSS .sav files, which may be compressed with zip for archiving and sharing more efficiently or utilized directly in SPSS as .sav files.

4.3.2. Process for validating NIBRS Extract Files

This section details the process of validating the production of the 2017–2023 extract files and the processes that led to the published versions. The logic behind this process was to use the available ICPSR extracts to validate the structure and content of the BJS-produced extracts created by RTI. Because NACJD only produced NIBRS extracts through 2016, the validation process itself occurred in two separate stages. The first stage used a reference dataset against which the new extracts could be compared for data or structural discrepancies. The second stage built on this initial work and applies the same principles to NIBRS data from 2017 onward. Because no comparison file exists, the validation process is more inferential and based on what values are expected to be seen in the aggregate. Through an iterative process of revising extracts and assessing them for internal and external consistency, the 2017-onward NIBRS extracts accurately reflect and organize the NIBRS data.

Step 1: Obtaining Data

As referenced, NIBRS extracts only existed through 2016. In developing protocols for new extract creation, it made the most sense to use a year with a direct ICPSR counterpart file for validation. This allowed direct incident- and victim-based comparisons between the published 2016 extract files and the new 2016 RTI extract files.

- The ICPSR NIBRS extracts were obtained directly from the ICPSR website.
- The RTI extracts were created and provided by staff from the RTI Center for Data Science. This was done iteratively, whereby observations and corrections from the validation process were integrated into each new file.
- For the purposes of both validation and submission of the data to BJS, the RTI and ICPSR extracts were converted to SPSS format for the validation and publishing process.

Step 2: Evaluating Data Dimensions

The first step in the validation process was to evaluate any differences in the structure and data dimensions between the RTI extract and the comparison file. In this flat file format, the discrepancies could affect either the number of variables or the number of observations.

Variable Alignment (Column): For each extract type, the variable lists between the RTI and ICPSR extracts were compared to identify missing variables. While limited in number, the missing variables were added to future iterations. Following a one-to-one matchup between the RTI and ICPSR variable lists, the focus turned to substantive decisions about what variables should be included in the final extracts. A small number of variables present in the ICPSR extract were defunct and excluded from the RTI templates. Any discrepancy in variable inclusion present in the final extracts is intentional.

Observation Alignment (Row): Across all extracts tested against the 2016 ICPSR file, there were large discrepancies in the number of records. RTI investigated these differences by creating a unique event identifier consisting of the ORI, incident number, and the victim sequence number (in the victim extract only). Merging the ICPSR and RTI-produced extracts by matched events helped to identify cases that were unique to either file.

- Throughout the validation process, there were no instances of records in the ICPSR that were omitted from the RTI extract.
- There were many instances of records in the RTI extract that exceeded what was in the ICPSR extract. These are largely attributed to the addition of new NIBRS reporting jurisdictions between the time ICPSR and RTI captured the data for the comparison time period. Additional cases were concentrated in ORIs that were new NIBRS reporters and reflect legitimate events that were retained in the final extracts. These discrepancies were less about problems with the production code and more to do with NIBRS being a data source that can change over time with new updates from agencies.
- The remaining discrepancy was explained by duplicate records. Duplicate records were easy to identify and could be explained by quirks in the production code that were fixed by the data science team.

Step 3: Evaluating Data Values

Now that the intended variables were included and there was reasonable alignment in the number of cases, the priority shifted to the values of the data points. The unique identifiers allowed for a crosswalk between the ICPSR and RTI-produced extracts. This allowed the two to be merged, where value differences were flagged and investigated.

- Because the RTI-produced and ICPSR extracts are sourced from the same data, most variables contained no discrepancies. The most common reason for value differences were formatting and transformation issues. These were easily identifiable and fixable across the multiple iterations of extract creation.

- Another source of error in early extracts was a difference in record order. To create a parsimonious file, both the ICPSR and RTI-produced extracts left room for three incident-level summary records of victims, offenders, and arrestees. A small percentage of incidents exceeded three records for any one of these categories. For those incidents with four or more, a fourth offender or arrestee may be omitted from the extract in which that segment is not the focus. For example, an incident with four victims may not have each victim sequence number represented in the incident extract, but all are present in the victim extract. The way these fields are ordered for inclusion was not consistent between ICPSR and RTI-produced extracts, leading to apparent discrepancies. The offender in position 1 in ICPSR and position 2 in the RTI extract may be the same person. Thus, the values would not align exactly but would be present in a different order. This was determined to be a non-issue, though the extract creation process was reformed to use sequence number order where relevant.
- In the remaining instances, the value differences reflected legitimate changes or updates to the NIBRS data. This never resulted in a loss of data values, instead indicating that additional information was made available between the time of the data downloads. These value changes were confirmed to be present in the raw data, not a result of faulty translation, and were determined to be a non-issue.

Step 4: Evaluating Missing Data

Missing data were handled similarly to valid data values; if a value was present in the ICPSR extract, the new extract was examined to determine any missing information. Conversely, the RTI-produced extract had slightly fewer missing values due to additional information about victims, offenders, or arrestees becoming available and added to the NIBRS dataset. It is important to note that ICPSR did report a wider range of details about why a value may be missing, where the data source was unable to be specified. Thus, for some instances in the ICPSR extract where a value might be “Not applicable”, the same information may be a system missing value in the RTI extract.

Step 5: Data Formatting and Publishing

The production of the extracts in R and their subsequent conversion to SPSS created several superficial differences in formatting and data attributes. Each of the extract types has a companion document with SPSS syntax to define and finalize variable types, labels, and properties.

Step 6: Data Distribution Testing (Unmatched Extracts)

The validation process for the extracts beyond 2016 differed because there was no direct ICPSR comparison against which to confirm the data values. For Steps 2, 4, and 5, the process remained largely the same; step 3 differed in its use of statistical inference.

- For step 3, a process was developed to systematically examine the distribution of values reported in the NIBRS data and compare those values to the distribution in the previous year of data to identify any extreme outliers. With the addition of new NIBRS reporting jurisdictions each year, in addition to expected changes in variable distributions over time, a visual assessment of the data and any outlying or reversed results would indicate a transformation or data fidelity issue. The mean and distribution differences were assessed for scale variables (e.g., victim age), while chi-square tests were conducted for categorical variables (e.g., offense type, offender race, etc.). Outliers can then be investigated to determine if anomalies are related to the extract creation process itself (i.e., some error in the data processing) or are an artifact of the data collection.

- One additional step with the newer extracts was accounting for new data values, data element formatting, and other changes to the NIBRS data that were not present in the ICPSR extracts on which the process was based. For each year of extracts, the full range of values was assessed for unmapped values or changes in how the data were recorded. As new states began reporting and additional value options were added, the extract creation and formatting process was similarly updated.

Overall, taking an iterative approach to replicating a known and valid extract architecture and then applying the same processes to newer years of data was successful in refining the extract creation process and generating the new extracts. Despite not having a comparison file, the validation process allowed for the verification of extracts beyond 2016.

4.4. Firearm Violence Extract File

NIBRS is the nation's most comprehensive source of information on firearm-involved crime incidents, as it covers both fatal and nonfatal violent offenses involving a firearm. Other data sources, such as the National Center for Health Statistics' National Vital Statistics System and the FBI's Supplementary Homicide Reports, provide homicide data but do not capture firearm possession for other offense types, such as robbery or aggravated assault. BJS's National Crime Victimization Survey (NCVS) also collects victimization information directly from crime victims about nonfatal firearm violence for the crime types of rape or sexual assault, aggravated assault, and robbery that have either been reported or not reported to law enforcement, but the NCVS does not distinguish between the type of firearm (e.g., handgun or rifle) and does not collect information on fatal firearm violence.

NIBRS captures firearm possession for many offense types and has advanced our understanding of the frequency and circumstances in which firearms are present in violent crime incidents. NIBRS is designed to connect weapon types, including firearms, to specific offenses within an incident, but not necessarily to a specific victim in the incident if there are multiple victims and multiple offenses. In violent crime incidents involving multiple victims, for example, the firearm may not have been used against some victims in the incident. Thus, incident-level counts of firearm violence potentially provide an incomplete picture of the number and characteristics of individual victims of firearm violence when violent crime incidents involve multiple victims and offenses.

BJS, with support from RTI, developed the NIBRS Firearm Violence Extract File to provide more detail on the number and characteristics of victims of firearm violence. The Firearm Violence Extract File is a specialized subset of violent victimizations involving a firearm based on data from the BJS NIBRS Victim Extract File. The first step in constructing the Firearm Violence Extract file was to restrict the NIBRS Victim Extract File to only those offense types for which NIBRS captures the presence of a weapon.[21] The second step was to identify victimizations involving a violent offense with a firearm present and connect the firearm-related offense(s) to the specific victim(s) involved.

To create an analytic file on offenses that included a firearm, we attached data on whether a weapon was present to the victim-level offense variables found in V4007–V4016. The victim-level offense variables were used for this process (as opposed to the incident-level offense variables) so that the

[21] NIBRS allows law enforcement agencies to provide information on up to three types of weapons or force used by the offender(s) in incidents involving the offenses of murder and nonnegligent manslaughter, negligent manslaughter, justifiable homicide, kidnapping, rape (including the offenses of rape, sodomy, and sexual assault with an object), fondling, robbery, aggravated assault, simple assault, extortion/blackmail, weapon law violations, human trafficking—commercial sex acts, and human trafficking—involuntary servitude. The type of weapon used can be a firearm or some other type of weapon. For the full list of weapon types recorded in NIBRS, see the FBI's ***NIBRS User Manual***.

weapon information would attach to the particular offense(s) related to each specific victim in the victim-level file. A code-based comparison of the victim-specific offense codes and incident-level offense codes was conducted to assign weapon information to each victimization. The program begins by examining each offense code listed in the victim-specific variables V4007–V4016 in the full NIBRS Victim Extract File. It then compares the offense code values to the incident-level offense codes found in V20061–V20063 ***(Exhibit 4-3)***. If there is a match between one or more of the victim-specific offense codes (V4007–V4016) and the incident-level offense codes (V20061–V20063), the program retrieves the corresponding weapon information stored in variables V20171–V20193. The programming code then copies the weapon information into new weapons variables that are specific to that victimization (i.e., victim).

In NIBRS, each offense can be connected to up to three weapons. The programming code checks and stores the weapon type codes for the incident-level offenses (V20061–V20063). For example, a new variable, V4007_w1, is the first weapon associated with the specific offense connected to the victim in V4007, which is the UCR Offense Code 1 for each victimization record. In the example in Exhibit 4-3, which is a screenshot of the Victim Extract File, the code checks V4007 for the first victim highlighted in blue in row 492, then checks each of the three incident-level offenses (V20061–V20063) to determine whether there is a match. If it matches, the program then pivots to V20173, which is the variable that stores the first weapon for the offense listed in V20063, captures that weapon code, and stores it in a new variable named V4007_w1. This process is repeated for the total number of victim-level offenses associated with each victimization.

Exhibit 4-3. Victim-specific offense matching to select weapons

	INCNUM	UNIQUEID2	victimIDsperUniqueID	V20061	V20062	V20063	V4007	V4008	V4009	V4010	V4011	V20171	V20172	V20173	V20181	V20182
487	Z8-JDVAB…	AR00800008Z8-JDVAB-F392.02e+07	2	131	.	.	131	.	.	.	.	400	.	.	.	.
488	3UAU4V8B…	AR00802003UAU4V8BLF392.02e+07	1	131	.	.	131	.	.	.	.	130	.	.	.	.
489	3UAU4V8B…	AR00802003UAU4V8BLF392.02e+07	2	131	.	.	131	.	.	.	.	130	.	.	.	.
490	3A-MQEFC…	AR00803003A-MQEFCMD I2.02e+07	1	131	.	.	131	.	.	.	.	990	.	.	.	.
491	3A-MQEFC…	AR00803003A-MQEFCMD I2.02e+07	2	131	.	.	131	.	.	.	.	990	.	.	.	.
492	NH-U73Z4…	AR00803000NH-U73Z4UWEM2.02e+07	1	100	120	131	131	100	.	.	.	990	990	200	.	.
493	NH-U73Z4…	AR00803000NH-U73Z4UWEM2.02e+07	2	100	120	131	120	100	.	.	.	990	990	200	.	.
494	NH-U73Z4…	AR00803000NH-U73Z4UWEM2.02e+07	3	100	120	131	120	100	.	.	.	990	990	200	.	.
495	O2-NJWNI…	AR00803000O2-NJWNINUF32.02e+07	1	131	.	.	131	.	.	.	.	900	.	.	.	.

4.5. NIBRS Extract File Codebook

To support the use of the NIBRS Data Extract Files, an Excel-based codebook has been created by BJS and is available on ICPSR. The codebook contains a worksheet for each of the NIBRS Data Extract Files: Victim, Incident, Arrestee, Offense, Batch Header, and Admin Segment. There is also a Common Values worksheet that includes commonly repeated values across the extract file variables (e.g., weapon type, victim injury), which is used to reduce scrolling on each extract file worksheet for variables with the same values. Each worksheet lists the variable name, variable type, variable label, values, value labels or a reference to the Common Values worksheet, notes and clarifications, the associated NIBRS data element, and the NIBRS segment origin for each variable (i.e., where does each variable come from within the NIBRS segment structure) for the respective extract file type. A separate Excel-based codebook is also available on ICPSR for the Firearms-Specific Victim Extract File. This codebook ***(Exhibit 4-4)*** is the same as the Victim Extract File, except it includes the offense-specific weapons attached to each victim.

Exhibit 4-4. NIBRS extract file codebook screenshot

Variable Name	Variable Type	Variable Label	Value	Value Label	Notes and Clarifications	NIBRS Data Element in NIBRS User Manual
V4018	Numeric	AGE OF VICTIM	0.1	Under 24 hours (neonate)	V4018 is the numeric age of the victim except in cases where the victim was under 1 year old or over 98 years old. In these circumstances, the value and label are listed in the values and value labels.	Data Element 26 - Age of Victim
			0.2	1-6 days old		
			0.5	7-364 days old		
			99	Over 98 years old		
			-5 to -9	See Value Reference 3 in Common Values tab for missing and unknown values		
V4019	Numeric	SEX OF VICTIM	-9	Undetermined	V4019 indicates the sex of an individual (person) victim in an incident.	Data Element 27 - Sex of Victim
			-7	Unknown/Missing/DNR		
			-6	Not applicable		
			0	Female		
			1	Male		
V4020	Numeric	RACE OF VICTIM	-9	Undetermined	V4020 indicates the race of an individual (person) victim in an incident.	Data Element 28 - Race of Victim
			-7	Unknown/Missing/DNR		
			-6	Not applicable		
			1	White		
			2	Black or African American		
			3	American Indian or Alaska Native		
			4	Asian		
			5	Native Hawaiian or Other Pacific Islander		
V4021	Numeric	ETHNICITY OF VICTIM	-9	Undetermined	V4021 indicates the ethnicity of an individual (person) victim in an incident. This is an optional data element. The ethnic designation of Hispanic or Latino includes persons of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish culture or origin, regardless of race.	Data Element 29 - Ethnicity of Victim
			-7	Unknown/Missing/DNR		
			-6	Not applicable		
			0	Not Hispanic or Latino		
			1	Hispanic or Latino		

4.6. Use Case Scenarios

Given the hierarchical structure of the NIBRS Data Extract Files (e.g., one crime incident can have multiple victims/offenders/arrestees), the complexity of the data can make it challenging to use for researchers interested in studying crime and crime patterns. The data can have multiple units of analysis; a row in the dataset can represent, for example, an incident, victim, offender, or arrestee. The purpose of this section is to provide three sample research questions and then demonstrate which NIBRS Data Extract File could be used to answer the specific question. Also included are example characteristics or variables that can be used from each dataset.

Scenario 1.

Research Questions: A researcher is interested in studying whether same-sex couples experience disparate arrest outcomes compared to opposite-sex couples during intimate partner violence victimizations.

Approach: In this example, the primary focus of the research question is on *victimizations*. Therefore, the victim-level extract could be used to study this topic. In the victim-level extract file, each row in the dataset represents one victim within a crime incident, though it is important to note that there may be multiple victims within the same incident.

Relevant Variables: The variable "RECSVIC" can be used to determine the number of victim records per ORI incident number. Although the primary interest is on the arrest of offenders, arrest information is available in the Victim Extract File, including all the demographic characteristics (age, race, and sex) of the arrestee (variables V60141–V60163), type of arrest (variables V60091–V60093), and the offense for which the individual is arrested (variables V60111–V60113), which may be different from the originating offense. Additional variables of interest would include the demographic characteristics (age, sex, and race) for the victim (V4018–V4020) and offender (V50071–V50093) as well as the victim/offender relationship (variables V4032–V4050). The victim/offender relationship has multiple categories that can be used to define "intimate partner," including victim was spouse, common-law spouse, offender (in the case of mutual combatants), boyfriend/girlfriend, ex-spouse, and ex-relationship. The offense variables (V4007–V4016) can be subset to, for example, murder, kidnapping/abduction, rape, sodomy, robbery, aggravated assault, simple assault, intimidation, and extortion/blackmail.

Scenario 2.

Research Question: A researcher is interested in examining the differences in the incident characteristics of female- versus male-perpetrated sexual violence.

Approach: In this example, the goal is to understand the differences between male and female offenders of sexual violence. In the *offender-level extract,* each row represents a single offender within a crime incident, with summarized information about the incident's offenses, victims, and arrests.

Relevant Variables: Like the victim-level file, in the offender extract there can be multiple offenders per crime incident, which can be identified with the variable "RECSOFR" in the extract file. The primary interest is to identify the demographic characteristics (age, sex, and race) of the offender (variables V5007–V5009) as well as whether the offense type was a sexual victimization (variables V20061–V20063). Sexual victimizations can include rape, sodomy, sexual assault with an object, fondling (indecent liberties/child molesting), incest, statutory rape, assisting or promoting

prostitution, and human trafficking—commercial sex acts. These can be used to compare differences in the incident characteristics, such as whether a weapon (e.g., firearm, knife, blunt object, or explosives) was involved (variables V20171–V20193), a victim sustained an injury (variables V40261–V40303) and the severity of that injury (e.g., apparent minor injury or other major injury), and whether an arrest occurred (variables V60091–V60093).

Scenario 3.

Research Question: For homicides involving multiple offenders, what are the incident characteristics that are associated with the presence of a female offender, and how do those characteristics differ from multiple or single male-only homicides?

Approach: In this example, the primary goal is to understand multiple-offender incidents where a homicide occurred. The incident-level extract file could be used to assess this research question. In the incident extract, each row represents a single incident with summarized information about victims, offenders, offenses, and arrests. Unlike the victim or offender extract files, each row is a unique crime incident where there can be multiple columns for each victim and each offender.

Relevant Variables: Multiple-offender incidents can be identified with the variable "RECSOFR" in the incident file. The demographics of the offenders can be identified using the variables V50071–V50093. Incident characteristics of interest could include the relationship type (e.g., victim was spouse, girlfriend/boyfriend, friend, acquaintance) between the offender(s) and victim (variables V40321–V40503), whether a weapon (e.g., firearm, knife, or blunt object) was present (variables V20171–V20193), and the circumstance of the homicide (variables V40231–V40243). For aggravated assault and homicide, NIBRS has variables that identify the circumstances of the offense(s). These include, for example, argument, drug dealing, gangland, lovers' quarrel, and mercy killing.

Appendix A: NIBRS Database Indicator Definitions

Table A-1. Victim and arrestee gender

Category	Row label	variable_for_link	variable_code_1	der_code
Victim sex	Male	der_victim_gender	1	victim_sex_code == "M"
Victim sex	Female	der_victim_gender	2	victim_sex_code == "F"
Victim sex	Unknown	der_victim_gender	3	victim_sex_code == "U", victim_sex_code == "X", victim_sex_code = Missing
Sex-specific victimization rate	Male	der_victim_gender	1	victim_sex_code == "M"
Sex-specific victimization rate	Female	der_victim_gender	2	victim_sex_code == "F"
Sex-specific victimization rate	Unknown	der_victim_gender	3	victim_sex_code == "U", victim_sex_code == "X", victim_sex_code = Missing
Arrestee sex	Male	der_arrestee_gender	1	arrestee_sex_code == "M"
Arrestee sex	Female	der_arrestee_gender	2	arrestee_sex_code == "F"
Sex-specific arrest rate	Male	der_arrestee_gender	1	arrestee_sex_code == "M"
Sex-specific arrest rate	Female	der_arrestee_gender	2	arrestee_sex_code == "F"

Table A-2. Victim and arrestee race

Category	Row label	variable_for_link	variable_code_1	der_code
Victim race	White	der_victim_race	1	victim_race_code == "W"
Victim race	Black	der_victim_race	2	victim_race_code == "B"
Victim race	American Indian or Alaska Native	der_victim_race	3	victim_race_code == "I"
Victim race	Asian	der_victim_race	4	victim_race_code == "A", victim_race_code == "AP", victim_race_code == "C", victim_race_code == "J"
Victim race	Native Hawaiian or Other Pacific Islander	der_victim_race	5	victim_race_code == "P"
Victim race	Unknown	der_victim_race	6	victim_race_code == "U", victim_race_code == "M", victim_race_code == "NS", victim_race_code == "O"
Race-specific victimization rate	White	der_victim_race	1	victim_race_code == "W"
Race-specific victimization rate	Black	der_victim_race	2	victim_race_code == "B"
Race-specific victimization rate	American Indian or Alaska Native	der_victim_race	3	victim_race_code == "I"
Race-specific victimization rate	Asian	der_victim_race	4	victim_race_code == "A", victim_race_code == "AP", victim_race_code == "C", victim_race_code == "J"
Race-specific victimization rate	Native Hawaiian or Other Pacific Islander	der_victim_race	5	victim_race_code == "P"

Category	Row label	variable_for_link	variable_code_1	der_code
Race-specific victimization rate	Unknown	der_victim_race	6	victim_race_code == "U", victim_race_code == "M", victim_race_code == "NS", victim_race_code == "O"
Arrestee race	White	der_arrestee_race	1	race_code_arrestee == "W"
Arrestee race	Black	der_arrestee_race	2	race_code_arrestee == "B"
Arrestee race	American Indian or Alaska Native	der_arrestee_race	3	race_code_arrestee == "I"
Arrestee race	Asian	der_arrestee_race	4	race_code_arrestee == "A", race_code_arrestee == "AP", race_code_arrestee == "C", race_code_arrestee == "J"
Arrestee race	Native Hawaiian or Other Pacific Islander	der_arrestee_race	5	race_code_arrestee == "P"
Arrestee race	Unknown	der_arrestee_race	6	race_code_arrestee == "U", race_code_arrestee == "M", race_code_arrestee == "NS", race_code_arrestee == "O"
Race-specific arrest rate	White	der_arrestee_race	1	race_code_arrestee == "W"
Race-specific arrest rate	Black	der_arrestee_race	2	race_code_arrestee == "B"
Race-specific arrest rate	American Indian or Alaska Native	der_arrestee_race	3	race_code_arrestee == "I"

Category	Row label	variable_for_link	variable_code_1	der_code
Race-specific arrest rate	Asian	der_arrestee_race	4	race_code_arrestee == "A", race_code_arrestee == "AP", race_code_arrestee == "C", race_code_arrestee == "J"
Race-specific arrest rate	Native Hawaiian or Other Pacific Islander	der_arrestee_race	5	race_code_arrestee == "P"
Race-specific arrest rate	Unknown	der_arrestee_race	6	race_code_arrestee == "U", race_code_arrestee == "M", race_code_arrestee == "NS", race_code_arrestee == "O"

Table A-3. Age category

Category	Row label	variable_for_link	variable_code_1	der_code
Victim age	Under 5	der_victim_age_cat_15_17	1	age_code == "NN", age_code == "NB", age_code == "BB", age_code == "01", age_code == "02", age_code == "03", age_code == "04"
Victim age	5–14	der_victim_age_cat_15_17	2	age_code == "05", age_code == "06", age_code == "07", age_code == "08", age_code == "09", age_code == "10", age_code == "11", age_code == "12", age_code == "13", age_code == "14"
Victim age	15	der_victim_age_cat_15_17	3	age_code == "15"
Victim age	16	der_victim_age_cat_15_17	4	age_code == "16"
Victim age	17	der_victim_age_cat_15_17	5	age_code == "17"
Victim age	18–24	der_victim_age_cat_15_17	6	age_code == "18", age_code == "19", age_code == "20", age_code == "21", age_code == "22", age_code == "23", age_code == "24"
Victim age	25–34	der_victim_age_cat_15_17	7	age_code == "25", age_code == "26", age_code == "27", age_code == "28", age_code == "29", age_code == "30", age_code == "31", age_code == "32", age_code == "33", age_code == "34"

Category	Row label	variable_for_link	variable_code_1	der_code
Victim age	35–64	der_victim_age_cat_15_17	8	age_code == "35", age_code == "36", age_code == "37", age_code == "38", age_code == "39", age_code == "40", age_code == "41", age_code == "42", age_code == "43", age_code == "44", age_code == "45", age_code == "46", age_code == "47", age_code == "48", age_code == "49", age_code == "50", age_code == "51", age_code == "52", age_code == "53", age_code == "54", age_code == "55", age_code == "56", age_code == "57", age_code == "58", age_code == "59", age_code == "60", age_code == "61", age_code == "62", age_code == "63", age_code == "64"

Category	Row label	variable_for_link	variable_code_1	der_code
Victim age	65+	der_victim_age_cat_15_17	9	age_code == "65", age_code == "66", age_code == "67", age_code == "68", age_code == "69", age_code == "70", age_code == "71", age_code == "72", age_code == "73", age_code == "74", age_code == "75", age_code == "76", age_code == "77", age_code == "78", age_code == "79", age_code == "80", age_code == "81", age_code == "82", age_code == "83", age_code == "84", age_code == "85", age_code == "86", age_code == "87", age_code == "88", age_code == "89", age_code == "90", age_code == "91", age_code == "92", age_code == "93", age_code == "94", age_code == "95", age_code == "96", age_code == "97", age_code == "98", age_code == "99"
Victim age	Unknown	der_victim_age_cat_15_17	10	age_code == "NS", age_code == "00"

Category	Row label	variable_for_link	variable_code_1	der_code
Victim age 2	Under 12	der_victim_age_cat_under18_2	1	age_code == "NN", age_code == "NB", age_code == "BB", age_code == "01", age_code == "02", age_code == "03", age_code == "04", age_code == "05", age_code == "06", age_code == "07", age_code == "08", age_code == "09", age_code == "10", age_code == "11"
Victim age 2	12–14	der_victim_age_cat_12_17_cat	1	age_code == "12", age_code == "13", age_code == "14"
Victim age 2	15–17	der_victim_age_cat_12_17_cat	2	age_code == "15", age_code == "16", age_code == "17"

Victim age 2	18+	der_victim_age_cat_2_uo18	2	age_code == "18", age_code == "19", age_code == "20", age_code == "21", age_code == "22", age_code == "23", age_code == "24", age_code == "25", age_code == "26", age_code == "27", age_code == "28", age_code == "29", age_code == "30", age_code == "31", age_code == "32", age_code == "33", age_code == "34", age_code == "35", age_code == "36", age_code == "37", age_code == "38", age_code == "39", age_code == "40", age_code == "41", age_code == "42", age_code == "43", age_code == "44", age_code == "45", age_code == "46", age_code == "47", age_code == "48", age_code == "49", age_code == "50", age_code == "51", age_code == "52", age_code == "53", age_code == "54", age_code == "55", age_code == "56", age_code == "57", age_code == "58", age_code == "59", age_code == "60", age_code == "61", age_code == "62", age_code == "63", age_code == "64", age_code == "65",

Category	Row label	variable_for_link	variable_code_1	der_code
				age_code == "66", age_code == "67", age_code == "68", age_code == "69", age_code == "70", age_code == "71", age_code == "72", age_code == "73", age_code == "74", age_code == "75", age_code == "76", age_code == "77", age_code == "78", age_code == "79", age_code == "80", age_code == "81", age_code == "82", age_code == "83", age_code == "84", age_code == "85", age_code == "86", age_code == "87", age_code == "88", age_code == "89", age_code == "90", age_code == "91", age_code == "92", age_code == "93", age_code == "94", age_code == "95", age_code == "96", age_code == "97", age_code == "98", age_code == "99"
Victim age 2	Unknown	der_victim_age_cat_2_uo18	3	age_code == "NS", age_code == "00"
Victim age 2	12–17	der_victim_age_cat_under18_2	2	age_code == "12", age_code == "13", age_code == "14", age_code == "15", age_code == "16", age_code == "17"

Category	Row label	variable_for_link	variable_code_1	der_code
Victim age 2	Under 18	der_victim_age_cat_2_uo18	1	age_code == "NN", age_code == "NB", age_code == "BB", age_code == "01", age_code == "02", age_code == "03", age_code == "04" age_code == "05", age_code == "06", age_code == "07", age_code == "08", age_code == "09", age_code == "10", age_code == "11", age_code == "12", age_code == "13", age_code == "14" age_code == "15" age_code == "16" age_code == "17"

Table A-4. Injury

Category	Row label	variable_for_link	variable_code_1	der_code
Injury	No	der_injury_no_yes	1	injury_code == "N"
Injury	Yes	der_injury_no_yes	2	injury_code == "B", injury_code == "I", injury_code == "L", injury_code == "M", injury_code == "O", injury_code == "T", injury_code == "U", If the offense of Murder and Nonnegligent Manslaughter (09A) is linked to the victim's record, then injury is automatically coded as yes.
Injury hierarchy	Murder and nonnegligent manslaughter, negligent manslaughter	der_injury_hierarchy	1	If the offense of Murder and Nonnegligent Manslaughter (09A) or Negligent Manslaughter (09B) is linked to the victim's record.
Injury hierarchy	Major injury (other major injury, severe laceration, possible internal injury)	der_injury_hierarchy	2	injury_code == "O", injury_code == "L", injury_code == "I"
Injury hierarchy	Unconsciousness, apparent broken bones, loss of teeth	der_injury_hierarchy	3	injury_code == "U", injury_code == "B", injury_code == "T"
Injury hierarchy	Apparent minor injury	der_injury_hierarchy	4	injury_code == "M"
Injury hierarchy	No injury	der_injury_hierarchy	5	injury_code == "N"

Category	Row label	variable_for_link	variable_code_1	der_code
Injury hierarchy 2	Murder and nonnegligent manslaughter, negligent manslaughter	der_injury_hierarchy2	1	If the offense of Murder and Nonnegligent Manslaughter (09A) or Negligent Manslaughter (09B) is linked to the victim's record.
Injury hierarchy 2	Other major injury	der_injury_hierarchy2	2	injury_code == "O"
Injury hierarchy 2	Severe laceration, possible internal injury	der_injury_hierarchy2	3	injury_code == "L", injury_code == "I"
Injury hierarchy 2	Unconsciousness, apparent broken bones, loss of teeth	der_injury_hierarchy2	4	injury_code == "U", injury_code == "B", injury_code == "T"
Injury hierarchy 2	Apparent minor injury	der_injury_hierarchy2	5	injury_code == "M"
Injury hierarchy 2	No injury	der_injury_hierarchy2	6	injury_code == "N"

Table A-5. Weapon involved

Category	Row label	variable_for_link	variable_code_1	der_code
Weapon involved	Personal weapons	der_raw_weapon_recode	1	weapon_code == "40", weapon_code == "85", weapon_code == "55", weapon_code == "75", weapon_code == "80"
Weapon involved	Firearms	der_raw_weapon_recode	2	weapon_code == "11", weapon_code == "12", weapon_code == "13", weapon_code == "14", weapon_code == "15", weapon_code == "11A", weapon_code == "12A", weapon_code == "13A", weapon_code == "14A", weapon_code == "15A"
Weapon involved	Knives and other cutting instruments	der_raw_weapon_recode	3	weapon_code == "16", weapon_code == "20"
Weapon involved	Blunt instruments	der_raw_weapon_recode	4	weapon_code == "17", weapon_code == "30"
Weapon involved	Other non-personal weapons	der_raw_weapon_recode	5	weapon_code == "35", weapon_code == "50", weapon_code == "60", weapon_code == "65", weapon_code == "70", weapon_code == "90"
Weapon involved	Unknown	der_raw_weapon_recode	6	weapon_code == "95"
Weapon involved - Yes 2	Firearms or explosives	der_weapon_yes_cat2	1	weapon_code == "11", weapon_code == "12", weapon_code == "13", weapon_code == "14", weapon_code == "15", weapon_code == "60", weapon_code == "11A", weapon_code == "12A", weapon_code == "13A", weapon_code == "14A", weapon_code == "15A"

Category	Row label	variable_for_link	variable_code_1	der_code
Weapon involved - Yes 2	Firearms	der_weapon_subset_firearm	1	weapon_code == "11", weapon_code == "12", weapon_code == "13", weapon_code == "14", weapon_code == "15", weapon_code == "11A", weapon_code == "12A", weapon_code == "13A", weapon_code == "14A", weapon_code == "15A"
Weapon involved - Yes 2	Another weapon other than firearms or explosives	der_weapon_yes_cat2	2	weapon_code == "16", weapon_code == "17", weapon_code == "20", weapon_code == "30", weapon_code == "35", weapon_code == "40", weapon_code == "50", weapon_code == "65", weapon_code == "70", weapon_code == "85", weapon_code == "90", weapon_code == "55", weapon_code == "75", weapon_code == "80"
Weapon involved - Yes 2	Knives and other cutting instruments	der_weapon_subset_knives	1	weapon_code == "16", weapon_code == "20"
Weapon involved - Yes 2	Unknown	der_weapon_yes_cat2	3	weapon_code == "95"
Weapon involved - Yes 3	Personal weapons	der_raw_weapon_recode_4_level	1	weapon_code == "40", weapon_code == "85", weapon_code == "55", weapon_code == "75", weapon_code == "80"

Category	Row label	variable_for_link	variable_code_1	der_code
Weapon involved - Yes 3	Firearms	der_raw_weapon_recode_4_level	2	weapon_code == "11", weapon_code == "12", weapon_code == "13", weapon_code == "14", weapon_code == "15", weapon_code == "11A", weapon_code == "12A", weapon_code == "13A", weapon_code == "14A", weapon_code == "15A"
Weapon involved - Yes 3	Other non-personal	der_raw_weapon_recode_4_level	3	weapon_code == "16", weapon_code == "20", weapon_code == "17", weapon_code == "30", weapon_code == "35", weapon_code == "50", weapon_code == "60", weapon_code == "65", weapon_code == "70", weapon_code == "90"
Weapon involved - Yes 3	Unknown	der_raw_weapon_recode_4_level	4	weapon_code == "95"
Weapon involved	No	der_raw_weapon_recode	7	weapon_code == "99"

Category	Row label	variable_for_link	variable_code_1	der_code
Weapon involved	Yes	der_raw_weapon_recode	1	weapon_code == "40", weapon_code == "85", weapon_code == "55", weapon_code == "75", weapon_code == "80", weapon_code == "11", weapon_code == "12", weapon_code == "13", weapon_code == "14", weapon_code == "15", weapon_code == "11A", weapon_code == "12A", weapon_code == "13A", weapon_code == "14A", weapon_code == "15A", weapon_code == "16", weapon_code == "20", weapon_code == "17", weapon_code == "30", weapon_code == "35", weapon_code == "50", weapon_code == "60", weapon_code == "65", weapon_code == "70", weapon_code == "90", weapon_code == "95"
Weapon involved hierarchy	Handgun	der_raw_weapon_hierarchy_recode	1	weapon_code == "12", weapon_code == "12A"
Weapon involved hierarchy	Firearm	der_raw_weapon_hierarchy_recode	2	weapon_code == "11", weapon_code == "11A"
Weapon involved hierarchy	Rifle	der_raw_weapon_hierarchy_recode	3	weapon_code == "13", weapon_code == "13A"
Weapon involved hierarchy	Shotgun	der_raw_weapon_hierarchy_recode	4	weapon_code == "14", weapon_code == "14A"
Weapon involved hierarchy	Other firearm	der_raw_weapon_hierarchy_recode	5	weapon_code == "15", weapon_code == "15A"
Weapon involved hierarchy	Knife/cutting instrument	der_raw_weapon_hierarchy_recode	6	weapon_code == "16", weapon_code == "20"

Category	Row label	variable_for_link	variable_code_1	der_code
Weapon involved hierarchy	Blunt object	der_raw_weapon_hierarchy_recode	7	weapon_code == "17", weapon_code == "30"
Weapon involved hierarchy	Motor vehicle	der_raw_weapon_hierarchy_recode	8	weapon_code == "35"
Weapon involved hierarchy	Personal weapons (hands, feet, teeth, etc.)	der_raw_weapon_hierarchy_recode	9	weapon_code == "40", weapon_code == "55"
Weapon involved hierarchy	Asphyxiation	der_raw_weapon_hierarchy_recode	10	weapon_code == "85", weapon_code == "75", weapon_code == "80"
Weapon involved hierarchy	Drugs/narcotics /sleeping pills	der_raw_weapon_hierarchy_recode	11	weapon_code == "70"
Weapon involved hierarchy	Poison (include gas)	der_raw_weapon_hierarchy_recode	12	weapon_code == "50"
Weapon involved hierarchy	Explosives	der_raw_weapon_hierarchy_recode	13	weapon_code == "60"
Weapon involved hierarchy	Fire/incendiary device	der_raw_weapon_hierarchy_recode	14	weapon_code == "65"
Weapon involved hierarchy	Other	der_raw_weapon_hierarchy_recode	15	weapon_code == "90"
Weapon involved hierarchy	No weapon	der_raw_weapon_hierarchy_recode	16	weapon_code == "01", weapon_code == "99"
Weapon involved hierarchy	Unknown	der_raw_weapon_hierarchy_recode	17	weapon_code == "95"
Weapon involved hierarchy	Not applicable	der_raw_weapon_hierarchy_recode	18	weapon_code = Missing
Weapon involved hierarchy within offense	Handgun	der_raw_weapon_hierarchy_recode	1	weapon_code == "12", weapon_code == "12A"

Category	Row label	variable_for_link	variable_code_1	der_code
Weapon involved hierarchy within offense	Firearm	der_raw_weapon_hierarchy_recode	2	weapon_code == "11", weapon_code == "11A"
Weapon involved hierarchy within offense	Rifle	der_raw_weapon_hierarchy_recode	3	weapon_code == "13", weapon_code == "13A"
Weapon involved hierarchy within offense	Shotgun	der_raw_weapon_hierarchy_recode	4	weapon_code == "14", weapon_code == "14A"
Weapon involved hierarchy within offense	Other firearm	der_raw_weapon_hierarchy_recode	5	weapon_code == "15", weapon_code == "15A"
Weapon involved hierarchy within offense	Knife/cutting instrument	der_raw_weapon_hierarchy_recode	6	weapon_code == "16", weapon_code == "20"
Weapon involved hierarchy within offense	Blunt object	der_raw_weapon_hierarchy_recode	7	weapon_code == "17", weapon_code == "30"
Weapon involved hierarchy within offense	Motor vehicle	der_raw_weapon_hierarchy_recode	8	weapon_code == "35"
Weapon involved hierarchy within offense	Personal weapons (hands, feet, teeth, etc.)	der_raw_weapon_hierarchy_recode	9	weapon_code == "40", weapon_code == "55"
Weapon involved hierarchy within offense	Asphyxiation	der_raw_weapon_hierarchy_recode	10	weapon_code == "85", weapon_code == "75", weapon_code == "80"
Weapon involved hierarchy within offense	Drugs/narcotics /sleeping pills	der_raw_weapon_hierarchy_recode	11	weapon_code == "70"
Weapon involved hierarchy within offense	Poison (include gas)	der_raw_weapon_hierarchy_recode	12	weapon_code == "50"

Category	Row label	variable_for_link	variable_code_1	der_code
Weapon involved hierarchy within offense	Explosives	der_raw_weapon_hierarchy_recode	13	weapon_code == "60"
Weapon involved hierarchy within offense	Fire/incendiary device	der_raw_weapon_hierarchy_recode	14	weapon_code == "65"
Weapon involved hierarchy within offense	Other	der_raw_weapon_hierarchy_recode	15	weapon_code == "90"
Weapon involved hierarchy within offense	No weapon	der_raw_weapon_hierarchy_recode	16	weapon_code == "01", weapon_code == "99"
Weapon involved hierarchy within offense	Unknown	der_raw_weapon_hierarchy_recode	17	weapon_code == "95"
Weapon involved hierarchy within offense	Not applicable	der_raw_weapon_hierarchy_recode	18	weapon_code = Missing
Weapon involved hierarchy collapse	Firearm	der_raw_weapon_hierarchy_recode_col	1	weapon_code == "12", weapon_code == "12A", weapon_code == "11", weapon_code == "11A", weapon_code == "13", weapon_code == "13A", weapon_code == "14", weapon_code == "14A", weapon_code == "15", weapon_code == "15A"

Category	Row label	variable_for_link	variable_code_1	der_code
Weapon involved hierarchy collapse	Other weapon	der_raw_weapon_hierarchy_recode_col	2	weapon_code == "16", weapon_code == "20", weapon_code == "17", weapon_code == "30", weapon_code == "35", weapon_code == "40", weapon_code == "55", weapon_code == "85", weapon_code == "75", weapon_code == "80", weapon_code == "70", weapon_code == "50", weapon_code == "60", weapon_code == "65", weapon_code == "90"
Weapon involved hierarchy collapse	No weapon	der_raw_weapon_hierarchy_recode_col	3	weapon_code == "01", weapon_code == "99"
Weapon involved hierarchy collapse	Unknown	der_raw_weapon_hierarchy_recode_col	4	weapon_code == "95"
Weapon involved hierarchy collapse	Not applicable	der_raw_weapon_hierarchy_recode_col	5	weapon_code = Missing
Firearm type	Handgun only	der_single_gun_cat	1	One type of firearm identified and (weapon_code == "12" or weapon_code == "12A") detected within offense
Firearm type	Long gun (rifle and shotgun) only	der_single_gun_cat	2	One type of firearm identified and (weapon_code == "13" or weapon_code == "13A") or (weapon_code == "14" or weapon_code == "14A") detected within offense

Category	Row label	variable_for_link	variable_code_1	der_code
Firearm type	Unknown firearm type (other firearm and firearm) only	der_single_gun_cat	3	One type of firearm identified and (weapon_code == "11" or weapon_code == "11A") or (weapon_code == "15" or weapon_code == "15A") detected within offense
Firearm type	Single gun type	der_single_multi_firearm_types	1	Only one of the following firearm types linked to the offense: handgun only, long gun (rifle and shotgun) only, unknown firearm type (other firearm and firearm) only
Firearm type	Multiple firearm types	der_single_multi_firearm_types	2	More than one firearm type detected within an offense.

Table A-6. Victim-offender relationship

Category	Row label	variable_for_link	variable_code_1	der_code
Victim–offender relationship	Intimate partner	der_relationship	1	relationship_code == "BG", relationship_code == "CS", relationship_code == "SE", relationship_code == "XS", relationship_code == "XR"
Victim–offender relationship	Other family	der_relationship	2	relationship_code == "CF", relationship_code == "CH", relationship_code == "GC", relationship_code == "GP", relationship_code == "IL", relationship_code == "OF", relationship_code == "PA", relationship_code == "SB", relationship_code == "SC", relationship_code == "SP", relationship_code == "SS"
Victim–offender relationship	Outside family but known to victim	der_relationship	3	relationship_code == "AQ", relationship_code == "BE", relationship_code == "EE", relationship_code == "ER", relationship_code == "FR", relationship_code == "NE", relationship_code == "OK", relationship_code == "CO", relationship_code == "FP", relationship_code == "FC"
Victim–offender relationship	Stranger	der_relationship	4	relationship_code == "ST"
Victim–offender relationship	Victim was offender	der_relationship	5	relationship_code == "VO"
Victim–offender relationship	Unknown relationship	der_relationship	6	relationship_code == "RU"

Category	Row label	variable_for_link	variable_code_1	der_code
Victim–offender relationship 2	Intimate partner plus family	der_relationship2	1	relationship_code == "BG", relationship_code == "CS", relationship_code == "SE", relationship_code == "XS", relationship_code == "XR", relationship_code == "CF", relationship_code == "CH", relationship_code == "GC", relationship_code == "GP", relationship_code == "IL", relationship_code == "OF", relationship_code == "PA", relationship_code == "SB", relationship_code == "SC", relationship_code == "SP", relationship_code == "SS"
Victim–offender relationship 2	Outside family but known to victim	der_relationship2	2	relationship_code == "AQ", relationship_code == "BE", relationship_code == "EE", relationship_code == "ER", relationship_code == "FR", relationship_code == "NE", relationship_code == "OK", relationship_code == "CO", relationship_code == "FP", relationship_code == "FC"
Victim–offender relationship 2	Stranger	der_relationship2	3	relationship_code == "ST"
Victim–offender relationship 2	Victim was offender	der_relationship2	4	relationship_code == "VO"
Victim–offender relationship 2	Unknown relationship	der_relationship2	5	relationship_code == "RU"
Victim–offender relationship hierarchy	Intimate partner	der_relationship_hierarchy	1	relationship_code == "BG", relationship_code == "CS", relationship_code == "SE", relationship_code == "XS", relationship_code == "XR"

Category	Row label	variable_for_link	variable_code_1	der_code
Victim–offender relationship hierarchy	Other family	der_relationship_hierarchy	2	relationship_code == "CF", relationship_code == "CH", relationship_code == "GC", relationship_code == "GP", relationship_code == "IL", relationship_code == "OF", relationship_code == "PA", relationship_code == "SB", relationship_code == "SC", relationship_code == "SP", relationship_code == "SS"
Victim–offender relationship hierarchy	Outside family but known to victim	der_relationship_hierarchy	3	relationship_code == "AQ", relationship_code == "BE", relationship_code == "EE", relationship_code == "ER", relationship_code == "FR", relationship_code == "NE", relationship_code == "OK", relationship_code == "CO", relationship_code == "FP", relationship_code == "FC"
Victim–offender relationship hierarchy	Stranger	der_relationship_hierarchy	4	relationship_code == "ST"
Victim–offender relationship hierarchy	Victim was offender	der_relationship_hierarchy	5	relationship_code == "VO"
Victim–offender relationship hierarchy	Unknown relationship	der_relationship_hierarchy	6	relationship_code == "RU"
Victim–offender relationship hierarchy	Unknown offender incidents	der_relationship_hierarchy	7	Victim is Individual or Law Enforcement Officers and from unknown offender incidents
Victim–offender relationship hierarchy	Missing from uncleared incidents	der_relationship_hierarchy	8	Any missing relationship that was not imputed is from uncleared incidents

Category	Row label	variable_for_link	variable_code_1	der_code
Victim–offender relationship hierarchy among known offenders	Intimate partner	der_relationship_hierarchy	1	relationship_code == "BG", relationship_code == "CS", relationship_code == "SE", relationship_code == "XS", relationship_code == "XR"
Victim–offender relationship hierarchy among known offenders	Other family	der_relationship_hierarchy	2	relationship_code == "CF", relationship_code == "CH", relationship_code == "GC", relationship_code == "GP", relationship_code == "IL", relationship_code == "OF", relationship_code == "PA", relationship_code == "SB", relationship_code == "SC", relationship_code == "SP", relationship_code == "SS"
Victim–offender relationship hierarchy among known offenders	Outside family but known to victim	der_relationship_hierarchy	3	relationship_code == "AQ", relationship_code == "BE", relationship_code == "EE", relationship_code == "ER", relationship_code == "FR", relationship_code == "NE", relationship_code == "OK", relationship_code == "CO", relationship_code == "FP", relationship_code == "FC"
Victim–offender relationship hierarchy among known offenders	Stranger	der_relationship_hierarchy	4	relationship_code == "ST"
Victim–offender relationship hierarchy among known offenders	Victim was offender	der_relationship_hierarchy	5	relationship_code == "VO"

Category	Row label	variable_for_link	variable_code_1	der_code
Victim–offender relationship hierarchy among known offenders	Unknown relationship	der_relationship_hierarchy	6	relationship_code == "RU"

Table A-7. Location type

Category	Row label	variable_for_link	variable_code_1	der_code
Location type	Residence/hotel	der_location_1_10	1	location_code == "14", location_code == "20"
Location type	Transportation hub/outdoor public locations	der_location_1_10	2	location_code == "01", location_code == "10", location_code == "13", location_code == "16", location_code == "18", location_code == "37", location_code == "42", location_code == "45", location_code == "50", location_code == "51"
Location type	Schools, daycares, and universities	der_location_1_10	3	location_code == "22", location_code == "44", location_code == "52", location_code == "53"
Location type	Retail/financial/ other commercial establishment	der_location_1_10	4	location_code == "02", location_code == "05", location_code == "06", location_code == "07", location_code == "08", location_code == "12", location_code == "17", location_code == "19", location_code == "23", location_code == "24", location_code == "40", location_code == "41", location_code == "46", location_code == "48", location_code == "55", location_code == "09"
Location type	Restaurant/bar/ sports or entertainment venue	der_location_1_10	5	location_code == "03", location_code == "21", location_code == "38", location_code == "39", location_code == "47"
Location type	Religious buildings	der_location_1_10	6	location_code == "04"

Category	Row label	variable_for_link	variable_code_1	der_code
Location type	Government/public buildings	der_location_1_10	7	location_code == "11"
Location type	Jail/prison	der_location_1_10	8	location_code == "15"
Location type	Shelter-mission/homeless	der_location_1_10	9	location_code == "54"
Location type	Other/unknown location	der_location_1_10	10	location_code == "00", location_code == "25", location_code == "49", location_code == "56", location_code == "57", location_code == "58"
Location type 2	Residence/hotel	der_location_1_11	1	location_code == "14", location_code == "20"
Location type 2	Transportation hub/outdoor public locations	der_location_1_11	2	location_code == "01", location_code == "10", location_code == "13", location_code == "16", location_code == "18", location_code == "37", location_code == "42", location_code == "45", location_code == "50", location_code == "51"
Location type 2	Schools, daycares, and universities	der_location_1_11	3	location_code == "22", location_code == "44", location_code == "52", location_code == "53"

Category	Row label	variable_for_link	variable_code_1	der_code
Location type 2	Retail/financial/ other commercial establishment	der_location_1_11	4	location_code == "02", location_code == "05", location_code == "06", location_code == "07", location_code == "08", location_code == "12", location_code == "17", location_code == "19", location_code == "23", location_code == "24", location_code == "40", location_code == "41", location_code == "46", location_code == "48", location_code == "55"
Location type 2	Restaurant/bar/ sports or entertainment venue	der_location_1_11	5	location_code == "03", location_code == "21", location_code == "38", location_code == "39", location_code == "47"
Location type 2	Religious buildings	der_location_1_11	6	location_code == "04"
Location type 2	Government/public buildings	der_location_1_11	7	location_code == "11"
Location type 2	Jail/prison	der_location_1_11	8	location_code == "15"
Location type 2	Shelter-mission/homeless	der_location_1_11	9	location_code == "54"
Location type 2	Drug store/doctor's office/hospital	der_location_1_11	10	location_code == "09"
Location type 2	Other/unknown location	der_location_1_11	11	location_code == "00", location_code == "25", location_code == "49", location_code == "56", location_code == "57", location_code == "58"
Location type 3	Residence	der_location_residence	1	location_code == "20"

Category	Row label	variable_for_link	variable_code_1	der_code
Location type 3	Not residence	der_location_residence	2	location_code == "14", location_code == "01", location_code == "10", location_code == "13", location_code == "16", location_code == "18", location_code == "37", location_code == "42", location_code == "45", location_code == "50", location_code == "51", location_code == "22", location_code == "44", location_code == "52", location_code == "53", location_code == "02", location_code == "05", location_code == "06", location_code == "07", location_code == "08", location_code == "12", location_code == "17", location_code == "19", location_code == "23", location_code == "24", location_code == "40", location_code == "41", location_code == "46", location_code == "48", location_code == "55", location_code == "09", location_code == "03", location_code == "21", location_code == "38", location_code == "39", location_code == "47", location_code == "04", location_code == "11", location_code == "15", location_code == "54", location_code == "00", location_code == "25", location_code == "49", location_code == "56", location_code == "57", location_code == "58"

Category	Row label	variable_for_link	variable_code_1	der_code
Location cyberspace	Cyberspace	der_location_cyberspace	1	location_code == "58"
Location type hierarchy within offense	Residence	der_location_residence	1	location_code == "20"

Category	Row label	variable_for_link	variable_code_1	der_code
Location type hierarchy within offense	Not residence	der_location_residence	2	location_code == "14", location_code == "01", location_code == "10", location_code == "13", location_code == "16", location_code == "18", location_code == "37", location_code == "42", location_code == "45", location_code == "50", location_code == "51", location_code == "22", location_code == "44", location_code == "52", location_code == "53", location_code == "02", location_code == "05", location_code == "06", location_code == "07", location_code == "08", location_code == "12", location_code == "17", location_code == "19", location_code == "23", location_code == "24", location_code == "40", location_code == "41", location_code == "46", location_code == "48", location_code == "55", location_code == "09" location_code == "03", location_code == "21", location_code == "38", location_code == "39", location_code == "47", location_code == "04", location_code == "11", location_code == "15", location_code == "54", location_code == "00", location_code == "25", location_code == "49", location_code == "56", location_code == "57", location_code == "58"

Category	Row label	variable_for_link	variable_code_1	der_code
Location Type 4	Residence	der_location_1_12	1	location_code == "20"
Location Type 4	Hotel	der_location_1_12	2	location_code == "14"
Location Type 4	Transportation hub/outdoor public locations	der_location_1_12	3	location_code == "01", location_code == "10", location_code == "13", location_code == "16", location_code == "18", location_code == "37", location_code == "42", location_code == "45", location_code == "50", location_code == "51"
Location Type 4	Schools, daycares, and universities	der_location_1_12	4	location_code == "22", location_code == "44", location_code == "52", location_code == "53"
Location Type 4	Retail/financial/ other commercial establishment	der_location_1_12	5	location_code == "02", location_code == "05", location_code == "06", location_code == "07", location_code == "08", location_code == "12", location_code == "17", location_code == "19", location_code == "23", location_code == "24", location_code == "40", location_code == "41", location_code == "46", location_code == "48", location_code == "55"
Location Type 4	Restaurant/bar/ sports or entertainment venue	der_location_1_12	6	location_code == "03", location_code == "21", location_code == "38", location_code == "39", location_code == "47"
Location Type 4	Religious buildings	der_location_1_12	7	location_code == "04"

Category	Row label	variable_for_link	variable_code_1	der_code
Location Type 4	Government/public buildings	der_location_1_12	8	location_code == "11"
Location Type 4	Jail/prison	der_location_1_12	9	location_code == "15"
Location Type 4	Shelter-mission/homeless	der_location_1_12	10	location_code == "54"
Location Type 4	Drug store/doctor's office/hospital	der_location_1_12	11	location_code == "09"
Location Type 4	Other/unknown location	der_location_1_12	12	location_code == "00", location_code == "25", location_code == "49", location_code == "56", location_code == "57", location_code == "58"

Table A-8. Agency indicator

Category	Row label	variable_for_link	variable_code_1	der_code
Agency indicator	City	der_agency_type_1_7	1	AGENCY_TYPE_NAME == "City"
Agency indicator	County	der_agency_type_1_7	2	AGENCY_TYPE_NAME == "County"
Agency indicator	University or college	der_agency_type_1_7	3	AGENCY_TYPE_NAME == "University or College"
Agency indicator	State police	der_agency_type_1_7	4	AGENCY_TYPE_NAME == "State Police"
Agency indicator	Other state agencies	der_agency_type_1_7	5	AGENCY_TYPE_NAME == "Other", AGENCY_TYPE_NAME == "Other State Agency"
Agency indicator	Tribal agencies	der_agency_type_1_7	6	AGENCY_TYPE_NAME == "Tribal"

Table A-9. Population group

Category	Row label	variable_for_link	variable_code_1	der_code
Population group	Cities and counties 100,000 or over	der_population_group	1	POPULATION_GROUP_ID == 3, POPULATION_GROUP_ID == 4, POPULATION_GROUP_ID == 5, POPULATION_GROUP_ID == 6, POPULATION_GROUP_ID == 13, POPULATION_GROUP_ID == 19
Population group	Cities and counties 25,000–99,999	der_population_group	2	POPULATION_GROUP_ID == 7, POPULATION_GROUP_ID == 8, POPULATION_GROUP_ID == 14, POPULATION_GROUP_ID == 20
Population group	Cities and counties 10,000–24,999	der_population_group	3	POPULATION_GROUP_ID == 9, POPULATION_GROUP_ID == 15, POPULATION_GROUP_ID == 21
Population group	Cities and counties under 10,000	der_population_group	4	POPULATION_GROUP_ID == 10, POPULATION_GROUP_ID == 11, POPULATION_GROUP_ID == 16, POPULATION_GROUP_ID == 22
Population group	State police	der_population_group	5	POPULATION_GROUP_ID == 17, POPULATION_GROUP_ID == 23

Table A-10. Clearance

Category	Row label	variable_for_link	variable_code_1	der_code
Clearance	Not cleared	der_clearance_cat	1	cleared_except_code == "N"
Clearance	Cleared through arrest	der_clearance_cat	2	There is an arrest in the incident
Clearance	Exceptional clearance	der_clearance_cat	3	cleared_except_code %in% c("A","B","C","D","E")
Clearance	Death of offender	der_exceptional_clearance	1	cleared_except_code == "A"
Clearance	Prosecution declined	der_exceptional_clearance	2	cleared_except_code == "B"
Clearance	In custody of other jurisdiction	der_exceptional_clearance	3	cleared_except_code == "C"
Clearance	Victim refused to cooperate	der_exceptional_clearance	4	cleared_except_code == "D"
Clearance	Juvenile/no custody	der_exceptional_clearance	5	cleared_except_code == "E"
Clearance	Not cleared through arrest	der_clearance_cat_1_2	1	cleared_except_code %in% c("A","B","C","D","E"), cleared_except_code == "N"
Clearance	Cleared through arrest	der_clearance_cat_1_2	2	There is an arrest in the incident
Clearance 2	Cleared incident	der_cleared_cat_1_2	1	There is an arrest in the incident or cleared_except_code %in% c("A","B","C","D","E")
Clearance 2	Not cleared incident	der_cleared_cat_1_2	2	cleared_except_code == "N"

Table A- 11. Time of day (incident time and reported time)

Category	Row label	variable_for_link	variable_code_1	der_code
Time of day-Incident time	Midnight–4am	der_time_of_day_incident	1	report_date_flag != "R" & incident_hour %in% c(0:3)
Time of day-Incident time	4–8am	der_time_of_day_incident	2	report_date_flag != "R" & incident_hour %in% c(4:7)
Time of day-Incident time	8am–noon	der_time_of_day_incident	3	report_date_flag != "R" & incident_hour %in% c(8:11)
Time of day-Incident time	Noon–4pm	der_time_of_day_incident	4	report_date_flag != "R" & incident_hour %in% c(12:15)
Time of day-Incident time	4–8pm	der_time_of_day_incident	5	report_date_flag != "R" & incident_hour %in% c(16:19)
Time of day-Incident time	8pm–midnight	der_time_of_day_incident	6	report_date_flag != "R" & incident_hour %in% c(20:23)
Time of day-Incident time	Unknown	der_time_of_day_incident	7	report_date_flag != "R"
Time of day-Report time	Midnight–4am	der_time_of_day_report	1	report_date_flag == "R" & incident_hour %in% c(0:3)
Time of day-Report time	4–8am	der_time_of_day_report	2	report_date_flag == "R" & incident_hour %in% c(4:7)
Time of day-Report time	8am–noon	der_time_of_day_report	3	report_date_flag == "R" & incident_hour %in% c(8:11)
Time of day-Report time	Noon–4pm	der_time_of_day_report	4	report_date_flag == "R" & incident_hour %in% c(12:15)
Time of day-Report time	4–8pm	der_time_of_day_report	5	report_date_flag == "R" & incident_hour %in% c(16:19)
Time of day-Report time	8pm–midnight	der_time_of_day_report	6	report_date_flag == "R" & incident_hour %in% c(20:23)
Time of day-Report time	Unknown	der_time_of_day_report	7	report_date_flag == "R"

Table A- 12. Property loss

Category	Row label	variable_for_link	variable_code_1	der_code
Property loss	None	der_property_loss	1	prop_loss_code == 1
Property loss	Burned	der_property_loss	2	prop_loss_code == 2
Property loss	Counterfeited/forged	der_property_loss	3	prop_loss_code == 3
Property loss	Destroyed/damaged/ vandalized	der_property_loss	4	prop_loss_code == 4
Property loss	Recovered	der_property_loss	5	prop_loss_code == 5
Property loss	Seized	der_property_loss	6	prop_loss_code == 6
Property loss	Stolen/etc.	der_property_loss	7	prop_loss_code == 7
Property loss	Unknown	der_property_loss	8	prop_loss_code == 8

Table A-13. Arrest type

Category	Row label	variable_for_link	variable_ code_1	der_code
Arrest type	On-view arrest	der_arrest_type	1	arrest_type_code == "O"
Arrest type	Summoned/cited	der_arrest_type	2	arrest_type_code == "S"
Arrest type	Taken into custody	der_arrest_type	3	arrest_type_code == "T"
Arrest type-specific arrest rate	On-view arrest	der_arrest_type	1	arrest_type_code == "O"
Arrest type-specific arrest rate	Summoned/cited	der_arrest_type	2	arrest_type_code == "S"
Arrest type-specific arrest rate	Taken into custody	der_arrest_type	3	arrest_type_code == "T"

Table A-14. Arrestee armed

Category	Row label	variable_for_link	variable_code_1	der_code
Arrestee armed	Firearm	der_raw_weapon_recode	1	weapon_code == "11", weapon_code == "12", weapon_code == "13", weapon_code == "14", weapon_code == "15", weapon_code == "11A", weapon_code == "12A", weapon_code == "13A", weapon_code == "14A", weapon_code == "15A"
Arrestee armed	Lethal cutting instrument	der_raw_weapon_recode	2	weapon_code == "16"
Arrestee armed	Club/blackjack/ brass knuckles	der_raw_weapon_recode	3	weapon_code == "17"
Arrestee armed	No	der_weapon_no_yes	1	weapon_code == "01"
Arrestee armed	Yes	der_weapon_no_yes	2	weapon_code == "11", weapon_code == "12", weapon_code == "13", weapon_code == "14", weapon_code == "15", weapon_code == "11A", weapon_code == "12A", weapon_code == "13A", weapon_code == "14A", weapon_code == "15A", weapon_code == "16", weapon_code == "17"

Table A-15. Gang involvement

Category	Row label	variable_for_link	variable_code_1	der_code
Gang Involvement	None/unknown gang involvement	der_gang_cat_no_yes	1	criminal_act_code == "N"
Gang Involvement	Juvenile or other gang	der_gang_cat_no_yes	2	criminal_act_code == "G", criminal_act_code == "J"
Gang involvement[a]	No	der_gang_cat_no_yes	1	criminal_act_code == "N"
Gang involvement[b]	Yes	der_gang_cat_no_yes	2	criminal_act_code == "G", criminal_act_code == "J"

[a] Same as None/unknown gang involvement.

[b] Same as Juvenile or other gang.

Table A-16. Multiple arrest

Category	Row label	variable_for_link	variable_ code_1	der_code
Multiple arrest indicator	Multiple	der_multiple_arrest	1	multiple_indicator == "M"
Multiple arrest indicator	Count	der_multiple_arrest	2	multiple_indicator == "C"
Multiple arrest indicator	Not applicable	der_multiple_arrest	3	multiple_indicator == "N"

Table A-17. Juvenile disposition

Category	Row label	variable_for_link	variable_code_1	der_code
Juvenile disposition	Handled within department	der_juvenile_disp	1	under_18_disposition_code == "H"
Juvenile disposition	Referred to other authorities	der_juvenile_disp	2	under_18_disposition_code == "R"
Juvenile disposition	Not applicable	der_juvenile_disp	3	Arrestee's age >= 18
Juvenile disposition	Unknown	der_juvenile_disp	4	under_18_disposition_code = Missing

Table A-18. LEOKA types of activity

Category	Row label	variable_for_link	variable_code_1	der_code
LEOKA Types of Activity	Responding to disturbance call	der_activity	1	activity_code == "01"
LEOKA Types of Activity	Burglary	der_activity	2	activity_code == "02"
LEOKA Types of Activity	Robbery in process	der_activity	3	activity_code == "03"
LEOKA Types of Activity	Attempting other arrest	der_activity	4	activity_code == "04"
LEOKA Types of Activity	Civil disorder	der_activity	5	activity_code == "05"
LEOKA Types of Activity	Handling, transporting, custody of prisoners	der_activity	6	activity_code == "06"
LEOKA Types of Activity	Investigating suspicious persons	der_activity	7	activity_code == "07"
LEOKA Types of Activity	Ambush	der_activity	8	activity_code == "08"
LEOKA Types of Activity	Mentally challenged	der_activity	9	activity_code == "09"
LEOKA Types of Activity	Traffic pursuits	der_activity	10	activity_code == "10"
LEOKA Types of Activity	All other	der_activity	11	activity_code == "11"

Table A-19. LEOKA type of officer assignment

Category	Row label	variable_for_link	variable_code_1	der_code
Type of Assignment	Two-officer vehicle	der_assignment	1	assignment_code == "F"
Type of Assignment	One-officer vehicle alone	der_assignment	2	assignment_code == "G"
Type of Assignment	One-officer vehicle assisted	der_assignment	3	assignment_code == "H"
Type of Assignment	Detective or special assignment alone	der_assignment	4	assignment_code == "I"
Type of Assignment	Detective or special assignment assisted	der_assignment	5	assignment_code == "J"
Type of Assignment	Other	der_assignment	6	assignment_code == "K", assignment_code == "L"

Table A-20. Completed drug-related offenses

Category	Row label	variable_for_link	variable_code_1	der_code
Completed drug-related offenses	Only drug/narcotic violations	der_drug_narcotic_equipment_cat	1	Incident has offense "35A" and attempt_complete_flag == "C" and does not have an incident with offense "35B" and attempt_complete_flag == "C"
Completed drug-related offenses	Only drug equipment violations	der_drug_narcotic_equipment_cat	2	Incident has offense "35B" and attempt_complete_flag == "C" and does not have an incident with offense "35A" and attempt_complete_flag == "C"
Completed drug-related offenses	Both drug/narcotic and drug equipment violations	der_drug_narcotic_equipment_cat	3	Incident has offense "35A" and attempt_complete_flag == "C" and incident has offense "35B" and attempt_complete_flag == "C"

Table A-21. Property seized

Category	Row label	variable_for_link	variable_code_1	der_code
Property description	Property seized	der_property_seized_any	1	If any property within an incident is associated with prop_loss_code == 6
Property description	Property not seized	der_property_seized_any	0	If all property within an incident is not associated with prop_loss_code == 6
Property seized	Drugs/ narcotics	der_property_seized_categories_short	1	prop_loss_code == 6 & prop_desc_code == "10"
Property seized	Drug/ narcotic equipment	der_property_seized_categories_short	2	prop_loss_code == 6 & prop_desc_code == "11"
Property seized	Chemicals	der_property_seized_categories_short	3	prop_loss_code == 6 & prop_desc_code == "45"
Property seized	Money	der_property_seized_categories_short	4	prop_loss_code == 6 & prop_desc_code == "20"
Property seized	Firearms	der_property_seized_categories_short	5	prop_loss_code == 6 & prop_desc_code == "13"

Table A-22. Suspected type of drug

Category	Row label	variable_for_link	variable_code_1	der_code
Suspected type of drug	Cocaine/crack cocaine (A, B)	der_suspected_type_of_drug	1	suspected_drug_code == "A", suspected_drug_code == "B"
Suspected type of drug	Marijuana/hashish (C, E)	der_suspected_type_of_drug	2	suspected_drug_code == "C", suspected_drug_code == "E"
Suspected type of drug	Opiate/narcotic (D, F, G, H)	der_suspected_type_of_drug	3	suspected_drug_code == "D", suspected_drug_code == "F", suspected_drug_code == "G", suspected_drug_code == "H"
Suspected type of drug	Hallucinogen (I, J, K)	der_suspected_type_of_drug	4	suspected_drug_code == "I", suspected_drug_code == "J", suspected_drug_code == "K"
Suspected type of drug	Stimulant (L, M)	der_suspected_type_of_drug	5	suspected_drug_code == "L", suspected_drug_code == "M"
Suspected type of drug	Depressant (N, O)	der_suspected_type_of_drug	6	suspected_drug_code == "N", suspected_drug_code == "O"
Suspected type of drug	Other (P)	der_suspected_type_of_drug	7	suspected_drug_code == "P"
Suspected type of drug	Unknown (U)	der_suspected_type_of_drug	8	suspected_drug_code == "U"
Suspected type of drug	More than 3 types (X)	der_suspected_type_of_drug	9	suspected_drug_code == "X"

Table A-23. Criminal activity

Category	Row label	variable_for_link	variable_ code_1	der_code
Criminal activity	Buying/receiving	der_crim_activity	1	criminal_act_code == "B"
Criminal activity	Cultivating/manufacturing/publishing	der_crim_activity	2	criminal_act_code == "C"
Criminal activity	Distributing/selling	der_crim_activity	3	criminal_act_code == "D"
Criminal activity	Exploiting children	der_crim_activity	4	criminal_act_code == "E"
Criminal activity	Operating/promoting/ assisting	der_crim_activity	5	criminal_act_code == "O"
Criminal activity	Possessing/concealing	der_crim_activity	6	criminal_act_code == "P"
Criminal activity	Transporting/ transmitting/importing	der_crim_activity	7	criminal_act_code == "T"
Criminal activity	Using/consuming	der_crim_activity	8	criminal_act_code == "U"
Criminal activity	Drug possession or trafficking	der_crim_activity _drug_poss_traff	1	criminal_act_code == "B", criminal_act_code == "C", criminal_act_code == "P", criminal_act_code == "U", criminal_act_code == "D", criminal_act_code == "O", criminal_act_code == "T"
Criminal activity	Drug possession for personal consumption	der_crim_activity _drug_poss_pc	1	criminal_act_code == "B", criminal_act_code == "C", criminal_act_code == "P", criminal_act_code == "U"
Criminal activity	Drug trafficking not for personal consumption	der_crim_activity _drug_poss_npc	1	criminal_act_code == "D", criminal_act_code == "O", criminal_act_code == "T"

Table A-24. Number of victims murdered

Category	Row label	variable_for_link	variable_code_1	der_code
Number of victims murdered	Yes	der_victim_murder_non_neg_manslaughter	1	The victim is associated with the offense code of "09A"
Number of victims murdered	No	der_victim_murder_non_neg_manslaughter	2	The victim is not associated with the offense code of "09A"

Table A-25. Victim ethnicity

Category	Row label	variable_for_link	variable_code_1	der_code
Victim ethnicity	Hispanic or Latino	der_victim_ethnicity	1	victim_ethnicity_code == "H"
Victim ethnicity	Not Hispanic or Latino	der_victim_ethnicity	2	victim_ethnicity_code == "N"
Victim ethnicity	Multiple/unknown/not specified	der_victim_ethnicity	3	victim_ethnicity_code == "M", victim_ethnicity_code == "U", victim_ethnicity_code == "X", victim_ethnicity_code is missing

Table A-26. Victimization performed by any offenders in an incident by age

Category	Row label	variable_for_link	variable_code_1	der_code
Number of victimizations performed by offenders aged 12 or older, or missing age, or from unknown offender incidents		der_offender_age_12_plus_missing_unk_inc	1	If offender is either: offender's age >= 12, offender's age is missing, or unknown offender incident (i.e. offender sequence number is 0)
Victimization performed by any offenders in an incident by age	12–17	der_offender_cat_12_17	1	12 <= offender's age & offender's age < 18
Victimization performed by any offenders in an incident by age	18 or older	der_offender_cat_18_plus	1	Offender's age >= 18
Victimization performed by any offenders in an incident by age	Known offenders and missing age	der_offender_age_missing	1	Missing offender's age and offender sequence number is not 0
Victimization performed by any offenders in an incident by age	Unknown offender incidents	der_unknown_offender_incident	1	Offender sequence number is 0