National Crime Statistics Exchange (NCS-X)
National Incident-Based Reporting System (NIBRS)
Estimation Project (NNEP)

Methods for Estimation Crime based on NIBRS
January 13, 2022

BUREAU OF JUSTICE STATISTICS - OFFICE OF JUSTICE PROGRAMS
FEDERAL BUREAU OF INVESTIGATION - CRIMINAL JUSTICE INFORMATION SERVICES
• Beginning January 1, 2021, the FBI no longer accepts crime data submitted in summary form and submissions can only be made using NIBRS

• The National Crime Statistics Exchange (NCS-X) is serving two purposes related to this change:
  1. Support states and a sample of law enforcement agencies transition to NIBRS compliant reporting systems
  2. Develop methodologies and procedures by which representative national, regional, and state-level crime estimates can be produced
Webinar Agenda

1. Provide overview of the estimation “system”
2. Discuss the impact that coverage has on producing estimates
3. Discuss key components of the estimation system
4. Discuss the indicators which will be estimated
5. Discuss data quality review conducted
6. Discuss the data editing/imputation rules
About the Speakers

▪ Marcus Berzofsky
  – Senior Statistician at RTI International
  – PI on the NCS-X NIBRS Estimation Project
  – 20 years experience working with implementing and analyzing surveys and related studies

▪ Alexia Cooper
  – BJS Unit Chief of the Law Enforcement Unit
  – To date has been project officer for the NCS-X NIBRS Estimation Project
  – Over 15 years experience conducting research and data collection in the field of criminal justice
Housekeeping Items

- **Questions**: Please enter questions into the chat; if the question can be immediately answered quickly, we will do that, otherwise, we will wait until the end of the presentation to answer questions and have discussion.

- **Schedule**: Presentation should last around 45 minutes providing another 45 minutes for questions and discussion.
Objectives, Accomplishments to Date, Schedule
NCS-X Estimation “System” Objectives

Primary Objectives

- Determine set of key NIBRS indicators for which accurate and reliable estimates can be produced.
- Develop population served estimates which can be split by key demographic characteristics such as age, sex and race.
- Use the NCS-X sample to produce national estimates.

Secondary Objectives

- Use NCS-X sample to produce subnational estimates (e.g., state-level estimates).
- Develop population served estimates for agencies traditionally classified as zero population agencies.
## Schedule of Key Milestones

<table>
<thead>
<tr>
<th>Estimated Completion Date</th>
<th>Key Milestone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2021</td>
<td>Develop 0.1 (Alpha) Version of Estimate Generation System prototype</td>
</tr>
<tr>
<td>Fall 2021</td>
<td>Produce estimates based on 2020 data year through estimation prototype system (0.1 Alpha Version)</td>
</tr>
<tr>
<td>Winter 2021</td>
<td>Begin to integrate estimation system into FBI and BJS infrastructures</td>
</tr>
<tr>
<td>Spring 2022</td>
<td>Develop 0.5 (Beta) Version of Estimation Generation System prototype which further finalizes methodology and optimizes system</td>
</tr>
<tr>
<td>Spring 2022</td>
<td>Update estimation methodology to account for additional transitioned agencies</td>
</tr>
<tr>
<td>Summer/Fall 2022</td>
<td>Disseminate information about the new estimation system, how it works, how it differs from the current UCR estimates, etc.</td>
</tr>
<tr>
<td>Summer/Fall 2022</td>
<td>Finalize 1.0 (production) Version of the estimation system to produce estimates based on the 2021 data year</td>
</tr>
</tbody>
</table>
Accomplishments to Date

**Indicators**
- Determined Indicators to be estimated in initial estimation system including LEOKA assaults
- Assessing additional measures such as detailed drug offense indicators

**Estimation**
- Developed weighting methodology weighting methodology based on expected transition of agencies
- Developed three types of weights: (1) national weight, (2) regional weight, and (3) state weight

**Quality**
- Determined set of quality checks to identify agencies with outlier responses to each data element
- Determined imputation methodology for victim, arrestee, and cleared offender demographics plus victim-offender relationship

**Population Estimation**
- Determined methodology to expand population estimates to be disaggregated by age, sex, and race
- Determining when it is possible to produce accurate population estimates for “zero population” agencies

**Variance Estimation & Suppression**
- Developed variance and suppression rules for 0.1 (Alpha version); suppression rules take bias into account
- Determining if alternative variance methodology or suppression rules are needed for 0.5 (Beta) Version

**Estimation System Development**
- Developed 0.1 Version (Alpha) of the Estimate Generation System (delivered on October 15, 2021)
- Working to further optimize the estimation system in anticipation of 0.5 Version (Beta) due for delivery in March 2022
Coverage – Impact on Estimation
Coverage Basics

What is “coverage”?  

• “Coverage” or the “Coverage Rate” is the percentage of population (e.g., agencies, population served) which can participate in the data collection (i.e., the agencies/states which have transitioned to NIBRS reporting)
• Example: State A has 300 agencies representing 5 million people; 200 agencies have transitioned to NIBRS serving 2 million residents of the state
  • Coverage by agency = 200/300 = 66.7%
  • Coverage by population served = 2 million/5 million = 40%

Why does the coverage rate matter?  

• The lower the coverage rate the more “uncertainty” there is in the estimate
• Uncertainty is the potential error in any estimate caused by not having responses from all eligible participants.
• To understand the amount of uncertainty, we need to calculate a confidence interval around the estimate
• Summary UCR estimates did not require a confidence interval as the program had extremely high coverage rates (~97%) making the amount of uncertainty negligible.
Expansion of NIBRS Coverage: 2018 to 2020

2018 NIBRS Coverage:
(43% with about 8,000 reporting LEAs)

2020 NIBRS Coverage
(57% with over 10,600 reporting LEAs)

Note: There are around 18,600 law enforcement agencies (LEAs) reporting to the UCR programs, either to NIBRS or SRS.
Currently, estimate that:
- 40 of largest 72 will be able to provide 12 months of 2021 data
- 13 of largest 72 will be able to provide 6-11 months of 2021 data
- 8 of largest 72 agencies will be able to provide 1-4 months of 2021 data
- 11 of largest 72 will not provide data for 2021 or it is unknown at this point
NIBRS Coverage Rate by Region: 2018-2020

Highlights

- NIBRS coverage is uneven across the country. In 2020, the Northeast had the lowest coverage rate of 26%. The South had the highest coverage rate of 69%.
- The expansion pace differs across the different regions.
- The low coverage rate in the Northeast hinders our ability to create the same weights that can be used to produce both the national and regional estimates.
- Aside from the weights for national level estimation, separate sets of weights are being created to produce regional-level estimates.

NR = Non-Reporter
Implications for National Estimation

Precision-based suppression rules will be developed to determine if an estimate can be released.

Some geographical areas and agency subgroups still have low coverage ratios.

- It is likely that estimates in some subgroups will be suppressed due to low precision.
- Separate weights are needed for regional estimation.
- Expected to improve over the next few years, given NIBRS’s expansion.

Using the sample of the 400 NCS-X agencies improves the precision of estimates.

- By 2020, 167 NCS-X agencies became NIBRS reporters. Most remaining agencies expected to complete transition by 2022.
- Large agencies and tribal agencies are difficult areas to expedite their transition.
# NIBRS Reporting States: with 80%+ Population Coverage

## States with 80% in 2018 (N= 22)

<table>
<thead>
<tr>
<th>Arkansas</th>
<th>Michigan</th>
<th>South Dakota</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorado</td>
<td>Montana</td>
<td>Tennessee</td>
</tr>
<tr>
<td>Connecticut</td>
<td>New Hampshire</td>
<td>Vermont</td>
</tr>
<tr>
<td>Delaware</td>
<td>North Dakota</td>
<td>Virginia</td>
</tr>
<tr>
<td>Idaho</td>
<td>Ohio</td>
<td>Washington</td>
</tr>
<tr>
<td>Iowa</td>
<td>Oregon</td>
<td>West Virginia</td>
</tr>
<tr>
<td>Kentucky</td>
<td>Rhode Island</td>
<td></td>
</tr>
<tr>
<td>Massachusetts</td>
<td>South Carolina</td>
<td></td>
</tr>
</tbody>
</table>

## 80% by 2021 (N= 8)

<table>
<thead>
<tr>
<th>Kansas</th>
<th>Minnesota</th>
<th>Missouri</th>
<th>Nevada</th>
<th>North Carolina</th>
<th>Texas</th>
<th>Utah</th>
<th>Wisconsin</th>
</tr>
</thead>
</table>
Implication for State-Level Estimation

NIBRS’s expansion progress varies greatly by state and are expected to change in the new future years.

• 26 states have more than 80% coverage, while some other states still have very low coverage rate.
• Alaska, California, Florida and New Jersey still have no NIBRS reporters in 2020.
• California, Florida and Alaska are expected to have significant expansion in 2021.

While not the sole determining factor, high coverage rates will be needed in order to publish estimates for a state

• Further assessment is needed to decide whether a state has sufficient coverage to produce state-level estimates. For example, all the large agencies in the state must be NIBRS reporter; tribal agencies can be critical for some states.
Other Estimation Considerations
Uncertainty Basics

What is uncertainty in the context of NIBRS estimates?

- Uncertainty in an estimate occurs when participation from the entire population is not obtained.
- Uncertainty potentially results in: (1) Variance and (2) Bias
  - Variance represents the uncertainty caused by random (sampling) error
  - Bias represents the uncertainty caused when the nonparticipants are not a random subset of the population

Why does uncertainty matter for NIBRS?

- Because of relatively low level of coverage at the national and regional levels and in many states, estimates need to be accompanied by a measure of uncertainty
- Typical probability-based surveys only account for variance because they are unbiased samples of the population
- Since the nonparticipating agencies are not random, bias needs to be taken into account

How will uncertainty be measured for NIBRS?

- The mean squared error (MSE) is a measure that accounts for both variance and precision
- The MSE can be used in the same way a typical standard error is to create a 95% confidence interval
Variance Estimation/Uncertainty and Suppression

Variance Estimation and Uncertainty Measure

- Variances will use Taylor Series Linearization
- Uncertainty will use the Mean Square Error (MSE) which will account for both variance and likely bias due to participating agencies being a nonrandom set
- Resulting confidence intervals will use the root MSE (RMSE) as the half-width rather than the SE (as is done in a traditional confidence interval)

Suppression

- All estimates will be produced for the estimation system – even those with low coverage or small sample sizes
- Suppression flag will identify those with large RMSE (>30%) or small agency sample size (n<10)
- Regions and states will be further suppressed if more than 50% of estimates suppressed and coverage below 80%
- No estimates will be suppressed if coverage for an area is nearly complete (> 95%)
Population Estimates

**Current approach**

- Sources used include of Population Estimates Program and Census Gazetteer Information
- Produces aggregate population served for general purpose city, county, and state police
- Does **not** produce population served estimates below the agency level
- Does **not** produce population served estimates for zero population agencies (those with no unique jurisdiction)

**Proposed Approach**

- Sources include the Population Estimates Program and 5-year American Community Survey
- Produces aggregate population served for general purpose city, county, and state police
- Produces population served estimates by age, sex and race
- Produces population served estimates for some zero population agencies (e.g., universities)
- Agency-level estimates will differ in some cases from current approach due to different methods and sources used to allow for splitting population estimates by age, sex, and race
Key Indicators to Be Estimated
**Key Indicator Basics**

**What is a “key” indicator?**
- A key indicator is an indicator being prioritized for estimation

**Why are key indicators needed?**
- NIBRS contains hundreds of indicators including interactions of multiple indicators
- To better focus the data quality review, certain estimates were prioritized for estimation

**How were the key indicators determined?**
- Indicators were reviewed among multiple stakeholders including members of FBI, BJS, law enforcement agencies, and academics
- The stakeholder group agreed upon the initial indicators based on a review of the quality of the larger variable set

**Can additional indicators be added to the estimation system?**
- Yes!
- As more agencies transition and data quality improves the ability to produce estimates for a larger set of indicators will be possible.
Key Indicators: Offense Types

<table>
<thead>
<tr>
<th>Crimes Against Persons</th>
<th>Crimes Against Property</th>
<th>Crimes Against Society</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Aggravated assault*</td>
<td>• Arson*</td>
<td>• Animal cruelty</td>
</tr>
<tr>
<td>• Simple assault</td>
<td>• Bribery</td>
<td>• Drug/narcotic offenses</td>
</tr>
<tr>
<td>• Intimidation</td>
<td>• Burglary*</td>
<td>• Gambling offenses</td>
</tr>
<tr>
<td>• Murder &amp; Nonnegligent Manslaughter*</td>
<td>• Counterfeiting/forgery</td>
<td>• Prostitution/obscene material</td>
</tr>
<tr>
<td>• Negligent Manslaughter</td>
<td>• Destruction, damage, and vandalism</td>
<td>• Prostitution offenses</td>
</tr>
<tr>
<td>• Kidnapping and abduction</td>
<td>• Embezzlement</td>
<td>• Weapon law violations</td>
</tr>
<tr>
<td>• Sex trafficking*</td>
<td>• Extortion/blackmail</td>
<td></td>
</tr>
<tr>
<td>• Labor trafficking*</td>
<td>• Fraud offenses</td>
<td></td>
</tr>
<tr>
<td>• Rape*</td>
<td>• Larceny/theft offenses*</td>
<td></td>
</tr>
<tr>
<td>• Sodomy</td>
<td>• Motor vehicle theft*</td>
<td></td>
</tr>
<tr>
<td>• Sexual assault with an object</td>
<td>• Robbery*</td>
<td></td>
</tr>
<tr>
<td>• Fondling</td>
<td>• Stolen property offenses</td>
<td></td>
</tr>
<tr>
<td>• Nonforcible sex offenses</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Indicates Part 1 offenses reported through summary UCR.
## Indicator Type: Characteristics of Individuals

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Categories</th>
<th>Reporting Unit of Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Incident</td>
</tr>
<tr>
<td>Age*</td>
<td>Under 5, 5-14, 15-17, 18-24, 25-34, 35-64, 65+</td>
<td>X</td>
</tr>
<tr>
<td>Sex*</td>
<td>Male, female</td>
<td>X</td>
</tr>
<tr>
<td>Race*</td>
<td>White, Black, American Indian or Alaskan Native, Asian, Native Hawaiian or Other Pacific Islander</td>
<td>X</td>
</tr>
<tr>
<td>Sex and race^</td>
<td>Same as above</td>
<td>X</td>
</tr>
<tr>
<td>Victim age category by offender age category^</td>
<td>Same as above</td>
<td>X</td>
</tr>
<tr>
<td>Victim sex by offender sex^</td>
<td>Same as above</td>
<td>X</td>
</tr>
<tr>
<td>Victim race by offender race^</td>
<td>Same as above</td>
<td>X</td>
</tr>
<tr>
<td>Juvenile disposition^</td>
<td>Handled within department, referred to other authorities, not applicable</td>
<td>X</td>
</tr>
<tr>
<td>Multiple arrest indicator^</td>
<td>Multiple, count, not applicable</td>
<td>X</td>
</tr>
<tr>
<td>Arrestee armed (yes/no)^</td>
<td>Firearm, lethal cutting instrument, club/blackjack/brass knuckles</td>
<td>X</td>
</tr>
</tbody>
</table>

^Counts and percentages will be calculated.
*Counts, rates, and percentages will be calculated.
# Indicator Type: Characteristics of Incidents

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Categories</th>
<th>Reporting Unit of Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Incident</td>
</tr>
<tr>
<td>Weapon involved^</td>
<td>Personal, firearms, knives/cutting instruments, blunt instruments,</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>other non-personal, unknown</td>
<td></td>
</tr>
<tr>
<td>Injury^</td>
<td>Yes, no</td>
<td>X</td>
</tr>
<tr>
<td>Multiple victims^</td>
<td>1 victim, 2+ victims</td>
<td>X</td>
</tr>
<tr>
<td>Multiple offenders^</td>
<td>1 offender, 2+ offenders, unknown offenders</td>
<td>X</td>
</tr>
<tr>
<td>Multiple offense incident^</td>
<td>1 offense, 2 offenses, 3+ offenses</td>
<td>X</td>
</tr>
<tr>
<td>Victim-offender relationship^</td>
<td>Intimate partner, other family, outside family but known to victim,</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>stranger, victim was offender, unknown relationship</td>
<td></td>
</tr>
<tr>
<td>Location type^</td>
<td>Residence/hotel, transportation hub/outdoor public locations, schools/</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>daycares/universities, retail/financial/other commercial establishment,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>restaurant/bar/sports or entertainment venue, religious buildings,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>government/public buildings, jail/prison, shelter-mission/homeless, other/unknown location</td>
<td></td>
</tr>
</tbody>
</table>

^Counts and percentages will be calculated.

*Counts, rates, and percentages will be calculated.
## Indicator Type: Characteristics of Incidents (cont.)

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Categories</th>
<th>Reporting Unit of Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time of day^</td>
<td>Midnight-4am, 4-8am, 8am-noon, noon-4pm, 4-8pm, 8pm-midnight, unknown</td>
<td>Incident: X, Offense: X</td>
</tr>
<tr>
<td>Population group (cities and counties combined by population size)^</td>
<td>100,000 and over, 25,000-99,999, 10,000-24,999, under 10,000, state police, possessions</td>
<td>Victim: X, Arrestee: X</td>
</tr>
<tr>
<td>Agency indicator^</td>
<td>City, county, university or college, state police, other state agencies, tribal agencies, federal agencies</td>
<td>Incident: X, Offense: X</td>
</tr>
<tr>
<td>Cleared through arrest^</td>
<td>Yes, no</td>
<td>Incident: X Offense: X</td>
</tr>
<tr>
<td>Exceptional clearance^</td>
<td>Death of offender, prosecution declined, in custody of other jurisdiction, victim refused to cooperate, juvenile/no custody</td>
<td>Incident: X</td>
</tr>
<tr>
<td>Property loss^</td>
<td>None, burned, counterfeited/forged, destroyed/damaged/vandalized, recovered, seized, stolen, unknown</td>
<td>Incident: X</td>
</tr>
<tr>
<td>Gang involvement^</td>
<td>None/unknown, juvenile or other gang involvement</td>
<td>Victim: X</td>
</tr>
</tbody>
</table>

^Counts and percentages will be calculated.
*Counts, rates, and percentages will be calculated.
Why is data quality important for estimation?

- Estimates are only as good as the data they are based on
- If data quality is poor the estimates will be biased (i.e., not accurate)

What is meant by “data quality” for NIBRS

- Each NIBRS incident must pass a set of edit checks to be accepted by the FBI
- Data quality review looks at unexpected values that may pass the edit checks but can lead to erroneous estimates

What measures are taken to address data quality?

- *Data review*. Data review looks for outlier responses at an aggregate level (e.g., agency level)
- *Imputation*. Imputation statistically fills in missing or unknown values
Data Quality Review

What is reviewed?

• All indicators being estimated
• Additional potential indicators for future inclusion

How is review conducted?

• Outlier agencies for each data element flagged for review
• Outlier agencies usually defined based on 95\textsuperscript{th} or 99\textsuperscript{th} percentile of distribution (e.g., agencies with the highest rate of assigning midnight for time of incident)
• Report built into prototype (beta version) listing flagged agencies for each data element

What happens to flagged agencies?

• Feedback loop with agencies (conducted by FBI)
• Identified agencies asked to correct data moving forward in future years (may not be time to correct for the current year)
Imputation Basics

**What is imputation?**
- Imputation is a statistical procedure to fill in missing data
- Imputation typically relies on similar cases (e.g., agencies) to determine what the missing value should be

**Why is imputation needed for estimation?**
- If excluded, cases with missing data can lead to undercounts of the true total
- Domain estimates (e.g., age, race) cannot be conducted when domain values are missing

**What types of imputation are done in the estimation system?**
- *Block imputation*: Imputes the number of crimes in a month for agencies that do not provide reports for all 12 months
- *Item imputation*: Imputes incident demographic characteristics to allow for estimates of victim and (some) offender characteristics
Imputation

- Block Imputation
  - If agency reports at least 2 months of data, it is counted as a responding agency
  - Missing months are imputed using hot deck
  - Donors are identified based on the predicated mean method for the total number of reported crimes by the agency

- Item Imputation: Hot Deck imputation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Victim</th>
<th>Offender (Cleared)</th>
<th>Offender (Uncleared)</th>
<th>Arrestee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Sex</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Race</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Victim-Offender Relationship</td>
<td>X</td>
<td>Can be linked to Arrestee</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Discussion