



Bureau of Justice Statistics

Establishing New Methods for Estimating Crime in the U.S. – The Transition to Incident-Based Crime Reporting through NIBRS

February 17, 2022
1:00pm – 2:30pm ET

About the Speakers

Cindy Barnett-Ryan

- FBI Data Scientist
- Primary representative from the FBI on the NCS-X NIBRS Estimation Project
- Over 25 years' experience working with Uniform Crime Reporting Program and law enforcement data collections

Marcus Berzofsky

- Senior Statistician at RTI International
- PI on the NCS-X NIBRS Estimation Project
- 20 years experience implementing and analyzing surveys and related studies



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

Establishing New Methods for Estimating Crime in the U.S.—The Transition to Incident-Based Crime Reporting through NIBRS

February 17, 2022

BUREAU OF JUSTICE STATISTICS - OFFICE OF JUSTICE PROGRAMS

FEDERAL BUREAU OF INVESTIGATION - CRIMINAL JUSTICE INFORMATION SERVICES



- At the beginning of 2021, the FBI UCR Program retired the SRS and fully transitioned crime reporting to NIBRS

- BJS partnered with the FBI on the National Crime Statistics Exchange (NCS-X) on the NIBRS transition by:
 1. Supporting state UCR programs and a scientifically selected sample of agencies transition to NIBRS reporting
 2. Developing methodologies and procedures by which 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

The background of the slide is a solid dark blue. It is decorated with various white geometric elements: thin curved lines, circles of different sizes, and small circular nodes. Some of these elements have arrows indicating a clockwise direction, suggesting a process or cycle. The overall aesthetic is clean, modern, and technical.

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.

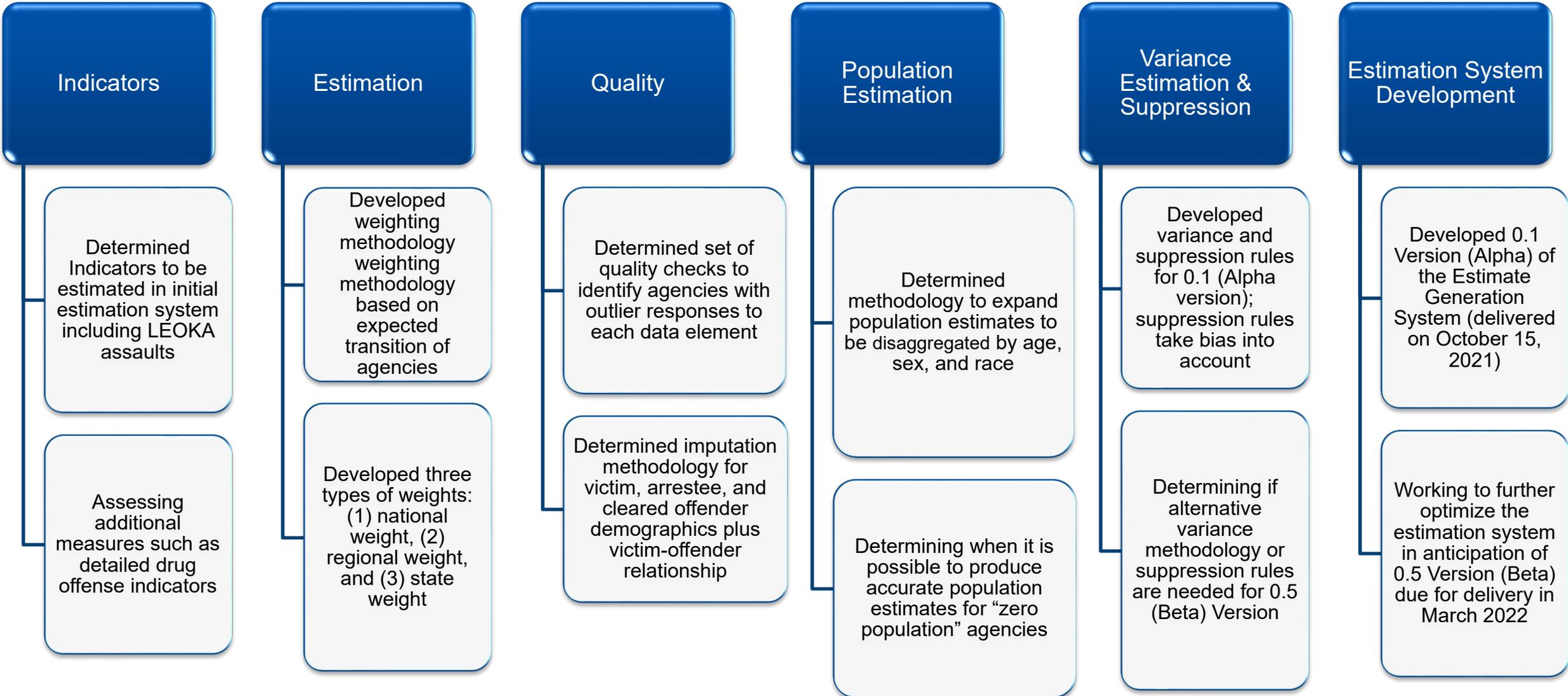
Secondary Objectives

- 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

Estimated Date	Key Milestone
Fall 2021	Develop 0.1 (Alpha) Version of Estimate Generation System prototype
Fall 2021	Produce estimates based on 2020 data year through estimation prototype system (0.1 Alpha Version)
Winter 2021	Begin to integrate estimation system into FBI and BJS infrastructures
Spring 2022 (early)	Develop 0.5 (Beta) Version of Estimation Generation System prototype which further finalizes methodology and optimizes system
Spring 2022 (mid)	Update estimation methodology to account for additional transitioned agencies
Spring 2022 (late)	Develop 0.75 (Gamma) Version of Estimation Generation System prototype, incorporating all final methodology and system modifications across the prototype testing lifecycle
Summer/Fall 2022	Disseminate information about the new estimation system, how it works, how it differs from the current UCR estimates, etc.
Summer/Fall 2022	Finalize 1.0 (production) Version of the estimation system to produce estimates based on the 2021 data year

Accomplishments to Date





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 = $2 \text{ million}/5 \text{ million} = 40\%$

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.

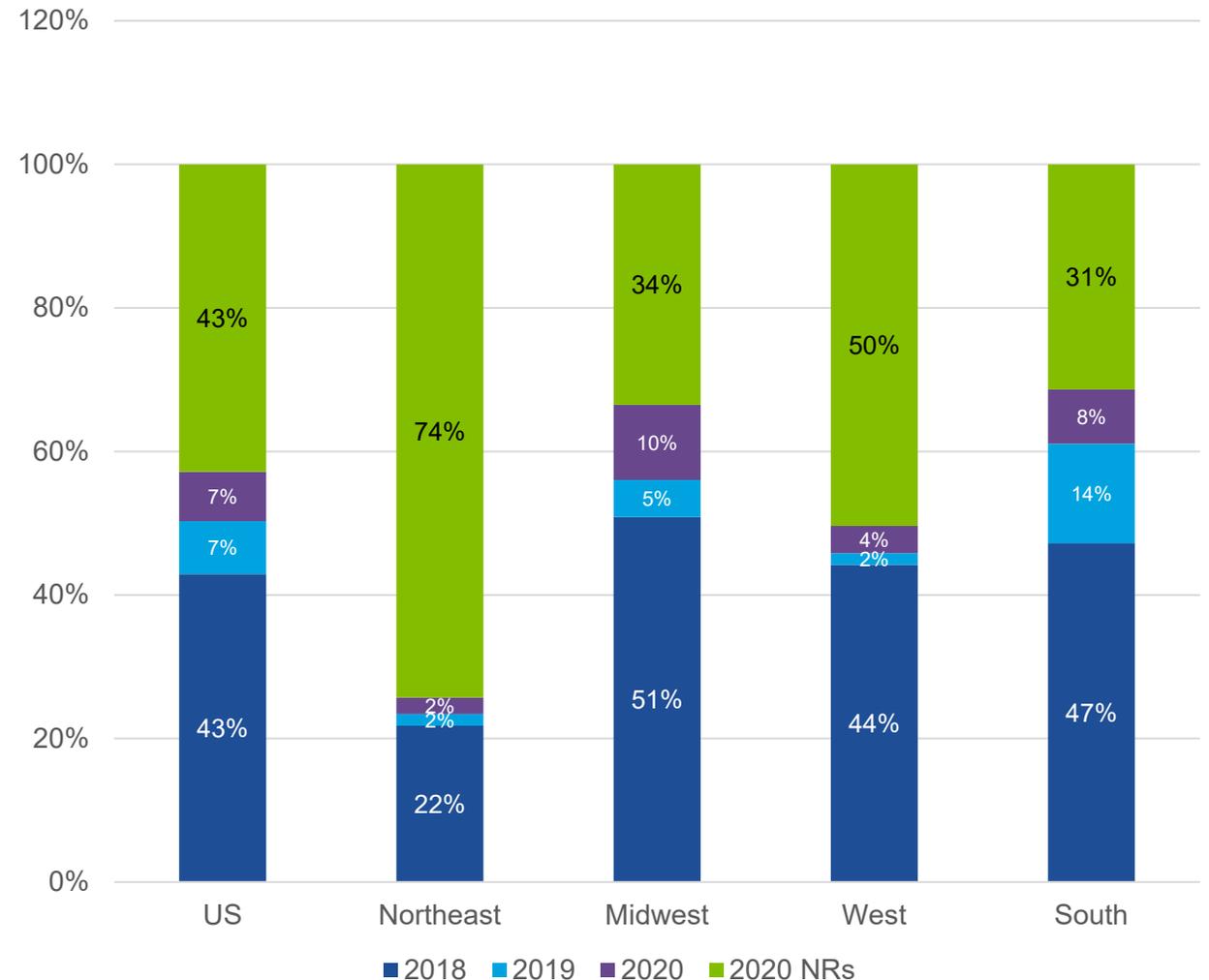
2021 Projected Data Availability for the 72 Largest NCS-X Sample Agencies

- Estimation sample includes large agencies that already reported to NIBRS, as well as the 72 largest sampled for NCS-X
- Among the 72 sampled agencies, we estimate that:
 - 40 will be able to provide 12 months of 2021 data
 - 13 will be able to provide 6-11 months of 2021 data
 - 8 will be able to provide 1-4 months of 2021 data
 - 11 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)			80% by 2021 (N= 8)
Arkansas	Michigan	South Dakota	Kansas
Colorado	Montana	Tennessee	Minnesota
Connecticut	New Hampshire	Vermont	Missouri
Delaware	North Dakota	Virginia	Nevada
Idaho	Ohio	Washington	North Carolina
Iowa	Oregon	West Virginia	Texas
Kentucky	Rhode Island		Utah
Massachusetts	South Carolina		Wisconsin

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

The background of the slide is a solid dark blue. It is decorated with several white, semi-transparent geometric elements: large overlapping circles, thin white arcs, and small circular icons with arrows indicating rotation or movement. These elements are scattered across the top half of the slide, creating a modern, technical aesthetic.

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 set 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

Crimes Against Persons

- Aggravated assault*
- Simple assault
- Intimidation
- Murder & Nonnegligent Manslaughter*
- Negligent Manslaughter
- Kidnapping and abduction
- Sex trafficking*
- Labor trafficking*
- Rape*
- Sodomy
- Sexual assault with an object
- Fondling
- Nonforcible sex offenses

Crimes Against Property

- Arson*
- Bribery
- Burglary*
- Counterfeiting/forgery
- Destruction, damage, and vandalism
- Embezzlement
- Extortion/blackmail
- Fraud offenses
- Larceny/theft offenses*
- Motor vehicle theft*
- Robbery*
- Stolen property offenses

Crimes Against Society

- Animal cruelty
- Drug/narcotic offenses
- Gambling offenses
- Prostitution/obscene material
- Prostitution offenses
- Weapon law violations

**Indicates Part 1 offenses reported through summary UCR.*

Indicator Type: Characteristics of Individuals

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

^Counts and percentages will be calculated.

**Counts, rates, and percentages will be calculated.*

Indicator Type: Characteristics of Incidents

Indicator	Categories	Reporting Unit of Analysis			
		Incident	Offense	Victim	Arrestee
Weapon involved [^]	Personal, firearms, knives/cutting instruments, blunt instruments, other non-personal, unknown	X	X	X	X
Injury [^]	Yes, no	X	X	X	
Multiple victims [^]	1 victim, 2+ victims	X	X		
Multiple offenders [^]	1 offender, 2+ offenders, unknown offenders	X	X		
Multiple offense incident [^]	1 offense, 2 offenses, 3+ offenses	X			
Victim-offender relationship [^]	Intimate partner, other family, outside family but known to victim, stranger, victim was offender, unknown relationship	X	X	X	
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	X	X		

[^]Counts and percentages will be calculated.

*Counts, rates, and percentages will be calculated.

Indicator Type: Characteristics of Incidents (cont.)

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

[^]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 95th or 99th 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)

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

- Agencies that report at least 3 months of data are counted as responding agencies
- 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

Variable	Victim	Offender (Cleared)	Offender (Uncleared)	Arrestee
Age	X	X		X
Sex	X	X		X
Race	X	X		X
Victim-Offender Relationship	X	Can be linked to Arrestee		X

Q&A Session

Please type your questions into Q&A
selecting All Panelists



Bureau of Justice Statistics