Module Objectives

- Describe the sampling procedures for SSOCS
- Explain the procedures for using weights and calculating standard errors
- Describe how imputation procedures were applied to missing data
Defines the population from which a sample is drawn for research purposes.
SSOCS sampling frame is based on the most recent Common Core of Data (CCD).
  For the 2016 SSOCS, this was the 2013-2014 CCD.

Certain types of schools from the CCD are excluded from the SSOCS sampling frame including:
  - Schools located outside of the 50 states and the District of Columbia
  - Bureau of Indian Education schools
  - Special education schools
  - Vocational schools
  - Alternative schools
  - Virtual schools
  - Ungraded schools
  - Schools with a high grade of kindergarten or lower

Regular schools, charter schools, and schools that have partial or total magnet programs are included in the frame.
Sample Design: Design Objectives

- The same general sample design was used to select a sample of schools for SSOCS in 2000, 2004, 2006, 2008, 2010, and 2016
- The objective of the 2015–2016 sample design was twofold
  - Obtain overall cross-sectional and subgroup estimates of important indicators of school crime and safety
  - Develop precise estimates of change in various characteristics relating to crime between the SSOCS administrations

Sample Design: Stratification

Stratified sample of regular public schools is drawn for each administration of SSOCS
- Ensures that different subgroups are adequately represented in the sample
- Involves dividing the sampling frame into relevant subgroups, or "strata," prior to sample selection
- Helps to increase accuracy when estimating population parameters for these subgroups by ensuring that different subgroups of a population are represented adequately in the sample
Sample Design: Defining the Strata

- SSOCS strata were defined by school level, locale, and enrollment size
  - School level indicates whether the school enrolls students in primary, middle or high school grades, or some combination
  - Locale – characterized as City, Suburb, Town, or Rural – is based on the Census-defined geographic area in which the school is located
  - Enrollment size is categorized into four ranges based on the number of students attending
- These three explicit stratification variables have been shown to be related to school crime and thus create meaningful subgroups for this survey

Sample Design: Defining the Strata (Continued)

Region and percent White enrollment were used as implicit stratification variables by sorting schools by these variables within each explicit stratum before sample selection

- Region is the Census region of the school, categorized as the Northeast, South, Midwest, and West
- Percent White enrollment represents the percentage of students enrolled in the school who are identified as White, non-Hispanic
SSOCS Sample Design: Drawing the Sample

Sample Design Process

- Define an optimal goal for the number of completed surveys to collect in order to meet their objectives
- Potential nonresponse must be taken into account
- Initial sample selected must be greater than the final goal

SSOCS Sample Design: Sample Weights

Weighting of the data is necessary to obtain population-based estimates, minimize nonresponse bias, and reduce sampling error

- Sample weights allow for inferences to be made about the total population of schools from which the sampled schools are drawn
- The method for calculating weights has not varied from year to year
Unit Response Rate and Weighting

- After the data are collected, a unit response rate is calculated for SSOCS.
- The unit response rate is the ratio of completed eligible respondents to the total count of eligible respondents.
- For SSOCS data, there are three measures to evaluate response.

Slide 12 of 26

Unit Response Rate and Weighting (Continued)

1. Completion Rate

\[
\frac{\text{C}}{\text{T}} = \frac{\text{2,092}}{\text{3,553}} = 58.9\%
\]

(Number of Completed Surveys)

(Total Sample Size)
Unit Response Rate and Weighting (Continued)

- Bias due to nonresponse
- NCES standards require analysis of nonresponse bias for surveys with base-weighted response rates less than 85%
  - SSOCS:2016 base-weight response was 62.9%
- Responding schools were compared to non-responding schools by
  - School level
  - Enrollment size
  - Locale
  - Percent White enrollment
  - Region
  - Number of Full-time Equivalent (FTE) teachers
  - Student-to-teacher ratio
  - Percentage of students eligible for free or reduced-price lunch

Unit Response Rate and Weighting (Continued)

- Significant differences between responding and nonresponding schools
- Underrepresented groups included
  - Schools with an enrollment of 1,000 students or more, urban schools, and schools in which less than 50 percent of students are White, non-Hispanic
- Correlations between these characteristics and survey variables were analyzed, and key estimates were compared between the lowest propensity respondents and other respondents
- CCD characteristics were found to be correlated with a number of survey variables
  - The observed bias in the CCD characteristics, if not adjusted for, would likely lead to bias in the SSOCS:2016 estimates
**Unit Response Rate and Weighting – CHAID Analysis**

- Chi-square Automatic Interaction Detector (CHAID) analysis was conducted to inform the selection of weighting classes
- Base weights were adjusted for potential nonresponse bias
- When the nonresponse-adjusted weights were applied, no significant bias remained in these characteristics
  - Weighting adjustments incorporated into the SSOCS:2016 weights help to mitigate nonresponse bias in key estimates
  - Survey variables are not observed for nonrespondents, so the exact amount of nonresponse remaining in key estimates cannot be known and may vary

**Item Nonresponse**

- Item nonresponse occurs when questions on the survey are not answered
- Item-level response rates are calculated and evaluated
- Unweighted item response rates are calculated
  - Example: 1,000 responses / 2,092 schools = 48% unweighted item response rate for that question
- After the unit base-weight is applied, weighted item-level response rates in SSOCS are generally high
  - SSOCS 2016 weighted item-level response rates ranged from 82% to 100%
Item Nonresponse (Continued)

SSOCS data set variables included in SSOCS 2016 Item Nonresponse Bias Analysis

- C0326 Number of physical attacks or fights with a weapon
- C0330 Number of physical attacks or fights without a weapon

Standard Error Calculation in SSOCS - Replication Techniques

- This method calculates appropriate SEs based on differences between estimates from the full sample and a series of created subsamples (replicates)
- Select replicate weights that are associated with the final sampling weight
- SSOCS replication weights use the Jackknife replication method
  - 2016 replicate weights are labeled 'REPFWT1' through 'REPFWT50'

Note: Estimates and standard errors for selected SSOCS variables are calculated and reported in web tables on the NCES website for your reference
Standard Error Calculation in SSOCS – Taylor-series Linearization

- This method uses primary sampling unit (PSU) and strata identifiers to calculate appropriate SEs
- Select the PSU and stratum variables (SCHID and STRATA) associated with the final sampling weight variable

Missing Data and Editing

- Files containing missing data can be problematic for research
  - Analysis of incomplete datasets may cause users to arrive at different conclusions
  - Certain groups of respondents may be more likely than others to leave some survey items unanswered, creating bias in the survey estimates
- Data editing included
  - Computer program-based editing and data retrieval follow-up interviewing were used in SSOCS to check for completeness and valid data ranges, consistency, and skip patterns
  - Data retrieval follow-up interviews were conducted for respondents missing critical survey items
Missing Data and Imputation

Even after computer program-based editing and data retrieval interviewing, completed SSOCS surveys still contain some level of item nonresponse

- Accordingly, imputation procedures are used to create values for all questionnaire items' to the existing first statement on the slide
- The base-weighted item response rates for SSOCS 2016 were generally high
  - The mean weighted item response rate was about 98 percent
  - After editing, 99 percent of items in the 2016 data file had weighted response rates of over 85 percent
  - Most of the data in the SSOCS file are the originally entered values from the respondents
- Imputation methods were tailored to the nature of each survey item using either aggregate proportions, best match, or clerical methods
- Imputation flags are provided for each variable where imputation was used

Imputation Methods - Aggregate Proportions

- Values estimated using data from respondents matched on key characteristics to the respondent with the missing data
- Rather than imputing counts from a single donor or a mean count from a group of donors, proportions were imputed using two methods
Imputation Methods - Best Match

Imputation Methods - Clerical

In some instances, missing data are available from other sources

- CCD sampling frame used to impute values for those schools missing student enrollment data for item 37
- CCD data were also available on school type for item 43 and the percentage of students eligible for free or reduced-price lunch, item 38a
- In other instances, research was done on school administrative records to estimate logical values for missing data
Imputation Methods - Order of Use

- Interrelationships between the SSOCS survey items require that a specific imputation order be followed
  - For example, because item 37 – student enrollment – is used in imputation for other variables, this item is imputed first
- In some cases, values for a particular variable are limited by the values of other variables
  - For example, the matrix of disciplinary actions by offenses in item 35 is related to the total number of offenses recorded and the total number of disciplinary actions recorded in earlier questions
  - If values must be imputed for these earlier items, this is done before item 35

Summary and Resources

Summary
- Described the sampling procedures for SSOCS
- Explained the procedures for using weights and calculating standard errors
- Described how imputation procedures were applied to missing data

Resources
- Common Core of Data (CCD)
- Non-responding schools (User's Manual)
- Standard Errors