

Slide 1 of 13

Title Slide: Getting Started with SSOCS Data

Slide 2 of 13

This module provides information needed to start using SSOCS data. In this module, users will learn which data are available for use, which data file formats are available, which data documentation files such as user's manuals and codebooks are provided, and how to access SSOCS data files and understand their naming conventions. There is also information on tables and reports already produced from the SSOCS data that are available to the public.

This module, and those that follow, will provide a solid foundation of knowledge for using the SSOCS data. However, the modules alone do not provide sufficient detail to analyze the data. Researchers should have a comprehensive understanding of school crime and safety and should carefully review the individual data elements of SSOCS before beginning an analysis.

First and foremost anyone using the data should read through the user's manuals. These provide a wealth of information about how NCES developed the various measures used in the questionnaire, what they indicate, and how data from the measures are reported or provided in the data file. The user's manual for the most recent release of SSOCS data can be accessed by clicking on the underlined screen text, 'user's manuals.'

All data files undergo a thorough review before release to the public. During this process, NCES identifies issues in the data that will be helpful to users. Including problems due to administration errors that cannot be fixed, or unusual patterns in the data that may or may not be explainable. These findings will be included in the user's manual.

It is also important for users to review the survey versions available online to understand how questions were asked and how they may have changed over successive years. As you review the instruments pay special attention to the paths through the questionnaire and the skip patterns. The SSOCS data collection instruments can be accessed by clicking on the underlined screen text, 'questionnaires.'

Lastly, check the SSOCS website, which can be accessed by clicking on the corresponding underlined screen text, for information about new data releases or issues identified in existing data files.

Slide 3 of 13

There are two types of SSOCS data files -- public-use data files and restricted-use data files.

The public-use file provides users with data from the survey without identifying information about each school. In this table we see hypothetical yes or no responses for three schools to seven questions. To produce a public-use data file, NCES staff conducts disclosure risk analyses to identify data elements that could disclose the identity of an individual survey respondent or school. Data elements that are identified as disclosure risks are masked or coarsened by recoding using top coding, bottom coding, or categorical recoding throughout. In some cases an entire variable is suppressed from the public-use file completely to make the public-use data file more manageable and to protect the confidentiality of sampled schools. Variables which are discussed in the survey documentation files, but which are not available in the public-use data files are denoted by “/R” after the variable name.

Slide 4 of 13

SSOCS is also available in a restricted-use file. Restricted-use files also are altered to protect the identity of any individual responding school, but they include more detailed information about the characteristics of the schools completing the survey. This information is provided because it is potentially important to researchers. NCES established a licensing system to provide qualified researchers access to restricted-use files. More information on the differences between public-use and restricted-use files and the licensing system for acquiring restricted-use files can be found in the module titled, ‘*Accessing NCES Micro-level Data*,’ which is accessible from the landing page.

Slide 5 of 13

Both the public-use and restricted-use SSOCS data files are available in several formats to accommodate commonly used statistical software packages. The file format options include a SAS data file with the associated SAS format library and SAS variable labels files; an SPSS for Windows data file; and a fixed-format ASCII (or text) file with an associated program to read the fixed-format file into SAS. The SSOCS 2000 and SSOCS 2002 data files are also available in Stata. Newer data can be converted to Stata from SPSS or SAS format. More information on how to convert files to Stata is included in the SSOCS user documentation.

Users can visit the SSOCS homepage to download public-use files for each year of SSOCS, or click on the underlined screen text ‘available.’ To receive a copy of the restricted-use files, users must apply for a restricted-use license.

The data file user’s manual and codebook are critical to using the SSOCS data. These two documents include important information about the methods of the study, the layout of the data, all variable names, and information on appropriate uses of the data. Separate codebooks accompany the public-use and restricted-use data files. The public-use data file documentation can be accessed from the SSOCS website, or by

Getting Started with SSOCS Data

clicking on the underlined screen text, 'available;' and the restricted-use documentation can be accessed directly from the restricted-use CD/DVD which accompanies the restricted-use data files that are given to licensees.

Slide 6 of 13

The SSOCS data file contains data from all completed questionnaires. The data file includes the unique school identifier (named SCHID); questionnaire item variables, including categorized versions of any open-ended response variables; composite (or created) variables, including the nesting variable (named STRATA); sampling frame variables; the final sampling weight (named FINALWGT) and associated jackknife replicate weights; and the imputation flags.

The first type of variable in the data file is the questionnaire item variable. The values assigned to these variables are taken directly from the responses to the questionnaire. For questionnaire item variables, a value of -1 indicates that the item is legitimately skipped. That is, based on an earlier item response, this item or subitem should be blank. In total, there are 34 items and 231 subitems corresponding directly to the items on the 2010 SSOCS questionnaire.

Composite Variables are those that are constructed based on responses to items in the questionnaire. Composite variables include summed variables and categorical variables derived from other numeric responses. These variables are included so that results from researcher to researcher using commonly needed constructed variables can easily be replicated. The steps used to create these variables are detailed in the SSOCS Codebook.

Sampling frame variables are also included in the data file. The values of sampling frame variables are taken from the Department of Education's Common Core of Data (or CCD). Both the school and district level CCD files were used. These variables contain information about school characteristics, including those that were used in stratifying the sample selected for SSOCS.

Weighting and variance estimation variables are included in the data file so that data can be appropriately weighted to generate estimates that are nationally representative of U.S. public schools.

Finally, the SSOCS data file contains imputation flags to alert users to data that have been imputed. Each questionnaire item variable – except for the introductory items and the open-ended items – has an imputation flag in this data file.

More information on each of these types of variables, including all variable names, labels, and locations in the data file, is available in the SSOCS codebook.

Slide 7 of 13

Each type of variable described earlier is identifiable based on the naming conventions for SSOCS.

Questionnaire item variables begin with the letter “C” and “zero” followed by the 3 digit source code assigned to the item. Source codes can be found next to the response fields for each item on the questionnaire. For example, let’s look at question number 2 on the 2009-2010 SSOCS questionnaire, “Does your school have a written plan that describes procedures to be performed in the following crises?” There are eight subitems to this question, labeled “a” through “h”. Look at item a, “shootings”. Look to the right to find the possible responses, which in this case are “yes” and “no”. Notice the small number preceding the boxes for the responses. For this subitem the source code is 154. In the data file, the variable containing the value of the response to this item is named C0154. The possible values for this variable are “1” yes, or “2” no.

Source codes allow data users to identify each unique question across multiple years of the survey. When items are reordered, the source code remains the same for the same question. Similarly, if questions are reworded or new questions added, they are assigned new source codes to avoid confusion with questions in previous versions. Variable names which include the designation “_R” after the source code indicate variables that have been recoded in the data file from the responses in the original survey – generally in order to group continuous variables into categories.

Slide 8 of 13

Composite variables are variables computed based on responses to other questions. These are identified in the data file by their unique descriptive names. Each composite variable includes the year of data collection, so that for the SSOCS 2010 data, the composite variable name includes the number ten. For example, the variable INCPOL10 is the total number of incidents reported to the police. To create this variable, the algorithm adds up the responses to the 15 items asking about types of incidents reported to the police.

Slide 9 of 13

Sampling frame variables contain values representing the characteristics used in the selection of the stratified sample for SSOCS. These are included in the data file because they can provide important descriptive information about the responding schools and can be used to compare responses between groups of interest. Sampling variable names begin with the prefix “FR” and an underscore. For example, the variable FR_HIGD is the highest grade taught at that school, as reported in the CCD. The “FR” and underscore highlight to you that this variable is a sampling frame variable. In this case, it is one that helps define the school level.

Not all frame variables, however, follow this naming on the restricted-use files. There are three frame variables on the restricted-use files that do not begin with “FR” and an underscore: the Census region of the schools (CENREG), the percent White enrollment

Getting Started with SSOCS Data

at the school (PERCWHT), and the number of full-time equivalent classroom teachers (FTE07). Refer to the codebooks for the restricted-use data files for more information on these variables.

Slide 10 of 13

There are a number of variables in the SSOCS data file that pertain to weighting. The 50 jackknife replicate weights and the final weight are all identifiable by the letters WGT in the variable name. For example, the 50 jackknife replicate weights are labeled REPWGT1 through REPWGT50. The final overall weight is labeled FINALWGT.

These weights must be used when estimating population statistics from the survey sample results and when generating standard errors for those estimates.

Slide 11 of 13

The final group of variables alerts users to imputation procedures that have been applied to other variables in the file. Imputation refers to the process of estimating values to replace missing or incorrect data. Imputation flag variables for each item are easy to identify in the SSOCS data file. Imputation flag variables begin with the letter “I” and are followed by the questionnaire item variable name to which they refer. Think back to the earlier example of variable C0154 – item 2, sub-item a. The imputation flag for this variable is IC0154 – the prefix I followed by the questionnaire item variable name.

Each questionnaire item variable has a corresponding imputation flag. If the value of the imputation flag is 0, then the value of the variable it refers to has not been imputed. Other non-zero values for imputation flags indicate the method used to impute the data.

Slide 12 of 13

It is not always necessary to access the data files directly to view information collected by SSOCS. In many cases, the information you need may be available in published tables and reports. SSOCS summary information is accessed through the SSOCS Table Library, which is also available on the NCES website and can be accessed by clicking on the underlined screen text, ‘[SSOCS Table Library](#).’ NCES generates and organizes tables by topic, including data summary tables on:

- Bullying and Cyber-bullying;
- Disciplinary Problems and Actions;
- Disruptions and Threats;
- Drugs, Alcohol, and Illegal Substances;
- Fear and Avoidance;
- Firearms, Explosives and Other Weapons;
- Gangs and Hate Crimes;
- Parent and Community Involvement;
- Physical Fights, Assaults, and Sexual Harassment;

Getting Started with SSOCS Data

- Prevention Programs and Training;
- Security Practices and Rules; and
- Violent Crimes and Theft

SSOCS data are also used in several reports published on a regular basis by NCES and the Bureau of Justice Statistics, or BJS. These reports include:

- Indicators of School Crime and Safety, published annually; and
- Crime, Violence, Discipline and Safety in U.S. Public Schools: Findings from the School Survey on Crime and Safety, published after each collection year of SSOCS.

A bibliography of academic papers and other publications using SSOCS data can be generated with the bibliographical search tool provided on the NCES website which can be accessed by clicking on the underlined screen text, 'bibliographical.'

It is possible that the answer to your research question can be found within an existing NCES product, including web tables and published reports.

Slide 13 of 13

This module has provided you with an overview of the available data and documentation files for SSOCS, the file formats and naming conventions, and existing SSOCS data tables and reports. Important resources that have been provided throughout the module are summarized in this slide along with the module's objectives for your reference.

You may now proceed to the next module in the series, or click the exit button to return to the landing page.