

Getting Started with the Data Collected Across the Secondary Longitudinal Studies from 1972-2000

Slide 1 of 16

Title Slide: Getting Started with the Data Collected Across the Secondary Longitudinal Studies from 1972-2000

Slide 2 of 16

This module introduces users to the data collected across the Secondary Longitudinal Studies from 1972-2000. The module describes the process of obtaining and accessing the NLS-72, HS&B, and NELS:88 data files and describes the resources that are available to learn more about the studies, the data, and the data files.

Information presented in this module will be helpful in understanding some of the more detailed information presented in subsequent modules. For this reason, users who are planning to proceed through the subsequent Secondary Longitudinal Studies Dataset modules and use the data for analytic purposes are strongly encouraged to complete this module first.

Slide 3 of 16

NCES designed the secondary longitudinal studies program to include data sets that are multipurpose, policy relevant, and pluralistic. The studies are multipurpose in that they are designed to obtain information on many topics for multiple audiences. The studies are policy relevant in that they identify alterable factors that are amenable to policy influence. The studies are pluralistic in that they satisfy the information needs of a diverse group of users.

Slide 4 of 16

Across the secondary longitudinal studies program, variable names within each study data set are designed to be consistent across time. For example base year data collections and data files are usually denoted as BY across secondary longitudinal studies datasets and studies. F1 indicates the first follow-up, F2 indicates second follow-up, and so on. BYS indicates base year student, S indicates student, P indicates parent, D indicates dropout, T indicates teacher, A indicates administrator, SC indicates school, TR indicates transcript, and BYI indicates base year ineligible. It should be pointed out, though, that variable names used across studies (NLS-72, HS&B, NELS:88, etc.) may not be consistent. Thus, researchers should become familiar with the variable naming structure for each of the data sets.

Cohorts that can be followed are specific to each study dataset. For example, NLS-72 has only a senior cohort; HS&B has a sophomore and senior cohort; while NELS:88 has 8th, 10th, and 12th grade cohorts.

It is also important to note that there are several common composite variables that are used across all the secondary longitudinal study datasets. Composite variables describing sex, race/ethnicity, urbanicity, socioeconomic status or SES, public vs. private school designation, and region of the country are among the most commonly

Getting Started with the Data Collected Across the Secondary Longitudinal Studies from 1972-2000

used. Composite variables should be used when they are available in place of raw data as they contain less missing data and are consistent across studies.

Slide 5 of 16

Data from the secondary longitudinal studies conducted from 1972-2000 are released in two formats. As of 2015, NLS-72 and HS&B data are only released in restricted-use format, while NELS:88 data are released in both public and restricted-use formats.

While NLS-72 and NELS:88 will include all cases ever surveyed, the HS&B sophomore and senior data files will be limited to the cohort samples: 14,825 cases in the HS&B sophomore cohort, and 11,995 cases in the HS&B senior cohort. The decision was made to limit the sample to cases that had been subsampled as of the last follow-up. The reason for this decision centered on technical issues that NCES identified in reviewing earlier archived releases of the HS&B files.

Slide 6 of 16

As discussed in the common module titled, 'Acquiring micro-level NCES datasets,' restricted-use data are only available to researchers who apply for and are granted a restricted-use license.

In general, the restricted-use file contains more data and a wider range of data values than are included in the public-use files. If you are interested in analyzing NLS-72 or HS&B data, you will need to be granted a restricted-use license.

In the event that you are interested in NELS:88 data, you should first examine the public-use data file to determine whether your specific analytical objectives can be met using public-use data. For most users, the public-use files provide all the data they will need for most analyses, though some users may find that only the restricted files have the specific data they need. The NELS:88 codebook identifies variables that have been altered or suppressed. Lastly, it is important to note that information regarding expanded sample, base year ineligible sample members, and high school and postsecondary transcript data are only available in the restricted-use files.

For more information regarding restricted-use data licenses, click on the corresponding underlined screen text.

Slide 7 of 16

NLS-72, HS&B sophomore cohort, HS&B senior cohort, and NELS:88 restricted-use data files are released in ASCII file format with accompanying SAS syntax (card) files. Most statistical analysis systems, including SPSS and Stata, can read SAS files and convert them into the necessary format.

The restricted-use data files will also include a codebook with frequencies and means for each variable in the data file. Frequencies are provided for discrete variables and means (with minimums, maximums and standard deviations) are provided for continuous variables. Analysts should use the ID and Universe variables, when

Getting Started with the Data Collected Across the Secondary Longitudinal Studies from 1972-2000

available, found on the restricted-use data files to understand the status of sample members across the data collection rounds. Universe variables describe sample member status in all five rounds of data collection. Value labels begin with BY status, followed by F1 and other follow-up statuses.

Lastly, analysts should carefully review the reserve codes that are provided on the restricted-use data files to understand missing data before recoding any data. Reserve codes are used to designate statuses of responses. Commonly used reserve codes include 6 (multiple response), 7 (refused critical item), 8 (missing), and 9 (legitimate skip).

Slide 8 of 16

The number of cases included on the secondary longitudinal studies data files varies by study. All cases that were ever part of the NLS-72 (22,652) are retained on the data file for analysis. Analysts should carefully consider the reserve codes provided on the data file to understand missing data.

The number of cases retained for analysis within HS&B varies by cohort. Cases for 14,825 members of the sophomore cohort who completed the fourth follow-up have been retained for analysis. Cases for the 11,995 members of the senior cohort who completed the third follow-up have been retained for analysis. Across both cohorts, survey items will be coded as missing for any year or years where sample members did not respond to a specific follow-up survey.

All cases that were ever part of the NELS:88 (27,805) are retained on the data file for analysis. Analysts should carefully consider the reserve codes provided on the data file to understand missing data.

Flags included with the data files facilitate the selection of specific samples (cross-section or panel) for analysis.

Because the subsampling used in these four studies varied, potential users should refer to the design and user manuals to get a better understanding of sub-sampling plans and the statuses at various follow-ups. For example, NELS:88 users must be aware of the statuses of base year non-respondents, base-year ineligible students, freshened students and the guidelines for subsampling students into follow-up data collection rounds.

Slide 9 of 16

All NELS:88 public-use data, base year (BY) through fourth follow-up (F4), are available for download from the eDAT online data access tool. The eDAT, can be accessed by clicking the corresponding underlined screen text.

Getting Started with the Data Collected Across the Secondary Longitudinal Studies from 1972-2000

Slide 10 of 16

The largest and most commonly used NELS:88 data files are the student-level data files which contain questionnaire, composite, and weight variables. Student-level data file questionnaire variables come from the student, parent, teacher, and school administrator questionnaires. School-level data files also contain questionnaire, composite, and weight variables.

Slide 11 of 16

NELS:88 has three sets of restricted-use student files. It is important to note that these three sets of NELS:88 student files do not contain the same number of cases. For example, NELS:88/92 contains approximately 27,000 cases; NELS:88/94 contains approximately 15,000 cases; and NELS:88/2000 contains approximately 12,000 cases. Weights are specific to each of these datasets. There is not a Base Year weight on the NELS:88/2000 dataset. It contains weights for the fourth follow-up only.

Slide 12 of 16

Even though there are major design features common across NLS-72, HS&B, and NELS:88, each of these data sets contain unique variations in methodologies used that require analysts to become familiar with the technical documentation and methodology reports that have been written. Researchers MUST review each study's documentation and understand how the data were collected, coded, and weighted to ensure accurate analyses. The slides that follow will detail the most important technical documentation and methodology reports that should be read to ensure accurate analysis of data from NLS-72, HS&B, and NELS:88.

Slide 13 of 16

The two most important documents that should be read to ensure accurate analyses of NLS-72 data are: National Longitudinal Study: Base Year (1972) through Fourth Follow-Up (1979) Data File User's Manual Volumes I through III, and The National Longitudinal Study of the High School Class of 1972 (NLS-72) Fifth Follow-Up (1986) Final Technical Report (NCES 87100). Both of these technical documents can be accessed by clicking the corresponding underlined screen text.

Volume I details the study background, sample, weights, and instrumentation (i.e., questionnaires and assessments). It is important to note that guidance regarding weight selection for the base year through the fourth follow-up is provided within this volume. Volume II contains Appendices K, L, and M, which detail composite score analysis, constructed education variables, and subject keyword index, respectively. Appendix P, found within Volume III, provides frequency distributions for the base year through fourth follow-up.

NCES 87100, the final NLS-72 technical report, details methods of collection, response rates, and other technical information.

Getting Started with the Data Collected Across the Secondary Longitudinal Studies from 1972-2000

Slide 14 of 16

The most important document that should be read to ensure accurate analysis of HS&B sophomore cohort data is High School and Beyond Fourth Follow-Up Methodology Report. Technical Report (NCES 195426), which can be accessed by clicking the corresponding underlined screen text.

This report describes and evaluates the methods, procedures, techniques, and activities that produced the fourth (1992) follow-up of the High School and Beyond (HS&B) study.

The most important document that should be read to ensure accurate analysis of HS&B senior cohort data is High School and Beyond Third Follow-Up (1986) Sample Design Report (NCES 88402), which can be accessed by clicking the corresponding underlined screen text.

This report provides information that documents major technical aspects of the third follow-up sample selection and implementation, describes the weighting procedures, documents major technical aspects of the third follow-up sample selection and implementation, examines the possible impact of nonresponse on sample estimates, and evaluates the precision of estimates derived from the sample.

Slide 15 of 16

The three most important documents that should be read to ensure accurate analysis of NELS:88 data are: User's Manual: NELS:88 Second Follow-Up: Student Component Data File (NCES 94376), National Education Longitudinal Study (NELS:88/94) Methodology Report (NCES 96174), and User's Manual: NELS:88 Base-Year to Fourth Follow-up: Student Component Data File (NCES 2002323). Each of these technical documents can be accessed by clicking the corresponding underlined screen text.

Appendix H, within NCES 94376, provides information on NELS:88 student data weights, flags, and composite variables for the base year, and first and second follow-ups. Additionally, socioeconomic status, or SES, variables are described. NCES 96174 provides information on the subsampling and data collection activities for the third follow-up. NCES 2002323 is the final NELS:88 technical report.

Slide 16 of 16

This module introduced users to the data collected across the Secondary Longitudinal Studies from 1972-2000. It described the process of obtaining and accessing the NLS-72, HS&B, and NELS:88 data files and also described the resources that are available to learn more about the studies, the data, and the data files.

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.