

---

# **Quick Guide to Using the NELS:88/2000 Data**

---

**Taken From: *Base-Year to Fourth Follow-up  
Data File User's Manual: Appendix A  
(NCES 2002–323)***

## Quick Guide to Using the NELS:88/2000 Data

The purpose of the “Quick Guide” is to orient potential users of the NELS:88/2000 data to suggested techniques for working with the data files. Special attention will be paid to topics that will help users avoid the most commonly made mistakes in working with NELS:88 data. This guide is meant to serve as an introduction, not a replacement for the *NELS:88 Base-Year to Fourth Follow-up Data File User's Manual*. The first two sections of this guide provide a brief overview of the NELS:88 survey and available data files. The third section provides general instructions on how to get started using the NELS data and an orientation on the software that can be used to manipulate the data. The final section contains a series of frequently asked questions (FAQs) that NCES receives from users. The questions are followed by responses from NCES.

### A.1 Introduction to NELS:88

#### A.1.1 Overview

During the spring term of the 1987-1988 school year, the National Center for Education Statistics (NCES) initiated a national longitudinal study of 8<sup>th</sup>-grade students attending 1,052 high schools across the United States. A total of 24,599 8<sup>th</sup>-graders were surveyed in the base year of NELS:88. Many of these same students were re-surveyed in 1990, 1992, 1994, and 2000. Depending on the year, data were also collected from parents, schools, and teachers, and from extant high school and postsecondary transcripts. In addition, cognitive tests were administered during the first three data collection waves in the 8<sup>th</sup>, 10<sup>th</sup>, and 12<sup>th</sup> grades. In total, the NELS:88/2000 data contain extensive information on a large subsample of the original 1988 sample (approximately 25,000) and include five waves of data (12,144 cases) which can be categorized into the following groupings:

- Student data (junior high/middle school and high school data: 1988, 1990, and 1992)
- Dropout data (1990 and 1992)
- Post-high school data (1994 and 2000) collected after scheduled high school graduation
- School administrator data (1988, 1989, 1990, and 1992)
- Teacher data (1988, 1990, and 1992)
- Parent data (1988 and 1992)
- High school transcript data (1989-92), collected in the fall of 1992<sup>1</sup>
- Postsecondary transcript data (1992-2000), collected in the fall of 2000

---

<sup>1</sup> Academic transcript data were systematically collected for grades 9, 10, 11, and 12 in the NELS:88 second follow-up. However, when school transcripts contained information about 8<sup>th</sup> grade or earlier course work, this information was preserved in the transcript file. Likewise, some students took one or more college-level courses while still in high school; this information, too, will appear on transcript files, and it is therefore possible for a student to have postsecondary courses recorded for a year prior to 1992.

### **A.1.2 Major Features**

The major features of NELS:88 include the integration of student, dropout, parent, teacher, and school data; the initial concentration on an 8<sup>th</sup>-grade student cohort with follow-ups at two-year intervals; the inclusion of supplementary components to support analyses of geographically or demographically distinct subgroups (e.g., regions of the country, Hispanics and Asian subgroups, private school students); and the design linkages to previous longitudinal studies and other current studies.

### **A.1.3 Research Issues that can be Addressed**

The longitudinal design of NELS:88 permits the examination of change in young people's lives and the role of schools in promoting growth and positive life outcomes. By design, the basic unit of analysis is the student, with the parental, school, and teacher components providing contextual information. In particular, data from NELS:88 can be used to investigate a multitude of research topics including:

- Completion of high school for those students who drop out of middle or high school;
- Access and choice to postsecondary schools;
- Persistence and completion of postsecondary degrees;
- The relationship between base-year demographic, academic, and family characteristics, and later educational and employment outcomes;
- The relationship between high school course taking and later postsecondary outcomes;
- The high school and postsecondary experiences and academic performance of language minority students;
- Students pursuit of the study of mathematics and science;
- Transitions from high school and postsecondary education into the world of work;
- Family formation, including marital status and children; and
- Trend analyses with previous longitudinal studies (e.g., NLS-72 and HS&B).

Given the number of issues that can be addressed by NELS:88, it is not surprising that some prospective users of the data have been heard to speculate that the number of potential analyses that can be conducted using the NELS:88/2000 data base is limited only by the imagination of the individual researchers. This view, while not totally defensible, does appear to have face validity when one first examines the numerous files available to analysts. Reality begins to emerge though, once analysts realize that NELS:88 does have both substantive and methodological limitations. There are design constraints (e.g., NELS did not sample regional or area vocational schools) and limitations of the data (e.g., small cell sizes for certain groups of individuals) that must be taken into consideration when planning analyses that use NELS:88.

### **A.1.4 Must Read Publications**

Before a researcher attempts to use the NELS:88 data files, it is strongly suggested that time be spent reading the NELS:88 user's manuals and design documents that reference the NELS:88 base-year and first three follow-up studies. The following list of documents will

provide researchers with much of the information that they will need to understand the complexities of the NELS:88 data files.

### ***Manuals/Technical Documentation***

- Base-year to Fourth Follow-up Data File User's Manual  
This manual documents the history of NELS:88 from the base-year through the fourth follow-up studies. It is the single most valuable document for working with the NELS:88/2000 data.
- Base-year Sample Design Report  
This report documents the procedures and results of data collection for the NELS:88 base-year survey of 8<sup>th</sup>-graders.
- Second Follow-up Student Component Data File User's Manual  
This manual documents the data collection activities of the second follow-up data collection and processing activities. It also includes information on the base-year and first follow-up data collections. In addition, it contains a copy of the third follow-up survey instrument.
- Third Follow-up Methodology Report  
This report documents the response rates for the study and the subsampling decisions that were made for the third follow-up study.
- Base-year Through Second Follow-up Psychometric Report  
This reports documents the base-year through second follow-up assessments in mathematics, reading comprehension, science, and social studies/history.

NOTE: Each of these manuals can be found on the NCES Web Site

<http://nces.ed.gov/pubsearch/getpubcats.asp?sid=023>

From the NCES web site, documents can be searched and downloaded.

### ***NCES Reports***

Beginning with the initiation of NELS:88, NCES has produced selected reports using the NELS:88 data. These reports can be found in electronic format on the NCES Web Site under <http://nces.ed.gov/pubsearch/getpubcats.asp?sid=023>.

### ***Other Reports***

To aid researchers in locating reports that have used NELS:88 data, NCES contracted with the Research Triangle Institute (RTI) to produce a comprehensive annotated bibliography of reports (including dissertations) that used NELS:88 data. This bibliography can be found on the NELS:88 Web Site at:

<http://nces.ed.gov/surveys/nels88/>

## A.2 Description of NELS:88 Files and Electronic Codebooks

The NELS:88 surveys are available in both public and restricted use versions. For both the public and restricted versions, the following waves of NELS:88 data are included:



Because of subsampling decisions (especially at the third follow-up study), the NELS:88 data are most efficiently utilized as three separate data sets representing three distinct populations of respondents.

These data sets are discussed below.

### A.2.1 Data Set # 1: NELS:88 Base Year (1988) through Second Follow-up (1992)

This file contains 27,394 cases and includes all respondents who participated in any of the first three waves (base year, first follow-up, or second follow-up). This includes base-year ineligible and freshened respondents (see user's manual for description of base-year ineligible and freshened students). Data for each student for the period 1988 (base year) through 1992 (second follow-up) can be thought of as one continuous record that contains the following sections:

1. Base-year student questionnaire and assessment data;
2. First follow-up student questionnaire and assessment data;
3. Second follow-up student questionnaire, assessment, and transcript data;
4. First follow-up dropout questionnaire and assessment data;
5. Second follow-up dropout questionnaire, assessment, and transcript data;
6. Base-year school administrator data;
7. First follow-up school administrator data;
8. Second follow-up school administrator data;
9. Base-year parent data;
10. Second follow-up parent data;
11. Base-year teacher data;
12. First follow-up teacher data; and
13. Second follow-up teacher data.

The first 3-4 characters of each variable name identify the section that the variable belongs to (e.g., BYS = Base-year Student; F2P = Second Follow-up Parent). At the end of the first and second follow-up student sections, the composite variables (and weights) are followed by responses for freshened students. At the end of the second follow-up freshened student variables, the record contains composite (summary) high school transcript variables.

Restricted-use version: This data set contains 27,805 cases and has not been subjected to the rigorous disclosure risk analysis to which the public-use version employed. (For example, variables have not been top- or bottom-coded.) The restricted version contains the following data sets (or megafiles):

- High school transcript course-level data (714,614 records);
- Base-year through second follow-up school-level data (2,451 records, including links to Common Core of Data, and Quality Education Data universe file; school zip code data; and middle grades school organization and reform practices); and
- Christian School Supplement data (817 records).

### A.2.2 Data Set # 2: NELS:88/94 (Base Year through Third Follow-up)

This file contains 14,915 cases subsampled from the base-year through second follow-up file. In addition to the sections listed above for the base-year through second follow-up data set, this file contains third follow-up student data and two separate institution files (institutions attended by NELS:88 students).

It should be noted that **this file does not contain cross-sectional weights for the base-year, first follow-up, or second follow-up files.** The 14,915 cases on this file should not be used for cross-sectional analysis of base-year, first follow-up, or second follow-up data.

Restricted-use version: This data set contains 14,915 cases and has not been subjected to the rigorous disclosure risk analysis to which the public-use version employed. In addition to the student-level data file, the restricted version contains:

- Postsecondary education attendance data (11,560 records)
- Institution information (2,771 records which include the IPEDS code of the institution)

### A.2.3 Data Set # 3: NELS:88/2000 (Base Year through Fourth Follow-up)

This file contains 12,144 cases subsampled from the sampling frame for the third follow-up study. In addition to the sections listed above (see BY through third follow-up), this file contains fourth follow-up student data.

Again, *this file does not contain cross sectional weights for the base-year, first follow-up, second follow-up, or third follow-up files*. The 12,144 cases on this file should not be used for cross-sectional analysis of base-year, first follow-up, second follow-up, or third follow-up data.

Table A-1 provides a summary of the data order and content for the NELS:88 data files by survey year. Note that this table describes both student-level files (including parent, teacher, and administrator data about the student) (items 1-15) and postsecondary institution-level data (items 16-18). The data user should be aware that the institution files may not be identified uniquely by student identification number or may not include student identification numbers. Information about these files and merging the student and school files is described in the fourth follow-up ECB (see especially the descriptions for STU\_ID, INCODE, and INSTNUM).

N0P comprises four data files (N0R comprises seven). The four N0P data files are: BYF4STU, PSEF3F4, INSTF3F4, and PSE1994. BYF4STU contains student level information across all rounds (1988-2000), including student, dropout, school, teacher, and parent data, and summary variables from the high school transcript component. PSEF3F4 contains information about postsecondary enrollment in 1994 and 2000. INSTF3F4 contains postsecondary institution-level data for the 1994 and 2000 rounds. Finally, PSE1994 contains data about postsecondary enrollment spells for the entire third follow-up sample (including individuals not represented in 2000).

#### **A.2.4 Description of NELS:88 Electronic Codebooks (ECBs) for Public Release Data**

Each of the three public use data sets described above is associated with a separate Electronic Codebook (ECB). The ECBs are tools that allow the user to browse through the lists of NELS:88 variables, variable descriptions, and frequencies.

**Table A-1.—NELS:88 student-level datafile content, by survey and ECB: 1988-2000**

List Order*	Base Year (1988)	1 <sup>st</sup> Follow-up (1990)	2 <sup>nd</sup> Follow-up (1992)	3 <sup>rd</sup> Follow-up (1994)	4 <sup>th</sup> Follow-up (2000)
1.	BY student data				
2.		F1 student data			
3.			F2 student data		
4.		F1 dropout data			
5.			F2 dropout data		
6.	BY school administrator data				
7.		F1 school administrator data			
8.			F2 school administrator data		
9.	BY parent data				
10.			F2 parent data		
11.	BY teacher data				
12.		F1 teacher data			
13.			F2 teacher data		
14.				F3 student data	
15.					F4 student data

  

The diagram illustrates the coverage of three different data file periods across the survey waves:

- N2P ECB (27,394 students):** Datafile=stmeg.pub. This period covers the Base Year (1988) and the 1st Follow-up (1990).
- N4P ECB (14,915 students):** Datafile=stmeg3.pub. This period covers the 1st Follow-up (1990) and the 2nd Follow-up (1992).
- N0P ECB (12,144 students):** Datafile=BYF4stu.dat. This period covers the 2nd Follow-up (1992), the 3rd Follow-up (1994), and the 4th Follow-up (2000).

\* List order refers to the order the variables appear on the data file and ECB.

Table A-1.—NELS:88 student-level datafile content, by survey and ECB: 1988-2000—continued

List	Base Year	1 <sup>st</sup> Follow-up	2 <sup>nd</sup> Follow-up	3 <sup>rd</sup> Follow-up	4 <sup>th</sup> Follow-up
16.				F3 attendance	F4 attendance
17.				F3 institution	F4
18.				F3 enrollment	

\* List order refers to the order the variables appear on the data file.

ECBs allow the user to search a list of variables based on key words or labels; tag (i.e., select) variables for analysis; generate SAS and SPSS syntax for system files; produce printed codebooks of selected variables; import tag files; and access database files for extraction. See the NELS:88/2000 ECB guide for a full description of the functions of the ECB.

Separate ECBs have been developed for the NELS:88 base year through second follow-up, NELS:88/94, and NELS:88/2000 data files. The three ECBs are:

***ECB # 1: N2P (NELS:88/92 Public-Use ECB)***

This ECB contains the software that will allow researchers to work with the base-year through second follow-up data. The base-year through second follow-up data can be used to examine 8<sup>th</sup>-graders, 10<sup>th</sup>-graders, 12<sup>th</sup>-graders, 8<sup>th</sup>-graders who make it as far as 10<sup>th</sup> grade, 8<sup>th</sup>-graders who make it as far as 12<sup>th</sup> grade, 10<sup>th</sup>-graders who make it as far as 12<sup>th</sup> grade, and dropouts. (27,394 cases – includes 24,599 original base-year students and additional freshened and base-year ineligible students).

***ECB # 2: N4P (NELS:88/94 Public-Use ECB)***

This ECB contains the software that will allow researchers to work with the base-year through third follow-up data. The base-year to third follow-up data can be used to examine students and dropouts who are still in the sample during the 1994 third follow-up (14,915 cases – subsampled during third follow-up).

**Note:** It should be pointed out that N4P should not be used to do cross-sectional analyses of 8<sup>th</sup>-, 10<sup>th</sup>-, or 12<sup>th</sup>-graders. The appropriate file for this purpose is the base-year through second follow-up data set (see N2P ECB above).

***ECB # 3: N0P (NELS:88/2000 Public-Use ECB)***

This ECB contains the software that will allow researchers to work with the base-year through fourth follow-up data. The base-year through fourth follow-up data can be used to examine students and dropouts who are still in the sample during the 2000 fourth follow-up (12,144 cases).

**Note:** It should be pointed out that this file should not be used to do cross-sectional analyses of 8<sup>th</sup>-, 10<sup>th</sup>-, or 12<sup>th</sup>-graders—the appropriate file for this purpose is the base-year through second follow-up data set (see N2P ECB above).

## **A.2.5 CD-ROMs**

The above NELS:88 data sets, ECBs, and supporting documentation for the NELS:88 base year through second follow-up, NELS:88/94, and NELS:88/2000 data files are located on two separate CD-ROMs.

***CD-ROM (NCES 2000-328)<sup>2</sup>***

This data product contains:

---

<sup>2</sup> This CD was initially released in March 1996 as NCES 96-128.

- NELS:88/92 data, including cross-section data from the base-year, first follow-up, and second follow-up studies;
- NELS:88/94 data (panel data from the third follow-up respondents);
- ECB software (discussed above);
- An ECB guidebook; and
- Electronic copies of relevant NELS:88 user's manuals.

**CD-ROM (NCES 2002-322)**

This newly released product contains the NELS:88 fourth follow-up data (NELS:88/2000) and other useful products. For example, the CD includes:

- Electronic Codebook (ECB) software (discussed in the next section);
- Copy of the ECB guidebook;
- Electronic copy of the *NELS:88 Base-Year to Fourth Follow-up Data File User' Manual*;
- Printable images of the base-year through fourth follow-up study data collection instruments; and
- Annotated bibliography of research employing NELS:88 data.

**A.3 Getting Started**

This section addresses three questions:

1. What you need to know to get started using NELS:88,
2. How to navigate through the data, and
3. How to generate program syntax to manipulate the data.

These three questions are addressed below.

**A.3.1 Question # 1: What do I need to know to get started using NELS:88?**

***Minimum requirements***

1. Obtain a CD-ROM with the NELS:88 data. Two CDs are currently available: NCES 2000-328 includes the cross-sectional and panel data for the base-year and first three follow-ups. NCES 2002-322 includes BY-F4 panel data from the fourth follow-up study.
2. Have access to a computer: The Windows-based ECB requires 4.0 MB of storage space.
3. Develop an analytical strategy for working with data. The sheer number of variables available in NELS:88, and the complex, longitudinal nature of it, make the need for an analytical strategy very important.

### **Loading and Using the ECB**

1. Install the ECB:
  - Place CD-ROM into CD-drive.
  - From Windows, click on “START” and then “RUN.”
  - Browse through CD-ROM Drive for “ecbw” folder and open “SETUP.EXE” file.
  - Setup will guide you through the installation of the ECB.
  - Click on ECB icon to run.
2. The user is ready to use the ECB once it is installed. By clicking on each “hot” key on the tool bar found at the top of the ECB screen, the user will quickly understand the structure of the file and the power provided by the ECB to produce data files. At this point, the user should consult the “Electronic Codebook Help Guide” available on the CD-ROM for a specific overview of the ECB functions. (This is a file named HELP.PDF.)
3. Examine the frequencies available for each variable on the ECB. By examining these data descriptions, the NELS:88 user will begin to appreciate the complexity of collecting data from human subjects (legitimate values, legitimate skips, refusals, etc.). It is important to realize that some respondents:
  - Did not respond to the entire instrument;
  - Skipped individual items;
  - Refused to complete selected items;
  - Did not reach the end of the questionnaire;
  - Completed abbreviated versions of the instrument;
  - Made illegal skips; and
  - Responded outside pre-defined valid ranges.

#### **A.3.2 Navigating through the NELS:88 ECB (Identifying a Model and Tagging Variables for Analysis)**

1. Define the base population for analysis and whether longitudinal or cross-sectional analysis is required.
2. Develop a conceptual model. What does prior research suggest is happening with the data (e.g., characteristics of students who are likely to drop out of school)?
3. Determine the *predictor variables* (e.g., high absenteeism, disadvantaged background, low test scores), *intervening processes* (e.g., courses completed, trouble with law, pregnancy), and *outcomes* (e.g., event of dropping out, completion of GED, degree attainment, income) that can be used to explain the model.
4. Determine which components (variables) of your model can be addressed with NELS:88 variables. If multiple sources of the same item are available on the data files (e.g., a parent’s report of family income versus a student’s report), choose the one believed to be most reliable and valid. If the variables that the

researcher needs are not available on the NELS:88 files, he/she should consider merging variables from other sources (e.g., Census, Common Core of Data) by working with the restricted-use files (versus the public-use files) through an NCES licensing agreement. A license is necessary to protect respondents since the restricted-use files contain individual level identifiers. Additional information can be located at:

<http://nces.ed.gov/pubsearch/licenses.asp>

5. Re-think original model. If the variables contained on the NELS:88 data files cannot be used to study original model, rethink the model and either modify the model or choose another data set.
6. The user can tag the variables of interest by clicking on the “tag box” next to each variable.
7. The analyst must also remember to choose the appropriate weights and flags for the population of interest. In each data file, flags can be selected to identify a particular part of the population. For example, flags are available to identify whether a student was a dropout at a particular point in time (e.g., third follow-up). Weights are variables placed on the dataset to compensate for the unequal probabilities of selection and to adjust for non-response. When used with flags, weights allow the analyst to make generalizations about the national populations represented by the various NELS:88 samples. When weights are not used and/or when a flag is used inappropriately, the estimates generated will not be representative of the population.

### **A.3.3 Generating SAS or SPSS Program Code and Codebook Text**

1. After tagging the variables of interest, go to “File” and then “Output.”
2. Select the program (e.g., SPSS to generate SPSS program code).
3. Specify directory and name of program code file.
4. Select appropriate button in “Confirmation” box.
5. To view the program code, select “File” and then “View Output.”
6. The program code can then be opened in the appropriate software (e.g., SPSS) to generate a working system file and run analyses. It may be necessary to modify the program slightly (check for “execute” statements, period locations and file names). The code should identify the ASCII data file location which will be the CD-ROM.

### **A.4 Frequently Asked Questions About NELS:88 (FAQs)**

Since the first release of NELS:88 data in 1990, NCES staff members have received many questions regarding “proper techniques” for working with the data. In this document, these questions (along with NCES responses) have been categorized into topical areas and presented as a guide. It is hoped that the responses will help users avoid the most commonly made mistakes in working with this important data source. This document is meant to serve as

an introduction or supplement, not a replacement, for the Base-Year to Fourth Follow-up Data File User's Manual.

To help the data user identify specific topics of interest, questions and responses have been grouped into the following categories:

**FAQs Table of Contents**

A.4.1	General and Background Questions.....	15
	What is NELS:88? .....	15
	What are some of the terms that I should be familiar with in dealing with NELS:88? .....	16
	What are some of the research issues that can be addressed with NELS:88?.....	16
	How were the NELS:88 data collection instruments designed? .....	17
	What are the interrelationships among the separate NELS:88 files?.....	18
	How is NELS:88 related to prior NCES longitudinal studies?.....	18
A.4.2	Sampling .....	19
	In simple terms, explain how the NELS:88 school and student samples were selected? .....	19
	Who do these schools and students represent? .....	20
	Did the NELS:88 core follow the same group of students through the first, second, third, and fourth follow-up studies? .....	21
	Why are there 27,394 cases on the NELS:88 public-use ECB when the base-year sample only contained 24,599 cases?.....	22
A.4.3	Weights .....	22
	What groups does the NELS:88 data set represent and how do I subset these groups? .....	22
	What are these flags and weights? .....	23
	Why do we need to use weights with the NELS:88 data? .....	23
	Why would unweighted estimates not be representative? .....	23
	Which weights and flags should I use in my analyses? .....	24
	Thanks for the description of the weights, but what does this mean in practical terms?.....	25
A.4.4	Design Effects.....	27
	Why do I need to use design effects when I do my significance testing?.....	27
	Is there another procedure that I can use to approximate the correct standard error?.....	27
	Will you please elaborate on this technique?.....	27
A.4.5	Electronic Codebooks .....	30
	Why does the NELS:88/94 CD contain two electronic code books? How are these ECBs different from the ECB on the NELS:88/2000 CD? .....	30
	What is the sequence of data files on the NELS:88 ECBs?.....	31
	When I receive my NELS:88 CD, what are some of the steps that I should follow to check out my CD? .....	32
A.4.6	Composite Variables.....	34
	What are the advantages of using composite variables in my analyses? .....	34
	How do I select variables for a working data file? .....	34
	How do I subset data files? .....	35

A.4.8	Transcripts .....	36
	How do I use the high school transcript file?.....	36
A.4.9	Privileged or Restricted-Use Data .....	36
	When do I need to use the restricted-use data file?.....	36

**A.4.1 General and Background Questions**

**Question: What is NELS:88?**

**Response:** During the spring of 1988, NCES initiated a longitudinal study of 8<sup>th</sup>-grade students attending 1,052 high schools across the fifty states and the District of Columbia. A subset of these students was resurveyed in 1990, 1992, 1994, and 2000, along with additional individuals who helped to form representative 10<sup>th</sup>-and 12<sup>th</sup>-grade cohorts. Data was collected from the students (even after some dropped out of school), their parents, schools, and teachers, and from extant high school and postsecondary transcripts. In addition, achievement tests (assessments in math, reading, science, and social studies) were administered during the first three waves. In total, almost 11,000 pieces of information were collected on a large segment (approximately 12,000) of the original sample of approximately 25,000 students. The NELS:88 survey system includes five waves of data which can be grouped into the following categories:

- 8<sup>th</sup>-grade/high school student data (1988, 1990, and 1992) including cognitive tests;
- Dropout data (1990 and 1992);
- Post-high school data (1994 and 2000) collected after scheduled high school graduation;
- Junior high/high school administrator data (1988, 1990, and 1992);
- Three waves of teacher data (1988, 1990, and 1992);
- Two waves of parent data (1988 and 1992);
- High school transcript data (1992); and
- Postsecondary transcript data (2000).

Given this breadth of coverage, it is not surprising that some prospective users of the data have been heard to speculate that the number of potential analyses that can be conducted using the NELS:88 database is limited only by the imagination of the individual researchers. This view, while not totally defensible, does have some face validity, especially when data users first examine the numerous files available to analysts. Reality begins to emerge though, once analysts begin to realize that NELS:88 does have some limitations. For example, there are design constraints (e.g., the NELS:88 sample did not sample area vocational schools) and data limitations (e.g., small cell sizes for certain groups of individuals) that must be taken into consideration when planning with the NELS:88 data.

The major features of NELS:88 include the integration of student, dropout, parent, teacher, and school studies; the initial concentration on an 8<sup>th</sup>-grade student cohort with follow-ups at 2-year intervals; the inclusion of supplementary components to support analyses of

demographically distinct subgroups (i.e., Hispanics, Asians, private school students); and the design linkages to previous longitudinal studies and other current studies.

**Question:** *What are some of the terms that I should be familiar with in dealing with NELS:88?*

**Response:** Knowledge of the following terms will help the user in reading through the following questions and responses. Additional information on these and other terms can also be found in appendix E (the NELS:88 glossary) in this data file user's manual.

<i>BYI:</i>	Base Year Ineligible students—students who were determined to be ineligible for the base year many of whom became eligible in subsequent follow-ups
<i>Bias:</i>	respondents differ from nonrespondents
<i>Cohort:</i>	factor in common (year of birth or grade)
<i>Cross-section:</i>	represent events at single point in time
<i>Design effects:</i>	a measure of design efficiency, typically related to the precision of estimates
<i>ECB:</i>	Electronic codebook
<i>Freshening:</i>	adding students to original sample during later waves of data collection to create new grade-representative cohorts
<i>IRT:</i>	Item Response Theory (permits vertical scaling of assessments)
<i>Longitudinal:</i>	similar measurements at multiple points in time
<i>Panel:</i>	surveying same individuals across time
<i>Weights:</i>	used to produce population estimates, or in other words, when one respondent represents a number of others in the population

**Question:** *What are some of the research issues that can be addressed with NELS:88?*

**Response:** The longitudinal design of NELS:88 permits the examination of change in young people's lives and the role of schools, teachers, community, and family in promoting growth and positive outcomes. In particular, data from NELS:88 can be used to investigate a multitude of issues in the context of the family, community, school, and classroom including:

- Students' academic growth over time;
- The transition from 8<sup>th</sup> grade to high school;
- The process of dropping out of school, as it occurs from 8<sup>th</sup> through 12<sup>th</sup> grades;

- The role of the school in helping disadvantaged individuals;
- The school experiences and academic performance of minority students;
- Students' pursuit of the study of mathematics and science;
- The features of effective schools;
- Access to and choice of postsecondary schools;
- Transitions to postsecondary education and the world of work;
- Educational outcomes and their relationships with workplace outcomes;
- Marriage, divorce, and other aspects of family formation among the 8<sup>th</sup> grade class of 1988; and
- Trend analyses with previous longitudinal studies (e.g., NLS-72 and HS&B).

This question is also addressed more fully in appendix B of the NELS:88/2000 descriptive report (Ingels, Curtin, Kaufman, Alt, and Chen 2002 [NCES 2002-321]).

**Question: How were the NELS:88 data collection instruments designed?**

**Response:** Instrument development was guided by the research objectives of NELS:88. Questionnaires and interviews were designed to meet the longitudinal goals of the study; items were chosen based on their utility in predicting or explaining future outcomes as measured in later survey waves. All of the instruments employed in the base-year through fourth follow-up studies were developed to provide continuity and consistency with earlier education longitudinal studies (NLS-72 and HS&B), as well as to address new areas of policy concern and to reflect recent advances in theory. In general, the process for each survey instrument consisted of the following steps:

1. NCES development of list of topics;
2. Contractor development of a content outline;
3. Content outline shared with other government agencies, policy groups, and interested parties;
4. Draft copy of survey instrument developed;
5. Review by the NELS:88 Technical Review Panel (a specially appointed, independent group of substantive, methodological, and technical experts);
6. Survey instrument revised based on reviewer comments;
7. Justification written for components of instruments;
8. NCES review of instruments;
9. Review of instruments by the federal Office of Management and Budget (OMB); and
10. Field testing of instruments, and revision based on field test results.

Scanned images of 15 data collection instruments from all waves of NELS:88 and from all respondent groups are included in the instrument folder on the NELS:88/2000 ECB.

**Question:** *What are the interrelationships among the separate NELS:88 files?*

**Response:** Using common IDs, the individual data files comprising NELS:88 have been merged with each other to form data files containing student, parent, school, and teacher data. By design, the basic unit of analysis for most NELS:88 analyses will be the student. Under this premise, the school administrator, parent, and teacher data can be thought of as providing contextual (e.g., background, school characteristics, “opportunity to learn”) data.

Because the base-year of NELS:88 involved the random selection of 1,000 schools from across the United States that contained 8<sup>th</sup>-graders, the 8<sup>th</sup>-grade school sample can be used (in conjunction with the 1988 school weight, BYADMWT) as a standalone file in which the school is the basic unit of analysis. The first and second follow-up school files, however, are not nationally representative of high schools with 10<sup>th</sup> or 12<sup>th</sup> grades, and therefore no school weight has been generated for them. These schools were not selected by a probabilistic method, but rather, entered NELS:88 by virtue of containing students who participated in NELS:88 during the base-year.

Universe variables have been constructed to provide researchers with a history of the involvement of each student over the base-year and four follow-up studies of NELS:88. These variables show the status of each student during the five data collection periods. For example, a given student may have been eligible for participation in the base-year study, a dropout during first follow-up data collection, back in school during the second follow-up, and subsampled for the third and fourth follow-ups. Universe variables can be used to subset cases to desired populations. Universe variables effectively limit the working data file to respondents who fit selected criteria (e.g., dropouts who are still part of the study at the fourth follow-up). The universe variables can be found at the beginning of the NELS:88 data files.

A special note on the organization and content of the data files in the NELS:88/2000 ECB (NOP or NOR) that relate to postsecondary education institutions might also be useful to data users. In contrast to the student-level files discussed in the previous paragraph, the NELS:88 postsecondary education data files are structured somewhat differently (e.g., at the institution or student-institution levels). Analysts will be able to merge this postsecondary information to the NELS:88 sample members with the careful use of the unique identifiers associated with each file. For example, the PSE attendance data in the NELS:88/2000 ECB (PSEF3F4.DAT) are organized on a sample member (STU\_ID) and school identification number (INCODE) basis. INSTF3F4.DAT is an institution-level file with INCODE as the unique identified. More information on merging these data files can be found in the variable descriptions for the items contained in the ECB.

**Question:** *How is NELS:88 related to prior NCES longitudinal studies?*

**Response:** All of the student and dropout questionnaires employed in the base-year, first follow-up, second follow-up, third follow-up, and fourth follow-up studies of NELS:88/2000 were designed to provide continuity and consistency with earlier education longitudinal studies.

(Note: Only student questionnaires were administered in the base-year study; separate student and dropout instruments were administered in the first and second follow-ups; a combined student/dropout interview was employed in the third and fourth follow-up data collection.) Specific items in the NELS:88 documents are replicates of items appearing in instruments from the National Longitudinal Survey of 1972 (NLS-72) and the 1980 and 1982 waves (i.e., the sophomore and senior cohorts) of High School and Beyond (HS&B). The comparability and consistency of items across these three data sets allow for (but are not limited to) the conduct of the following trend analyses:

- NELS:88 1990 sophomores can be compared to HS&B 1980 sophomores;
- NELS:88 1990 sophomores two years later (that is, in 1992) can be compared to HS&B 1980 sophomores two years later in 1982);
- NELS:88 1990 sophomore cohort dropouts (as of 1992) can be compared to HS&B 1980 sophomore cohort dropouts (as of 1982);
- NELS:88 1992 seniors can be compared to HS&B 1980 seniors, and NLS-72 1972 seniors; and
- NELS:88 1992-collected high school transcripts can be compared to HS&B 1982-collected high school transcripts (the 1980 and 1990 sophomore cohorts, two years later).

Comparisons are also possible using high transcript data collected for 1992 (NELS:88 senior cohort members); 1982 (HS&B seniors); and 1987, 1990, 1994, and 1998 high school graduates in NAEP schools. Analyses of postsecondary transcript data are also possible between NELS:88, NLS-72 and the HS&B cohorts.

#### **A.4.2 Sampling**

**Question:** *In simple terms, explain how the NELS:88 school and student samples were selected?*

**Response:** Base Year: The NELS:88 schools were selected from a universe file of approximately 40,000 public and private 8<sup>th</sup>-grade schools across the United States. For the approximately 1,000 public and private schools with 8<sup>th</sup> grades that were sampled and agreed to participate in NELS:88, complete 8<sup>th</sup>-grade rosters were produced for each school. From this roster, approximately 24 students were randomly selected. The remaining students on the roster were then grouped by race and ethnicity, and additional 2-3 Asian and Hispanic students were then selected for each school.

First Follow-up: Prior to the first follow-up data collection period, approximately 90 percent of the students moved from a K-8/junior high school/middle school setting to high school. Because of these transitions, students had to be traced to their new schools. In addition, school dropouts needed to be identified, contacted and convinced to participate in the follow-ups. New (freshened) students needed to be added to the sample so that the first follow-up data would be representative of high school sophomores. Without these freshened students, the file would not be representative of 10<sup>th</sup>-grade students who did not have the opportunity to participate during the base year (e.g., students who were out of the country in 1988 or who were 1990 10<sup>th</sup>-graders but not in 8<sup>th</sup> grade in spring term 1988). Also, 8<sup>th</sup>-graders who had been considered ineligible

(e.g., students with physical or mental disabilities or limited-English proficiency) during the base year but whose eligibility status had since changed were added in. These base year ineligible students were given the opportunity to participate during the 10<sup>th</sup> grade. Because of the wide dispersal of students, the base-year 8<sup>th</sup> grade cohort was subsampled.

Second follow-up: Prior to data collection, students needed to be traced. Many stayed in the same high school in which they were surveyed as 10<sup>th</sup>-graders. Others transferred to new high schools, graduated early, or dropped out of high school. It was also necessary to freshen the sample with 12<sup>th</sup> graders who did not have the opportunity to be surveyed in prior waves. While no students were subsampled out of the 1992 round, some components (full contextual data from school, transcripts, parent) are based on subsamples.

Third follow-up: Prior to data collection, the decision was made to subsample the NELS:88 respondent population to 14,000 respondents. Selected groups were selected with certainty (that is *all* were selected).

Fourth follow-up: Both respondents and nonrespondents for the third follow-up sample were selected for the 2000 survey. Subsampling was then performed upon sample members who had proved difficult to interview, producing a overall sample of 15,237 members of the NELS:88 population at the start of data collection.

**Question: *Who do these schools and students represent?***

**Response:** The 1,000 schools represent the approximately 40,000 public and private schools in the United States in 1988 that had 8<sup>th</sup>-grade students. The nearly 25,000 students sampled in NELS:88 represent the 3,000,000 8<sup>th</sup>-graders attending schools in 1988, with the exception of Bureau of Indian Affairs (BIA) schools, special schools for students with disabilities, area vocational schools that do not enroll students directly, and schools for dependents of U.S. personnel serving overseas.

NELS:88 data can be used to represent the following groups:

- 8<sup>th</sup>-grade schools in 1988 (cross-section): use N2P
- 8<sup>th</sup>-grade students in 1988 (cross-section): use N2P
- 10<sup>th</sup>-grade students in 1990 (cross-section): use N2P
- 8<sup>th</sup>-grade respondents who are still in the study in 1990 (panel): use N2P
- 12<sup>th</sup>-grade students in 1992 (cross-section): use N2P
- 8<sup>th</sup>-grade students who participate in 1<sup>st</sup> and 2<sup>nd</sup> follow-ups (panel): use N2P
- 10<sup>th</sup>-grade respondents who are still in study in 1992 (panel): use N2P
- NELS:88 respondents two years after scheduled high school graduation (cross-section): N4P
- 8<sup>th</sup>-grade students who are still in study in 1994 (panel): use N4P
- 10<sup>th</sup>- and 12<sup>th</sup>-grade students who are still in study in 1994 (panel): use N4P

- NELS:88 respondents eight years after scheduled high school graduation (cross-section): NOP
- 8<sup>th</sup>-grade students 12 years later (panel): use NOP
- 10<sup>th</sup>-grade students 10 years later (panel): use NOP
- 12<sup>th</sup>-grade students 8 years later (panel): use NOP

**Question:** *Did the NELS:88 core follow the same group of students through the first, second, third, and fourth follow-up studies?*

**Response:** Although the major objective of NELS:88 was to follow a group of 8<sup>th</sup>-graders, there were both additions and deletions to the sample as it progressed between 1988 and 2000. The additions included the augmentations of the base-year sample with freshened and Base Year Ineligible (BYI) students. The deletions included subsampling performed several times during the study. These augmentations and deletions are described below.

Freshened students: The NELS:88 sample was freshened with additional 10<sup>th</sup>-graders in 1990 and additional seniors in 1992. These students were added so that the sample would be nationally representative of sophomores and seniors in those years. Students who were freshened into the sample did not have the opportunity to be selected into the sample during the 8<sup>th</sup> grade (e.g., they may have been out of the country or out of grade sequence).

BYI students: In addition to freshened students, some base-year ineligible (BYI) students were also added to the first and second follow-up samples. BYI students were the individuals (approximately 5%) who were deleted from the 8<sup>th</sup>-grade sampling process by the school principal or headmaster who determined that these students' lack of English language proficiency, or physical or mental disability, would make it unduly difficult for them to complete self-administered questionnaires or cognitive tests, or would not produce a valid assessment of their abilities. These students were re-evaluated during the first and second follow-up studies. Those whose status had changed (e.g., they had become proficient in English) such that they could now participate were returned to the study.

Subsampling: The NELS:88 sample was subsampled at several points in the study. The initial subsampling occurred during the first follow-up. Because students moved from 1,000 junior high school settings (8<sup>th</sup> grade) to almost 5,000 high school settings (by 10<sup>th</sup> grade), a decision was made to subsample those individuals who moved to schools that enrolled few NELS:88 students (i.e., fewer than 10 NELS:88 students). Major subsampling also took place during the third and fourth follow-up study.

In addition, the researcher also should be aware that for budgetary reasons, high school transcripts were not collected from all students (17,100 during second follow-up). Thus, there is a separate weight for students with these transcripts.

**Question: Why are there 27,394 cases on the NELS:88 public-use ECB when the base-year sample only contained 24,599 cases?**

**Response:** The base-year to second follow-up data set included in the public-use ECB (N2P) and restricted use ECB (N2R) contain all cases that were ever part of NELS:88. This includes individuals who were ineligible to participate in the base year, as well as those who were freshened into the study during the first or second follow-up studies. It is thus important to use flags and weights to create a working data set to delete cases that may be extraneous to your planned analysis. For example, as illustrated below (from N2R), for the first variable on the base-year 8<sup>th</sup>-grade student file ("BYS2A"), you might not want to keep the 3,206 cases that were blank for this variable (includes freshened and BYI students) if, for instance, you are analyzing only base-year data. These freshened and BYI cases were not in the sample during the base year.

BYS2A. Is your mother or female guardian living? (MARK ONE)

Code	Freq	Percent	Label
1	23,967	99.3	YES
2	160	0.7	NO
{blank}	3,206		{NONR/NOT IN SAMPLE THIS WAVE}
8	472		{MISSING}
TOTAL	27,805		

To select cases where there are data for the base year, you can either use the base-year weight "BYQWT" (BYQWT > 0) or choose on variable BYS2A (BYS2A NE "blank").

NOTE: For data users of the public-use ECB produced for the NELS:88 fourth follow-up study (N0P), the respondent population contained on this ECB (N = 12,144) reflects the NELS:88 sample members who were retained in 2000 and responded to the fourth follow-up interview. Thus, the frequencies and percentages provided in the example above (excerpted from the N2R ECB [N = 27,805]) will not be reproduced by the N0P ECB (which is based upon a subset [12,144] of sample members included in the N2P ECB [N = 27,394]).

#### A.4.3 Weights

**Question: What groups does the NELS:88 data set represent and how do I subset these groups?**

**Response:** The NELS:88 data represent many different populations (e.g., 8<sup>th</sup>-graders in 1988; sophomores in 1990; seniors in 1992; 8<sup>th</sup>-graders who were still in school at 12<sup>th</sup> grade; 8<sup>th</sup>-graders who dropped out of school by 10<sup>th</sup> grade; etc.). These groups can be identified through use of flags and analysis weights.<sup>3</sup>

<sup>3</sup> Analysis weights are also known as *nonresponse-adjusted weights*, and as *final weights*. They are to be distinguished from *raw weights* (or *design weights*), which have not been adjusted to compensate for patterns of nonresponse. Only analysis weights appear on the NELS:88 data files.

**Question: What are these flags and weights?**

**Response:** Flags are variables that were put onto the NELS:88 files to indicate status at a given point in time (e.g., dropout status) or a condition (e.g., handicapped). A series of useful flags (universe variables) can be found at the beginning of the data set. These universe variables give the status of each individual for each data collection (e.g., eligible during base-year, dropout during first follow-up, in-school during second follow-up). Flags can be used by the researcher to select cases for analyses. Most of the flags can be found at the end of each file (e.g., fourth follow-up student/dropout file) with composite and derived variables. For example, F4PNLFL is the base-year through fourth follow-up panel flag that indicates the sample member responded at each of the five waves of NELS:88 data collection.

Weights are variables that are put onto the file to compensate for unequal probabilities of selection and to adjust for the effects of nonresponse. Using weights allows a researcher to make generalizations to the national populations represented by NELS:88. On the NELS:88 student files for the base-year through second follow-up studies (N2P/N2R), there are 12 different analysis weights (these weights are described below). Each of these weights is specific for a given population. Depending on the group to whom the data are designed to generalize, the individual weights have positive values (>0) for respondents who are members of that particular group and zero (0) for all others. (There are nine different weights for the data presented in the NELS:88/2000 ECBs NOP and NOR.)

**Question: Why do we need to use weights with the NELS:88 data?**

**Response:** If we do not use weights, the estimates that we produce will not be representative of the population about which we are attempting to estimate.

**Question: Why would unweighted estimates not be representative?**

**Response:** In the base year of NELS:88, approximately 25,000 students were sampled from across the nation. These 25,000 8<sup>th</sup>-grade students represent the 3,000,000 students who attended 8<sup>th</sup> grade in the United States in 1988. Thus, each student represents approximately 120 students (3,000,000/25,000=120). But because some policy relevant groups (e.g., Asians, Hispanics, private school students) were over-sampled (greater than their proportion in the population), they are over-represented in the file. Depending on the sampling ratio, the weights for these students would be smaller than the average student. By the same token, other students may represent more than 120 students because they were under-sampled during the study. Nonresponse adjustment must also be taken into consideration because the weights of questionnaire nonrespondents are distributed among the respondents with similar characteristics. Thus, weights reflect both unequal probabilities of sampling and nonresponse adjustments. It is not unusual for a specific weight on a follow-up file to have a range of over 20,000 (e.g., F4QWT ranges from 8.71 to 20,898.71—yes, a single student represents 20,899 other students). Therefore, it is incumbent upon the researcher to use appropriate weighting variables.

**WARNING: The researcher should avoid breaking down the sample into such small categories that the analysis is questionable. For example, if a crosstabulation table has a single cell with fewer than 30 cases when the data are not weighted, NCES recommends that the results not be displayed or that the cell be combined with another cell (if appropriate).<sup>4</sup>**

**Question: Which weights and flags should I use in my analyses?**

**Response:** Table A-4 presents the various weights that can be used with the base-year to fourth follow-up public-use data for NELS:88. Included also in the table are the weights and sample flags that should be used with the analyses; and importantly, the study ECB that contains the data. This includes the fourth follow-up ECB (N0P) and the base-year to second follow-up ECB (N2P), released in 1996. The unweighted sample sizes for the analysis groups and the flags required to subset the data are also included (the required analysis [nonresponse-adjusted] weights for the sample are shown in parenthesis.) (Note: Some of the groups below must be further subset to represent meaningful analysis populations.)

**Table A-4.—Descriptions of important NELS:88 analysis groups with unweighted sample sizes and related weights and flags for the groups**

Analysis Group	ECB	Number (unwgt)	Flag and weight
1988 8 <sup>th</sup> -graders	N2P	24,599	IF BYQFLG=1; ( <b>BYQWT</b> )
1990 10 <sup>th</sup> -graders	N2P	17,753	IF F1SEQFLG=0; ( <b>F1QWT</b> )
1990 cross section (includes dropouts)	N2P	19,394	IF F1STAT=0; ( <b>F1QWT</b> )
1988 8 <sup>th</sup> -graders in 1990 (panel)	N2P	7,424	IF F1PANFLG=1; ( <b>F1PNLWT</b> )
1992 12 <sup>th</sup> graders	N2P	16,114	IF F2SEQFLG=0; ( <b>F2QWT</b> )
1992 cross section (includes dropouts)	N2P	19,220	IF F2STAT=00; ( <b>F2QWT</b> )
1988 8 <sup>th</sup> -graders in 1992 (panel)	N2P	16,489	IF F2PNLFLG=1; ( <b>F2PNLWT</b> )
1992 cross sectional analysis of students with high school transcript data	N2P	17,100	IF <b>F2TRSCWT</b> > 0
2000 cross sectional analysis with F4 respondents	N0P	12,144	( <b>F4QWT</b> )
1988-2000 panel analysis of 8 <sup>th</sup> grade class of 1988	N0P	10,827	IF F4PNLFL=1 ( <b>F4PNLWT</b> )
2000 cross sectional analysis of students with high school transcript data	N0P	10,310	IF <b>F4TRSCWT</b> > 0

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Longitudinal Study of 1988 (NELS:88), 1988-2000.

<sup>4</sup> Note that a row can be calculated by determining a mean of 0 and 100 values given to a response where the denominator is the sum of the crosstabulation variable. This frequency can also be shown as a percent of the class variable response by a crosstab response; in this case it is the sum of the row, not a single cell. In this example, no single cell can have a frequency less than 3.

**Question:** *Thanks for the description of the weights, but what does this mean in practical terms?*

**Response:** Perhaps the following examples can help clarify how the analysis weights can be used to help define your sample:

Example # 1: You are interested in examining the gains in math between the 8<sup>th</sup> and 12<sup>th</sup> grades.

Decisions that you need to make:

- 1. Which ECB should I use?** In this case, you are examining the BY to F2 period, so you would use N2P. If you wanted to examine the math gains and the impacts this might have had on future earnings, you would want to use N0P, the ECB that contains data from all five waves of NELS:88.
- 2. Should I use a cross-sectional or panel weight?** A panel weight would be more appropriate because you are following a group of students over time. In fact, there are two panel weights that would be appropriate (F2PNLWT and F2TRP1WT).
- 3. Which of these two panel weights do I use?** You can use either. The choice depends on whether you want to limit your study to those with transcripts or not. If you want to control for math course taking, the logical choice would be F2TRP1WT (the subset of the panel sample who have transcripts).
- 4. What do I do with the students who drop out of school?** If you are only interested in students who graduate on schedule, you need to drop high school non-completers from the sample.
- 5. What variable do I use to specify cases I want to delete?** The following variable "F2RTROUT" is a composite variable created from high school transcripts showing high school completion status in 1992.

Unweighted

Code	Freq	Percent	Label
01	13,471	77.9	SPRING 1992 GRAD
02	272	1.6	OTHER 1992 GRAD
03	154	0.9	PRE-1992 GRAD
04	22	0.1	DIPLOMA/SPEC ED
05	9	0.1	CERT OF ATTEND
06	420	2.4	STILL ENROLLED IN HIGH SCHOOL
07	2,003	11.6	DROPPED OUT
08	424	2.5	TRANSFERRED
09	25	0.1	AGED OUT
10	4	0.0	DIED
11	10	0.1	LEFT/HEALTH RSN
12	35	0.2	RECEIVED GED
13	87	0.5	OTHER
14	349	2.0	STATUS UNDETERM
{blank}	10,520	.0	{NONR/NOT IN SAMPLE THIS WAVE}

If you just wanted those who were high school graduates, you may want to select those with values of 1, 2, or 3 on **F2RTROUT**. Alternately, your final decision may be to only keep those who met the following criteria:

$F2TRP1WT > 0$  and  $(1 \leq F2RTROUT \leq 3)$ : These include only students with transcript data who graduated from high school some time during 1992 (values of 1, 2, or 3 on F2RTROUT). You would use F2TRP1WT to weight this sample.

Example # 2: You are interested in describing the sophomore class of 1990.

Decisions that you need to make:

1. **Which ECB should I use?** In this case, you are examining only the NELS:88/90 data period (i.e., the NELS:88 first follow-up study) so you would use the N2P ECB. Analysts who wish to track the 10<sup>th</sup>-grade cohort through 2000 must use the N0P ECB.
2. **Do I need a cross-sectional or panel weight?** In this example, a cross-sectional weight would be appropriate because it will also include students who were added (freshened) to the sample during the first follow-up to represent sophomores in 1990. Without these freshened students, the sample would only represent 8<sup>th</sup>-grade students who made it as far as the 10<sup>th</sup> grade. It would not have students who did not have the opportunity as 8<sup>th</sup>-graders of being in the study (e.g., out of the country during base-year sampling in 1988).
2. **What are the cross-sectional weights that I could choose?** The most logical choice is F1QWT. This weight represents a cross-section of respondents in 1990 (both students and dropouts regardless of grade).

**3. How do I limit my analysis to respondents who were 10<sup>th</sup>-graders in 1990?**

There are several flags that can be used to subset your sample. One possibility is to use variable "F1SEQFLG" from the first follow-up. Analysts should also become familiar with the Universe variables—the first variables list on each NELS:88 ECB.

Code	Freq	Percent	Label
0	17,753	84.5	ENROLLED IN 10 <sup>th</sup> GRD
1	726	3.5	ENROLLED IN OTHR GRD
2	2,540	12.1	NOT APPLICABLE
{blank}	6,786	.0	{NONR/NOT IN SAMPLE THIS WAVE}

Thus, for your analyses you would subset for those individuals with F1SEQFLG =0 and F1QWT > 0.

**A.4.4 Design Effects**

**Question:** *Why do I need to use design effects when I do my significance testing?*

**Response:** Because the NELS:88 sample design involved stratification, disproportionate sampling of certain strata (e.g., oversampling of Hispanics), and clustered (e.g., students within a school) probability sampling, the resulting statistics are more variable than they would have been had they been based on data collected from a simple random sample of the same size. A number of statistical packages (SUDAAN and STRATTAB are two of several possible examples) take account of complex sampling designs in the calculation of standard errors. The AM software does so as well (available for free download: [am.air.org](http://am.air.org)).

**Question:** *Is there another procedure that I can use to approximate the correct standard error?*

**Response:** Use of variance estimation software is recommended. However, the data file user's manuals for each wave of NELS:88 include tables of average design effects that can be used to adjust standard errors. Citations for these documents can be found in appendix D of the BY–F4 data file user's manual.

**Question:** *Will you please elaborate on this technique?*

NOTE: NCES recommends using a software package that produces adjusted standard errors. The following technique is only an approximation to the correct standard errors.

**Response:** The following example demonstrates how this technique may be used.

Example: The following question is posed: Do 12<sup>th</sup>-grade boys spend more time on school days playing video games than 12<sup>th</sup>-grade girls?

Procedure: Produce a cross-tabulation (using SAS or SPSS) that crosses gender (male vs. female) by self-reported time playing video games.

Step 1—Run this cross-tabulation both weighted and unweighted.

Table A - F2SEX by VIDEO

F2SEX(COMPOSITE SEX) VIDEO - WEIGHTED

Frequency				
Percent				
Row Pct				
Col Pct	DON'T PLAY VIDEO	LESS 1HR PER DAY	1 OR MORE HRS /DAY	Total
MALE	669986	399120	255511	1324616
	25.48	15.18	9.72	50.38
	50.58	30.13	<b>19.29</b>	
	40.26	64.66	73.42	
FEMALE	994082	218150	92514	1304746
	37.81	8.30	3.52	49.62
	76.19	16.72	<b>7.09</b>	
	59.74	35.34	26.58	
Total	1664068	617270	348025	2629363
	63.29	23.48	13.24	100.00

Table B - F2SEX by VIDEO

F2SEX(COMPOSITE SEX) VIDEO - UNWEIGHTED

Frequency				
Percent				
Row Pct				
Col Pct	DON'T PLAY VIDEO	LESS 1HR PER DAY	1 OR MORE HRS /DAY	Total
MALE	4107	2480	1507	<b>8094</b>
	25.10	15.16	9.21	49.47
	50.74	30.64	18.62	
	39.22	64.55	73.62	
FEMALE	6366	1362	540	<b>8268</b>
	38.91	8.32	3.30	50.53
	77.00	16.47	6.53	
	60.78	35.45	26.38	
Total	10473	3842	2047	16362
	64.01	23.48	12.51	100.00

Frequency Missing = 2858

NOTE: VIDEO RECODED FROM F2S34A

Step 2—Calculate standard errors for both males and females who play video games more than one hour per school day using the following formula.

St. Error =  $\text{SQRT} [( \text{weighted \%} * (100 - \text{weighted \%} ) ) / \text{unweighted } n]$ . Compute Standard Errors for males and females separately.

Male St. Error =  $\text{SQRT} [(19.3 * 80.7) / 8094] = \mathbf{0.44}$

Female St. Error =  $\text{SQRT} [(7.1 * 92.9) / 8268] = \mathbf{0.28}$

Step 3—For significance testing, pool the standard errors for males and females.

Pooled Standard Error =  $\text{SQRT} [( \text{St. Error for males} )^2 + ( \text{St. Error for females} )^2 ]$ .

Pooled St. Error =  $\text{SQRT} [(0.44)^2 + (0.28)^2] = \mathbf{0.52}$

The pooled standard error is used when you are making a comparison between two groups.

Step 4—Correct the pooled standard error for design effect.

Corrected Pooled Standard Error =  $\text{root design effect (DEFT)} * \text{Pooled St. Error}$

Mean Root Design Effect from table 3.3.1-12, page 55 = **1.89** (Second Follow-up Student Data File User's Manual (NCES 94-374))

Corrected Pooled Standard Error =  $1.89 * 0.52 = \mathbf{0.98}$

Step 5—Check for significance.

T-test =  $(\% \text{ males} - \% \text{ females}) / \text{corrected Pooled St. Error}$

T =  $(19.3 - 7.1) / 0.98 = \mathbf{12.44}$  {significant t test}

Males report playing videos for more than 1 hour almost three times the rate as females (19.3% vs. 7.1%).—See table A for weighted percents and table B for sample sizes (numbers in bold).

**SUMMARY – Use weighted percentages (estimates) from table A, unweighted sample sizes from table B, and mean design effect from data file user's manual. See individual user's manuals for more detailed descriptions of design effects.**

#### A.4.5 Electronic Codebooks

**Question:** *Why does the NELS:88/94 CD contain two electronic code books? How are these ECBs different from the ECB on the NELS:88/2000 CD?*

**Response:** The two ECBs on the NELS:88/94 CD (released in 1996) represent two samples. The first, **N2P** (public-use file) or **N2R** (restricted-use file), includes those respondents who participated in the base year, first follow-up, or second follow-up surveys (approximately 27,000 cases). The second ECB, **N4P** or **N4R**, includes only those cases who were subsampled and responded to the NELS:88 third follow-up study (approximately 14,000 cases). This second set of ECBs (N4P/N4R) has been superseded by the NELS:88/2000 ECB, which expanded data collection to 2000. In other words, N2P/N2R includes the full NELS:88 sample through 1992, or when most of the NELS:88 respondents were completing high school. NOP/NOR follows a subset of these students and dropouts through 2000, when most sample members would have been approximately 26 years old.

N2P = NELS:88 second follow-up public-use file

N2R = NELS:88 second follow-up restricted-use file

N4P = NELS:88 third follow-up public-use file (superseded by NOP)

N4R = NELS:88 third follow-up restricted-use file (superseded by NOR)

NOP = NELS:88 fourth follow-up public-use file

NOR = NELS:88 fourth follow-up restricted-use file

NOT = NELS:88 fourth follow-up postsecondary transcript file (restricted)

N2P or N2R should be used when the investigator is interested in the following populations:

- 1988 8<sup>th</sup>-graders
- 1990 10<sup>th</sup>-graders
- 1992 12<sup>th</sup>-graders
- Base-year to first follow-up panel
- Base-year to second follow-up panel
- First follow-up to second follow-up panel
- Dropouts during first and second follow-ups
- 9<sup>th</sup>- to 12<sup>th</sup>-grade transcript studies

NOP or NOR should be used when the investigator is interested in the following populations:

- 1991-92 school year high school graduates 8 years later
- 12<sup>th</sup> grade to fourth follow-up panel (spring 1992 seniors in 2000)
- 10<sup>th</sup> grade to fourth follow-up panel (spring 1990 sophomores in 2000)
- Base-year to fourth follow-up panel (spring 1988 8<sup>th</sup>-graders in 2000)

**WARNING: The base-year, first follow-up, and second follow-up cross-sectional weights are not on the N0P or N0R ECBs. If the proposed analysis is examining middle school or secondary school issues, then the researcher should be using N2P or N2R.**

**Question:** *What is the sequence of data files on the NELS:88 ECBs?*

**Response:** The files containing the N2P and the N0P data are arranged in a specific order. For example, the NELS:88 N2P ECB begins with a series of five universe variables that describe how the respondent entered the study and the student's status at each wave of data collection (i.e., base-year, first follow-up, and second follow-up studies). These five universe variables are followed on N2P by:

1. Universe status of respondents at times of interviews (1988, 1990, and 1992).
2. Base-year student variables followed by base-year composite variables and weights.
3. First follow-up student variables followed by
  - a. first follow-up composite variables and weights; and
  - b. first follow-up freshened student responses
4. Second follow-up student variables followed by
  - a. second follow-up composite variables and weights;
  - b. second follow-up freshened student responses; and
  - c. composite transcript variables
5. First follow-up dropout variables followed by first follow-up composite variables and weights.
6. Second follow-up dropout variables second follow-up composite variables and weights.
7. Base-year school variables followed by base-year composite variables and weights.
8. First follow-up school variables followed by first follow-up composite variables.
9. Second follow-up school variables followed by second follow-up composite variables.
10. Base-year parent variables.
11. Second follow-up parent variables.
12. Base-year teacher variables.
13. First follow-up teacher variables.
14. Second follow-up teacher variables.

Since the fourth follow-up data include postsecondary education information presented at institution and student by institution levels, the order of files in N0P data is slightly different:

*STUDENT DATA*

1. Universe status of individuals at times of interviews (1988, 1990, 1992, 1994, and 2000)
2. Base-year student variables
3. First follow-up student variables
4. Second follow-up student variables
5. First follow-up dropout variables
6. Second follow-up dropout variables
7. Base-year school variables
8. First follow-up school variables
9. Second follow-up school variables
10. Base-year parent variables
11. Second follow-up parent variables
12. Base-year teacher variables
13. First follow-up teacher variables
14. Second follow-up teacher variables
15. F3 student/dropout variables (multiple sections from CATI)
16. F4 student/dropout variables (multiple sections from CATI)
17. F4 derived variables and weights

*POSTSECONDARY EDUCATION DATA*

18. F3 and F4 student\*institution-level postsecondary education information
19. F3 and F4 institution-level postsecondary education information
20. F3 student\*institution\*episode postsecondary education information

**Question:** *When I receive my NELS:88 CD, what are some of the steps that I should follow to check out my CD?*

**Response:** The following steps may help you get a better understanding of the NELS:88 data.

Step 1—Make sure that you have the right file

A general rule that should be followed by all researchers when they receive data from the government or any other source is to check the file for accuracy. Does this file include what you think it does? The following questions should be answered for the NELS:88 CD.

1. Does the NELS:88/2000 CD contain the files listed in the documentation? Check directory and sub-directories.

2. After running frequencies on selected variables on the data file (e.g., first variable, last variable, and five at random in between first and last), do the frequencies agree with those shown for ECB or user manuals? If not, did you receive the correct version?
3. Do the analysis weights (final nonresponse-adjusted weights) contained on the data files allow you to replicate weighted frequencies found in the user manuals? You may want to run weighted frequencies on a single variable using each of the weights contained on the file.

### Step 2—Understanding the Data

Assuming that you performed the above steps and you are confident that the files appear to contain what you hoped they would, it is now time to start learning about the files that you will be working with. Start by asking the following questions:

What were the processes involved in getting data from students via questionnaires and cognitive tests to the medium (CD) that you now possess? Just because you did not collect the data does not mean that you do not need to know the procedures that were involved in collecting and processing the data. You also need to understand the quality control checks that were performed by the contractors in processing the data. It is important to realize that some respondents did not respond to entire instruments; other respondents skipped individual items. For example,

- a) some refused to complete selected items,
  - b) some did not reach the end of the questionnaire,
  - c) sometimes abbreviated versions of instruments were used in data collection,
  - d) some respondents made illegal skips, and
  - e) some respondents responded outside valid ranges.
2. What can I do to further my understanding of the cases and variables that I plan to use? You can perform your own quality control procedures by answering the following questions:

Are the cases that I selected representative of the population to which I wish to generalize? How do the various breakdowns of the data compare to known population numbers? Is my sample biased—do nonrespondents look different from respondents?

#### A.4.6 Composite Variables

**Question:** *What are the advantages of using composite variables in my analyses?*

**Response:** Composite variables (also called derived variables) were developed for NCES by NORC and RTI to help the researcher analyze the NELS:88 data. These variables were usually created from two or more variables and are often considered to be more accurate measures of the underlying concept than the individual variables that were used to create them. For example the base-year socioeconomic status variable (BYSES) is a composite variable made up of five separate variables from the base-year parent questionnaire representing both parents' education levels, both parents' occupations, and family income (e.g., BYP30, BYP31, BYP34B, BYP37B, and BYP80).

#### A.4.7 Model Building

**Question:** *How do I select variables for a working data file?*

**Response:** The following sequence of steps will help you to produce your own working data file.

##### Model Building

After a researcher understands 1) how the NELS:88 data were collected and processed, 2) limitations of the data, and 3) research issues that can be addressed, he/she is ready to begin selecting variables for his/her working data files. The working data file will be used to test the models that are derived from previously developed conceptual models. Before a working data set is created though, the following steps are suggested:

- 1) Develop a conceptual model—What does prior research suggest is happening with the data (e.g., characteristics of students who are likely to drop out of school)?
- 2) Determine the predictor variables (e.g., high absenteeism, disadvantaged background, low test scores), intervening processes (e.g., courses completed, trouble with law, pregnancy), and outcomes (e.g., event of dropping out, completion of GED) that can be used to explain the model.
- 3) Determine which components (variables) of your model can be addressed with NELS:88 variables—If there are multiple sources of data (e.g., student, parent, school, teacher) available on the NELS:88 data files, choose the ones that the researcher believes are most reliable and valid (e.g., parent report of family income is more reliable than student report). If the variables that the researcher needs are not available on the NELS:88 files, he/she should consider merging variables from other sources (Census, Common Core of Data) through licensing agreement.

- 4) Rethink original model—If the variables contained on the NELS:88 data files cannot be used to study original model, rethink the model and either 1) modify the model or 2) choose another data set.

Once the above steps have been completed, it is time to subset the NELS:88 data files into working data files. The following steps are suggested:

- 1) Determine which variables are needed from each of the NELS:88 data files. For example, the model may specify that the following variables are needed from selected files. For example, base-year student data on aspirations, TV exposure, and hours of homework per week can be used to predict fourth follow-up outcomes like highest education attainment, satisfaction with occupation, or income.
- 2) Determine the analysis population that you wish to work with. This will need to be known so that correct survey questions, filters, and weights can be tagged and included in the extracted files.
- 3) Use the ECB (N0P or N2P) to tag variables and then create a SAS or SPSS system file.
- 3) Check log of computer runs to determine if program is doing what you want it to do rather than the directions provided by computer program.
- 4) Run frequencies and/or means on all variables in working data file to serve as codebook and documentation.
- 5) Document all cases that are excluded from the working data files (e.g., who is being deleted from the analysis?).

**Question: How do I subset data files?**

**Response:** It is very important for the user of NELS:88 data files to learn the proper techniques for subsetting the data. If a user does not correctly subset the files, there will be extraneous cases on the working data file that can potentially complicate the analyses and result in erroneous findings.

Why subset? The NELS:88 data were designed to serve many different audiences. Because of this role, the data can be subset to represent many different populations (e.g., the 8<sup>th</sup>-grade class of 1988; the 12<sup>th</sup>-grade class of 2002; the panel of 8<sup>th</sup>-graders who participated in the study from the base-year through the four subsequent waves). By applying the analysis weights, the user can specify the population that is to be examined. (Of course, some flags [e.g., F1BYQFLG] can also be used for subsetting.) The following examples demonstrate how the researcher can use weights to subset the NELS:88 data.

Example # 1: Using the NELS:88 base-year through second follow-up ECB (N2P), the user can specify by using the subsetting statement (in SAS or SPSS) IF BYQWT > 0 that he/she only wants those cases that have valid responses for the base-year survey. Instead of 27,000 cases, the file will contain 24,599.

Example # 2: If a user is interested in postsecondary education outcome variables, he/she will want to use the base-year through fourth follow-up ECB (N0P). By specifying use of this ECB, the user has already limited the number of valid cases to 12,144 (instead of 27,000). Thus 15,000 invalid cases have been eliminated.

NOTE: Analysis (final nonresponse-adjusted) weights such as BYQWT have positive values for cases that should be included in the population of interest and values of "0" for all other cases. Thus, programs that subset the N2P ECB with "IF BYQWT > 0" will only keep cases that have positive values, or in other words, those cases with responses in the base year from eligible sample members.

#### **A.4.8 Transcripts**

**Question:** *How do I use the high school transcript file?*

**Response:** The transcript file is a separate large file on the N2R restricted CD-ROM that contains separate records for each course that a student took while attending high school. Each separate course is assigned a course ID (F2RCSSC) that can be used to group courses (e.g., F2RCSSC codes ranging from 270000 to 279999 represent math courses).

Information on postsecondary education transcripts can be found in another ECB.

#### **A.4.9 Privileged or Restricted-Use Data**

**Question:** *When do I need to use the restricted-use data file?*

**Response:** You should consider using the NELS:88 restricted files when you need data that are more detailed (e.g., individual transcript course data; characteristics of student's neighborhood) than data contained on the public release files. For example, if you need transcript data or zip code data, you should think about obtaining a NCES license agreement.

Contact Cynthia Barton at (202) 502-7307 or e-mail at [Cynthia.Barton@ed.gov](mailto:Cynthia.Barton@ed.gov) for information on how to obtain a license.