



## Getting Started with the NHES Data

### Module Objectives

- Introduce users to the data collected through the NHES and the resulting micro-level data files
- What data are available
- How to obtain the data
- Considerations to ensure quality in your NHES analyses
- Types of variables available on the data files
- Data ordering and naming conventions
- Resources to learn more about the data files

**Current NHES Data**

Surveys on early childhood programs and family involvement in schools

Collection Year	Unit of Analysis	Respondent	Collection Method
2012	Child	Parent or knowledgeable adult in the household	20 – 25 page paper-based mail survey

**Past NHES Data**

Survey topics	NHES Survey Administration									
	1991	1993	1995	1996	1999 <sup>1</sup>	2001	2003	2005	2007	2012
Early childhood education/program participation	✓		✓		✓	✓		✓		✓
Adult education	✓		✓		✓	✓	✓	✓		
School readiness		✓			✓				✓	
School safety and discipline		✓								
Parent and family involvement in education				✓	✓		✓		✓	✓
Civic involvement				✓	✓					
After-school programs and activities			✓ <sup>2</sup>		✓	✓ <sup>3</sup>		✓		
Household and library use				✓						

<sup>1</sup>NHES:1999 was a special end-of-decade administration that measured key indicators from the surveys fielded during the 1990s.

<sup>2</sup>These items were only asked about children in first through third grades.

<sup>3</sup>The NHES:2001 survey about after-school programs and activities (ASPA) also included before-school programs.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 1991, 1993, 1995, 1996, 1999, 2001, 2003, 2005, 2007, and 2012.

### Data File Basics

- NCES public-use data are available for free online along with comprehensive user's manuals
- Most NHES data and manuals can be downloaded directly from the [study's website](#)
- NHES data are also available in the [Education Data Analysis Tool \(eDAT\)](#)
- Restricted-use data are available to qualified users who are approved for a [restricted-use license](#)
- More information about restricted-use data is available in the Common Module [Acquiring Micro-level NCES Data](#)

### Getting NHES Data Files

- 1999 to 2012 Data Files are available online at the [study's website](#)
- Syntax code necessary to import the data into
  - SPSS
  - SAS
  - Stata
- User's Manuals
- Codebooks

**NHES Data Files Available Online**

<u>Surveys</u>	Data Collection Years
Adult Education	2001, 2003, 2005
After-School Programs and Activities	1999, 2001, 2005
Early Childhood Program Participation	1999, 2001, 2005, 2012
Parent and Family Involvement in Education	1999, 2003, 2007, 2012
School Readiness	1999, 2007

**NHES Data Files Available from the eDAT**

1991-2012 Data Files

- 1991-1996 data files are only available through the [eDAT](#)

### Types of Variables on the NHES

Most variables are numeric but some are character variables (also called string variables)

- The NHES variables can be categorized as
- Case identifiers (BASMID)
- Questionnaire item variables
- Derived variables
- Household characteristics variables
- Appended variables
- Weighting variables
- Imputation flag variables

### Variable Naming Conventions

- All variables in the NHES have meaningful names that reflect what the variable describes, for example RACEETH2 for race and ethnicity
- Variables with the same name in multiple years are usually directly comparable. Variables with similar but different names are not directly comparable
- Variable names can be found in the annotated questionnaires in the data file user's manual and in the codebooks
- Variable descriptions (labels) include the survey question number or D, for "derived variable"
- Variables that appear on more than one survey have corresponding letters in the variable description, for example, E for the survey about enrolled students and H for the survey about homeschooled students. The naming conventions for imputation flag variables, weight variables, replicate weights, and stratum and cluster identification variables are described in [NHES Sample Design, Weights, Variance, and Missing Data](#)

### Variable Naming Conventions (Continued)

- Typically, BASMID is used as the case identifier
- The variables PATH or MAINRSLT denote the main populations of interest (e.g., elementary school student, middle school student, high school student, homeschooled student)

### Appended Data

- NHES data files contain data appended from external data sources maintained by the Census Bureau and from NCES's school databases
  - These variables provide information about sampled children's schools and school districts and the surrounding communities
- Appended data are described in the data file user's manual
- Some appended data are only available to NHES restricted-use data license holders

### Imputed Data

- All [missing data](#) are imputed
- Imputation flags note when data have been imputed
- It is recommended that imputed data be used, but imputation flags are provided to facilitate recoding
- Missing responses are different from "not applicable" responses, which are denoted in the data file with -1

### Using the NHES Data

- When using a CD, make sure the data CD is in the drive
- When using downloaded data, make sure the read-in code is correctly specified for the downloaded file location
- Start with an original, write-protected data file
- Separate data coding/preparation and analysis
- Document your syntax

### Considerations for Quality Analyses

- Read the user's manuals
- Check the survey website regularly
- Understand the unit of analysis
- Understand mode differences between 2012 and past NHES survey
- Review questionnaires
  - Check skip patterns
- Locate and examine all derived variables before you begin your analyses

### More Considerations for Quality Analyses

- Examine data quality of variables you are using in analysis
  - Run descriptive statistics/frequencies
- Run results with unweighted data
  - Helps identify potential data coding errors
  - Avoid small cell estimates

### More Considerations for Quality Analyses (Continued)

- Use the appropriate [weights](#) to account for the complex survey design when producing estimates and standard errors
  - All NHES data files contain replicate weights for calculating standard errors
- Check recodes through crosstabs, frequencies, etc.
- Check results for plausibility
  - Verify against external sources when possible
  - If your results are surprising, check for coding or programming errors

### Module Summary and Resources

#### Summary

- Introduced users to the data collected through the NHES and the resulting micro-level data files
- What data are available
- How to obtain the data
- Considerations to ensure quality in your NHES analyses
- Types of variables available on the data files
- Data ordering and naming conventions
- Resources to learn more about the data files

#### Resources

- [NHES data available from the Study's website](#)
- [Education Data Analysis Tool \(EDAT\)](#)
- [Restricted-use license](#)
- [Acquiring Micro-level NCES Data](#)
- [NHES Topical Surveys/Questionnaires](#)
- [NHES Sample Design, Weights, Variance, and Missing Data](#)