



NHES Sample Design, Weights, Variance, and Missing Data

Module Objectives

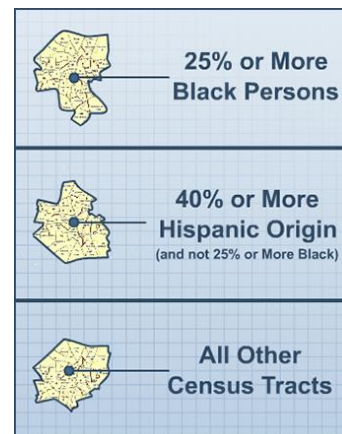
- Summarize sample design as it relates to study weights
- Describe the weights that must be applied to assure data are representative of the target population(s)
- Explain the procedures for calculating appropriate standard errors
- Describe the missing data codes used in the NHES and explain how to handle missing data to ensure accurate analysis

NHES:2012 Sample Design

- Two-stage stratified address-based sample (ABS)
- First-stage Screener questionnaire used to enumerate household members and determine eligibility for an extended Topical survey in the second stage
- Screener and Topical surveys completed through self-administered paper-based mail questionnaires

Sample Stratification: Stage 1

- Addresses were stratified into three race/ethnicity strata
 1. Census tracts with 25% or more Black persons
 2. Census tracks with 40% or more persons of Hispanic origin (and not 25% or more Black)
 3. All other Census tracts
- Households in minority strata were oversampled



Sample Stratification: Stage 1 (Continued)

- NHES:2012 sample contained all types of residential addresses for best possible coverage
- Addresses divided into two groups
 - Group 1: Post Office (PO) Boxes not designated as household's "only way to get mail" (OWGM)
 - Group 2
 - All other residential address types including street and city-style addresses
 - High rises
 - Rural routes
 - OWGM PO Boxes
 - Seasonal, vacant, and educational addresses
 - Drop points and PO Box throwbacks
- Addresses were then subsampled at different rates from each of the two groups

Within-household Sampling: Stage 2

- One child selected per household for Topical survey
- Differential sampling rates used to ensure sufficient sample size for the ECPP
- Each household was pre-designated to receive either the ECPP or PFI Topical Survey if they had eligible children in both domains
 - If household only had children eligible for ECPP, one child was randomly selected for the ECPP
 - If household only had children eligible for PFI, one child was randomly selected for the PFI
 - If household had children eligible for both, one child was randomly selected in pre-designated domain to receive the appropriate survey

NHES Weights and Sample Design

- NHES data cannot be treated as if they are from a simple random sample
- Weight variables are provided on the data files to account for characteristics of the NHES design
- Weights are also adjusted for undercoverage and unit nonresponse
- Weights must be used for computation of population estimates
 - ECPP data are representative of children ages 0-6, not yet enrolled in school
 - PFI data are representative of children age 20 or younger, enrolled in kindergarten through 12th grade or homeschooled for equivalent grade

How Are NHES Weights Created?

Household-Level Weights

- Basis for person-level weights
- Household base weight is the probability of selection of the address from the address
 - Product of the weight associated with differential sampling of addresses based on the race/ethnicity stratum of the address and differential sampling of PO Box addresses that were designated as OWGM
- Base weight is adjusted for screener nonresponse and address ineligibility using auxiliary data from the frame
- Chi-Square Automatic Interaction Detection (CHAID) analysis was used to identify characteristics associated with Screener nonresponse
 - Decision-tree technique that sequentially identifies the most likely response predictors

How Are NHES Weights Created? (Continued)

Person-Level Weights

- Household-level weight used as base weight
- Account for probability of selection within household
- Adjusted for nonresponse to the Topical questionnaire using characteristics from Screener and methodological variables
- Raked to population control totals for key person characteristics (race/ethnicity, income, household size, age, home tenure, and parent's education) using American Community Survey (ACS) data

Final NHES Weight

The final NHES weight takes into account the complex sample design at both the household and person levels

- These final weights are put on the data file
 - FEWT (ECP)
 - FPWT (PFI)

ECPP Sample Counts and Distributions

Table 1. Sample sizes and population counts of children aged 0-5, not yet enrolled in kindergarten, by child characteristics: 2012

Characteristics	Sample	Sample Percentage	Population	Population Percentage
Total	7,892	100.0	21,673,765	100.0
Sex				
Male	4,084	51.8	11,273,798	52.0
Female	3,808	48.3	10,399,967	48.0
Race/ethnicity				
White, non-Hispanic	4,290	54.4	10,892,603	50.3
Black, non-Hispanic	787	10.0	2,889,532	13.3
Hispanic	1,827	23.2	5,468,531	25.2
Asian or Pacific Islander, non-Hispanic	497	6.3	1,108,587	5.1
Other race, non-Hispanic	491	6.2	1,314,512	6.1
Attends weekly center-based care				
Yes	3,181	40.3	7,327,733	33.8
No	4,711	59.7	14,346,032	66.9

NOTE: Percentages may not add to 100 due to rounding
 SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Program Participation Survey of the National Household Education Surveys Program, 2012

PFI Sample Counts and Distributions

Table 2. Sample sizes and population counts of children aged 5 – 20, enrolled in kindergarten through 12th grade, by child characteristics: 2012

Characteristics	Sample	Sample percentage	Population	Population percentage
Total	17,120	100.0	51,773,221	100.0
Sex				
Male	8,891	51.9	26,753,759	51.7
Female	8,229	48.1	25,019,462	48.3
Race/ethnicity				
White, non-Hispanic	9,376	54.8	26,712,212	51.6
Black, non-Hispanic	2,036	11.9	7,355,263	14.2
Hispanic	3,782	22.1	12,003,955	23.2
Asian or Pacific Islander, non-Hispanic	983	5.7	2,879,157	5.6
Other race, non-Hispanic	943	5.5	2,822,634	5.5
School Level				
Elementary School	6,753	39.5	25,440,362	49.1
Middle School	3,949	23.1	11,697,409	22.6
High School	6,418	37.5	14,635,450	28.3

NOTE: Percentages may not add to 100 due to rounding
 SOURCE: U.S. Department of Education, National Center for Education Statistics, Parent and Family Involvement in Education Survey of the National Household Education Surveys Program, 2012

Key Reminders About NHES Weights

- Always weight data when producing estimates
- Unweighted data
 - Will not account for sample design, nonresponse, or undercoverage
 - Will not represent the population

Standard Error Calculation in the NHES:2012 [Replication Techniques](#)

- These methods calculate appropriate standard errors based on differences between the full sample and a series of created subsamples (replicates)
- Replicate weight names match the main sample weight name
 - FEWT1-FEWT80 for the ECPP
 - FPWT1-FPWT80 for the PFI
- NHES:2012 replicate weights use the jackknife 1 (JK1) method

Standard Error Calculation in the NHES:2012 Taylor Series Linearization

- This method uses primary sampling unit (PSU) and strata identifiers to calculate appropriate standard errors
- PSU and strata variable names are
 - EPSU and ESTRATUM for the ECPP
 - PPSU and PSTRATUM for the PFI

Missing Data in the NHES:2012

- For the most part, the NHES:2012 has high item response rates
- Reasons for item nonresponse
 - Respondent did not know answer
 - Respondent preferred not to answer
 - "Short form" version of survey
 - Inconsistent responses set to missing during data editing

Imputed Data in the NHES:2012

- Three [imputation methods](#) were used in the NHES:2012
 - Unweighted sequential hot deck imputation
 - Weighted random imputation
 - Manual imputation
- All imputed variables have corresponding imputation flags on the data file
 - Allows users to identify imputed values
 - Value of imputation flag variable denotes imputation method used

Imputed Data in the NHES:2012 (Continued)

There are four [exceptions](#) where missing data are not imputed

- "Other, specify" text strings
- "Hispanic, race not reported"
- Appended non-questionnaire data (such as sampling frame and derived variables) were not imputed
- Questions that were skipped based on previous responses are coded as -1 (not applicable)

"Not Applicable" Codes in the NHES:2012

- Two PFI questionnaires were used in the NHES:2012 data collection
 - PFI-enrolled
 - PFI-homeschool
- Children whose parents responded to the PFI-enrolled questionnaire will have -1 (not applicable) codes for any variables that correspond to items that were only asked on the PFI-homeschool questionnaire
- The reverse will be true for children whose parents responded to the PFI-homeschool questionnaire

"Not Applicable" Codes in the NHES:2012 (Continued)

Analysts may want to recode some "not applicable" codes to valid values, depending on the research question

- For example, children not attending center-based child care on a regular basis are coded "not applicable" for the hours per week spent in center-based care

Module Summary

Summary

- Summarized sample design as it relates to study weights
- Described the weights that must be applied to assure data are representative of the target population(s)
- Explained the procedures for calculating appropriate standard errors
- Described the missing data codes used in the NHES and explained how to handle missing data to ensure accurate analysis

Resources

- [Analyzing NCES Complex Survey Data](#)
- [Statistical Analysis of NCES Datasets Employing a Complex Sample Design](#)
- [2012 Data File User's Manual](#)