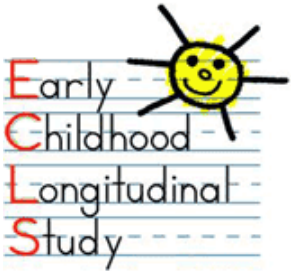
Slide 1 of 31



ECLS-B Sample Design, Weights, Variance, and Missing Data

Slide 2 of 31

Module Objectives

- Describe the ECLS-B weights that must be applied to assure data are representative of the target population
  - Summarize the ECLS-B sample design as it relates to study weights
- Describe the procedures for calculating standard errors
  - Identify the ECLS-B study design variables that must be used in statistical software to compute the correct standard errors
- Describe the missing data codes used in the ECLS-B data and how to handle missing data to ensure accurate analysis

Slide 3 of 31

ECLS-B Sample Design: Clustering

- The sample for the ECLS-B is not a simple random sample (SRS) of the target population
- Cases were clustered within primary sampling units to reduce field costs
- Multistage, stratified, clustered design
  - Stage 1: Country separated into primary sampling units (PSUs), which were counties or groups of contiguous counties
  - Stage 2: Birth certificates were sampled from within sampled PSUs
- Supplementary sample of American Indian/Alaska Native (AIAN) PSUs
  - Counties with a relatively higher number of AIAN births were identified for inclusion in the supplemental AIAN sample to increase the number of AIAN children selected

Slide 4 of 31

ECLS-B Sample Design: Clustering (Continued)

- Children who were selected lived in closer geographical proximity to one another than would be the case in an SRS
- Children living within a PSU tend to be more similar to one another on many characteristics than children living in different PSUs
- Therefore, variation in the clustered sample is lower than it would be in an SRS

Slide 5 of 31

ECLS-B Sample Design: Oversampling

The ECLS-B includes oversamples of

- Children born with low birth weight and very low birth weight
- Chinese American, other Asian American/Pacific Islander, and American Indian/Native Alaskan children
- Twins

Slide 6 of 31

Purpose of Weights: Review

- Weights are used to make sample data representative of the target population
- Weights account for differential selection probabilities and differential patterns of response/nonresponse

Slide 7 of 31

Sampling Weights for the ECLS-B

- Multiple weights are provided for analysis using data from each round of data collection
- ECLS-B weights sum to the population of children born in the United States in 2001
- For components that have additional eligibility requirements, the sum of the sampling weights is the number of individuals in the target population who meet the criteria for eligibility for the given study component
- Every ECLS-B weight is adjusted for nonresponse to one or more of the data collection components used to collect information about these children and their experiences over time
    - Estimates produced using these nonresponse adjusted weights are representative of the characteristics and experiences of the population of children born in the United States in 2001 even though some of the sampled children did not participate in the assessments and some of their parents and care providers did not participate in the study

Slide 8 of 31

Sampling Weights for the ECLS-B (Continued)

Development of the ECLS-B weights reflects the longitudinal design of the study and the requirements for participation

- To be eligible for participation in a round after the 9-month data collection, a parent interview must have been completed in the prior round
- Also, to be eligible for the different components within a round of data collection (for example, the child assessment, the father questionnaires, the child care provider interview, and the child care observation in the 2-year data collection), a parent interview must have been completed for that round
- All weights for given round of data collection are adjusted for nonresponse to the parent interview at that round as well as the parent interview in prior rounds

Slide 9 of 31

Sampling Weights for the ECLS-B (Continued)

- Selection of a weight is driven primarily by
  - the sources of data about the children and their experiences that are being used in that analysis
  - consideration of which weight is adjusted for nonresponse to those sources
- Ideally there would be a nonresponse adjusted weight available for every component of every round of data collection
  - It is neither economical nor practical to create nonresponse adjusted weights for every combination of components across every round of data collection
- Researchers must decide which weight is the best one to use, given their research question

Slide 10 of 31

ECLS-B Nomenclature for Weight Variables: Component

W	Weight
R	Parent Interview (R refers to “Respondent”)
C	Child Assessment
J	Early Care and Education Provider Interview Child Care Provider (CCP) Early Care and Education Provider (ECEP) Wrap-Around Early Care and Education Provider (WECEP)
P	Early Care and Education Setting Observation
F	Father (resident and/or nonresident)
D	Father (resident only as no nonresident father questionnaire was fielded in the preschool round of data collection)

Slide 11 of 31

ECLS-B Nomenclature for Weight Variables: Study Rounds

1	9 months
2	2 years
3	Preschool
4	Kindergarten 2006
5	Kindergarten 2007
K	Kindergarten 2006/Kindergarten 2007

Slide 12 of 31

Examples of ECLS-B Weight Names

W1R0	Weight that is adjusted for nonresponse to the parent interview data from round 1
WKR0	Weight that is adjusted for nonresponse to the parent interview in the round in which children entered kindergarten (either kindergarten 2006 or kindergarten 2007)
W3D0	Weight that is adjusted for nonresponse to the resident father questionnaire in round 3
W523J0	Weight that is adjusted for nonresponse to the care and education provider telephone interviews in rounds 2 or 3 (and parent interview data through wave 5)

Slide 13 of 31

Selection of Weight for Analysis

How to decide which weight to choose

- Look at the round(s) of data in the analysis
    - Choose a weight that corresponds with the latest round in the analysis. For example, if the analysis includes any preschool round data (round 3) and no data from later rounds, choose a preschool weight (W3)
- Look at the components providing the data in the analysis
    - Choose a weight that adjusts for nonresponse to the greatest number of components providing data for the analysis

Slide 14 of 31

Selection of Weight for Analysis (Continued)

How to decide which weight to choose

- It may not be possible to find the “perfect weight,” or one that is adjusted for nonresponse to every component at every round from which data are being used
- If no weight corresponds exactly to the combination of components included in the analysis, researchers might prefer to use a weight with more components included
- Although such a weight may result in a slightly smaller analytic sample, it will adjust for the sample and nonresponse associated with each of the components it covers
- If a researcher chooses a weight with fewer components included, then missing data should be examined for potential bias

Slide 15 of 31

Selection of Weight for Analysis (Continued)

How to decide which weight to choose

Weight	9-month components		
	Parent	Child	RFSAQ NRFSAQ
W1R0	Yes		
W1C0	Yes	Yes	
W1F0	Yes		Yes
W1FC0	Yes	Yes	Yes

Note: RFSAQ= Resident Father Self-Administered Questionnaire and NRFSAQ= Nonresident Father Self-Administered Questionnaire.

Slide 16 of 31

Selection of Weight for Analysis (Continued)

Weight	9-month component			2-year component				
	Parent	Child	RFSAQ NRFSAQ	Parent	Child	RFSAQ NRFSAQ	CCP	CCO
W2R0	Yes			Yes				
W2C0	Yes	Yes		Yes	Yes			
W2F0	Yes		Yes	Yes		Yes		
W2FC0	Yes	Yes	Yes	Yes	Yes	Yes		
W2C2J0	Yes	Yes		Yes	Yes		Yes	
W2C2P0	Yes	Yes		Yes	Yes		Yes	Yes
W22J0	Yes			Yes			Yes	
W22P0	Yes			Yes			Yes	Yes
W22F0	Yes			Yes		Yes		
W2C1F0	Yes	Yes	Yes	Yes	Yes			

Note: RFSAQ= Resident Father Self-Administered Questionnaire and NRFSAQ= Nonresident Father Self-Administered Questionnaire.

Slide 17 of 31

Weight Considerations When Using Child Assessment Data

- In the 2-year, preschool, and kindergarten rounds of data collection, the response patterns for the child assessment component were determined to be so close to the response patterns for the parent interview that no separate ("C") weights needed to be developed for use in analyses that include child assessment data from just one of those rounds
- When analyzing child assessment data from just one of these later rounds, use the "R" weight
- The response pattern for the 9-month child assessment did differ significantly enough from that of the parent interview at 9 months to warrant development of a separate weight adjusted for nonresponse to the 9-month child assessment (W1C0)
    - When analysis uses 9-month child assessment data, choose a "C" weight
    - Consider using a "C" weight when conducting analysis using multiple rounds of assessment data, even if 9-month data are not included

Slide 18 of 31

Weight Considerations When Analyzing AIAN Children

- To preserve the sample of American Indian/Alaska Native (AIAN) children, all AIAN children were eligible for the preschool and kindergarten data collections if they had a complete 9-month parent interview, regardless of whether they had a complete parent interview in other rounds
- Specific analytic weights were developed for the AIAN cases that are not contingent on response to every data collection after the 9-month parent interview

Slide 19 of 31

A Look at Weights in the ECLS-B Data

I_ID	W1R0	W1C0	W1F0	W1FC0	Y1CHRACE
100001	1002.61	1017.01	1135.81	1139.76	HISPANIC, RACE SPECIFIED
100002	575.47	601.66	.	.	HISPANIC, RACE SPECIFIED
100003	704.30	723.16	.	.	BLACK OR AFRICAN AMERICAN NON-HISPANIC
100004	54.39	57.35	79.63	79.77	WHITE, NON-HISPANIC
100005	666.79	719.40	741.46	756.68	MORE THAN 1 RACE, NON-HISPANIC
100006	62.83	.	.	.	ASIAN, NON-HISPANIC
100008	657.05	671.55	762.45	746.93	WHITE, NON-HISPANIC
100009	611.87	.	761.58	.	WHITE, NON-HISPANIC

NOTE: The period (.) indicates data that are system missing.

Slide 20 of 31

Weighted Data Sum to the Population Total

	W1R0	W1C0	W1F0	W1FC0	W2R0
Valid n	10,700	10,200	7,000	6,800	9,850
Mean	373.99	391.07	517.76	530.83	403.22
Minimum	4.08	4.72	5.53	5.95	4.38
Maximum	1,854.91	1,981.78	3,850.44	4,050.91	2,488.89
Sum (N)	3,997,169	3,997,169	3,618,138	3,618,138	3,965,681

Note: Unweighted sample sizes rounded to the nearest 50.

Slide 21 of 31

Standard Error Calculation in ECLS-B: Replication Techniques

- This method calculates appropriate SEs based on differences between estimates from the full sample and a series of created subsamples (replicates)
- Select replicate weights that are associated with your main sampling weight (e.g., W1R1 to W1R90 for weight W1R0 and W1C1 to W1C90 for W1C0)
- ECLS-B replication weights use the jackknife 2 (JK2) method

Slide 22 of 31

Standard Error Calculation in ECLS-B: Taylor Series Linearization

- This method uses primary sampling unit (PSU) and strata identifiers to calculate appropriate SEs
- Select the identifiers that are associated with your main sampling weight (e.g., W1PSU and W1STR for weight W1R0)

Slide 23 of 31

Missing Data Values Used in the ECLS-B Data

- Missing data are retained for most variables
- Standard ECLS-B missing data codes

(blank)	System missing (unit nonresponse)
-1	Not applicable, including legitimate skips
-4	Data suppressed due to administration error
-7	Refused (a type of item nonresponse)
-8	Don't know (a type of item nonresponse)
-9	Not ascertained (a type of item nonresponse)

Slide 24 of 31

Exceptions to Standard Usage for Missing Data Values

- Bayley Short Form – Research Edition
  - Missing values for t-scores are coded as -99
- Birth Certificate Data
  - Several missing data codes and labels
  - Codes associated with different labels vary among birth certificate variables
  - For more information, see the 2001 Natality Technical Appendix available through the NCHS website
- Fine and Gross Motor Item Data
  - Codes of 95, 96, and 97 were used to identify cases for which the item was uncodeable, not administered, or had no response respectively

Slide 25 of 31

Example of Recoding Missing Data

9-month Parent Interview questions

- P1 CH010 Please think back to when [CHILD/TWIN] was born. As a newborn, did [CHILD/TWIN] have to stay longer in the hospital because of medical problems?

  1 YES
  2 NO (skip to CH035)
- P1 CH015 How many days did [CHILD/TWIN] stay in the hospital because of medical problems?

Slide 26 of 31

Example of Recoding Missing Data (Continued)

- Question #1: What is the average number of days that infants stayed in the hospital after birth because of medical problems?
- Question #2: Of those infants who had to spend some time in the hospital after birth because of medical problems, how many days on average did they spend in the hospital?

Slide 27 of 31

Example of Recoding Missing Data: Q#1

Question #1: What is the average number of days that infants stayed in the hospital after birth because of medical problems?

- Recoding steps
  - If the answer to CH010 (spent time in hospital after birth?) is 'No,' then recode values of -1 on CH015 (# days in hospital) to 0
  - Then recode all remaining missing data values for CH015 (-1, -7, -8, -9) to missing

Slide 28 of 31

Example of Recoding Missing Data: Q#2

Question #2: Of those infants who had to spend some time in the hospital after birth because of medical problems, how many days on average did they spend in the hospital?

- Recoding steps
  - Only those cases with a response of 'Yes' to CH010 (spent time in hospital due to medical problems) should be included in the analysis
  - All missing data values for CH015 (-1, -7, -8, -9) should be recoded to missing

Slide 29 of 31

Example of Recoding Missing Data

Weighted Results for Research Questions (using W1R0 weight)

Research Question	Weighted Mean
1. What is the average number of days that infants stayed in the hospital after birth because of medical problems?	1.9 days
2. Of those infants who had to spend some time in the hospital after birth because of medical problems, how many days on average did they spend in the hospital?	14.7 days
INCORRECT – NO RECORDING OF MISSING VALUES FOR CH015	1.0 days

SOURCE: U.S. Department of Education, National Center for Education Statistics. Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) 9-month – Kindergarten 2007 Longitudinal Restricted-Use Data File. Data are for training purposes only. Please do not cite.

Slide 30 of 31

Module Summary

- Described the ECLS-B weights that must be applied to assure data are representative of the target population
  - Summarized the ECLS-B sample design as it relates to study weights
- Described the procedures for calculating standard errors
  - Identified the ECLS-B study design variables that must be used in statistical software to compute the correct standard errors
- Described the missing data codes used in the ECLS-B data and how to handle missing data to ensure accurate analysis

Module Resources

- Analyzing NCES Complex Survey Data
- Statistical Analysis of the NCES Datasets Employing a Complex Sample Design
- 2001 Natality Technical Appendix