

Slide 1 of 44

Title Slide: Considerations for Analysis of ECLS-B Data

Slide 2 of 44

The ECLS-B data file contains a wealth of data collected from different sources. This module describes analytic considerations that should be kept in mind when using data from some of these sources.

In the first section, data from the household roster are described. The household roster is a matrix reflecting who lived in the household with the ECLS-B study child at each round of data collection. This section of the module explains how to use the roster data available in the ECLS-B data file to identify respondents to the parent interview, parents and parental figures in the household, and other household members, as well as how to identify changes in household composition over time.

In the ECLS-B an emphasis was placed on obtaining information directly from both resident fathers and nonresident fathers. In the second section of the module, considerations for using data collected from fathers, for example the way in which fathers were identified to be study respondents and how that method of identification may affect the analysis of data from and about fathers, are discussed.

The next section of this training module discusses considerations for using data from twin pairs. The special oversample of twins in the ECLS-B is discussed, as well as particular data considerations analysts will need to be aware of when examining data collected about twins.

Lastly, as discussed in the introductory module, the ECLS-B collected kindergarten information for the study children in two different two school years, 2006-07 and 2007-08, because children born later in 2001 were not age-eligible for kindergarten in 2006-07 (round 4) and some children experienced a delayed entry (that is, they did not enter kindergarten when they were age-eligible to do so). Due to the differences in the samples included in each kindergarten collection, data from only one round, or from both rounds combined, are better suited to answer different types of research questions. Analysis of data collected in the kindergarten 2006 and kindergarten 2007 rounds requires careful consideration from users about the group of children to which they want to generalize and whether their outcomes of interest may be related to school exposure. The last section of the module provides information about the data considerations researchers must be aware of when analyzing data from these two rounds, in particular which round or rounds of data to use, given a particular research question.

There are two ways to access the information within this module. You can click on one of the links on the screen, which will take you directly to detailed information about the household roster, fathers, twin pairs, or kindergarten rounds. Or you can click the 'next' button to advance to the next page of the module and view all of the analysis considerations presented in sequential order. At the end of each portion of the module,

Considerations for Analysis of ECLS-B Data

you will be provided with a button that will return you to this objectives page, from which you may either select another portion of the training module, or exit the module.

Throughout this module, underlined blue screen text indicates a link to additional resources.

Slide 3 of 44

The household roster includes information about every individual living in the household with the study child at each round. The information collected includes the household member's relationship to the child, both generally (for example, mother, father, sister, other relative or nonrelative) and specifically (for example biological, step-, adoptive, or foster mother), and the household member's sex, race/ethnicity, and age, as well as whether that person was residing in the household in a specific round of data collection. For household members who were living in the household in one round and no longer living in the household in a later round, the roster information indicates why the person was no longer in the household.

Slide 4 of 44

Household composition information was collected in every round of data collection as part of the parent interview.

At 9 months, parents were asked to identify everyone living in the household with the child.

Slide 5 of 44

In the 2-year, preschool, and kindergarten data collections, information about household composition was updated by collecting information about new individuals who had joined the household since the previous round of data collection and documenting household members who were no longer living in the household. In some cases, questions were asked to obtain information that was missing for household members who lived in the household in previous rounds.

Slide 6 of 44

In order to be able to use household roster information effectively, it is important to understand how the roster data are structured in the data file. Each person in the household occupies a position in the roster and is assigned a position number that remains the same across rounds. Data on specific individuals are retained in the matrix of roster information even if the individual does not remain in the household. The data file includes variables that indicate whether each individual was living in the household at any given round. Roster variable names begin with the letter "Q" and a number indicating the round of data collection. The names end with a double-digit roster position number. For example, the round 1 variable indicating the relationship of the first person in the roster to the sampled child is named Q1REL_01. The variable indicating the age of the 15th person in the roster in round 4 is named Q4AGE_15.

Slide 7 of 44

This grid shows the naming conventions for the roster variables across the rounds of data collection.

Variables with information on static characteristics, or characteristics that generally do not change over time, such as sex and race are included in the matrix of roster information only once and have an underscore as the second character, rather than a number indicating the round of data collection.

In the first two rounds of data collection, race/ethnicity information was collected only for a subset of household members. Specifically, it was collected for the child, the mother, and the father. If the parent interview respondent was not the child's mother or father, race/ethnicity information was also collected for the parent interview respondent and the parent interview respondent's spouse. In the last three rounds of data collection, race/ethnicity information was collected for any person who was new to the household in that round. Note that there are many variables associated with race/ethnicity in the data file but only two are presented here as examples. These are the variables indicating whether a person was White and whether a person was Black or African American.

The Q variables with CURR, spelled C-U-R-R, in their name are the ones indicating whether a specific household member was living in the household in that round of data collection. For instance, the variable Q2CURR_01 indicates whether or not the person in roster position one was living in the household in the second round of data collection.

Slide 8 of 44

These next few slides provide an example using a hypothetical family in order to elucidate how information on household composition was collected in the study and how such data are provided in the household roster information on the data file.

In this example, the study child's biological mother is the respondent to the 9-month parent interview. She indicates that, in addition to the sampled child, the child's step-father and an older brother also live in the household.

Slide 9 of 44

This grid shows the values that would be stored in the various roster variables describing each person in this example. A double pound sign corresponding to the roster position number is used in the variable names for ease of presentation here, but in the data file the person's roster position number would appear in place of the double pound sign.

As can be seen in the grid, Person 1 is the mother. Looking down the column of information for person 1, we can see that she is the biological mother. Information about her age and sex, as well as her race/ethnicity is available. Q1CURR equals "Yes" because she is living in the household at the time of the 9-month data collection. Person

Considerations for Analysis of ECLS-B Data

2 in the roster is almost always the sampled child. For this reason, the Q2REL_02 variable is set to -1 (not applicable) since this would be indicating the child's relationship to him or herself. Person 3 is the step-father and Person 4 is the step-brother. Notice that race/ethnicity information is not available for the step-brother in this round because he is not one of the people for whom race/ethnicity information was collected; he is not the sampled child, a parent, the parent interview respondent, or the respondent's spouse/partner. Q1CURR equals "Yes" for persons 3 and 4 because they are living in the household at the time of the 9-month data collection.

For all the variables associated with other roster positions, that is positions 5 through 21, all data are coded as -1, "not applicable," with one exception. For roster positions that are filled in later rounds with a new member of the household, the Q CURR variable is coded as 2-no. So, in this example, the fact that Q1CURR equals "no," rather than "not applicable," for roster position 5 indicates that a fifth person will be listed as a new member of the household in a later round of data collection. Any roster position for which Q1CURR is set to -1 will never have information for any individual in that position at any round.

Slide 10 of 44

The data file includes a set of identification, or ID, variables for certain household members at every round of data collection. These variables are provided to help researchers use the household roster information. As will be shown in a later slide, they can also be used in conjunction with other household roster information to track changes in household composition over time. These ID variables indicate the person's position number in the household roster.

Variable Y1RESPID is the position number of the respondent at round one. This is the person to whom the relationship described in the variable X1RSPREL, spelled X-1-R-S-P-R-E-L, applies and about whom the information in the Respondent Information (RI) section of the parent interview is collected.

Variable Y1MOMID is the roster position of the sampled child's mother in the household and is the person to whom the specific relationship described in the variable Y1MOMTYP, spelled Y-1-M-O-M-T-Y-P, applies. In households with more than one female adult, a series of hierarchical rules was used to determine which female household member would be the person identified in Y1MOMID, with the biological mother given highest preference. That is, if the biological mother lived in the household at the time of the interview, she was identified as the child's mother in Y1MOMID. If the biological mother did not live in the household at the time of the interview, a mother of another type, for example an adoptive mother, or another female household member such as a grandmother, would have been selected according to the pre-established selection rules. If no household member was identified through this process as a mother/female guardian, the composite variable Y1MOMTYP was set to 9 ("No resident mother"), and Y1MOMID was set to -1 ("Not applicable"). More detail on the hierarchical rules is available in the technical documentation.

Considerations for Analysis of ECLS-B Data

Similarly, variable Y1FTHID, spelled Y-1-F-T-H-I-D, is the roster position of the sampled child's father in the household and is the person to whom the specific relationship described in the variable Y1FTHTYP, spelled Y-1-F-T-H-T-Y-P, applies. A process similar to the one used to identify mothers was used to identify fathers.

Variable P1SPON_R is the roster position of the respondent's spouse/partner when the respondent has a spouse or partner and is the person about whom the information collected in the Spouse Information (SI) section of the parent interview applies.

Finally, variable X1CHID is the roster position of the sampled child. In most instances, the sampled child occupies roster position 2, so X1CHID is coded as 2. The exception is that for cases where the sampled child is a twin, X1CHID will be 3 for one member of each twin pair.

In this example, Y1RESPID is 1 because it corresponds to the parent respondent who is in position 1 of the roster. Because the parent respondent is the child's mother, Y1MOMID is also 1. The father ID variable is 3 because the step-father occupies the third position in the roster, and spouse ID is also 3 because the step-father is the respondent's spouse. The sampled child occupies roster position 2, so X1CHID is 2.

Slide 11 of 44

Following the example of the hypothetical family into round 2, during the 2-year parent interview we learn that the mother has left the household and the child's grandfather is now living with the sample child as well.

Since the mother is no longer in the household, the child's step-father completes the parent interview, making him the parent respondent for round 2.

Slide 12 of 44

Looking at the roster information again, this time for round 2, we see information that has changed. The highlighted cells indicate the differences between the round 1 and round 2 roster variables for this hypothetical household.

Though the child's biological mother is not living in the household in the current round, roster position 1 still contains information for her. For example, the static information about her remains in the round 2 roster variables. However, variable Q2AGE_01 has been coded -1 since this information changes across time but is updated only for current household members. Q2CURR01 is coded as 2, indicating that the mother is not currently living in the household. Since the mother is not currently living in the household, the respondent was asked to indicate why the mother is no longer in the household. This information is provided in the variable Q2REAS01, which notes that the mother is living elsewhere for employment reasons.

The round 2 roster variables for persons 2 through 4 are identical to the round 1 roster variables in this example, with the exception of age, which was increased. Additionally, now the child's grandfather occupies position 5 in the roster. Information about age and

Considerations for Analysis of ECLS-B Data

sex, but not race/ethnicity, was collected for the grandfather. Here we see that the information in the Q CURR variable has changed from “no” to “yes” to indicate that he is a current member of the household.

Slide 13 of 44

Codes on the ID variables for this round of data collection change a bit because the household composition has changed. The respondent is no longer identified as Person 1 because the mother is not in the household and is not the respondent to the parent interview. In this round, the step-father, who occupies roster position 3, is the respondent, so the code on the respondent ID variable, Y2RESPID, is 3.

Also, since the mother from round 1 is no longer in the household and there is no new person in the household who can be identified as a mother or mother figure, the mother ID variable, Y2MOMID, is -1, inapplicable. In situations where there is no mother in the household, Y2MOMTYP is coded 9.

Since the respondent has changed, and the step-father does not currently have a spouse/partner in the household, spouse ID is now set to -1, inapplicable.

Again, the sample child remains in roster position 2, so the child ID variable, X2CHID, is set to 2.

Slide 14 of 44

This grid presents the codes for the different identification variables side-by-side to highlight the changes to the variables between round 1 and round 2. Any change in codes indicates a change in the respondent to the parent interview or a change in household composition or, as is the case with this hypothetical family, in both. Thus, these variables can be used to track individuals across time, in particular to determine if the parents in the household are the same people across the different rounds of the study.

For example, by comparing the code for variable Y1MOMID to the code for variable Y2MOMID, you can see that the biological mother does not remain in the household across rounds. The identification variables also tell the data user that the father figure is the same from round 1 to round 2 since the code for variable Y1FTHID is the same as the code for variable Y2FTHID.

Keep in mind, however, that changes to the identification variables do not necessarily mean that the individual identified in one round is no longer in the household. For example, if at round 1, the child lived with her grandmother and grandfather, and neither the mother nor the father also lived in the household, the child's grandmother would be identified in variable Y1MOMID as the mother figure and the grandfather would be identified as the father figure in variable Y1FTHID. If, however, at round 2, the child's biological mother joined the household, she would be identified in variable Y2MOMID as the child's mother even though the grandmother still lived in the household.

Considerations for Analysis of ECLS-B Data

Therefore, while these identification variables can be used as an indication of possible household composition changes, the Q CURR variables need to be consulted in order to determine who is still in the household.

Slide 15 of 44

Composite variables related to the parents living in the household with the child, including mother's and father's race/ethnicity, age, education level, and employment status, provide information about the people who are identified in the MOMID and FTHID ID variables. The data for these variables come mainly from the parent and spouse sections of the parent interview. Because of the hierarchical rules used to determine who was identified as the child's mother or mother figure for MOMID and the child's father or father figure for FTHID that were discussed earlier, the respondent and his or her spouse are not always the people identified in MOMID and FTHID. This would be the case if, for example, a child lived with her biological parents, grandmother and grandfather, and the grandmother completed the parent interview. In this instance, the information collected in the respondent section would have been about the grandmother and the information collected in the respondent spouse section would have been about the grandfather. In cases where the respondent and his or her spouse are not identified as the child's mother or mother figure and father or father figure, the information collected in the respondent and respondent spouse sections of the parent interview are not used in the computation of the parent composite variables because they do not pertain to the people identified in MOMID and FTHID. In this instance, data for the parent composite variables will be missing or imputed.

This concludes the discussion of the household roster. To return to the objectives page to make another selection, click the "Return to List" button.

Slide 16 of 44

As discussed in the section of this module about the household roster, all permanent household members were listed, or rostered, during the Family Structure section of the parent interview. The household member who was a male spouse or partner of a female respondent to the parent interview was identified as the resident father. This resident father was the person to whom the self-administered questionnaire was given in the 9-month, 2-year, and preschool data collections. As a result of this method of identification, the person who completed the resident father questionnaire was not necessarily a biological or an adoptive father. For example, some resident fathers were non-married partners of the parents. In cases where the grandmother was responding to the parent interview, the person identified as the resident father was the child's grandfather. This design was chosen so that information could be obtained about the person in the household who was functioning as the father figure in the child's life. Analysts whose research question is specifically about certain types of fathers will need to consider the exact relationship of the father figure to the study child.

Another consideration when analyzing data from resident fathers is that there was a small number of cases in each round for which the resident father was the parent

Considerations for Analysis of ECLS-B Data

interview respondent. In these cases, fathers were not supposed to complete the resident father questionnaire. The questions from the resident father questionnaire were included in the parent interview and asked just of fathers. The father was also asked to complete the self-administered questionnaire for the respondent to the parent interview. In order for analyses on fathers to include all the information collected from fathers, analysts must use data from both the resident father self-administered questionnaire and the parent interview.

Slide 17 of 44

Nonresident fathers were identified in the parent interview, just as the resident father was. For the purposes of this study, nonresident fathers were identified as biological fathers who did not reside with the child. There were several eligibility criteria that had to be met before a child's biological nonresident father would be contacted and asked to complete a questionnaire.

First, the child had to reside with his or her biological mother and the biological mother had to be the respondent to the parent interview. This criterion was put in place because the study had to obtain permission from the biological mother to contact the biological father. In many cases this was necessary because the only way contact information for the nonresident father could be obtained was through the biological mother.

To be eligible for the questionnaire the nonresident father also had to meet some minimum criteria for contact. He had to have seen the child at least once in the last month, at least 7 days in the last 3 months, or been in touch with the child's birth mother at least once a month in the 3 months preceding the parent interview. These screening criteria were used for nonresident fathers in order to obtain reliable information on caregiving and involvement activities from these fathers. Also, these criteria were put in place before the first round of data collection. Nonresident fathers who did not have at least this minimum level of contact with their infant children were likely to be completely uninvolved or even unlocatable.

Slide 18 of 44

If the contact criteria were met and the biological mother provided contact information for the nonresident biological father, the father was sent a self-administered questionnaire to complete. If the mother did not have an address for the father but could provide a phone number, the nonresident father was contacted by phone. Due to low response rates, as well as budgetary constraints, data were not collected directly from nonresident fathers after the 2-year data collection.

Slide 19 of 44

Because the ECLS-B is a sample of children born in the United States in 2001, the data generalize to these children, and the child is always the unit of analysis, even when the focus of analysis is on fathers. The father data are not representative of all fathers, so it would not be appropriate to use these data make statements about fathers generally. Statements can be made about the fathers of children born in the United States in 2001, though even those statements need to be caveated. It must be kept in mind that the data are representative of children with fathers who met specific criteria in terms of knowledge of and involvement with their child. This is a particularly important consideration when looking at the nonresident father data. There can be important differences between nonresident fathers who have regular contact with their children, like those fathers included in the ECLS-B, and nonresident fathers who do not, especially on some of the topics that can be examined with the ECLS-B data, such as payment of child support. Researchers using the father data should carefully consider how the identification and eligibility of fathers may impact the findings of their analysis.

Slide 20 of 44

The data file contains variables that can be used to identify children with father data and the specific relationship the person identified as the resident father or father figure had to the study child. For the 9-month and 2-year rounds of data collection, there are variables that are coded 1 if data for resident or nonresident fathers are available in the data file. There is one such variable for the preschool collection that indicates whether data for resident fathers are available in the file.

As mentioned previously, in some cases, the parent interview was completed by the resident father. These cases can be identified using the variables that indicate the respondent relationship to the child, which have RSPREL in their names. Note that while this table only provides variable names through the preschool round, such variables are also available for the two kindergarten collections.

There are two sets of variables that help you identify a father's specific relationship to the child. The first set of variables includes those that have FTHTYP in their name and identify the type of resident father in the household, specifically whether there was a biological, adoptive, step, foster, or other kind of father in the household. These variables are available for all rounds of data collection. The second set of variables, which are only applicable for the 9-month, 2-year, and preschool rounds and have RELCH in their names, indicate the type of father who completed the resident father self-administered questionnaire.

Finally, there are two composite variables that indicate whether the type of resident father identified in the household in one round was the same type of resident father identified in the household at a later round. For example, these variables can be used to determine whether there was a biological father in the household at 9 months and a biological father in the household at 2 years. However, except in the case of a biological father, these variables do not necessarily indicate that the resident father is the same

Considerations for Analysis of ECLS-B Data

person across rounds or whether the same person completed the resident father questionnaire in multiple rounds. For example, a step-father in the household in the 2-year data collection may not be the same step-father in the household in the preschool collection. Household roster information can be used to determine whether the resident father is the same person across rounds.

Slide 21 of 44

In addition to the effects that the identification and eligibility of fathers may have on research findings, there are some other considerations to think through when using data on fathers over time. There is movement of fathers in and out of the children's households. Some children with a resident father in an earlier round do not have a resident father in later rounds. Other children were not living with a resident father in the 9-month collection but were in a later round. For those children who were living in households with resident fathers in multiple rounds, the person identified as the father or father figure in one round is not necessarily the person identified as the father or father figure in a different round. Any longitudinal analysis for which it is important that the father or father figure be present and/or the same person across rounds should take steps to identify and exclude cases for which the father changes.

Not only is it possible to have data on multiple fathers across the duration of the study, it is possible to have information on multiple fathers within one round of data collection. Specifically, if a child is living with a resident father and also has a nonresident biological father, there may be data from both the resident father self-administered questionnaire and the nonresident biological father questionnaire. Analysts will need to decide which of those data to use in analyses, depending on their research question.

One other thing to keep in mind is that the data file has some composite variables with information pertaining to the resident father. These variables draw on information collected in multiple places, including the father questionnaires and the respondent and spouse/partner information sections of the parent interview. Researchers should review the composite variables that are available to see if any of them meet their analytic needs.

This concludes the discussion of fathers. To return to the objectives page to make another selection, click the "Return to List" button.

Slide 22 of 44

The ECLS-B study included an oversample of twin pairs in order to facilitate analyses of twin characteristics. The oversample allows for better estimates of twin characteristics than would be possible if twins were sampled at the rate they occur in the population. In round 1, there are approximately 1,600 twins, constituting 800 twin pairs, with data on the data file. Round 2 includes approximately 1,500 twins, round 3 includes approximately 1,400 twins, round 4 includes about 1,200 twins, and round 5 includes approximately 400 twins. As a reminder, and as discussed in the introductory module, only certain groups of children were eligible for and included in the last round of data

Considerations for Analysis of ECLS-B Data

collection. Twins of eligible children were included even if they, themselves, were not eligible.

Slide 23 of 44

During the parent interview, parents were asked questions that pertained to both twins and for which responses were not expected to vary across twins only once. For instance, questions about family structure and community and social support, as well as information about the respondent and his or her spouse were asked just once.

Questions for which the answers could vary for the two twins were asked twice, once for each twin. For example, questions about child development, literacy, school readiness, health, and individual school experiences were asked separately for each twin.

Other study components handled the twin-specific questions differently. Please refer to the Data File User's Manual and the specific survey instruments for more details about the twin sample.

Slide 24 of 44

Since the ECLS-B data file is a child-level file, each twin is included on the data file as a separate child case with a unique value on the identification variable, I_ID, spelled I-underscore-I-D. There are several variables that can be used to identify which cases are twins and which cases constitute twin pairs.

The composite variables X1TWIN through X5TWIN indicate whether, at each round, there is a sampled twin case on the data file for any given child case. For example, if a specific case has X1TWIN equal to 1, yes, this means that for the 9-month data collection round, there are data on the file for another case that is the twin of this particular case.

The variable I_TWID, spelled I-underscore-T-W-I-D, is the I_ID (the child identification number) for the particular child's twin case.

Additionally, the variable I_TWIDPR, spelled I-underscore-T-W-I-D-P-R, is a unique identifier for every twin pair. It is set to the I_ID for one of the two twins in the twin pair and can be used to identify pairs of twins.

Finally, X1CHID through X5CHID are equal to the roster position (either 2 or 3) for each child. For twin pairs, one twin will always occupy roster position 2 and the other will always occupy roster position 3.

The next screen provides an illustration of these variables in the data file and how they can be used to identify twins and twin pairs.

Slide 25 of 44

Considerations for Analysis of ECLS-B Data

This table provides information for three hypothetical child cases in the data file, two of which are twins in the same twin pair and one of which is a non-twin. The first two rows correspond to the children who are twins.

Looking at the data for the case in the first row, the case has the child identification number 398474. X1TWIN is equal to 1, indicating that this case has a twin on the file with data for the 9-month data collection. The I_ID for this case's twin is given in the variable I_TWINID, which is equal to 427038. I_TWINPR is set to the I_ID for the first twin in the twin pair, which happens to be the case in this row, or 398474. Thus, in any given twin pair, there will be one twin for whom I_ID and I_TWINPR will be the same. Also, looking at the data for the case in the second row, the value for I_TWINPR is identical to the value of I_TWINPR for the child's twin in the first row. Both twins in each twin pair have the same value for I_TWINPR. Finally, X1CHID, which indicates the child's roster position, is set to 2 for twin 1 and 3 for twin 2.

The third case listed in this illustration shows the data on the file for a child case that is not a twin. In non-twin cases, I_TWINID and I_TWINPR are set to system missing and X1TWIN through X5TWIN are equal to 0, indicating there is no twin case with data associated with the case on the data file.

Researchers interested in comparing the characteristics of twins in pairs may wish to restructure the data file in order to create a twin-level file. There are multiple ways of doing this, depending on the statistical software you are using to analyze the data.

Slide 26 of 44

Researchers studying twins often want to know the zygosity of the twins. Zygosity is defined as the degree of identity in the genome, or whether the twins are identical or fraternal. Fraternal twins are dizygotic; they developed from two separate eggs. Identical twins are monozygotic; they developed from a single egg that split into two.

The only definitive evidence of zygosity is a genetic test. Since the ECLS-B did not include the collection of biological samples, children's DNA was not collected. As a result, the study does not have a definitive indicator of zygosity on the data file.

However, data were collected from parents and the field interviewers to provide some indication of the zygosity of the ECLS-B twin pairs.

Slide 27 of 44

The parent interview conducted in round 2 (when the children were approximately 2 years old) included a series of questions drawn from existing research to help determine the zygosity of same-sex twin pairs. These questions were not asked of parents of opposite-sex twin pairs, because such pairs are always fraternal.

The series of questions included prompts such as:

- Are {CHILD} and {TWIN} identical or fraternal (non-identical) twins? (This question was asked at both 9-months and 2-years.)
- Are there any differences between your twins in the following physical characteristics... (texture of your twins' hair, shades of your twins' hair colors, eye color, complexion, facial appearance, shapes of your twins' ear lobes)?
- Did the twins' first teeth begin to come in at about the same time?
- Are twins often mistaken for each other... (by respondent, by relatives, by strangers)
- Would you say that child and twin are as physically alike as "two peas in a pod"...?.

Slide 28 of 44

Similar questions were also included in an interviewer remarks questionnaire, which was completed at the end of the home visit, in order to obtain information independent of the parents' perception of the similarity of their children.

Responses from these questions were used to create a derived zygosity variable. Using published research about identifying zygosity based on parental report as a guide, a zygosity variable was created that identifies ECLS-B twins' zygosity based on the parent and field-interviewer survey items. This variable, along with a detailed description of how it was derived, is included in a supplemental data file included on the ECLSB 9-month through Kindergarten 2007 Restricted Use DVD in a folder titled "Zygosity."

All of the variables used in the derivation of zygosity are included in the main ECLS-B data file, so researchers could use this information in any way they feel would best identify the zygosity of the ECLS-B twins.

This concludes the discussion of twin pairs. To return to the objectives page to make another selection, click the "Return to List" button.

Slide 29 of 44

A primary purpose of the last two rounds of data collection was to collect information about children's kindergarten experiences. There are two features of the study design that must be kept in mind in analyses of kindergarten experiences, because these features have implications for the kinds of statements that can be made about the findings.

Considerations for Analysis of ECLS-B Data

First, it is important to remember that the ECLS-B data are representative of children born in the United States in 2001 who were born to mothers 15 years of age or older, were not adopted before they were 9 months old, and had not died or moved out of the country prior to the data collection round. These data are not representative of all kindergartners. A key group that would be included in a cohort of kindergartners, but that is not included in the ECLS-B, are children who came to the United States after they were born. The kindergarten round data can be used to make statements about the children born in the United States in 2001 when they were in kindergarten. The ECLS-B's study design supports research on children born in the U.S. during 2001, their development across the early years, their outcomes and experiences as preschoolers, and how their experiences as preschoolers relate to their experiences when they were followed up as kindergartners.

Second, the timing of data collection in the kindergarten rounds is an important consideration. Exposure to school can be an important characteristic to consider in many analyses. For most children, information was collected about kindergarten experiences in the fall of the academic year, before children had much exposure to school. Therefore, the data in the ECLS-B are representative of the characteristics and experiences of children in the cohort at kindergarten entry. The ECLS-B data are not well-suited for research questions about children's experiences over the course of kindergarten or about how kindergarten experiences, for example the characteristics of their teachers or classrooms, are related to kindergarten outcomes, because the sample children did not have a full year of kindergarten experience at the time of assessment and no data were collected toward the end of the children's kindergarten year.

Slide 30 of 44

The next few slides discuss how to decide which round or rounds of kindergarten data should be used in analyses concerned with children's experiences and characteristics in kindergarten. Before moving into that discussion, it is useful to quickly review the kindergarten data collection rounds and children's kindergarten enrollment status in each round.

The kindergarten 2006 data collection, referred to as round 4, included all the ECLS-B children regardless of their kindergarten enrollment at the time. Approximately 75 percent of the sample was in kindergarten or higher, either enrolled in school or being homeschooled.

For the kindergarten 2007 data collection, referred to as round 5, only a subsample of the ECLS-B children were included. This subsample comprised children who were not yet in kindergarten in the 2006 collection, children who were in kindergarten in the 2006 collection and were identified by parents as repeating kindergarten in 2007-08, and twins of children in each of these groups.

Approximately 4 percent of the children who were enrolled in kindergarten in 2006 repeated kindergarten in 2007.

Slide 31 of 44

Which round or rounds of kindergarten data collection are included in an analysis depends entirely on the research question being addressed, in particular whether school enrollment or exposure is likely to be related to the outcomes of interest.

Data collected in round 4 (the kindergarten 2006 round) alone are best suited to answer questions about children born in 2001 when they were approximately 5 years old and, specifically, questions for which school enrollment and school exposure are not likely to be related to the outcome of interest.

For example, the kindergarten 2006 data may be used to answer questions such as

- What percentage of children born in the U.S. in 2001 were overweight or at risk for being overweight when they were about 5 years old? What child and family characteristics are related to the likelihood that a child born in 2001 was overweight when he or she was about 5 years old? or
- What percentage of children born in the U.S. in 2001 had been diagnosed with certain illnesses, such as asthma or repeated ear infections, by the time they were about 5 years old? What are the factors related to whether a child has received one of these diagnoses?

Slide 32 of 44

Some research questions may be best answered by analyzing data collected only in the kindergarten 2007 round of data collection, though the types of questions that can be addressed with the 2007 data alone are more limited due to the nonrandom subsample from which data were collected in that round. Specifically, the kindergarten 2007 data are most useful for studying children who repeated kindergarten.

Examples of research questions that can be addressed by analysis of just the kindergarten 2007 data include

- What percentage of children born in 2001 attended kindergarten in 2006 and repeated kindergarten in 2007? or
- How is the likelihood of repeating kindergarten related to child, family, and school characteristics? Who is most likely to repeat kindergarten?

Slide 33 of 44

In order to answer research questions about children born in 2001 when they are first enrolled in kindergarten, specifically, questions for which school enrollment and exposure are related to the outcome of interest, data from both rounds 4 and 5, that is kindergarten 2006 and kindergarten 2007, should be used.

Researchers need to combine data from rounds 4 and 5 to answer questions such as

- What child and family characteristics relate to children's early reading and math ability in the beginning of kindergarten? or
- Is participation in early care and education prior to kindergarten entry related to children's cognitive and/or socioemotional development as they start kindergarten?

The next section of slides discusses in more detail how to combine data from the two kindergarten rounds of data collection to conduct an analysis of ECLS-B children at kindergarten entry.

Slide 34 of 44

For the majority of questions that examine the ECLS-B children at kindergarten entry, a new set of "kindergarten round" data will need to be constructed in order to create a set of kindergarten variables with data from the round of data collection in which the child was first in kindergarten.

This slide details the steps required to create this new set of data. The first step is to identify the round in which the child was first in kindergarten. The second step is to identify both the 2006 and 2007 versions of the variables to be used in analyses. Third, new variables need to be created from the separate 2006 and 2007 variables that use data from the round in which the child was first in kindergarten. As with any analyses, an appropriate sampling weight needs to be selected. A special set of weights that begin with the letters "WK" were developed specifically for analyses of children at kindergarten entry that use data from both kindergarten collections. The final step is to analyze the data using the new "kindergarten round" variables and the selected sampling weight.

The next five slides discuss each of these steps in more detail.

Slide 35 of 44

There are several variables on the data file that can be used to identify the round in which a given child was first in kindergarten. However, it should be noted that such identification was not always straightforward, due to the variety of types of kindergarten programs that the children attended. For example, some children attended a 1-year kindergarten program in a public or private school, some attended a 2-year kindergarten program in which attendance in the second year should not be considered repeating, some attended kindergarten in an early learning center where they also attended preschool, and some were homeschooled for kindergarten. Data obtained from parents or teachers, or both, were consulted to determine if and when a child was first in kindergarten, and children were classified using the information available.

The variable XKWHENK provides information on when children were first enrolled in school or being homeschooled for kindergarten or higher. The variable takes into account data from both kindergarten collections to identify when children were first enrolled in school or homeschooled and, for non-homeschooled children, their grade level. The majority of children were identified as enrolled in kindergarten (and not homeschooled) for the first time in 2006 or enrolled for the first time in 2007. This variable also identifies children who skipped kindergarten (that is, those who went directly into first grade), children who were enrolled in school for the first time in 2006 or 2007 but whose grade level was unknown or who were in ungraded or multigrade classrooms, and children who were not enrolled in school at the time of either the 2006 or 2007 parent interview. Researchers must decide how they want to treat homeschooled children and children who were not in traditional kindergarten classrooms when conducting analysis of children as they entered kindergarten for the first time.

Note that children identified in category 10, “unknown” are those children who were not yet enrolled in kindergarten in 2006 and were nonrespondents for the parent interview and child assessment in 2007. Thus, their status in 2007 is unknown.

The second variable listed here, XKKDATA, simply indicates in what round the child was first in formal school or homeschooled for kindergarten or higher. Thus, this variable indicates which round of data should be used in the creation of the new “kindergarten round” variables.

Slide 36 of 44

The next step is to select variables of interest. In most circumstances, the outcome variables and covariates such as demographic variables, family characteristics, and other independent variables of interest will all come from the round in which the child was first enrolled in school or being homeschooled for kindergarten or higher.

For example, if an analysis is looking at early reading scores at kindergarten entry by mother’s education level, the reading score and the mother’s education data would be taken from the round in which the child was first in kindergarten or higher.

Considerations for Analysis of ECLS-B Data

However, there are some exceptions to this general rule. Obviously, if a variable of interest is only available in one round, either because it is a static characteristic (such as sex, race or birth weight) or because the data were collected only in an earlier round, for example the duration of breastfeeding, then the analysis would use the variable from the round for which it is available.

Or, if the analysis is looking at the effect of an experience or characteristic in the year prior to kindergarten entry, for example nonparental care in the year before kindergarten, then data would come from the round prior to the round in which the child was first enrolled in school or being homeschooled for kindergarten or higher. The example provided in the choosing variables resource document illustrates an example of this.

Slide 37 of 44

The next step involves creating new variables that combine data from both kindergarten rounds of data collection. While referred to here as “kindergarten round” variables, it should be kept in mind that these data still describe children born in 2001 when they were in kindergarten.

The variable XKKDATA can be used to determine if the new “kindergarten round” variables will pull data from round 4 or from round 5. If XKKDATA is equal to 1, then the child was first enrolled in school or homeschooled in 2006 and the new “kindergarten round” variable will use data from round 4. If XKKDATA is equal to 2, then the child was first enrolled in school or homeschooled in 2007 and the new “kindergarten round” variable will use data from round 5.

The variable XKWHENK can be used to limit the analytic sample if so desired. For example, it can be used to exclude homeschoolers and/or children who went directly into a grade higher than kindergarten in any analyses focusing on children enrolled in kindergarten.

Slide 38 of 44

As mentioned earlier, there is a series of weight variables designed specifically for use in analyses of ECLS-B children at kindergarten entry, whether that was in the kindergarten 2006 or the kindergarten 2007 collection.

These weights, with the prefix “WK,” can also be used in analyzing parent data, as well as parent data in conjunction with data from other ECLS-B components.

Information presented in the ECLS-B Sample Design, Weights, Variance, and Missing Data can be used in conjunction with the technical documentation to determine which WK weight is most appropriate for your analysis.

Slide 39 of 44

Once the new “kindergarten round” variables have been created and a weight has been selected, you will conduct your analysis as you normally would using your new variables.

Results should be reported in terms of the ECLS-B target population. For example, an appropriate statement about findings from an analysis of early reading scores could begin “The average early reading score for children born in the U.S. in 2001 as they first enter kindergarten is...”

Remember, most kindergarten assessments were conducted in the fall; therefore, children had only had a few months of formal schooling at the time of the assessment.

An example using the steps outlined here is provided in the Analysis Example resource document.

Slide 40 of 44

There are also some considerations that must be kept in mind when analyzing data about the children who repeated kindergarten in 2007.

During the summer of 2007, a screener questionnaire was mailed to the homes of sampled children who were in kindergarten during the 2006-07 school year. This included children who were enrolled in kindergarten, those who were homeschooled for kindergarten, and children who were enrolled in ungraded or multi-graded classes. Parents were asked to indicate what grade their child would be enrolled in in the upcoming 2007-08 school year.

In addition to the children who had not yet started school in 2006, the kindergarten 2007 round included children whose parent had indicated on the screener that they would be in kindergarten again in 2007. Twins of children eligible for the 2007 round were also included even if they were not eligible themselves.

Slide 41 of 44

The variable X5RPTR, spelled X-5-R-P-T-R, can be used to identify children who were most likely repeating kindergarten in 2007. This variable has three categories: (1) Yes, child repeating kindergarten; (2) No, child not repeating kindergarten; and (3) child in 2-year, transitional, or ungraded program. X5RPTR is set to “not applicable” for those children who were not yet enrolled in school in 2006 or were homeschooled in either 2006 or 2007.

Multiple sources of information were used to classify children on X5RPTR, including the teacher’s report of grade level and class type and the parent’s indication of whether children were repeating kindergarten in 2007. The study documentation describes how these different reports were used to derive this kindergarten repeater indicator. All the

Considerations for Analysis of ECLS-B Data

data used to create this indicator are included on the data file so users may decide for themselves how to define repeaters if they desire to do so.

To analyze just children born in 2001 who entered kindergarten for the first time in 2006 and repeated kindergarten in 2007, the analytic sample should be limited to just those cases for which X5RPTR is equal to 1.

Slide 42 of 44

It is not necessary to combine data from multiple rounds to do analyses focused on kindergarten repeaters. Only data from the kindergarten 2007 round are needed, but the analytic sample must be limited, as just discussed.

Analyses of children repeating kindergarten should be weighted by one of the “W5” weights. Again, information presented in the module titled, “ECLS-B Sample Design, Weights, Variance, and Missing Data” can be used in conjunction with the technical documentation to determine which W5 weight is most appropriate for your analysis.

It must be noted that the ECLS-B data set only contains data for a certain group of repeaters, that is, those who were in kindergarten for the first time in 2006 and repeated kindergarten in 2007. Because there was no kindergarten 2008 collection, the study does not have information about the children who entered kindergarten for the first time in 2007 and repeated kindergarten in 2008. Thus, it is important to keep in mind that these data are NOT representative of all kindergarten repeaters or even of all children born in 2001 who repeated kindergarten. Therefore any analysis of kindergarten repeaters in the ECLS-B should be presented as “children born in the US in 2001 who were in kindergarten for the first time in 2006 and repeated kindergarten in 2007.”

This concludes the discussion of the kindergarten rounds. The module summary is presented next. To return to the objectives page to make another selection, click the “Return to List” button.

Slide 43 of 44

This module has described the analytic considerations that should be kept in mind when using data from some of these multiple sources of information in the ECLS-B. Specifically, considerations for using data collected from the household roster, fathers, twin pairs, and the two kindergarten data collection rounds of the ECLS-B were discussed.

Slide 44 of 44

Additionally, important resources have been provided throughout the module and are summarized here for your reference.

You have now completed this series of modules. Click the “Exit” button to return to the landing page.