Child Care & Early Education
RESEARCH CONNECTIONS

ICPSR 03804

Early Head Start Research and Evaluation (EHSRE) Study, 1996-2010: [United States]

United States Department of Health and Human Services. Administration for Children and Families

User Guide: Pre-Kindergarten (PreK) Follow-up Phase

ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

About *Research Connections*

These data are made available by the Child Care and Early Education *Research Connections* project. *Research Connections* promotes high quality research in child care and early education and the use of that research in policymaking.

Research Connections is operated by the National Center for Children in Poverty at the Mailman School of Public Health, Columbia University and the Inter-university Consortium for Political and Social Research at the Institute for Social Research, University of Michigan, through a cooperative agreement with the Office of Child Care, Office of Family Assistance and the Office of Planning, Research, and Evaluation, Administration for Children and Families in the U.S. Department of Health and Human Services.

About the Data

These data are based on research conducted as part of the Early Head Start Research and Evaluation Project. Mathematica Policy Research, Inc. was responsible for the creation of the public use files.

Terms of Use

Information about Copyrighted Content

OVERVIEW OF THE EHSRE PUBLIC USE DATA COLLECTION

As an introduction to this data collection, the following brief summary of the Early Head Start Research Evaluation (EHSRE) Study is provided to help better acquaint you with the data and documentation that are available for this collection. The EHSRE study is comprised of both public use and restricted use data. The Child Care & Early Education *Research Connections* data archive at ICPSR distributes the public use data and documentation only. Users interested in obtaining the restricted use data and documentation associated with this collection should visit the Henry A. Murray Research Center at Harvard University (link).

Please note, most of the documentation provided by the data producer, Mathematica Policy Research, was originally written to be used with the Restricted Use Data and has been adapted for use with the Public Use file. As such, not all references found within the documentation pertain to the Public Use file. These materials may refer to elements of the study that have been updated, are no longer in use for this release of the study, or that pertain to the Restricted Use Data.

The EHSRE project consists of three phases: (1) Birth to Three ["0-3"], (2) Pre-Kindergarten Follow-up ["PreK"] and (3) Elementary School Follow-up ["G5"]. Previously, *Research Connections* distributed the Birth to Three data only and now distributes all three phases. Content is presented in a different way than how it was with the original data release (June 2004). A summary of the data and documentation is provided below.

Summary of Data File Formats

The EHSRE data is provided as all phases combined into one data file, with multiple format options:

- **ASCII** (column-delimited, text) file with corresponding setup/syntax files
  - SAS
  - SPSS
  - Stata
- **ASCII** (tab separated) for use with other programs (e.g. Excel, etc)
- "**Ready-to-Go**" datasets that are ready to use without using setup/syntax file
  - SAS
  - SPSS
  - Stata

Special instructions for SAS formats

To optimize SAS capabilities when working with EHSRE files, Mathematica recommends that users make certain adaptations, related to missing values and SAS formats, to the default SAS program setup provided by *Research Connections* for reading in data from the ASCII text file. The special SAS formats are provided for those interested in applying labels to the special SAS missing values. Information about using/applying the SAS formats is provided below.

Missing Values

Most missing values in the text data files are in the form of negative numbers. Negative numbers are not recognized as missing values in SAS, but the provided SAS program includes code to convert them to the "special" missing value codes used in EHSRE data files (see table below). SAS users should ensure that the block of code (based on an array of all _numeric_ variables) that "changes the -[number]

missing values to the .[character] missing values" is active (not commented out), and that the code that changes negative values in each variable to "." is inactive (commented out), in order to have all of the missing value codes appear in their SAS data files.

Formats (value code labels)

The value statements in the Proc Format that are included in the *Research Connections* SAS program are based on the negative number missing value codes (as in the text data file). If you convert the missing values to special SAS codes (as described in the previous paragraph), and use the formats provided, then there will be no labels associated with any of the special SAS missing values in your SAS data file. You can either rely on the standard definitions of the missing value codes (see table below), or use the additional file provided, "Program-fmt.sas", to apply missing value labels. A separate FORMAT statement (for use in a SAS Data step or in a Proc such as Proc Freq) is provided for the data file, with the file name "Program-fmt_stmt.sas". If you are familiar with the use of %INCLUDE statements in SAS, you can use such statements to refer to the added formats files as needed, or you can copy all (or selected) text from the added files directly into your own SAS program files. You generally cannot combine the use of formats provided in the *Research Connections* SAS program with those in the separate files, because they use different names (in the value statements) for the same list of codes.

Missing Category	SPSS and Text File Code	SAS Code
Don't Know	-1	.A
Logical Skip	-2	.B
Refused	-3	.C
Not Applicable	-4	.D
Item Missing	-5 or SYSMIS	.E or .
Section Missing (may be due to phone interview, child not present)	-6	.F
Special Missing - CAPI problem; child data unscorable / uncodable / procedural problem	-7	.G
Not in Version (was added after data collection began)	-8	.H

Summary of Documentation

There are 4 PDF documents provided for use with the EHSRE data collection: 3 User Guides and 1 Codebook. A description of each follows:

- **User Guides** –provide information about the EHSRE study design, execution, and data to inform and assist researchers interested in using the one or more phases of the data for analyses. There are 3 User Guides in total –one for each phase. Each user guide contains information with a focus on the corresponding phase and includes the following topics:
  - Introduction to EHSRE Study
  - Study Design/Methodology
  - Data Collection
  - Data File Preparation
  - Naming and Coding Conventions
  - Data File Content
  - References

- **Codebook** - provides variable descriptions including labels, frequencies, summary statistics, question/derivation text information, and the data collection instruments. The following items are included in the User Guide:
  - Public Use File Variable List – list of the variables available in the public use data
  - BR Parent Interview Crosswalk - details the correspondence among questions and variables across the three parent interviews
  - Variable Descriptions and Frequencies
  - **Codebook Appendices, Volume I** –additional materials to provide a more detailed description of the entire set of EHSRE data including information on the source variables that went into the constructs that are in the public use file.
  - **Codebook Appendices, Volume II** – information about the survey instruments used during data collection
    - Baseline forms
    - Parent interviews**
    - Child assessments and video protocols**

**The instruments associated with the Grade 5 phase are provided as an appendix in the Grade 5 User Guide.

I. INTRODUCTION TO EHSRE STUDY

The national Early Head Start Research and Evaluation (EHSRE) study began in 1995, at about the same time that the first Early Head Start programs were funded. It was designed to provide a rigorous, comprehensive evaluation of the impacts of Early Head Start programs on low-income families and their young children through the children's third birthdays. The initial study followed children through age 3. The EHSRE then was extended to follow the original families and children at the children's PreKindergarten year when they were approximately five years of age. We refer to the original study as the "0 – 3" study wave and the PreKindergarten follow-up as the "PreK" wave. A 5th-grade follow-up wave is currently being conducted and data and documentation from the "G5" will eventually become part of this data archive.

The following sections provide some background on the EHSRE study design, data collection, and public use data file. More detailed information on study design and data is available in Early Head Start Research and Evaluation (EHSRE) Study, 1996-2010: [United States]: Codebook Appendices, Volume I, while copies of the questionnaires are available in Early Head Start Research and Evaluation (EHSRE) Study, 1996-2010: [United States]: Codebook Appendices, Volume II. Additional detailed information on the study design, as well as evaluation results when children were 3 years old can be found in Volumes I and II of the final evaluation report, *Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start* (2002a). Information on the PreKindergarten survey is found in*:* Love, J.M., Chazan-Cohen, R, Raikes, H.H., Faldowski, R.A., Vogel, C., Klute, M., Kisker, E.E., Brooks-Gunn, J., Martin, A. *What Makes a Difference: Analyses from the Prekindergarten Followup of the Early Head Start Study* (unpublished manuscript).

The EHSRE public use data now being released in the Child Care & Early Education *Research Connections* (CCEERC) contains variables used in analyses to produce the EHS (0 – 3) final evaluation report, *Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start* (2002a) plus variables used in PreKindergarten analyses in Love, J.M., Chazan-Cohen, R, Raikes, H.H., Faldowski, R.A., Vogel, C., Klute, M., Kisker, E.E., Brooks-Gunn, J., Martin, A. *What Makes a Difference: Analyses from the Prekindergarten Followup of the Early Head Start Study* (unpublished manuscript). A few items that would enable identification of families or sites have been excluded, and some items in this file may have only been used in exploratory analyses but not reported on. An expanded set of Early Head Start data files, including source data from direct interviews and assessments, data from additional sources, and qualitative data, is also available to qualified researchers as "restricted-use files (RUF)" through the Henry A. Murray Research Center at Harvard at http://www.murray.harvard.edu/ . The restricted-use files include technical support enabling better access to the complex network of files included within the Early Head Start data set. Limited technical support is available for the public use file in the CCEERC.

This User Guide documentation provides information needed to use the EHSRE public use file available through the CCEERC. The appendices to this documentation, available as appendices in the codebook, include additional materials to provide a broader understanding of the entire set of EHSRE data, including information on the source variables that went into the constructs that are in the public use file. Among the materials in the appendices are much more detailed information on the sources, content, and preparation of the full set of EHS data files, and copies of certain baseline and follow-up data collection forms (the baseline Head Start Family Information System "HSFIS" forms and four waves of parent interviews which were administered based on the child's age). Additional documentation on other data sources is available for the restricted-use data files. Two PreK documents have been added to the appendices: Parent Interview and the Child Assessment Protocol.

The following sections give an overview of the design and data collection for the EHSRE study, and information for users of the public use data file, including data file preparation, variable naming conventions, missing value codes, and the relationship between the contents of the public use file and the restricted-use files available through the Murray Center.

II. THE STUDY DESIGN

Funded by the Administration on Children, Youth and Families (ACYF), Early Head Start is a comprehensive, two-generation program that focuses on enhancing children's development while strengthening families. Designed for low-income pregnant women and families with infants and toddlers up to age 3, the programs provide a wide range of services through multiple program options. Services include child development delivered in home visits, child care, case management, parenting education, health care and referrals, and family support. Early Head Start programs meet families' and communities' needs through one or more official program options: (1) home-based, (2) center-based, (3) combination (in which families receive both home visits and center experiences), and (4) locally designed.

Site Selection. Seventeen Early Head Start programs participated in the national EHSRE study. They were drawn mostly from the Wave I Early Head Start programs that were funded in late 1995, and they had agreed, as a condition of funding, to participate in local and national research if selected to do so. ACYF purposely selected the research sites, based on proposals submitted to the Head Start Bureau by university research teams in partnership with EHS grantees. Their selection criteria included: (1) programs had to be able to recruit twice as many families as they could serve; (2) programs had to have a viable research partner; and (3) in aggregate, programs had to provide a national geographic distribution that represented the major programmatic approaches and settings and reflected diverse family characteristics thought to be typical of Early Head Start families nationally.

Because a program may offer multiple programmatic approaches, we characterized programs for research purposes according to the options they offered families. Thus, programs were grouped according to three approaches: (1) home-based, (2) center-based, and (3) mixed-approach.

Because the 17 research programs were not randomly selected, findings based on the data cannot be formally generalized to all Early Head Start programs. However, the features of the 17 programs, as well as the characteristics of their enrolled families and children, are similar to those of all Early Head programs funded in 1995 and 1996. Thus, to the extent that the quality and quantity of services offered in the 17 programs are similar to those offered nationwide, findings from these data are likely to pertain to Early Head Start programs more broadly[1].

Sample Enrollment. As soon as the programs were selected, beginning in spring 1996, Mathematica Policy Research, Inc. (MPR) staff began working with their staffs to implement the random assignment process in conjunction with each program's regular enrollment procedures. Although poverty-level families with children up to age 3 are eligible for Early Head Start, only low-income families with children up to 12 months old at the time of enrollment were eligible for the evaluation. Programs participating in the research were asked to focus recruiting and enrollment during the sample enrollment period on these families. Some sites also enrolled pregnant women. Overall, one-fourth of the families enrolled while pregnant with the focus child. Except for recruiting about twice as many families as they could serve, programs were expected to recruit as they would in the absence of the research, with special instructions to be sure to include all the types of families that their program was designed to serve (including those whose babies had disabilities).

Random Assignment. As soon as programs determined through their application process that families met the Early Head Start eligibility guidelines, they sent the names to MPR, and MPR staff entered the names and identifying information into a computer program that randomly assigned the families either to the program or to the control group (with equal probabilities). Program staff then contacted the program group families, while representatives of the local research partners notified the control group families of their status.

[1] Site codes are not included in the data file due to confidentiality concerns, but a variable indicating program approach is included.

Control group families were not allowed to receive Early Head Start services until their applicant child reached the age of 3. They could, however, receive any other services available in their community.

Sample enrollment and random assignment began in July 1996 and was completed in September 1998. In most sites, sample intake occurred over a two-year period, although some took less time.

During the sample intake period, 3,001 families were randomly assigned to the program (1,513) and control (1,488) groups. The samples in most sites include between 150 and 200 families, divided fairly evenly between the two research groups.

As expected, random assignment yielded program and control groups whose average baseline characteristics are very similar. Therefore, the only difference between the two research groups at random assignment was that the program group was offered Early Head Start services and the control group was not. Subsequent analyses showed that the baseline characteristics of program and control group members who actually completed subsequent interviews and assessments were comparable. Thus, differences in the subsequent outcomes of the two groups can be attributed to the offer of Early Head Start services with a known degree of statistical precision. A large number of explanatory variables based on baseline characteristics of sample members were included in the regression models used to estimate program impacts, which also gave equal weight to each site regardless of sample sizes.

For more details on study design and references to further information in project reports, see Early Head Start Research and Evaluation (EHSRE) Study, 1996-2010: [United States]: Codebook Appendices, Volume I.

The PreKindergarten Study Design

The PreKindergarten (PreK) follow-up study of EHSRE children and families was conceived and designed to document the long-term consequences of receiving either Early Head Start services or other community services up until age 3 combined with subsequent Head Start or

other formal early care and education programs on children's school readiness and parent functioning (ACF, 2007a).

The PreK follow-up study was designed to follow former Early Head Start and control group children and families from the time children turned 3 years old until the summer preceding their scheduled kindergarten entry.

III. DATA COLLECTION

Birth to Three Phase Data Collection

Data Sources. Comprehensive data from multiple sources were collected for the Early Head Start Research and Evaluation study. Baseline data were collected when participants applied to participate in Early Head Start. Follow-up interviews and assessments were conducted based on two different collection schedules. Each family's use of services and progress toward self-sufficiency were seen as likely to be a function of the amount of time since the family applied for Early Head Start services, so the interview schedule for these data was based on the number of months since random assignment. Other data—particularly those related to child and family development—were more likely to be a function of the increasing age of the focus child over time, so the data collection schedule for these interviews and assessments was tied to children's birth dates.

The data sources are summarized in Table III.1. Those from which constructed variables are included in the public use file are:

1. Baseline Data from the Head Start Family Information System (HSFIS) Program Application and Enrollment Forms and the MPR Tracking System. Most baseline data came from Head Start Family Information System (HSFIS) Program Application and Enrollment Forms that were completed by families at the time of program application. Some information on research status (program/control group assignment) and some characteristics of the applicant, mother, and focus child is from MPR's sample tracking system.

2. Parent Services Follow-Up Interviews (PSI) Targeted for 6, 15, and 26 Months After Random Assignment. These data contain information on (1) the use of services both in and out of Early Head Start (such as the receipt of home visits, and of services related to case management, parenting, health, employment, and child care); (2) progress toward economic self-sufficiency (such as employment, welfare receipt, and participation in education and training programs); (3) family health; and (4) children's health.

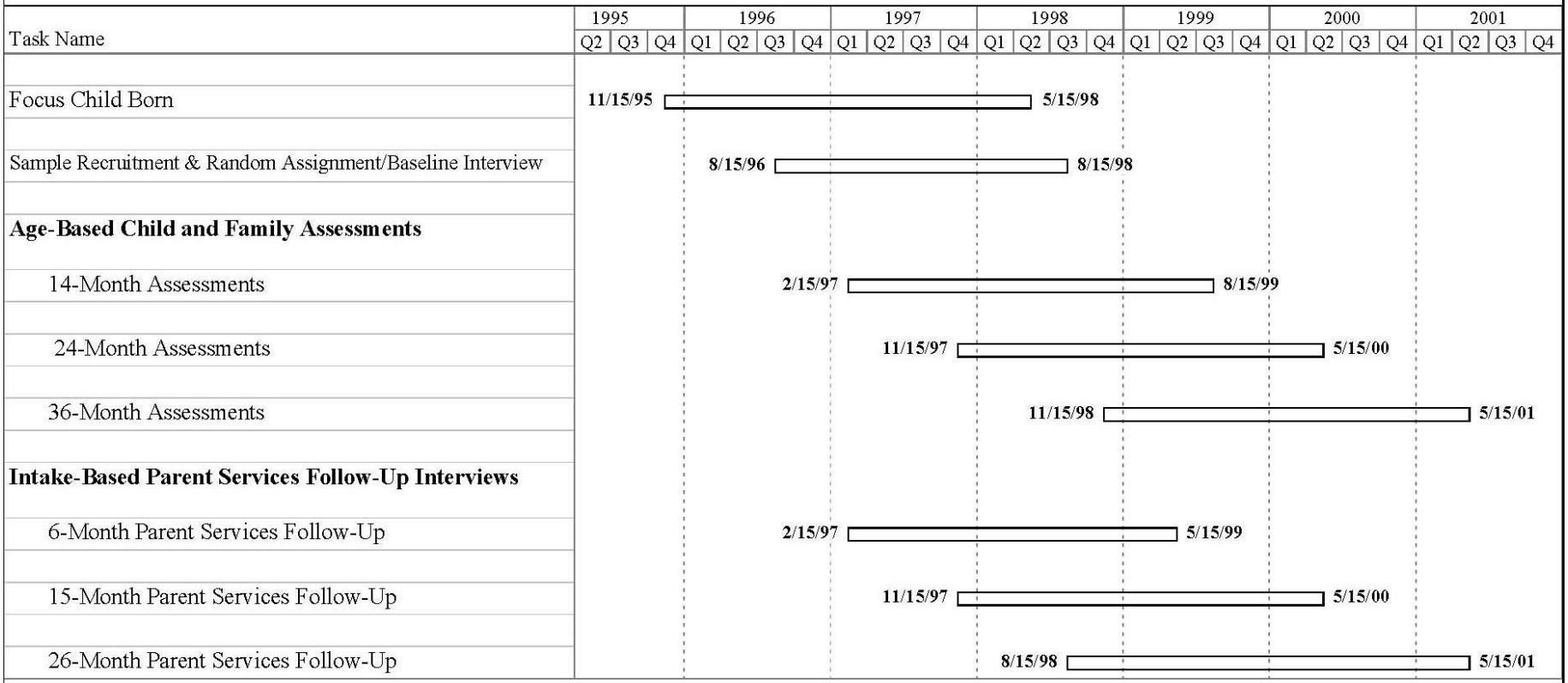
3. Parent Interviews (PI) Targeted for Completion When Children Were 14, 24, and 36 Months Old. These interviews obtained a large amount of information from the primary caregivers about their child's development and family functioning.

4. Child and Family Assessments Targeted for Administration When Children Were 14, 24, and 36 Months Old. Field interviewers conducted direct child assessments (such as Bayley assessments and Peabody Picture Vocabulary Tests [PPVTs], the latter only at 36 months) and recorded information from their observations of children's behavior and home environments for child and family assessments in Child Record Booklets (CRBs). They also videotaped semistructured parent-child interactions. Codes were developed from the videotapes and are included in the data files.

FIGURE 1

SAMPLE RECRUITMENT AND DATA COLLECTION FOR THE EARLY HEAD START EVALUATION

Date Ranges Represented in the Public Use Files



Dates in public use files were rounded and some outliers were bottom- or top-coded.

TABLE III.1 EARLY HEAD START RESEARCH AND EVALUATION (EHSRE) STUDY DATA SOURCES

Data Source	Content	Data Collection Schedule				
		At baseline	6, 15, and 26 months after random assignment	When child was 14 months old	When child was 24 months old	When child was 36 months old
Head Start Family Information System (HSFIS) Application and Enrollment Forms and MPR Tracking System	Baseline information on: Each family member, including applicant, father, and focus child (if born) Family circumstances The mother's pregnancy (if child unborn)	X				
Parent Services Follow-Up Interviews (PSI)	Information at intervals after random assignment on: The use of Early Head Start and other services (including home visits, case management, child care, and services related to parenting, health, and employment) Progress toward economic self-sufficiency (such as employment, welfare receipt, and education and training program participation) Family and children's health		X			
Parent Interviews (PI)	Information, at intervals based on the child's age, on many aspects of child development and family functioning, including: Parenting, discipline Home environment, health and safety The child's father/father-figure Child's developmental characteristics Child care arrangements Family & parent routines & activities			X	X	X
Child and Family Assessments	Collected separately: Bayley Scales of Infant Development (BSID) Mental Development Index (MDI) Codes developed from videotaped semistructured parent-child interactions In Child Record Booklet (CRB): Observations of children's behavior and home environments, including Bayley Scales of Infant Development (BSID) Behavioral Rating Scales Peabody Picture Vocabulary Test-III (PPVT), and/or Test de Vocabulario en Imagenes Peabody (TVIP), (only at 36 months)			X (not PPVT / TVIP)	X (not PPVT / TVIP)	X

Data Source	Content	Data Collection Schedule				
		At baseline	6, 15, and 26 months after random assignment	When child was 14 months old	When child was 24 months old	When child was 36 months old
Child Care Provider Interviews and Observations	Interview and observation data were collected from child care providers (both child care centers and family child care providers or relatives) for children who were in child care arrangements that met particular criteria.			X (if child in care that met criteria)	X (if child in care that met criteria)	X (if child in care that met criteria)
Father Interviews[2]	In 12 sites biological fathers and father figures, as identified by the children's mothers, were interviewed directly about fathering issues				X (selected sites)	X (selected sites)
Videotaped Father-Child Interactions and Father Child Record Booklets	In seven sites, videotaped semistructured father-child interactions were coded, and information from observations of the children and fathers is recorded in Father Child Record Booklets.				X (selected sites)	X (selected sites)
Exit Interviews	Final updated summary information on the use of services (analogous to questions in the PSIs, done only if at least one month past 26-month PSI) Questions about experiences in Early Head Start, for program group members only					X (selected circum-stances)

Note: There is no information from the shaded data sources in the public use data file.

5. Child Care Provider Interviews and Observations Targeted for Administration When Children Were 14, 24, and 36 Months Old. Interview and observation data were collected from child care providers for children who were in child care arrangements that met particular criteria when they were approximately 14, 24, and 36 months old. Different data collection instruments were used for children in child care centers and children cared for by family child care providers or relatives. However data from both types of providers may be used together for some types of analyses.

6. Father Interviews Targeted for Collection When Children Were 24 and 36 Months Old. In addition to asking mothers about their child's father, biological fathers and father figures in 12 sites were interviewed directly about fathering issues at the time of the 24- and 36-month birthday-related interviews (but not when children were 14 months old).

[2]The father study was supported with funding from the National Institute of Child Health and Human Development, the Ford Foundation, and the Office of the Assistant Secretary for Planning and Evaluation, as well as ACYF.

7. Videotaped Father-Child Interactions When Children Were 24 and 36 Months Old. In seven of the sites where fathers and father figures were interviewed, interviewers also videotaped semistructured father-child interactions when children were 24 and 36 months old, which were then coded and recorded information from their observations of children and fathers in Father Child Record Booklets. No constructs were created based on this data.

8. Exit Interviews When Children Reached 36 Months of Age. The exit interview obtained final updated summary information on the use of services (analogous to questions in the PSIs) and, for program group members only, included questions specifically about their experiences in Early Head Start. The Exit Interviews were generally conducted in conjunction with the 36-month PI; the general service use questions were only to be administered if the 26-month PSI had been conducted more than a month earlier. No constructs were created based on this data.

0 – 3 Survey Response. Response rates were higher for the PSIs and the PIs than for the Bayley and video assessments. Furthermore, as expected, response rates decreased somewhat over time. The rate for the PSI ranged from 82 percent for the 6-month to 70 percent for the 26-month; for the PI from 78 percent for the 14-month to 70 percent for the 36-month; and for the Bayley and video assessments from 63 - 66 percent at 14-months to 55 percent at 36-months. In general, the same families responded to the different interviews. For example, among those who completed a 36-month PI, 87 completed a 24-month PI, and 81 percent completed both a 14- and 24-month PI. Similarly, among those who completed a 36-month video assessment, 99 percent also completed a 36-month PI, and 92 percent also completed a 36-month Bayley assessment.

Importantly, response rates were similar for program and control group members for all data sources. However, they were consistently 2 to 6 percentage points higher for the program group.

Interviews were completed on average about one month after their target date (based on time since random assignment or age of focus child), and generally over 90% were completed between one month before and three months after the target date. The distributions of interview and assessment completion ages were similar for program and control group families.

See Early Head Start Research and Evaluation (EHSRE) Study, 1996-2010: [United States]: Codebook Appendices, Volumes I and II for more detail on the EHSRE data sources, the response

rates to the interviews and assessments, and the timing of interviews relative to when they were scheduled.

PreK Data Collection

The PreK data collection was targeted for the spring or summer before kindergarten entry. Children with spring and summer birth dates would have about a 2-year interval between their 3-year-old and prekindergarten data. However, depending on the age criteria set by states or local school districts for kindergarten entry (usually the date that a child turned 5 years old), children with fall birthdates might not be eligible for kindergarten until they were almost 6 years old. Thus, children's ages at the prekindergarten data collection varied more widely than they did on birthday-related data collected during the birth to three phase of the project. Child-related measures were adjusted for these wide age spans by using age-based scoring algorithms and norms where applicable. During the interval between the 0-3 data collection and the PreK interview, three brief tracking interviews were scheduled at varying intervals from the time of the child's 3rd birthday until just prior to the PreK data collection to maintain contact with the families and collect some information on types of childcare, continued participation in some form of Head Start program, parent jobs and income.

PreK data were collected in the spring and summer of 2001, 2002, and 2003. Data collectors scheduled time in each parent's home to interview the child's primary caregiver (usually the mother), assess the child, and videotape parent-child interactions. Windows for all prekindergarten data collection components opened in February preceding children's scheduled kindergarten entry and, with some variation, closed two months after kindergarten entry. Overall, 78.5% of parent interviews and 79.5% of child assessments were completed prior to kindergarten entry; while 14.5% of parent interviews and 14.6% of child assessments were completed within two months of kindergarten entry.

Scheduling and Procedures

Materials were designed so that parent interviews, child assessments, and videotaped parent-child interactions would be conducted first; then any eligible providers identified from the parent interviews. In a number of sites where many children attended 9-month early care and education programs, however, initial effort focused on identifying providers for teacher/director interviews and program observations before they closed for the summer. As a result, among the 1,123 families with program observations or teacher interviews, 57.9% were completed before the parent interview, 37.4% after the parent interview, and 4.7% without a parent interview.

Telephone parent interviews were an option for families who had moved out of the area, or for whom circumstances precluded in-person interviews. Spanish-language versions of the parent interview and child assessment materials were prepared for Spanish language dominant families; while videotaped parent-child interactions were conducted in whatever language the parent and child preferred to use.

Prekindergarten Data Collection

The prekindergarten data collection included four main components: (1) parent interviews provided a comprehensive assessment of parent and family functioning; (2) child assessments summarized child kindergarten readiness in key areas of attention, preliteracy and emerging numeracy skills, and receptive vocabulary; (3) videotaped parent-child interactions were used to characterize parent-child interaction styles and dyadic relationships; and (4) early care and education provider observations and interviews captured the quality and other characteristics of prekindergarten program environments that children attended.

1. Parent Interviews. Parent interviews, conducted with a child's primary caregiver (94.3% with the child's biological mother) obtained data on a number of important parental and child domains. They included the following: (1) child enrollment in Head Start, preschool, and other child care arrangements; (2) ratings of children's social competence, school readiness, and behavior; (3) parent-child activities in the home and community; (4) parenting and disciplinary practices; (5) parent health, mental health, and stressors; (6) child health; (7) services received;

(8) household composition; (9) the child's father/father figure; (10) parent's employment and household income; and (11) the home environment.

2. Child Assessments. The child assessment battery was designed to provide direct measures of key areas in child school-related achievement and behavioral domains considered critical for success in kindergarten. Intended to maximize comparability with direct child assessments conducted in the Head Start Family and Child Experiences Survey (FACES; ACF, 2003, 2006, 2007c) and other national studies of child development and school readiness, the Early Head Start prekindergarten child assessment battery included a series of tasks predictive of later school achievement. Specific domains covered in the assessment included sustained/focused attention, preliteracy skills (such as recognizing letters, words, and knowledge about books and stories), emerging numeracy (such as symbolic representation, counting, simple addition and subtraction), receptive vocabulary, and assessor ratings of child approaches to challenging and repetitive tasks. All measures were drawn from instruments with well-established psychometric properties that had previously been used to assess preschool-aged children in other national studies, such as FACES. Measures included the Leiter-R Attention Sustained Task (Roid & Miller, 1997); Woodcock-Johnson Tests of Achievement–Revised Letter-Word Identification and Applied Problems Tests (Woodcock & Johnson, 1990); Peabody Picture Vocabulary Test III (PPVT; Dunn & Dunn, 1997); *Goodnight Moon* (Brown & Hurd, 1975) story and print concept questions; and Leiter-R: Examiner Ratings (Roid & Miller, 1997).

Direct child assessment measures used with Spanish-speaking children included Spanish versions of standardized measures, when available, as well as Spanish translations of all other measures, scripts, and instructions. Standardized Spanish versions of corresponding English-language measures included the Woodcock-Muñoz Pruebas de Aprovechamiento-Revisada (Woodcock- Muñoz-R; Woodcock & Muñoz-Sandoval, 1996) for the Woodcock-Johnson Tests of Achievement – Revised, the Test de Vocabulario en Imagenes Peabody (TVIP; Dunn, Padilla, Lugo & Dunn, 1986) for the Peabody Picture Vocabulary Test III, and the book *Buenas Noches, Luna* (Brown & Hurd, 1995) for *Goodnight Moon* in the story and print concept questions (ACF, 2003, p.141; ACF, 2006, pp. A19-A20).

Across all sites, 187 (6%) of the 1,877 child assessments included Spanish measures. Out of these 101 (54%) were Spanish-only assessments; while the remaining 86 (46%) contained both English and Spanish measures. In 71 (83%) of the mixed language cases, the child assessment appears to have begun in English then switched to Spanish. Two children were assessed with both Woodcock-Johnson-R and Woodcock-Muñoz-R tests; and three children were assessed with both the PPVT-III and TVIP.

3. Videotaped Parent-Child Interactions. A standard protocol for obtaining videotaped parent-child interaction data was included in the prekindergarten follow-up to provide an estimate of the quality of parent-child communication and characteristics of parent's individual and dyadic behavior during a semistructured parent-child play interaction. Based on the 54-month parent-child structured interaction task developed for the National Institute of Child Health and Human Development Study of Early Child Care (NICHD SECC, 1995), these tasks were intended to capture critical correlates of child language development, social-emotional development, and school-related behavior problems. Also, in contrast to parent-, teacher-, and observer-rated aspects of child behavior, videotaped parent-child interactions allow important

types of parent and child behavior to be objectively rated by trained coders whose opinions and ratings are unaffected by supplemental knowledge of the children and their caregivers. The ratings are also relatively free of biases that may be introduced from interviewing and observing caregivers and their children in their homes.

Two tasks were recorded, a Play-Doh play task and a parent-child conversation that provided a speech sample. Videotapes of the Play-Doh task were coded by the research team at Columbia University.

4. Early Care and Education Program Interviews and Observations. For those children enrolled in a center- or home-based early care and education program, observers conducted interviews with the director and the child's primary teacher, as well as observations of the program environment. The goal of the provider interviews was to collect information on program characteristics; staff experience, education, and training; as well as provider knowledge and beliefs about child development and educational activities. It also included provider perceptions of the focus child's behavior and skills. Program observations obtained information on setting structure, program processes, and provider-child interactions. For most measures, parallel versions with minor wording changes and omissions of irrelevant sections were employed. The Family Day Care Environment Rating Scale (FDCRS; Harms & Clifford, 1989) was substituted for the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford & Cryer, 1998) to capture information on the quality of home-based programs.

Table III.2. Outcome Domains and Data Sources for PreKindergarten Study

Domain	Outcome	Items	Source
Child Negative Social-Emotional Development			
	CBCL Aggressive Behavior Problems		Parent Report
	FACES Aggression		Parent Report
	FACES Hyperactivity Scale		Parent Report
	FACES Withdrawn Behavior Scale		Parent Report
	FACES Total Problem Behaviors Scale		Parent Report
	Child Negativity during Play		Observer Rating/ Parent-Child Play Video
Child Positive Approaches Toward Learning			
	FACES Social Skills and Positive Approaches to Learning		Parent Report
	Leiter-R Cognitive Social Standard Score, This score includes Leiter-R Examiner Ratings for: Attention Scaled Score, Organization/Impulse Control,Scaled Score; Activity Level Scaled Score; and Sociability Scaled Score.		Interviewer Ratings/ Child Assessments
	Leiter-R Emotion Regulation Standard Score. This score includes Leiter-R Examiner Ratings for: Energy and Feelings Scaled Score; Mood and Regulation Scaled Score; Anxiety Scaled Score; and Sensory Reactivity Scaled Score[a]		Interviewer Ratings/Child Assessments

Domain	Outcome	Items	Source
	Leiter-R Attention Sustained[b]		Child Assessment
	Leiter-R Attention Sustained Total Correct		Child Assessment
	Leiter-R Attention Sustained Total Errors		Child Assessment
	Child Engagement of Parent during Play		Observer Rating/ Parent-Child Play Video
Child Pre-Academic Skills			
	English Woodcock-Johnson-Revised Letter-Word Identification		Child Assessment
	Spanish Woodcock-Muñoz-Revisada Identifcación de Letras y Palabras		Child Assessment
	English Receptive Vocabulary (Peabody Picture Vocabulary Test III)		Child Assessment
	Spanish Receptive Vocabulary (Test de Vocabulario en Imagenes Peabody)		Child Assessment
	Woodcock-Johnson-Revised Applied Problems		Child Assessment
	Woodcock-Muñoz-Revisada Problemas Aplicados		Child Assessment
	Child Individualized Education Plan		Parent Report
Child Health			
	Parent Rating of Child's Health		Parent Report
	Speech Problems		Parent Report
Parenting and Home Environment			
	HOME Total Score		Parent Report + Interviewer Observation
	HOME Learning Environment		Parent Report + Interviewer Observation
	HOME Warmth		Parent Report + Interviewer Observation
	Parent Supportiveness during Play		Observer Rating/ Parent-Child Play Video
	Parent Negative Regard during Play		Observer Rating/ Parent-Child Play Video
	Parent Detachment during Play		Observer Rating/ Parent-Child Play Video
	8 Teaching Activities		Parent Report
		Told child a story?	
		Taught child letters, words, or numbers?	
		Taught child songs or music?	
		Worked on arts and crafts with child?	
		Played with toys or games indoors?	
		Played a game, sport, or exercised together?	
		Took child along while doing errands?	
		Involved child in household chores?	
	Reads to Child Daily		Parent Report
	Children's Books (26 or more)		Parent Report
	Child Spanked within Past Week		Parent Report
	Parent Attended Meetings/Open Houses		Teacher Interview
Family Well-Being and Mental Health			
	Someone in household had alcohol/drug problem		Parent Report

Domain	Outcome	Items	Source
	Child Witnessed Violence		Parent Report
	Parent Health Status		Parent Report
	Depressive Symptoms (Center for Epidemiologic Studies – Depression scale short form)		Parent Report
	Parent Witnessed or Was Victim of Violence		Parent Report
	Parent Abused in Past Year		Parent Report
	Parent Self-Sufficiency		
	Time Employed in Past 6 Months		Parent Report
	Monthly Household Income (dollars)		Parent Report
	Early Care & Education Program Quality		
	Howes' Peer Play		Center Observer
	Early Childhood Environment Rating Scale-Revised		
	ECERS-R Total Score		Center Observer
	ECERS-R Space and Furnishings		Center Observer
	ECERS-R Personal Care Routines		Center Observer
	ECERS-R Language-Reasoning		Center Observer
	ECERS-R Activities		Center Observer
	ECERS-R Interaction		Center Observer
	ECERS-R Program Structure		Center Observer
	ECERS-R Parents and Staff		Center Observer
	Arnett Caregiver Interaction Scale		
	Arnett CIS Total Score		Center Observer
	Arnett CIS Positive Interactions		Center Observer
	Arnett CIS Punitiveness		Center Observer
	Arnett CIS Detachment		Center Observer
	Arnett CIS Permissiveness		Center Observer
	Early Head Start Age Participation Indicators		
	Ever in Head Start		Constructed variable: B4EVERHS
	In any formal program at age 3, imputed		Constructed variable: B4AFP36
	In any formal program at age 4, imputed		Constructed variable: B4AFP48
	In any formal program at both ages 3 and 4, imputed		Constructed variable: B4AFBOTH
	PreK: Parent supportiveness-average supportive & cognitive stimulation		Constructed variable: B4PSUP_R

[a] Individual ratings are included in the Leiter Emotional Regulation Standard Score as well.
[b] The Leiter Attention Sustained Score equals correct responses minus errors.

Prekindergarten Survey Response

In calculating the Prekindergarten survey response rates, nonrespondents are defined as any sample member for whom a designated data element was not collected. They include sample

members whose children died (and who were not, by design, followed-up), withdrew from the study, could not be located, or refused an interview/assessment/observation. At least one element of tracking or prekindergarten data was collected for 77.6% of the original 3,001 sample members. Prekindergarten data were collected on 2,142 (71.4%) of the sample.

Table III.3 presents response rates for the major tracking and prekindergarten data elements, for the total sample and separately for the Early Head Start and control groups. While the differences in response rates between the Early Head Start and control groups are statistically significant, they are not substantial (i.e. about 5 percentage points or less). Moreover, the response rate patterns are similar to those found at the 3-year data collection two years earlier (ACF, 2002, pp. 46-47 & Table II.2). Specifically, response rates are somewhat higher for the Early Head Start program group than for the control group (80.4% vs. 74.7% overall) and response rates are higher for parent interviews than for direct child assessments and videotaped parent-child interaction data.

Table III.3. Prekindergarten Survey Response Rates Overall and by Treatment Status for Key Follow-Up Study Data Elements[3]

Data Element	Overall	EHS	Control	Difference[4]
Any Tracking or Prekindergarten Data	2,329 (77.6%)	1,217 (80.4%)	1,112 (74.7%)	<.01
Any Tracking Interview	2,016 (67.2%)	1,047 (69.2%)	969 (65.1%)	<.01
Any Prekindergarten Data	2,142 (71.4%)	1,110 (73.4%)	1,032 (69.4%)	<.02
Prekindergarten Parent Interview	2,063 (68.7%)	1,071 (70.8%)	992 (66.7%)	<.01
Any Direct Child Assessment Data	1,877 (62.6%)	974 (64.4%)	903 (60.7%)	<.03
Any Parent-Child Video Data	1,808 (60.3%)	946 (62.5%)	862 (57.9%)	<.01
Sample Size	3,001	1,513	1,488	

[3]The percentages shown in these tables exclude missing values. A case was considered to be a respondent in the pre-kindergarten year if we had a completed parent interview, child assessment, video data, center data, or Howes data. These percentages may differ from those shown in other publications because of: the way missing values were treated, how respondents were defined, and the use of slightly different versions of the baseline characteristic variables.

[4] *p*-values refer to likelihood ratio chi-square tests of differential response rates between the EHS and control groups.

For additional information on data sources and response rates see Love, J.M., Chazan-Cohen, R, Raikes, H.H., Faldowski, R.A., Vogel, C., Klute, M., Kisker, E.E., Brooks-Gunn, J., Martin, A. *What Makes a Difference: Analyses from the Prekindergarten Followup of the Early Head Start Study* (unpublished manuscript).

Response Bias Analysis. Effects that sample attrition exerts on the composition of the follow-up sample expressed as response bias are as important as response rates. Internal validity of findings is maximized when there are few differences in baseline characteristics between

treatment and control respondents (Cook & Campbell, 1979; Shadish, Cook & Campbell, 2002). If this is the case, then tests of program impacts are valid for the sample of respondents included in the follow-up study. Internal validity is statistically assessed by comparing follow-up treatment to follow-up control group respondents. Attrition bias, an especially important threat to internal validity in longitudinal studies, is minimized when the follow-up sample still resembles, on baseline characteristics, the original sample that was randomized. Attrition bias is tested by comparing treatment group respondents to non-respondents and control group respondents to non-respondents.

Table III.4 presents data relevant for examining both internal validity and attrition bias. Overall, it suggests that the treatment and control groups are comparable at the prekindergarten follow-up point maintaining internal validity of the results. It also suggests that, despite some statistical differences between respondents and nonrespondents, on many important characteristics, the prekindergarten sample remains comparable to the baseline sample and the magnitude of attrition bias may be modest.

Table III.4[5]. Response Biases at Prekindergarten and 3-year Follow-Ups on Selected Baseline Characteristics

Characteristic	PreK Respondents			PreK Nonrespondents				3-year Respondents		
	EHS	Comp	(a)	EHS	Comp	(b)	(c)	EHS	Comp	(d)
Baseline Program and Contextual Characteristics										
Rural Setting	43.60	43.90		34.24	35.96	**	**	43.42	43.81	
Program Approach						***				
Center-Based	22.70	20.74		13.40	20.18			22.74	20.91	
Home-Based	44.95	45.54		51.61	45.61			45.21	44.80	
Mixed	32.34	33.72		34.99	34.21			32.05	34.29	
Implementation Pattern						***	**			
Early	34.41	34.30		34.74	35.96			35.18	35.78	
Late	38.74	37.79		24.57	29.17			37.69	36.08	
Incomplete	26.85	27.91		40.69	34.87			27.13	28.15	
Random Assignment Date (●)						***				
Before 10/1996	34.86	34.98		39.21	39.91			35.36	35.68	
10/1996 to 6/1997	28.65	31.01		34.49	30.48			28.65	31.32	
After 6/1997	36.49	34.01		26.30	29.61			35.99	33.00	
Baseline Sociodemographic Characteristics										
Highest Grade Completed (●)										
< 12	46.55	46.96		50.38	49.43			45.24	45.26	
12 or GED	28.54	29.55		24.55	30.57			29.23	29.48	
12+	24.91	23.48		25.06	20.00			25.53	25.26	
Race / Ethnicity (●)						**				
White	39.19	38.17		31.99	34.70			39.40	39.98	
African American	32.23	34.69		39.55	35.62			32.48	33.77	
Hispanic	24.63	22.86		21.66	24.66			24.02	21.67	
Other	3.94	4.27		6.80	5.02			4.09	4.58	
English Language Ability (●)						*	+			
English is primary language	80.24	79.55		79.39	75.00			80.35	80.10	
Speaks English well	8.53	10.17		12.47	10.38			8.76	10.21	
Does not speak English well	11.22	10.27		8.14	14.62			10.89	9.69	
Primary Occupation (●)										
Employed	23.72	24.47		20.76	22.25			25.35	23.87	
In school or training	22.33	21.55		21.27	21.33			21.92	20.68	
Neither	53.95	53.98		57.97	56.42			52.73	55.45	

Characteristic	PreK Respondents			PreK Nonrespondents				3-year Respondents		
	EHS	Comp	(a)	EHS	Comp	(b)	(c)	EHS	Comp	(d)
Living Arrangements (●)						+	***			
Living with spouse	25.41	26.29		23.63	23.33			25.96	27.09	
Living with other adults	40.00	41.29		33.58	34.22			38.50	38.75	
Living with no other adults	34.59	32.42		42.79	42.44			35.54	34.16	
Adult Male Present in Household (●)	39.64	41.19		33.83	34.22	+	*	40.02	40.94	
Number of Adults in Household						**	***			
1	35.62	33.30		43.78	44.22			36.74	35.06	
2	51.04	52.39		46.27	47.11			50.18	51.39	
3 or more	13.35	14.31		9.95	8.67			13.08	13.55	
Number of Children 0-5 in Household (●)										
0	63.60	64.95		66.17	65.56			64.82	63.35	
1	27.75	26.39		24.88	27.78			26.68	28.29	
2 or more	8.65	8.67		8.96	6.67			8.50	8.37	
Number of Children 6-17 in Household (●)										
0	63.15	64.85		67.41	69.78			63.12	65.74	
1	23.51	22.01		21.89	19.78			24.26	21.22	
2 or more	13.33	13.15		10.70	10.44			12.62	13.05	
Number of Moves (●)						+	+			
0	51.16	51.86		43.92	44.47			50.76	51.08	
1	27.75	26.86		33.33	30.59			28.29	27.70	
2 or more	21.10	21.27		22.75	24.94			20.95	21.23	
Owns Home	12.88	12.59		5.74	7.71	***	**	12.37	12.28	
Household Income as a Percentage of Poverty Level (●)										
Less than 33	29.87	30.17		31.02	29.54			29.26	27.91	
33 to 67	31.06	28.15		36.45	31.71			31.30	30.46	
67 to 99	25.32	27.32		20.48	24.66			25.19	27.43	
100 or more	13.74	14.37		12.05	14.09			14.26	14.20	

Characteristic	PreK Respondents			PreK Nonrespondents				3-year Respondents		
	EHS	Comp	(a)	EHS	Comp	(b)	(c)	EHS	Comp	(d)
Prior Enrollment in Head Start or other Child Development Program (●)	12.91	13.37		12.47	13.46			13.02	14.08	
Baseline Entitlement Receipt										
Received AFDC/TANF (●)	33.30	33.50		42.06	37.29	**		32.74	33.33	
Received Food Stamps (●)	45.21	46.40		55.94	50.94	** *		45.79	46.40	
Received Medicaid	75.19	75.63		80.74	72.41	*		75.70	74.66	
Received SSI (●)	7.33	6.80		6.07	7.55			6.82	7.30	
Received WIC (●)	87.50	86.60		87.34	84.43			87.38	86.13	
Received Public Housing	10.34	7.72	*	7.12	11.56	+	*	10.00	8.65	
Baseline Resource Inadequacies										
Food (●)	4.60	6.21		5.82	6.53			4.74	6.83	+
Housing (●)	12.49	12.10		11.61	15.99		+	11.92	12.04	
Money for Necessities (●)	19.58	20.28		24.46	25.00	*	+	19.96	20.31	
Medical Care (●)	14.06	13.70		13.66	17.09			13.43	13.98	
Transportation (●)	20.71	21.65		21.49	24.35			21.01	22.25	
Childcare	33.00	32.43		38.72	39.71	+	*	33.37	34.08	
Money for Supplies	26.45	28.32		29.08	31.87			25.03	30.18	*
Support from Family and Friends	12.84	12.10		13.04	18.37		**	12.23	11.86	
Parenting Information	36.72	37.74		31.37	40.66	+		36.07	38.31	
Baseline High Risk Indices										
Teen Mother (< 20 yo) (●)	37.52	38.76		40.15	37.72			37.05	37.56	
Received Welfare	51.69	51.47		60.69	55.42	**		52.15	51.30	
Not Married/Cohabiting	74.59	73.71		76.37	76.67			74.04	72.91	
Less than HS Diploma or GED	46.55	46.96		50.38	49.43			45.24	45.26	
Not Employed/School/Training	53.95	53.98		57.97	56.42			52.73	55.45	
Maternal Risk Index						*				
0-2 risks	44.07	43.64		35.79	41.12			44.98	44.66	
3 risks	29.68	32.07		36.34	29.68			30.63	30.96	
4-5 risks	26.25	24.29		27.87	29.20			24.39	24.38	
Baseline Child Characteristics										
Focus Child's Age at Randomization (●)										
Unborn	24.50	27.33		23.33	24.78			23.99	25.77	
0-4 months	35.95	34.50		38.96	35.31			35.27	34.99	
5-12 months	40.45	38.18		37.72	39.91			40.73	39.25	

Characteristic	PreK Respondents			PreK Nonrespondents				3-year Respondents		
	EHS	Comp	(a)	EHS	Comp	(b)	(c)	EHS	Comp	(d)
Focus Child is Male (●)	51.31	50.97		52.71	49.07			50.94	50.00	
Focus Child was Firstborn Child (●)	60.73	62.85		66.75	62.70	*		61.87	60.71	
Child Born more than 3 Weeks Early (●)	15.01	11.74	*	14.64	13.01			13.80	11.84	
Birthweight less than 2500 grams (●)	9.19	7.60		10.34	10.34			8.54	7.49	
Stay in Hospital after Birth	17.81	15.94		17.76	16.57			17.14	16.33	
People Concerned about Child's Health	12.45	13.88		14.54	11.84			12.74	14.37	
Child Received Evaluation (●)	5.25	7.08		8.27	6.62	+		5.44	6.36	
Child has Estab/Bio.Med/Environ Risks	42.03	44.88		41.81	45.02			41.16	45.01	
Child has Established Risks (●)	11.29	10.25		12.54	11.29			11.74	10.14	
Child has Biological or Medical Risks (●)	17.31	16.16		20.91	18.39			17.43	16.81	
Child has Environmental Risks (●)	33.38	36.65		29.97	35.69			32.01	36.47	+
Child Covered by Health Insurance	90.86	91.04		88.01	86.47		*	91.02	92.33	

[5] The percentages shown in these tables exclude missing values. A case was considered to be a respondent in the pre-kindergarten year if we had a completed parent interview, child assessment, video data, center data, or Howes data. These percentages may differ from those shown in other publications because of: the way missing values were treated, how respondents were defined, and the use of slightly different versions of the baseline characteristic variables.

(●) Denotes impact analysis control variable

(a) Significance levels are from tests comparing preK program and control group respondents (preK internal validity)
(b) Significance levels are from tests comparing program group respondents and non-respondents (preK external validity)
(c) Significance levels are from tests comparing control group respondents and non- respondents (preK external validity)
(d) Significance levels are from tests comparing 36-month program and control group respondents (3-year internal validity)

*** Contrast is statistically significant at the .001 level
** Contrast is statistically significant at the .01 level
* Contrast is statistically significant at the .05 level
+ Contrast is statistically significant at the .1 level

The second and fourth sets of columns in Table III.4 (labeled "PreK Respondents" and "3-year Respondents") display data on baseline characteristics of respondents useful for assessing internal validity at the prekindergarten and 3-year follow-up interviews, respectively. The indicators in columns (a) and (d) show very few significant differences between the program and

control respondents on any of 51 baseline characteristics measured, suggesting that prekindergarten follow-study findings will have the same high degree of internal validity that the 3-year findings had. Among prekindergarten respondents, only two of the differences are statistically significant, even at the liberal significance level of .10. This is less than half of the 5 differences that would be expected by chance. Therefore, we conclude that an impact analysis would represent a valid assessment of the effects of Early Head Start treatment.

The second and third sets of columns in Table III.4 (labeled "Prek Respondents" and "Prek Nonrespondents") present information relevant for assessing attrition bias, or whether the results of the prekindergarten follow-up can be generalized to the original Early Head Start sample recruited between 1996 and 1998. The primary question is whether respondents (who were observed during the prekindergarten follow-up) systematically differ on baseline characteristics from nonrespondents (sample members not observed at the prekindergarten follow-up). Any characteristic on which they differ suggests divergence of the follow-up sample from the baseline sample recruited; however, if the magnitude of the difference is small, possible bias may still be considered minimal. In the second and third sets of columns in Table 3, the comparisons of greatest interest are between Early Head Start PreK Respondents (1st "EHS" column) and Early Head Start PreK Nonrespondents (2nd "EHS" column), and between Control Group PreK Respondents (1st "Comp" column) and Control Group PreK Nonrespondents (2nd "Comp" column). The indicators in columns (b) and (c) describe characteristics on which respondents differ from nonrespondents within the Early Head Start and control groups, respectively. The following are the important trends in this analysis:

- On a number of important baseline demographic characteristics, the prekindergarten sample does not measurably differ from the full baseline sample either overall or within group. Specifically, the baseline and prekindergarten samples are similar on such key characteristics as age of mother at birth of focus child; mother's educational attainment; whether employed or in school/training; number of children in the household; income; age of focus child at

enrollment; focus child with pre-term birth or low birth weight; or having established, biological, or environmental risks.

- The Early Head Start sample at prekindergarten over-represents center-based programs (22.7% of respondents were in center-based programs vs. 13.4% of nonrespondents) and programs that were later implementers of the Head Start Program Performance Standards; while under-representing home-based and incomplete implementers.

- The Early Head Start sample at prekindergarten is somewhat less disadvantaged than the baseline sample: Respondents were more likely to own their own home, less likely to be receiving TANF or Food Stamps, more likely to be low risk, and more likely to be living with a spouse or other adults.

- The control group prekindergarten sample shows fewer differences from its baseline sample than the Early Head Start group does.

These differences suggest that attrition bias may be manifest in a number of characteristics and some caution is warranted when drawing inferences from the prekindergarten analyses and applying them to the original Early Head Start research sample. The magnitude of the statistically significant differences, however, are not large and values on the characteristics remain well with the ranges of variability observed among low-income families eligible for Early Head Start services nationally, suggesting that practical implications of the potential bias may be small. We conclude, therefore, that although the prekindergarten sample may have diverged from the original sample randomized on a few characteristics, on many others it remains comparable and findings from the prekindergarten sample in the Early Head Start Research and Evaluation Project continue to apply to families served by Early Head Start programs.

IV. DATA FILE PREPARATION

Data Cleaning. The primary goal in checking and cleaning the survey data was to minimize data errors and create data files that are as accurate as reasonably possible. In general, we took a conservative approach to data cleaning. That is, data reflect the information recorded on the hard-copy instrument or in CAPI, except to correct any obvious recording errors (such as transposed digits, or incorrect dates that could be verified from other sources) and any errors that may have crept in during the data entry process. We double-checked hard-copy documents as needed to resolve apparent discrepancies involving critical data items, but were not able to check all questionable data against hard-copy instruments.

Scale Scores. In computing a score for a multi-item scale for inclusion in a data file, analysts consulted scale developers and experts regarding how to handle missing items for particular scales. In the absence of guidelines for a specific scale, however, analysts implemented the following convention for handling missing items: If more than 25 percent of the items were missing for an individual case, or the scale was not based on a simple calculation, the scale was coded as missing. If 25 percent or fewer of the items were missing from a multi-item scale and the scale was based on a simple calculation such as a mean or sum, the score was calculated using imputed values for the missing items. For example, if the total score was the sum of the item scores or the average item score, we assigned the average item value across nonmissing items to the missing items. The actual item responses are included in the source data file and documentation unless copyright or agreements with test publishers prohibit it.

Confidentiality Issues. To ensure respondent confidentiality and reduce the risk of individual disclosure, a few cases and some variables were omitted from the public and restricted-use data files, and changes were made to other variables in these data files, as follows:

- We excluded all data from 24 families for whom our records indicate a miscarriage or death or adoption of the focus child within three years after random assignment. Depending on the date of death, few of these families were eligible for any data collection, and very few completed any interviews.

- Because sample sizes in individual study sites are relatively small, no information is provided on families' site location. We have created new case IDs (IDnum) that are not assigned by site, and excluded from the files the ID numbers of child care providers and data collectors, both of which could provide the means to link cases by site.

- Due to confidentiality concerns, dates, including the date of random assignment, dates of data collections, and the focus child's date of birth, have been rounded to the midpoint of the calendar quarter, and some have also been bottom and/or top-coded.[6] Restricted-use files include the focus child's age in months at the time of the data collection.

- Dates of birth for persons other than the focus child have been omitted from the files, and in their place, the person's age in years at the time of the data collection (and at random assignment, for the applicant) is provided in restricted-use files. Some ages were also bottom- and/or top-coded; other household members were top-coded at age 66, and most other limits are specified in variable labels.

For additional information on our preparation of the data files see Early Head Start Research and Evaluation (EHSRE) Study, 1996-2010: [United States]: Codebook Appendices, Volume I.

[6] We bottom-coded variables with small frequencies in the lowest categories by collapsing those categories into the highest of the low-incidence categories. Similarly, we top-coded variables with small frequencies in the highest categories by collapsing those categories into the lowest of the high-incidence categories. Variable labels and/or notations in the codebook note when this was done and how it was implemented for particular variables.

V. NAMING AND CODING CONVENTIONS

To facilitate the use of the data, variable names have been created based on a set of naming conventions. The names include a document-specific prefix combined with further specifying information. The prefixes indicate the type of document and the "wave" from which the variable came, as described below. Most variable names for items taken from survey instruments and similar documents are based on the document prefix plus the item number in the hard-copy instrument. Some constructed variables also make use of the document prefixes, plus mnemonic information to complete the name, although other constructed variables use strictly mnemonic names. The naming conventions and document prefixes, as well as standard codes used in the data files are described in more detail below, and in further detail in Early Head Start Research and Evaluation (EHSRE) Study, 1996-2010: [United States]: Codebook Appendices, Volume I.

A. DOCUMENT RELATED PREFIXES/NAMING CONVENTIONS

Baseline/HSFIS Data. Constructed baseline subgroup variables and items from MPR's tracking system have mnemonic names, while variables directly from the HSFIS include a prefix to indicate from which of the forms they were taken:

- Prefix (first digit): H
- Form indicator (second digit): A=Application form, E=Enrollment form

Parent Services Interview (PSI) and Exit Interview Data. Some PSI constructs have mnemonic names, while the PSI and Exit Interview source variables and some constructs are identified by 2- or 3-digit prefixes:

- Prefix (first digit): P
- Wave indicator (second digit): 0 = 06-month PSI, 1 = 15-month PSI, 2 = 26-month PSI, 3=Exit Interview

- "Cumulative" indicator (third digit, constructs only): V (from eVer) indicates that the variable is cumulative up to the wave specified in the second digit.
- Examples: P0_HEAR is whether the focus child had received hearing testing, according to the 6-month PSI, P1V_HEAR is whether the focus child had ever received hearing testing by the 15-month PSI, using information from both the 6-month and 15-month PSIs.

Birthday-Related Data. The multiple elements of the birthday-related data collection are identified with prefixes consisting of the following digits:

- Prefix (first digit):
  - B = Birthday Related (for all 0-3 variables, and PreK parent interviews and videos)
  - C = Child assessment and observations (PreK)
  - T = Caregiver ("Teacher")interviews and observations (PreK)
- Wave indicator (second digit): 1 = 14-month data collection, 2 = 24-month data collection, 3 = 36-month data collection, 4=Pre-Kindergarten data collection, V = "eVer" used for constructed variables combining information from the first 3 waves (14-, 24-, and 36-month)
- Document indicator (third digit, and sometimes fourth digit):
  - P = **P**arent interview
  - R = child **R**ecord booklet
  - B = **B**ayley
  - V = **V**ideotaped interactions codes (fourth digit indicates type of task: 3=**3** bag (the parent-child semistructured play task), H=**H**igh chair, T=**T**eaching, P=**P**uzzle)
  - D = father (**D**ad) interview
- Examples: B2V3PINT is a variable from the 24-month video 3-bag task coding, BVP_MAL1 is constructed using information from the first 3 waves of parent interviews.

Constructed Variables. The majority of constructed variable names include a prefix indicating the data source in the first digits, as described above, then mnemonic characters for the remainder of the names. Some based on baseline/HSFIS, PSI, and child care observations data have fully mnemonic names.

B. VARIABLE CODING CONVENTIONS

Most variables in the EHSRE public use file are binaries, coded: 0 = No, 1 = Yes. The units for continuous variables and codes for many categorical variables are indicated in the variable labels; please refer to the detailed codebook for further information on the codes for specific variables. (Some code information is also provided for SPSS and SAS users through value labels and format value statements.) The same set of standard codes is used to indicate reasons for missing data in all types of numeric variables.

The standard codes for missing numeric data[7] for all EHSRE data files differ depending on the file format, as shown in the chart below. In all text files and SPSS files, and all files available through the CCEERC, various types of missing data are represented by negative number codes. In SAS files available through the Murray Center a corresponding set of SAS "special" missing value codes are used. The file of SPSS data definition statements (provided for users who wish to create an SPSS file from the text file) includes SPSS syntax to assign the negative numbers as missing values for all numeric variables. The file of SAS data definition statements (provided for users who wish to create a SAS file from the text file) includes a "commented out" block of SAS code to convert the negative numbers all to the basic "." missing code. The creation of a SAS file from a text file should include SAS statements to convert the negative numbers to the special SAS missing codes as listed in the following chart.

7 The single exception in the public use file to the negative missing value codes is the variable CMTHS, which is the age in months of the Focus Child at the random assignment date, and includes valid negative numbers and NO missings.

Conversion of Missing Categories to SPSS Negative Codes and SAS Code

Missing Category	SPSS and Text File Code	SAS Code
Don't Know	-1	.A
Logical Skip	-2	.B
Refused	-3	.C
Not Applicable	-4	.D
Item Missing	-5 or SYSMIS	.E or .
Section Missing (may be due to phone interview, child not present)	-6	.F
Special Missing - CAPI problem; child data unscorable / uncodable / procedural problem	-7	.G
Not in Version (was added after data collection began)	-8	.H

VI. THE PUBLIC USE AND RESTRICTED-USE MICRODATA FILES

The EHSRE public use data in the CCEERC contains variables used in analyses to produce the EHS final evaluation report, *Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start* (2002a) plus variables used in PreKindergarten analyses in Love, J.M., Chazan-Cohen, R, Raikes, H.H., Faldowski, R.A., Vogel, C., Klute, M., Kisker, E.E., Brooks-Gunn, J., Martin, A. *What Makes a Difference: Analyses from the Prekindergarten Followup of the Early Head Start Study* (unpublished manuscript). Most variables in the public use file are constructed variables but some source variables are included because they were suitable for analyses as defined. Among all the EHSRE data sources (as listed above in Section III), all variables in the Public Use File are drawn from the following:

- Baseline Data from HSFIS Forms and the MPR Tracking System
- Parent Services Follow-Up Interviews (PSI) (3 waves, scheduled related to assignment date)
- Parent Interviews (PI) (4 waves, schedule related to focus child birth date)
- Child and Family Assessments (4 waves, schedule related to focus child birth date)
- Child Care Provider Interviews and Observations (4 waves, schedule related to focus child birth date)
- Father Interviews (3 waves, schedule related to focus child birth date)

The public use variables are listed in the next section of this codebook, grouped according to their data source and research topic. In some cases related groups of variables are listed together on a single line (indicated by a checkmark in the Array column), with a "wildcard" representation of the names of the variables. In most cases the "wildcard" is an "n" in the second digit of the variable names, which represents the data waves (see Section V above on variable naming conventions). Information on the specific waves for which the variable is defined is in

the Data Waves column. For other PSI based constructs, there may be a “wildcard” “n” or “j” at the end of the variable name (e.g. P2V_HWn, AFDCQj), where the varying number at the end of the variable names may represent calendar quarters or other time periods, as explained in the codebook entries for the variables.

The source data from which these variables were derived, as well as additional data collected for the EHSRE study, are available to researchers through a restricted-use archive at the Murray Center. The source data are generally stored in a separate data file for each data source, including separate files for each of the 4 waves of most types of follow-up data. However, data from certain closely related sources have been put into merged files to reduce the total number of separate data files: at each of the 3 “birthday related” age points and at PreK data from various “child and family assessments” (Bayley assessments, video interactions coding, PPVT/TVIP, and Child Record Booklet) are in one file and information from all of the various instruments used to collect information on childcare providers has been merged into one file for data from centers and one for data from family providers. Each data file contains a record only for families with data for that source (or from one of the several sources in the merged files). Data from all files can be merged using a common family ID variable, IDnum. Each file has only one (or no) record per IDnum (with the exception of two files with specialized baseline data on family members).

The codebook appendices provide much more detailed information on the EHS data sources and data files, including copies of certain baseline and follow-up data collection forms (the baseline Head Start Family Information System “HSFIS” forms and four waves of parent interviews). Additional documentation on other data sources are available for the restricted-use data files.