Early Head Start Family and Child Experiences Survey (Baby FACES) Design Options Report

September 2015

Baby FACES 2009

OPRE Report 2015-99

Early Head Start Family and Child Experiences Survey (Baby FACES) Design Options Report

OPRE Report 2015-99

September 2015

Submitted to:

Amy Madigan, *Project Officer*
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Submitted by:

Yange Xue
Kimberly Boller
Cheri A. Vogel
Jaime Thomas
Pia Caronongan
Nikki Aikens

Mathematica Policy Research

Project Director:

Cheri A. Vogel
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393

Contract Number: HHSP23320072914YC

Mathematica Reference Number: 06432.140

Suggested citation:

Xue, Y., K. Boller, C.A. Vogel, J. Thomas, P. Caronongan, and N. Aikens (2014). Early Head Start Family and Child Experiences Survey (Baby FACES) Design Options Report. OPRE Report 2015-99. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

This report is in the public domain. Permission to reproduce is not necessary. This report and other reports sponsored by the Office of Planning, Research and Evaluation are available at http://www.acf.hhs.gov/programs/opre/index.html.

Disclaimer:

The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services

TECHNICAL WORK GROUP MEMBERS

We would like to thank the following members of the Baby FACES Technical Work Group. The views expressed in this publication do not necessarily reflect the views of these members.

Catherine Ayoub
Harvard University

Lisa Berlin
University of Maryland

Jeanne Brooks-Gunn
Teachers College
Columbia University

Margaret Burchinal
Frank Porter Graham Institute
University of North Carolina, Chapel Hill

Judith Carta
University of Kansas

Michael Foster*
University of Alabama at Birmingham

Carol Hammer
Temple University

Brenda Jones Harden
University of Maryland

Judith Jerald
Save the Children

Ellen Eliason Kisker
Twin Peaks Partners, LLC

Tammy Mann
Campagna Center

Helen Raikes
University of Nebraska

Lori Roggman
Utah State University

Martha Zaslow
Society for Research in Child Development
Child Trends

*We are very sad to note that Michael Foster died before publication of this report.

CONTENTS

OVERVIEW vii

EXECUTIVE SUMMARY viii

I THE FUTURE OF EARLY HEAD START RESEARCH AND EVALUATION 1

A. What Are the Key Purposes of an Investment in a Descriptive Study of Early Head Start? 2

B. What May Stakeholders Need to Know about Early Head Start over the Next 10 Years? 3

C. How Often and in What Formats Do Stakeholders Need Early Head Start Information? 5

D. Roadmap of the Report 5

II LESSONS LEARNED FROM BABY FACES 2009 AND IMPLICATIONS FOR FUTURE STUDY DESIGNS 7

A. Overview of Baby FACES 2009 Design 7

B. Gaps in Information Available from the Baby FACES 2009 Longitudinal Cohort Design 10

C. Baby FACES 2009 Measurement Challenges 11

D. Informing Future Research Questions and Designs 13

III CONCORDANCE OF MEASUREMENT STRATEGIES WITH RESEARCH QUESTIONS AND STUDY DESIGNS 18

A. Program Features, Service Inputs, and Service Quality 18

B. Child Outcomes 20

C. Family and Home Characteristics, Parent–Child Relationships, and Family Outcomes 23

D. New and Emerging Measurement Areas 24

IV POTENTIAL OPTIONS FOR FUTURE STUDY DESIGNS 26

A. Cross-Sectional Design 26

B. Longitudinal Design 29

C. Combination Designs and Special Study Options..... 32

D. Considerations for Analyses, Reporting and Dissemination For All Design Options 38

REFERENCES 40

APPENDIX A 45

TABLES

IV.1. Summary of Example Data Collection Components and Schedule for Cross-Sectional Design 29

IV.2 Summary of Example Data Collection Components and Schedule for Short-Term Longitudinal Design.....32

IV.3 Summary of Example Data Collection Components and Schedule for Long-Term Longitudinal Design 32

IV.4 Summary of Advantages and Disadvantages of Different Design Options 38

FIGURES

II.1 Early Head Start Framework for Programs Serving Infants and Toddlers and their Families 15

OVERVIEW

Early Head Start has a long tradition of evaluation and descriptive research, beginning with the Early Head Start Research and Evaluation Project that was launched in 1995 when the program began. More recently, the Early Head Start Family and Child Experiences Survey (Baby FACES 2009), funded by the Office of Planning, Research and Evaluation (OPRE), followed a nationally representative sample of 89 programs and two cohorts of children enrolled in the programs. This report summarizes lessons learned from the design of Baby FACES 2009 and describes ways that future descriptive studies of Early Head Start may be designed to build on past work and address the questions and information needs for research, policy, and practice.

The Baby FACES 2009 study had several strengths, including the breadth of information collected over time, assessment across multiple domains of child development, and detailed information about overall classroom and home-visit quality as well as service receipt and service options. However, there were several limitations to the Baby FACES 2009 design. For example, Baby FACES 2009 had no true baseline assessment of child or family well-being. Additionally, the sampling approach limited generalizability (e.g., sample included only newborns and 1 year olds; centers, classrooms, and home visitors were not sampled) and the ability to conduct subgroup analyses (i.e., due to sample size). Baby FACES 2009 also had several challenges in the area of measurement design. For example measures of infant/toddler development with strong psychometric properties and that can be used longitudinally were difficult to identify, as were valid and reliable measures of classroom and home visit quality.

A critical consideration for future descriptive studies of Early Head Start is identifying the research questions of interest and how different designs may be better equipped to answer different questions. Baby FACES 2009 addressed a range of research questions aimed at describing the Early Head Start population and services, as well as relating program services to child and family outcomes. Future descriptive studies of Early Head Start might expand on these questions to address specific topics in a more complex way, such as a more in-depth examination of program implementation and innovation or how quality of services, relationships, and contexts are associated with children's well-being and competence. Furthermore, future descriptive studies of Early Head Start will need to consider a measurement strategy that is aligned with the research questions, while addressing measurement challenges and the need for continuity across data collection waves.

The report concludes with a discussion of options for future study designs and advantages and challenges for each type of design. Options include (1) a cross-sectional design, which would provide a comprehensive snapshot of Early Head Start at one point in time; (2) a longitudinal design, which would follow a particular cohort and could examine progress over time, and; (3) combination designs, such as a Basic Add-On Design. The Basic Add-On design has two components. The Basic component provides regular data on a key set of program, child, and family indicators—dashboard indicators—in a representative sample of programs and children. The Add-On elements of the design complement the Basic with information on important topics and may use a range of methods depending on the research questions of interest.

EXECUTIVE SUMMARY

The Future of Early Head Start Research and Evaluation

Early Head Start has a long tradition of evaluation and descriptive research, beginning with the Early Head Start Research and Evaluation Project (EHSREP) that was launched in 1995 when the program began. More recently, the Early Head Start Family and Child Experiences Survey (Baby FACES), funded by the Office of Planning, Research and Evaluation (OPRE), followed a nationally representative sample of 89 programs and two cohorts of children enrolled in the programs. The end of the Baby FACES 2009 study provides an opportunity to step back and consider what is known about Early Head Start, what can be known or understood better, and how to best go about answering current questions and anticipating future research needs.

To discuss the current needs of the Office of Head Start (OHS) and the Administration for Children and Families (ACF) more broadly and how future studies of Early Head Start might be shaped to flexibly address questions, OPRE convened a technical work group (TWG) meeting in February 2013. The meeting reviewed findings from Baby FACES 2009 and then solicited input from workgroup members about the research questions that are most important to consider going forward and the design options for answering those questions. Overall, TWG members confirmed that there is value in providing a periodic, nationally representative view of the program that also includes flexibility to address emerging issues and new research questions that inform decision-making at the national and local levels.

This report provides an overview of the purposes of continued investment in Early Head Start research, summarizes lessons learned from Baby FACES 2009, and describes ways that future descriptive studies of Early Head Start could be designed to address the questions and information needs highlighted by the TWG members.

A. What Are the Key Purposes of an Investment in Early Head Start Research and Evaluation?

Key purposes of future descriptive studies of Early Head Start will be to document program performance, inform policy, and inform training and technical assistance (T/TA) to support program improvement.

- Questions about performance measurement help understand what programs are doing and how they are doing it. These types of questions require data on program and staff characteristics, aspects of services such as frequency and quality, and child and family outcomes. These data can provide insight on the relationships among program and staff features, service quality, and child and family outcomes.
- A national study also can provide information that will help with policy development, for example, by describing the characteristics, strengths and needs of the population served by Early Head Start as well as the services they receive.
- A national study can also help identify the topic areas and issues where T/TA support is most needed from a national perspective.

B. What May Stakeholders Need to Know about Early Head Start over the Next 10 Years?

Four main areas that seem likely to be of ongoing interest to stakeholders about Early Head Start in the coming decade include (1) *Service quality*: Research has shown that quality matters for child outcomes (Burchinal et al. 2008, 2009) yet much work is still needed to identify tools that effectively capture quality of classrooms serving infants and toddlers, of home visits, and of the array of comprehensive services provided to Early Head Start families. (2) *Priorities for T/TA*: While understanding various aspects of program performance, such as service quality, is important for informing a national T/TA agenda, it will be important to understand how programs identify their own training needs and how training and professional development activities are provided to staff. (3) *Transitions out of Early Head Start*: An important function of Early Head Start is not only to intervene early in the lives of children and families who are at risk of adverse outcomes, but also to help them transition to other appropriate early childhood programs as their time in Early Head Start ends. (4) *How specific subgroups of children and families are faring*: As the demographic makeup of the U.S. changes, there will likely be increased interest in dual language learners (DLLs), children of recent immigrants, and children at high risk for suboptimal development by age 3.

Lessons Learned from Baby FACES 2009 and Implications for Future Study Designs

Baby FACES 2009 was the first national study of Early Head Start to collect information on program operations and management; characteristics and educational background of program staff (teachers and home visitors); quality of the classrooms and home visits that children participate in; services offered to and received by families; family characteristics, functioning, and well-being; and children's developmental progress. The wealth of information on many different aspects of the program allows for investigation of a multitude of questions about the experiences of the children and families served. Nevertheless, the study was not without its limitations which might be improved upon in future research.

A. Overview of Baby FACES 2009 Design

Baby FACES 2009 used a longitudinal cohort design to answer its research questions (Box E.1). Specifically, it took a census of children in two birthday windows from a nationally representative sample of 89 programs. The sample of programs was selected to ensure heterogeneity in terms of program size, percentage of DLLs served, service approach, urbanicity, and ACF region. Children and their families were followed longitudinally until they left or aged out of the Early Head Start program (age 3).

The study collected data through annual interviews with parents, teachers, home visitors, and program directors. Staff also provided reports on child outcomes and relationships with families. Direct child assessments and video-recorded parent-child and assessor-child interactions were also collected, along with weekly staff reports on services offered to and received by each child/family.

Baby FACES 2009 also included an implementation component built around the cohort design. Program directors were asked to complete a self-administered questionnaire that included the implementation rating scales adapted from the Survey of Early Head Start Programs (SEHSP; Vogel, et al 2006) in the first

round of data collection. In subsequent rounds of data collection, they were asked to provide similar program implementation information in a telephone interview.

Box E.1. Baby FACES 2009 Research Questions

Describing Early Head Start and Program Services

- What is Early Head Start? What are the program models employed, the qualifications of staff, and other important program features and characteristics?
- What is the overall status of program implementation and quality?
- What specific services are delivered to families and how are these services individualized to meet the needs of each child and family?

Describing the Population Served

- What are the characteristics of the families Early Head Start serves (includes demographic, household, and family characteristics; needs; and risk factors)?
- How are Early Head Start children and families faring over time?

Relating Program Services to Child and Family Outcomes

- How are child and family needs and outcomes associated with services received over time? Are there relationships between program features, quality, and outcomes?
- What are the characteristics of and services for special populations and subgroups? Examples of subgroups include children with identified special needs, highest-risk families, mothers with depression, DLLs, and mothers pregnant at program enrollment.
- What family and child characteristics are linked to services received? What characteristics are linked to outcomes?

Assessing Measures Used in Baby FACES

- Compared with the measures used in research projects, what are the psychometric properties (including reliability and validity) of measures routinely used by Early Head Start programs?
- What can researchers learn from fielding these instruments that can help inform their use at a local program level?

The major strengths in Baby FACES 2009 study include:

- Information on children and families across multiple domains, and longitudinally over time.
- In-depth information about staff and their experience, education, and professional development activities.
- Overall quality of classrooms and home-based services for families and children and attributes of classrooms and home visits that relate to quality.
- Details on service receipt and service options, and patterns of entry and exit from the program.

B. Gaps in Information Available from the Baby FACES 2009 Longitudinal Cohort Design

Despite its strengths, some aspects of the Baby FACES 2009 study design make it difficult to answer certain types of questions:

- There is no true baseline assessment, which limits analysis of changes between program enrollment and program exit
- The sample included only newborns and 1-year-olds, which limits generalizability of findings to the whole Early Head Start population.
- Small sample sizes limit the extent to which subgroup analysis can be conducted.
- The study sampled neither centers, nor classrooms, nor teachers/home visitors and cannot describe the quality of services at the program or center level.
- Similarly, home visit quality cannot be generalized to all children and families receiving home-based services.
- The measure of implementation used limits what the study can say about quality of program implementation.

C. Baby FACES 2009 Measurement Challenges

Although measures for Baby FACES were chosen with care and with the input of many experts in the field, they were subject to limitations. These limitations are not unique to Baby FACES and most are common to any effort to measure infant/toddler development and services geared toward them.

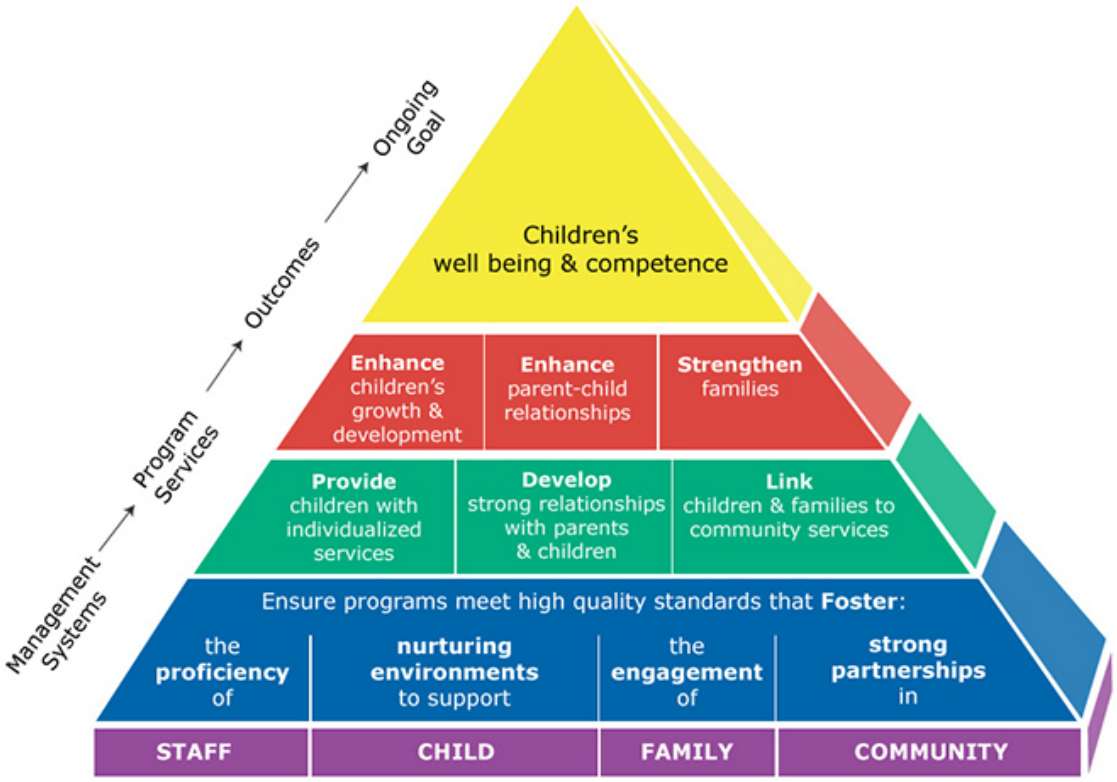
- It is challenging to find measures of infant/toddler development with strong reliability and predictive validity because developmental changes occur quickly and are not strongly predictive of later functioning.
- It is difficult to measure children's development longitudinally when the same instrument cannot be used across developmental periods.
- Assessment of DLLs is difficult and requires thoughtful decision rules about which language to assess children's language and communication skills (in addition to needing measures in languages other than English).
- There are few classroom quality observation tools for settings serving infants and toddlers and few of those have well established psychometric properties.
- Home visit process quality needs to be better understood, with measures that capture the most important components of visits, and procedures that gather an adequate sample of visits.

D. Informing Future Research Questions and Designs

Informed by lessons learned in Baby FACES 2009 and guided by the conceptual framework (Figure E.1), future descriptive studies of Early Head Start could assess program implementation, the quality of services, relationships, and contexts that are associated with children's well-being and competence, and

the relationships among them in a more complex way. This includes continuation and expansion of some research questions from Baby FACES 2009 as well as addressing some new research questions as suggested by TWG members.

Figure E.1. Early Head Start Framework for Programs Serving Infants and Toddlers and Their Families



Source: Framework for Programs Serving Infants and Toddlers and Their Families. Head Start Approach to School Readiness. HHS/ACF/OHS. 2012.

Describing Early Head Start Program Services

- Program implementation (Newly developed for future studies of Early Head Start)
  - What are the characteristics of program implementation?
  - Are Early Head Start program strategies implemented with fidelity? What factors (for example, external systems and implementation input) support implementation with high fidelity?

  - What innovations do programs make? What are the mechanisms for program improvement?

- Program quality (Expanded from Baby FACES 2009)
  - What does Early Head Start quality look like (in a representative sample)?
  - What are the factors that shape quality in Early Head Start programs?

- Professional development (Newly developed for future studies of Early Head Start)
  - How is staff professional development delivered in programs, and what are its influences on staff?
  - What are the indicators of effective professional development in Early Head Start?
  - How do both the quality of supervision and staff competencies change over time?

- Service delivery (Expanded from Baby FACES 2009)
  - How do Early Head Start programs deliver comprehensive services to all families? What is the nature of the partnerships they participate in to meet family and child needs across all of the outcome domains?
  - How are services individualized to meet the needs of each child and family?

- Program features (Addressed in Baby FACES 2009)
  - What are the qualifications of staff, and other important program features and characteristics?

Describing the Early Head Start Population

- Characteristics of children and families served (Expanded from Baby FACES 2009)
  - What are the characteristics of enrolled children and families (overall and by key subgroups)?
  - What are the needs of families? Are programs meeting those needs?

- Child and family functioning over time (Baby FACES addressed this question, but the sample of children and families were not representative of the Early Head Start population)
  - How do children and families fare over time during Early Head Start program enrollment?

Relating Program Services to Child and Family Outcomes

- Associations of services and outcomes (Expanded from Baby FACES 2009)
  - How are child and family characteristics and outcomes associated? How are they associated with services received over time?
  - Are there relationships between program features and outcomes? How do they vary by subgroups?

Concordance of Measurement Strategies with Research Questions and Study Designs

Future descriptive studies of Early Head Start will need to consider updates and improvements to existing measures while balancing needs for keeping constructs fairly constant over time. This will allow comparisons to prior waves while also considering new policy questions and exploring the utility of new measures. The research questions can guide the selection of data collection instruments and provide examples of key constructs to be assessed at each level of the Early Head Start performance framework.

- At the program level, key aspects could include: (1) program implementation; (2) the types of services that programs provide (directly or through referrals) and the frequency of service receipt for individual families, as well as the degree to which services are individualized to meet families' needs; (3) program service quality; and (4) staff characteristics, beliefs/attitudes, and well-being.
- At the child level, constructs to assess should be guided by the Head Start Approach to School Readiness that includes five essential domains for learning and development: (1) language and literacy development, (2) cognition and general learning, (3) approaches to learning, (4) physical development and health, and (5) social and emotional development. These are the areas of child development that Early Head Start is working to support.[1]
- At the family level, the usual measures of family background and characteristics, such as family income, parental education and employment, DLLs, and immigrant status are important. Additionally, measures of families could cover the areas of parenting, parent well-being (including psychological and physical health and risky behaviors), and the home environment and routines.
- New and emerging areas that might be promising to explore in future descriptive studies of Early Head Start include executive functioning, toxic stress, and biological and neuro-psychological measurement in infants/toddlers.

Potential Options for Future Study Designs

Future descriptive studies of Early Head Start will require a design that is guided by the Early Head Start conceptual framework. The design will also address gaps in the Baby FACES 2009 design and align with the study's primary research questions. The particular design selected will depend on the research questions prioritized for the study and the levels of data and frequency of reporting required to answer those questions.

Questions to address before selecting a design include:

---

[1] See http://eclkc.ohs.acf.hhs.gov/hslc/hs/sr/approach. Note: In 2015, the Office of Head Start released a newly revised *Head Start Early Learning Outcomes Framework: Ages Birth to Five* designed to represent the continuum of learning for infants, toddlers, and preschoolers and to replace the earlier frameworks. This new framework includes five central domains which align with the five domains discussed here. See http://eclkc.ohs.acf.hhs.gov/hslc/hs/sr/approach/pdf/ohs-framework.pdf.

- What has greater priority—tracking characteristics over time or providing a comprehensive snapshot at a single point in time?
- Would ACF and its stakeholders want to provide information across all families enrolled in centers/home visits at a given point in time or to be able to represent children and families at key developmental ages?
- Should the study measure children's growth across their time in the program? As children transition to new settings and enter kindergarten?
- What levels of data are important to stakeholders—for example, do stakeholders think having a representative sample at the staff or classroom level is necessary?

The answers to these questions will determine whether a cross-sectional design, a longitudinal design, or some combination of the two is most appropriate.

A. Cross-Sectional Design

A cross-sectional design for future descriptive studies of Early Head Start could provide a comprehensive snapshot at one point in time and provide information that would address a wide range of questions. A cross-sectional design could be used to describe the characteristics of Early Head Start programs, teachers and home visitors, and classrooms and home visits, and the population of Early Head Start children and families (including demographic, household, and family characteristics, family needs, and risk factors). This design could be used to assess the functioning of children and families at a given point in time.

Sample Design. In a cross-sectional design, individuals can be selected to be representative of a population. The TWG members highlighted the importance of a representative sample at the classroom/home visitor and program level and broader coverage of the entire Early Head Start population. Baby FACES 2009 selected a nationally representative sample of Early Head Start programs, and surveyed all children in each of two age ranges within sampled programs. Therefore, it cannot offer descriptions of children and families served by Early Head Start who were outside of those age ranges. Teachers/home visitors and classrooms were studied only when they were linked to the study children and thus are not representative of all Early Head Start staff. This sampling strategy limits the ability to aggregate the data to generate program- or center-level indicators of quality. In addition, small sample sizes limit the child-level subgroups that can be analyzed.

The key element of a cross-sectional design for future descriptive studies of Early Head Start is nationally representative samples of Early Head Start programs, teachers/home visitors, and children. In a cross-sectional study of Early Head Start, researchers could select a nationally representative sample of children that is large enough to enable subgroup analysis, as well as a representative sample of teachers/home visitors in Early Head Start that permits aggregating teacher/home visitor level data to the program level. One way to achieve a sample that addresses the limitations in Baby FACES 2009 is through multi-stage sampling at the program, center and classroom (or home visitor), and child levels.

Data Collection Approach. The cross-sectional study could be launched as frequently as every two years, assuming data collection in the first year and analyses and reporting in the second year. The data collection could occur at one point in time in a program year (e.g., in the fall or spring). In other words,

children and families would only participate in the study at one point in time, although the same programs could be sampled over multiple data collection waves if desired.

Advantages and Challenges of the Cross-Sectional Design. The key advantage of a cross-sectional design is that—for a relatively low cost and with relatively low burden on programs and families—it can provide comprehensive snapshots of Early Head Start programs, centers, teachers/home visitors, classrooms, and/or children, depending on the target populations chosen. A cross-sectional study could also provide nationally representative estimates of outcomes for children of different ages, as well as for key child and family subgroups. It is important to note, however, that ensuring sufficient sample size to support studying such subgroups would add to the costs of a cross-sectional survey. Data from a repeated cross-sectional survey, in which the same population is sampled at different points in time, can be used to measure changes over time in the aggregate. For example, researchers could use data from a repeated cross section to measure the change over time in the proportion of 2-year-olds in Early Head Start who are at or above national norms on an assessment of interest. If the priority of the study is to provide nationally representative snapshots of Early Head Start, then a cross-sectional design would be sufficient.

Though a cross-sectional design can capture changes over time in aggregate, a limitation is that the study would not follow the *same* children over time. Thus, a cross-sectional design would not be the optimal choice if examining changes in individual children's and families' outcomes over time is a top priority. This design also could not address questions about how staff or program characteristics, program implementation, service quality, or other factors might predict future child and family outcomes because there are no baseline outcomes to be used as control variables in the analyses.

B. Longitudinal Design

The defining feature of a longitudinal study is that it follows the same respondents over time. Like the cross-sectional design discussed above, longitudinal studies can survey representative samples of individuals, though maintaining the representativeness of the sample over time can be a challenge due to loss of participants via attrition. The longitudinal design for future descriptive studies of Early Head Start could track children and families over time and answer questions about child and family functioning and progress over the years in the program, in addition to the questions that could be answered in the cross-sectional design, described above.

Sample Design. As with the sampling strategies in a cross-sectional design, it is important for a longitudinal design to have nationally representative samples of Early Head Start programs, teachers/home visitors, and families/children. The sampling strategies for the cross-sectional design are applicable for the longitudinal design as well. Multi-stage sampling at the program, center and classroom (or home visitor), and child levels could achieve representative samples at different levels. Sample sizes need to take into account attrition rates over time.

In Baby FACES 2009, the age cohort design limited the representativeness of the sample for all children and families and teachers/home visitors and classrooms. Moreover, Baby FACES 2009 provides no true baseline data at the start of Early Head Start participation. The new longitudinal design could address these issues by having a nationally representative sample of children and families and/or sampling children and families at program entry (i.e., newly entering children/families at the start of the program

year). Decisions on which approach to use will depend on the questions of interest and focus of the study.

Data Collection Approach. We provide two options for the data collection schedule for a longitudinal design: (1) short-term longitudinal data collection, and (2) long-term longitudinal data collection. A short-term longitudinal study could be conducted every two years, with the first year for data collection, and the second year for analysis and reporting. The data collection could occur over a single program year, for example, with data collected in the fall and spring[2]. The long-term longitudinal study could be conducted every five years, with data collection occurring over three years, and final reporting in the last two years. The data on child/family outcomes could be collected at baseline (or program entry) in the fall and then in the spring of each program year (for a maximum of three years). Children would be followed until they are within a pre-defined window around their third birthday. Under either the short- or long-term option, before the data collection for the first cycle of the longitudinal study, there could also be an additional year for planning or piloting work on the measures.

Advantages and Challenges of the Longitudinal Design. The key advantage of a short-term longitudinal design is that—for a relatively low cost and with relatively low burden on programs and families[3]—it collects "baseline" data in the fall, enabling researchers to examine how program features such as program quality are associated with changes in outcomes by controlling for baseline scores. In addition, attrition problems in a short-term longitudinal design will not be as severe as in a long-term longitudinal design. However, the short-term follow-up would not track child progress and family functioning over their years in the program.

By obtaining data on the same children and families throughout their program experiences and transitions out of the program, a long-term longitudinal design confers the ability to track child and family outcomes over time, as well as the program experiences that support children's development and transition to preschool. However, because of program attrition, unlike a repeated cross-sectional design, longitudinal designs do not describe how child and family outcomes in aggregate are changing in Early Head Start nationally over time.

Following individual children over time involves costs to locate and track respondents and burdens families and program staff who respond to repeated survey waves. In addition, if survey items or assessments differ across survey waves, the ability to assess change over time at the level of the individual is lessened. This is a particular concern with the assessment of infants and toddlers. Finally, attrition from the study sample complicates analyses of longitudinal data, making the assessment of the experiences of a representative sample of children more difficult. Following Early Head Start children who leave their programs could mitigate issues around attrition and provide useful information but would be costly. Replacing sample members who leave their programs is also possible but threatens comparability across survey waves.

---

[2] Although Early Head Start is a year-round program, most grantees set a program year start/end date in the summer.

[3] The burden on families is still considerably higher than in a cross-sectional design if we are conducting a full battery twice within one year.

C. Combination Designs and Special Study Options

To provide flexibility and timely findings for program improvement and policy, ACF may be interested in considering learning from an approach selected for the FACES redesign project—the Core Plus Design (West et al. 2012). This approach combines ongoing data collection on programs and children to ensure comparability across waves coupled with additional studies that meet potentially changing information needs during the life of the study. Future descriptive studies of Early Head Start could employ a similar design, which we call the "Basic Add-On" design. The Basic Add-On design has two components. The Basic component provides regular data on a key set of program, child, and family indicators—dashboard indicators—in a representative sample of programs and children. The Add-On elements of the design complement the Basic with information on important topics and may use a range of methods depending on the research questions ACF wants answered.

The elements of the dashboard that are collected and reported could be re-evaluated periodically to determine if or when new measures should be added. There is flexibility so the design can address a range of questions about children, families and programs, and potentially, a need for quick turnaround of data. Future descriptive studies of Early Head Start should be designed to easily respond to rapidly evolving research and policy questions.

There are limitations to relying on simple indicators of performance measurement. Often there is a tension between collecting indicators for the dashboard and the need to contextualize the data. The more context is included, the more the effort becomes like a full study rather than a dashboard. The design should also be able to provide in-depth information on topics of particular interest, for example, program implementation, growth/change in child and family outcomes, and associations among classroom/home visit quality and family and child outcomes. With the Basic Add-On design, the Basic Option can be used to provide data on dashboard indicators, and the Add-On Option can be used to address in-depth special topics.

The Basic Add-On design lends itself to turning the data around more quickly and informing program management and policy decisions. While some questions may take longer to answer (for example, questions that involve collecting data over time), it is important to share and disseminate findings as quickly as possible and to respond to emerging issues in a timely fashion.

The Basic Options: Repeated Nationally Representative Short-term Longitudinal Implementation and Child/Family Outcome Study. This approach supports reporting on dashboard indicators using a repeated, nationally representative sample of programs, classrooms/home visitors, and children/families that provides sufficient sample sizes to study key child and family subgroups (defined by, for example, family risk, immigrant status). Either the short-term longitudinal or cross-sectional sampling approaches described earlier could be utilized; here we suggest using the same sampling approaches as the short-term longitudinal design. This design would allow for longitudinal analysis of program-level data as well as analysis of short-term (fall to spring) changes over time at the child level. The analysis could identify trends in the population served, children's progress from fall to spring of the program year, and program services and quality. The Basic study could be conducted regularly, for example, every two years, and could focus on any or all of the different levels of data: programs and centers, staff/classrooms/home visits, and children and families.

The Add-On Options: Longitudinal Studies of Growth/Change and Rapid Cycle Studies of Specific Program Features and Innovations. The Add-On Options, which could supplement the Basic study, offer opportunities to collect information on a wider range of topics and increase the flexibility of studies of Early Head Start. They include (a) more in-depth cross-sectional descriptive studies at less frequent periodicities (rotating or one-time studies, supplements, or topical modules); (b) longitudinal studies that follow children across their time in their Early Head Start programs to investigate growth/change in child/family outcomes and the associations of program, classroom, family, and child characteristics and child outcomes; (c) as-needed rapid cycle special studies to explore the relationships between program initiatives, practices, and other topics and child and/or family outcomes.

Here we discuss several topics for Add-On studies that may be of interest to ACF and other stakeholders, including topics raised by Baby FACES TWG members.

- **Studying leavers and understanding program attrition.** Understanding reasons for leaving Early Head Start before eligibility ends is important, especially given the high and not well understood rate of program attrition identified in Baby FACES 2009.
- **Curriculum, assessment, and professional development systems.** This special study could assess the curriculum, assessment, and professional development systems that are currently in place, how they align with staff professional development needs (or how these needs are identified), and investigate how programs are measuring change in connection with these systems.
- **Home visit quality.** A special study on this topic could involve the use of video-recording technology and other methods to better understand home visit quality. As part of this special study, it would be important to learn more about tools and resources used by home visitors, supervision provided to them, how home visitors individualize the services they provide based on children's developmental stages, and how to help parents understand and support development. The special study could follow family-home visitor pairs over time.
- **Program implementation.** Lessons learned from Baby FACES 2009 and other research about Early Head Start implementation point to the need for a new approach to studying implementation in Early Head Start, which could be a special study topic for future studies of Early Head Start. As a mature program, Early Head Start implementation research at this stage should focus on implementation strategies (structure and processes) at multiple levels to support and sustain high-fidelity implementation and improve child and family outcomes.

In summary, each of the design options described above could provide valuable information about Early Head Start and each has its advantages and challenges. The selection of a particular design would depend on the priority of ACF and its stakeholders. Table E.1 summarizes the key advantages and disadvantages of each option.

Table E.1. Summary of Advantages and Disadvantages of Different Design Options

Research Objectives	Cross-Sectional Design	Short-Term Longitudinal Design	Long-Term Longitudinal Design	Basic Add-On Design
Describe Early Head Start Services				
Programs	All	All	All	All
Classrooms / Home Visitors	All	All	All	All
Describe Population Served				
Children and Families	All	All if sampled regardless of entering time; enrollment cohort only if sampled among newly entering children and families	All if sampled regardless of entering time; enrollment cohort only if sampled among newly entering children and families	All
How individual children and families change over time	No	Yes	Yes	Yes
How programs and the population of children and families change over time (in aggregate)	Yes, if using a repeated cross-sectional design	No	No	Yes
Relationships between services and outcomes for children and families	No	No	Yes	Yes
Other Considerations				
Flexibility of adding new research questions that would be of interest to ACF and stakeholders	Yes, new questions can be added with each new cohort	Yes, new questions can be added with each new cohort	No	Yes
Quick turn-around of data	Yes	Yes	No	Yes
Cost	Lowest	Middle of range	Highest	Highest

CHAPTER I: THE FUTURE OF EARLY HEAD START RESEARCH AND EVALUATION

Early Head Start has a long tradition of evaluation and descriptive research, beginning with the Early Head Start Research and Evaluation Project (EHSREP) that was launched in 1995 when the program began. The EHSREP provided evidence of the effectiveness of the initial Early Head Start programs, as well as insight into their implementation and quality (ACYF 2001; ACF 2002). More recently, the Early Head Start Family and Child Experiences Survey (Baby FACES), funded by the Office of Planning, Research and Evaluation (OPRE), followed a nationally representative sample of 89 programs. It enrolled two cohorts of children who were either prenatal/newborns or about 1 year old in spring 2009. The study followed children annually until age 3 or until they left the program, conducting interviews with parents, staff, and program directors; classroom and home visit observations; direct child assessments at ages 2 and 3; and staff ratings of children's development and weekly service receipt. The conclusion of the Baby FACES 2009 study provides an opportunity to step back and consider what is known about Early Head Start, what can be known or understood better, and how to best go about answering current questions and anticipating future research needs. In other words, it is a good time to consider the Early Head Start research and evaluation that will be needed in coming years.

To discuss the current needs of the Office of Head Start (OHS) and the Administration for Children and Families (ACF) more broadly and how future descriptive studies of Early Head Start might be shaped to flexibly address questions, OPRE convened a technical work group (TWG) meeting in February 2013. The meeting reviewed findings from Baby FACES 2009 and then solicited input from TWG members about the research questions that are most important to consider going forward and the design options for answering those questions. The Head Start Family and Child Experiences Survey (FACES) redesign effort (West et al. 2012) informed this process. Meeting attendees came from multiple backgrounds and perspectives and included researchers, practitioners, and policymakers. The TWG members noted that there is a need for data about program performance and practices as they are occurring so that stakeholders can better understand program implementation, assess what is happening with current initiatives, and support continuous program improvement. In addition, members highlighted several research priorities that future Early Head Start studies should consider, including: drawing a more complete picture of the Early Head Start population; describing key subgroups of children and families such as demographic groups defined by immigrant or high-risk status, for example; and understanding quality and quality improvement at the program level. Members also suggested the importance of having a flexible design that lends itself to answering questions that are important to a range of audiences and turning the data around more quickly to inform program improvement and policy decisions. While some questions may take a while to answer (for example, because of the need to collect data over time), there is a need to share and disseminate findings as quickly as possible and to respond to emerging issues in a timely fashion.

Overall, TWG members confirmed that there is value in providing a periodic, nationally representative view of the program that also includes flexibility to address emerging issues and new research questions that inform decision-making at the national and local levels. This chapter draws on those discussions to provide an overview of the purposes of continued investment in Early Head Start research. It highlights what stakeholders may want to know about the program in the next decade, and how that information would be best provided. It emphasizes that future efforts should be timely, topical, rigorous, and flexible.

A. What Are the Key Purposes of an Investment in a Descriptive Study of Early Head Start?

The design of future descriptive studies of Early Head Start should be driven by OHS and OPRE evaluation priorities, goals, and needs. This will ensure that the design is optimized to address the most important questions. This section describes three primary purposes of Baby FACES 2009 and future descriptive studies of Early Head Start—documenting program performance, informing policy, and informing training/technical assistance to support program improvement—and discusses the types of information required.

1. Performance Measurement

Since the program's inception, research has played a key role in understanding the performance of Early Head Start programs and in informing continuous quality improvement. In recent years, the focus of performance measurement has shifted toward more process measures. In addition to documenting the number of children and family served, Early Head Start program performance is examined in terms of what programs are doing and how they are doing it. These types of questions require data on program and staff characteristics, aspects of services such as frequency and quality, and child and family outcomes. Most programs already collect data in all of these areas, although not in a uniform way or using the same measures. Nor are they required to report service quality or child outcome data at the national level. A national study can complement data collected at the local and national levels by assessing relationships among program and staff features, service quality and child and family outcomes and answering critical questions about how programs are helping to meet the needs of the families they serve.

Understanding program implementation is also key to understanding the performance of Early Head Start programs and to informing continuous quality improvement. Traditionally, implementation in Early Head Start has been measured in terms of adherence to the Head Start Program Performance Standards (hereafter "the performance standards"). Measuring implementation using compliance with the performance standards was most useful when programs were young and levels of compliance with the standards in many areas varied widely among programs. Now that programs are more uniformly implementing the performance standards well, a systematic implementation framework is needed. It should be flexible enough to allow for individualization of services so that programs can best serve their families. Analyses of implementation ratings in Baby FACES 2009 suggest the need for more comprehensive measures of implementation. Chapters II and V describe performance and implementation in more detail.

2. Policy Development

A second purpose of investment in Early Head Start research and evaluation is to inform policy. For example, currently there is a uniform set of expectations for services that are made available to families that vary by service option (home-based, center-based, and combination). Policies related to those expectations could be informed by national data on service dosage and the service delivery options preferred by families, potentially tailoring services to reflect those preferences while remaining true to the core elements of the program model. Another possibility is that a national study could provide information that will help shape new or emerging policy initiatives, such as those related to the new Early Head Start-child care partnerships program (see http://www.acf.hhs.gov/programs/ecd/early-

learning/ehs-cc-partnerships). Future descriptive studies of Early Head Start could include an add-on study in this area, or the study could develop research partnerships with the grantees.

3. Training and Technical Assistance (T/TA) to Support Continuous Program Improvement

Another purpose for a national study—and a critical area for Early Head Start programs generally—is informing T/TA to support program improvement. Among the questions that could be addressed by a national study are (1) How is training delivered in programs, and what are its influences on staff? (2) What are program needs around choice of curricula and assessments to individualize services? (3) At what level of quality are curricula being delivered? (4) Ultimately, how are all of these factors associated with improvements in outcomes for families and children? Answers to these questions can help identify particular issues for the T/TA System to consider when developing new resources or support mechanisms for programs.

Another way that a national study can support continuous improvement is to use the in-depth information gathered to identify high performing grantees. What makes a program excellent? Are those characteristics reflected in observational measures (for example, the Infant Toddler Environment Rating Scales [ITERS]), or are other characteristics of classrooms, staff, or programs also critical? Can the attributes of excellence be broken down and understood so that a T/TA system can help all programs achieve high levels of performance?

B. What May Stakeholders Need to Know about Early Head Start over the Next 10 Years?

There are a host of different topics that could be examined to inform performance measurement, policy, and T/TA. Stakeholders in need of information include OHS, OPRE, the research community, and national T/TA centers as well as program directors and practitioners. Based on conversations with the TWG and other stakeholders, it is clear that different types of stakeholders need different kinds of information. Broadly, they want to be able to (1) Describe Early Head Start program services, (2) Describe the population served (3) Relate program services to child and family outcomes, and (4) Address cross-cutting issues, or issues that link the three aforementioned objectives. Within these broad objectives, four main areas seem likely to be of ongoing interest to stakeholders about Early Head Start in the coming decade: service quality, priorities for T/TA, transitions out of Early Head Start, and how specific subgroups of children and families are faring . This section briefly discusses each area.

1. Service Quality

Research has shown that service or process quality matters for child outcomes (Burchinal et al. 2008, 2009) and this is reflected in the Head Start Designation Renewal System (DRS). Specifically, the DRS uses an assessment of the quality of teacher-child interactions in Head Start preschool classrooms as one of seven indicators of Head Start program performance. While the DRS does not use a similar measure for programs serving infants and toddlers, there may be interest in doing so in the future. Quality measures for infant/toddler classroom settings show some promise. In addition to the toddler version of the Classroom Assessment Scoring System (CLASS-T) that was used in Baby FACES 2009, an infant version has been developed (LoCasale-Crouch 2014), as well as the Quality of Caregiver-Child Interaction for Infants and Toddlers (Q-CCIIT) measure (Atkins-Burnett et al. 2014; Roggman 2014). A national study provides an opportunity to explore the utility and functioning of these measures.

Additionally, providing a national picture of Early Head Start service quality can support programs in understanding their own data. In Baby FACES 2009, programs reported that although they collected a great deal of data on all manner of topics, few were able to use the data in a flexible way (Vogel et al. 2011; Vogel et al. forthcoming). By providing national estimates of quality on measures being used at the local level (e.g., ITERS, CLASS-T) programs can get a sense of how they are doing relative to other Early Head Start programs—where they seem to have their biggest strengths and areas in need of improvement. And a national study can answer questions about how the quality and intensity of services relate to staff and program characteristics and to outcomes for children and families (as was examined in Baby FACES 2009).

Early Head Start has the additional need for measuring quality of home visits—a field that is still in development. Measures of home visit quality in Baby FACES 2009 had limited relationships with staff characteristics and with child outcomes (Aikens et al. forthcoming). A national study focused on understanding the characteristics of visits and home visitors associated with enhanced outcomes could increase parity with what is known about factors that affect classroom quality. Furthermore, little is currently known about the quality and intensity of other types of services such as socializations, family support, parent education, mental health, physical health, and prenatal services. Another area for additional exploration is the referrals that programs provide and whether and how they follow up on referrals that are made. Finally, little is known about how the characteristics and practices of program leaders (supervisors and directors) relate to quality and support quality services. Future descriptive studies of Early Head Start can examine these issues and fill an important gap in knowledge.

2. Training and Technical Assistance Priorities

While understanding various aspects of program performance, such as service quality is important for informing a national T/TA agenda, it will be important to understand how programs identify their own training needs and how training and professional development activities are provided to staff. There is a great deal of evidence to show that 'one-shot' workshops have very little if any impact on learning or changes in practice (and this evidence spans the gamut of early childhood programs to continuing medical education) (Boller et al. 2014; Forsetlund et al. 2009; Wasik et al. 2013). A national study can provide information to help programs use the most reliable means of identifying training and professional development needs, and can develop training modalities on those topics that have been shown to affect outcomes.

3. Transitions and Continuity of Care

An important function of Early Head Start is not only to intervene early in the lives of children and families who are at risk of adverse outcomes, but also to help them transition to other appropriate early childhood programs as their time in Early Head Start ends. Baby FACES 2009 provided information about program practices around expected transitions (when children neared age 3). The study also provided insight about children and families who left Early Head Start before their eligibility ended (approximately 37 percent of the Baby FACES sample did so). Analyses highlighted the reasons for leaving and explored whether there were characteristics that were associated with higher rates of leaving, but because the study was not designed to follow leavers, this information was limited (Caronongan et al. 2014). These questions can be examined more thoroughly in future descriptive studies of Early Head Start.

Additionally, Baby FACES 2009 was able to provide the first in-depth information on the frequency and timing of transitions *within* the program (between service options and changes to different teachers and

home visitors). These types of data will continue to be valuable moving forward given the growing evidence of the importance of continuity of care.

4. Experiences of Key Subgroups of Children and Families

Another purpose of a national study can be to focus on populations of interest. As the demographic makeup of the U.S. changes, there will likely be increased interest in dual language learners (DLLs),[4] children of recent immigrants (two highly interrelated groups), and children at high risk for suboptimal development by age 3 (such as, children exposed to high levels of psychosocial and economic risk/adverse experiences/toxic stress). These early adverse experiences can lead to school failure as early as kindergarten if not ameliorated. As a two-generation program, Early Head Start also needs to be able to address the needs of parents who themselves may face adverse experiences and/or trauma, which can affect their ability to parent their children in a supportive way.

C. How Often and in What Formats Do Stakeholders Need Early Head Start Information?

Different audiences have different needs for the type and timing of information about Early Head Start programs, families, and children. Future descriptive studies of Early Head Start could address these different needs through careful planning. It will be important for future studies of Early Head Start to find ways to provide information rapidly and to disseminate that information in a way that is accessible to various stakeholders. It is likely that OHS, OPRE, the research community, and national T/TA centers will be primary audiences for the work, along with federal policymakers and the public. It will also be important to provide information that is accessible to program directors and practitioners, particularly if encouraging programs' own use of data is of interest.

Baby FACES 2009 has provided information in a variety of formats, and it is likely that these will continue to be useful in future work. Reports are generally longer, include detailed information, and are often more technical. Future studies could consider shorter turnaround tables and figures of key indicators to provide a quick snapshot of new data, as in FACES 2014-2018. Baby FACES 2009 produced a number of short reports focused on specific topics aimed at a research/policy audience with companion nontechnical briefs that are directed at a practitioner audience (Bandel et al. 2014). Presentations at conferences have also been useful for disseminating data. The study team presented at both research and practitioner conferences, including the Society for Research in Child Development (SRCD), the Association for Public Policy Analysis and Management (APPAM), the Head Start Leadership Institute (HSLI), the Head Start Research Conference (HSRC), and the Birth To Three Institute (BTT). One outlet that the study has not yet used is webinars; these might be a cost-effective alternative for future work and could be made accessible to multiple types of audiences. Infographics and focused brief documents (for example, Bandel et al. 2014) are another user-friendly alternative that could be useful for practitioners.

D. Roadmap of the Report

The chapters that follow elaborate on ways that future descriptive studies of Early Head Start could be designed to address the questions and information needs described here. Specifically, subsequent

[4] A new Center for Early Care and Education Research-Dual Language Learners has been funded by ACF (http://cecerdll.fpg.unc.edu/).

chapters describe lessons learned from Baby FACES 2009 (II); measurement issues and potential research questions for consideration (III); and potential options for study design in future work (IV). Chapter II proposes a conceptual framework for future descriptive studies of Early Head Start to support the recommendations outlined in the report. Although the report lays out design options for future studies, it does not discuss the sample sizes or costs associated with each proposed option.

CHAPTER II: LESSONS LEARNED FROM BABY FACES 2009 AND IMPLICATIONS FOR FUTURE STUDY DESIGNS

Baby FACES 2009 was the first national study of Early Head Start to collect information on program operations and management; characteristics and educational background of program staff (teachers and home visitors); quality of the classrooms and home visits that children participate in; services offered to and received by families; family characteristics, functioning, and well-being; and children's developmental progress. The wealth of information on many different aspects of the program allows for investigation of a multitude of questions about the experiences of the children and families served. Nevertheless, the study was not without its limitations. In this chapter, we recount key aspects of the Baby FACES 2009 study design and identify features that could potentially be enhanced in future Early Head Start research to permit examination of an even wider range of research questions. We also describe the conceptual framework that should be used to guide the design of future studies. Finally, we identify potential research questions for further study.

A. Overview of Baby FACES 2009 Design

In 2007, the Office of Planning, Research & Evaluation (OPRE) in the Administration for Children and Families (ACF), U.S. Department of Health and Human Services, contracted with Mathematica Policy Research and its partners to implement a six-year longitudinal study in 89 Early Head Start programs around the country. Baby FACES 2009 used a longitudinal cohort design to answer its research questions (Box II.1)(Vogel et al. 2011). Specifically, it took a census of children in two birthday windows from a nationally representative sample of 89 programs. The sample of programs was selected to ensure heterogeneity in terms of program size, percentage of dual language learners served, service approach, urbanicity, and ACF region. Baby FACES 2009 followed two cohorts of children through their time in Early Head Start. Annual data collection began in spring 2009 and ended in spring 2012 (when the Newborn Cohort turned 3). The Newborn Cohort included 194 pregnant mothers and newborn children. The 1-year-old Cohort included 782 infants who were approximately 1 year old (age 10 to 15 months at the start of the study).

Baby FACES used a multi-method measurement approach to collecting data about Early Head Start programs, staff, services offered and received, service quality, and the children and families served (Vogel et al. 2015) (Table A.1 in Appendix A provides details on the key measures used). Program characteristics and implementation measures included program approach at the program and family levels, overall program implementation, and characteristics of the population served. Staff characteristics and program quality measures included staff depression symptoms, quality of the staff-parent relationship, home visit quality and content, classroom quality, and classroom child-adult ratio. Measures of services offered and received included family demographic, economic, and psychological risk, parenting quality, home environment quality, parent support for learning, neighborhood characteristics, parent depression symptoms, parenting stress, and family conflict. Child measures assessed communication, motor, personal-social/social-emotional, behavior problems, and problem solving skills. We used multiple measures of children's communication and language skills, including a video-coded child-assessor play interaction, direct child assessment, and parent and staff ratings. Assessors also rated children's behavior during the child assessment for their engagement and ability to regulate their emotions. Child measures were collected by phone and survey from parents and Early Head Start Staff when children were 1. At ages 2 and 3, we also conducted direct assessments during data collection visits to the family home.

Box II.1. Baby FACES 2009 Research Questions

Describing Early Head Start and Program Services

- What is Early Head Start? What are the program models employed, the qualifications of staff, and other important program features and characteristics?
- What is the overall status of program implementation and quality?
- What specific services are delivered to families and how are these services individualized to meet the needs of each child and family?

Describing the Population Served

- What are the characteristics of the families Early Head Start serves (includes demographic, household, and family characteristics; needs; and risk factors)?
- How are Early Head Start children and families faring over time?

Relating Program Services to Child and Family Outcomes

- How are child and family needs and outcomes associated with services received over time? Are there relationships between program features, quality, and outcomes?
- What are the characteristics of and services for special populations and subgroups? Examples of subgroups include children with identified special needs, highest-risk families, mothers with depression, DLLs, and mothers pregnant at program enrollment.
- What family and child characteristics are linked to services received? What characteristics are linked to outcomes?

Assessing Measures Used in Baby FACES

- Compared with the measures used in research projects, what are the psychometric properties (including reliability and validity) of measures routinely used by Early Head Start programs?
- What can researchers learn from fielding these instruments that can help inform their use at a local program level?

The study collected data through annual interviews with teachers (267 in 2010, the first year staff from both cohorts were included in the teacher interviews), home visitors (323 in 2009), and program directors (89 each year of the study). Staff also provided reports on child outcomes and relationships with families. Direct child assessments and video-recorded parent-child and assessor-child interactions were also collected along with weekly staff reports on services offered to and received by each child/family (for 793 total children from July 2009 through June 2011 for the 1-year-old Cohort and from July 2010 through June 2012 for the Newborn Cohort) (Vogel et al. 2015).

Baby FACES 2009 also included an implementation component built around the cohort design. Program directors were asked to complete a self-administered questionnaire that included the implementation rating scales adapted from the Survey of Early Head Start Programs (SEHSP; Vogel, et al 2006) in the first round of data collection. In subsequent rounds of data collection, they were asked to provide similar program implementation information in a telephone interview.

Baby FACES 2009 had many strengths in term of its design and, more specifically, the information collected. The major strengths are highlighted below.

- **The design allowed for the study of children and families across multiple domains.** For example, Baby FACES 2009 data indicate that, at age 3, children in Early Head Start fared well in terms of physical health and general development. They were approaching norms on some measures of development, although not all. The measure of auditory comprehension

(Preschool Language Scale, Fourth Edition, PLS-4) indicated children were performing similar to norms, but lagging on receptive vocabulary (measured by the Peabody Picture Vocabulary Test, Fourth Edition, PPVT-4). In terms of family and home environments, Baby FACES 2009 data indicated that Early Head Start children lived in emotionally supportive and cognitively stimulating homes, but also in neighborhoods in poor condition.

- **Baby FACES 2009 data provide a snapshot at a particular age and also allow for examination of change over time within children and families**. For example, over the course of the study, children's receptive and expressive vocabulary skills as reported by staff increased. On average, DLL children scored lower on English vocabulary than children from English-speaking homes, but they grew at a rate similar to children from English-speaking homes. Parents reported reductions in parenting stress and improved mental health.
- **Baby FACES 2009 provides in-depth information about staff and their experience, education, and professional development activities**. The study documented the characteristics of staff overall (such as turnover rates reported by program directors) with more in-depth staff-reported experience for the teachers and home visitors serving children in the study sample. Staff overall were well-educated and experienced with 37 percent of teachers and 58 percent of home visitors with a bachelor's degree or higher (Vogel et al. 2015). Staff reported on a range of professional experiences and supports, from the training they received to receipt of coaching or mentoring. They also reported on their health and depressive symptoms.
- **Baby FACES 2009 provides important information about the overall quality of classrooms and home-based services for families and children and about attributes of classrooms and home visits that relate to quality.** Early Head Start home visit and classroom quality was in the moderate or mid-range, with similar patterns as in the broader literature using similar measures. Quality scores were relatively stable over time, with one exception – the quality of instructional practices decreased over time. Measures of classroom quality were associated with many staff characteristics and child outcomes, while measures of home visit quality were found to have limited relationships with staff characteristics and with child or parenting outcomes.
- **Baby FACES 2009 also provides details on service receipt and service options, and patterns of entry and exit from the program.** Rich data from the family services tracking (FST) system provided important information not previously available from a nationally representative sample. For example, we used FST data to determine whether and how common it was for families to change service options (6 percent of children from the 1-year-old cohort in multiple approach programs changed service option between age 1 and 2), and how common it was for children and families to experience a change in teacher or home visitor (27 percent of children in the 1-year-old cohort changed teacher or home visitor between ages 1 and 2). Most importantly, it allowed a more accurate estimate of how often and when families left the program, and the characteristics of these families. Thirty-seven percent of families who were enrolled in spring 2009 left the program before their period of eligibility ended. The data also lends itself to examination of seasonal patterns in program offerings and family service take-up.

B. Gaps in Information Available from the Baby FACES 2009 Longitudinal Cohort Design

Despite its strengths, some aspects of the Baby FACES 2009 study design make it difficult to answer certain types of questions:

- **There is no true baseline assessment, which prevents drawing conclusions about the effect of program enrollment on outcomes.** Because children and families were already enrolled in the program (with varying lengths of enrollment) when the study began, it is not possible to fully account for differences in families' experiences prior to the study. Of course, because Baby FACES 2009 was a descriptive study, there is no comparison group that would tell us what would have happened to children in the absence of children enrolling in Early Head Start.
- **Sample design approach limits generalizability of findings to the whole Early Head Start population.** Baby FACES 2009 followed two birth cohorts of children (newborn and 1-year-old) longitudinally throughout their experience in Early Head Start. However, the cohort design might not be the best approach for Early Head Start because children were not sampled from the whole Early Head Start population. A cohort design may not provide the best nationally representative data at the child/family level, because (1) programs may offer both home visiting and center-based services, and families may change service options during their enrollment, and (2) families may enroll throughout the entire year.
- **Small sample sizes limit the extent to which subgroup analysis can be conducted.** Although there is great interest in comparing the experiences and outcomes of children and families with different demographic characteristics or risk factors, for most subgroups there is not sufficient sample size and therefore not enough statistical power to conduct such analyses using Baby FACES 2009 data.
- **The study cannot be used to describe the quality of services at the program or center level.** Although Baby FACES 2009 was designed to describe a nationally-representative sample of Early Head Start programs, the study children from each program are not necessarily representative of all children enrolled at that program or center. Similarly, the program staff members included in the study are staff members who were working with study children and are not necessarily representative of all program staff.
- **Data on home visit quality cannot be generalized to all children and families receiving home-based services.** For logistical reasons, home visit observations were conducted only once per year for each home visitor in the study. It was not always feasible to observe a home visitor with the same family on multiple occasions. This presents challenges in studying home visit quality over time, particularly if there is reason to believe that the quality of visits is driven to some extent by the family with whom the home visitor is working at a particular time.
- **There are limitations to what the study can say about the association between length of program enrollment and child outcomes.** There are two primary reasons for this limitation. First, the design does not allow for descriptions of children who enrolled after age 1. Second, children who left the program early were no longer eligible to be part of the study. As such, it was not possible to compare the age 3 outcomes of children who received 3 years

of program services versus children who left early, or children whose families enrolled at pregnancy versus children who enrolled after age 1.

- **The measure of implementation used limits what the study can say about quality of program implementation.** Analyses of implementation ratings in Baby FACES 2009 indicated less variation in implementation ratings than those in the EHSREP and weak or nonexistent associations with other program characteristics and data on program quality or child outcomes. It is likely that two factors contributed to the lack of associations: (1) less comprehensive and in-depth data were collected about implementation than in the EHSREP, and (2) Early Head Start programs were more mature and most programs were now meeting the minimum thresholds defined in the performance standards and reflected in the implementation ratings that were developed to focus mainly on compliance to the standards.

C. Baby FACES 2009 Measurement Challenges

In addition to not being able to answer particular questions, Baby FACES also struggled with some of the same measurement issues that affect many studies of early childhood programs and populations. Baby FACES 2009 used a variety of measures from multiple sources. These measures helped answer the research questions about program features and services; staff characteristics; quality of services; family characteristics, functioning, and well-being; and children's developmental progress. The results based on these measures and how they performed in Baby FACES also reinforced broader infant/toddler measurement concerns and highlighted the need to continue to engage measurement experts in future Early Head Start research.

1. Measurement Concerns

Psychometric properties of the measures. Baby FACES 2009 provided rich information about measures for infants and toddlers, and raised some concerns about measurement issues. It is very challenging to identify measures of children's development with strong reliability and predictive validity. Measuring change for young children during the infant/toddler period is difficult – developmental changes occur quickly and are not strongly predictive of later functioning. During infancy, skills are not sufficiently manifested to be measured with precision and development is not a smooth linear process. The psychometric properties of existing measures for infants and toddlers are poor (Xue et al. submitted; Xue et al. forthcoming). Deciding how to deal with these measurement issues is something future studies will have to confront. Ideally, future studies would focus as much as possible on constructs that predict school readiness. Is collecting data using measures that do not have strong reliability and predictive validity worth the cost? If measuring growth for subgroups of children is an area of interest, future studies will need to make sure the measures have similar psychometric properties across different subgroups

Another key issue in measurement is identifying the sources of variability in the measure used to determine whether one respondent or measurement mode is better than another. In addition, Baby FACES 2009 found that within-program variation is often greater than across-program variation, especially with regard to implementation quality. This is a challenge because of the measurement error introduced.

Child progress over time. Measuring child development constructs longitudinally when the same instrument cannot be used across developmental periods is also an important issue to consider. Given

the rapid changes in infants and toddlers during their first three years of life, it is challenging to identify assessments that span that entire period. To address this, Baby FACES 2009 tried bridging the instruments at different ages by adding overlap items in different forms and used Item Response Theory (IRT) methods to generate an equal-interval developmental scale for analysis looking at growth. OPRE may want future studies of Early Head Start to continue to use this promising approach in selecting and designing child outcome measures or explore new, promising measure that can better capture these constructs longitudinally.

Assessment of children who are dual language learners (DLLs). Another measurement issue is related to the assessment of DLLs. Researchers need ways to decide which language to use to assess children's language and communication skills (in addition to needing measures available in languages other than English). Considerations include how to set up decision rules for which language to use. These rules could include selecting the language of assessment based on the language spoken in the home and encompass whether children who are DLLs should be assessed in both languages; and whether language assessments should be conducted if children are mainly exposed to a language other than English or Spanish in the home. In Baby FACES 2009, the PLS-4 was used to assess children's language skills, and the assessment was conducted in Spanish and English if a parent reported that the child was exposed to Spanish in the home. Future studies of Early Head Start could consider the possibility of using a language screener in combination with parent reports to decide whether children should be assessed in both languages.

Classroom observation. It is also challenging to select what classroom observation tools to use in Early Head Start. First, there are few measures of classroom quality appropriate for use in settings serving infants and toddlers, and still fewer with well-established psychometric properties. There is little consensus about what quality should look like for infants/toddlers, and what tool(s) best measure quality. Second, given the wide age range of children in Early Head Start, researchers likely would have to use different measures over time. Having two different measures limits the study's ability to compare quality over time. In general it is desirable to use the same measures over time but again, children's environmental needs change rapidly in the first three years of life. Some consideration should be given to how practices and measurement of quality may differ by age. The ITERS-R and the CLASS-T were used in Baby FACES 2009. The Q-CCIIT measure of caregiver-child interaction quality in infant and toddler settings shows adequate psychometric properties and might be a promising candidate for classroom observation (Atkins-Burnett et al. 2014). It is appropriate for use in both center-based programs and family child care homes (FCCs), as well as single- and mixed-age infant and toddler classrooms. For future studies of Early Head Start, the new CLASS-Infant may be available for use, which if selected could be used in conjunction with the CLASS-T to assess classrooms serving infant through age 3 service period (LoCasale-Crouch 2014).

Home visit observation. Home visit quality measurement is a field still in development (Roggman et al. 2014). Baby FACES 2009 was the first large scale study to date in Early Head Start that has used home visit process quality measures, the Home Visit Rating Scale-Adapted (HOVRS-A; Roggman et al. 2009). However, it found limited associations between home visit quality and staff characteristics and with child outcomes. It will be a challenge for future studies to identify, modify, or develop an innovative measure that captures associations with enhanced child and family outcomes.

2. Engaging Measurement Experts

Measure selection for studies of infant and toddler development requires great care and careful review of psychometric properties and understanding of the tradeoffs inherent in measuring development at young ages. Advisory groups can provide consultation to ACF and study teams on measures in future descriptive studies of Early Head Start and help to reach consensus on measurement strategies. The Baby FACES 2009 TWG members included measurement experts, but in addition to those experts we also engaged a number of other experts in small group or in one-on-one conversations. By building in additional resources to engage more measurement experts, future studies would have broader access and engagement of the leading measurement experts in each domain of inquiry. The types of measurement issues the Baby FACES 2009 team grappled with in selecting measures and weighing their pros and cons included whether they:

1. Have demonstrated predictive validity and internal consistency reliability;
2. Are appropriate for use with low-income families and their children, who are racially, ethnically, and linguistically diverse and who might include children with disabilities;
3. Allow comparisons to Baby FACES 2009 and other studies (for example, the EHSREP and the Early Childhood Longitudinal Study-Birth Cohort [ECLS-B]);
4. Have the least burden to programs and families; and
5. Bring the best approaches from smaller scale research to assessing challenging constructs on a large scale, such as home visit quality.

To address these and the other domain-specific issues that arise during measures selection, it will be helpful for future studies of Early Head Start to consult with an advisory group at the beginning of the study and have resources available to engage additional experts throughout the life of the study. In fact, it was useful to reengage some of the measurement experts during the analysis phase of Baby FACES as questions arose.

D. Informing Future Research Questions and Designs

Baby FACES 2009 provided descriptive evidence that can be used to inform the research questions and design of future studies. In particular, the findings of Baby FACES 2009 may help identify priorities for future research and weigh the tradeoffs between various design options, such as following children and families over time to examine multiple development and family outcomes vs. taking an in-depth look at program and service quality. This section describes a conceptual framework that can inform future research questions and poses potential questions for a future descriptive study of Early Head Start that attempt to address the needs of key stakeholders identified in Chapter I.

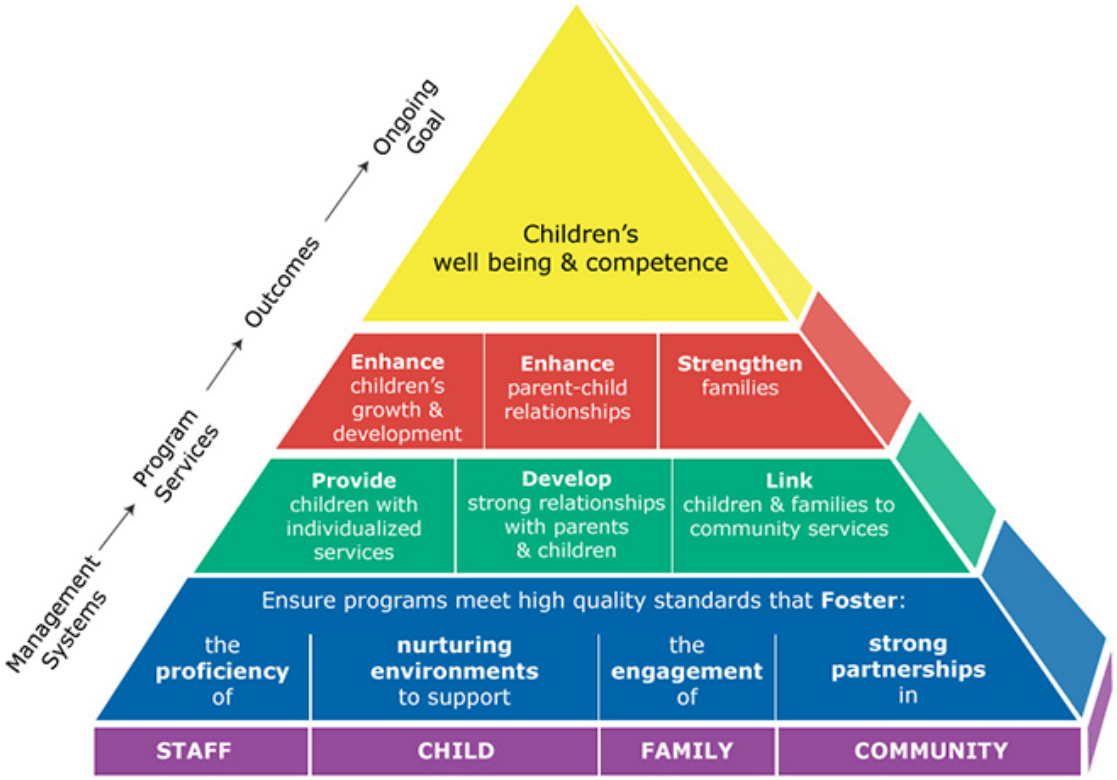
1. Early Head Start Conceptual Framework

Early Head Start programs provide a wide range of services, including child development services, child care, parenting education, case management, health care and referrals, and family support. In addition to delivering many services directly, programs also form partnerships with other community service providers. To ensure the quality of their offerings, Early Head Start programs adhere to two key institutional benchmarks: (1) the performance standards and (2) the Framework for Programs Serving Infants and Toddlers and Their Families. The performance standards are the rules and regulations that

explicitly identify what programs must do to ensure high quality services (for example, they specify child-to-adult ratios in child care centers, educational requirements for staff, and the types of services that must be offered) (ACF 1996). The framework is a conceptual model that describes the mechanisms by which high quality programs are thought to affect children's outcomes. The framework is structured as a pyramid that rests on a foundation of four cornerstones (community, staff, family, and child development) that the Advisory Committee on Services for Infants and Toddlers deemed essential for quality Early Head Start programs (ACF 2003). On this foundation, the framework builds four layers, with management systems as the base that supports program services. These services are expected to bring about positive family and child outcomes and contribute to the ongoing goal of children's well-being and competence (Figure II.1). Baby FACES 2009 was guided by this framework. This framework can still serve as a starting point for future studies of Early Head Start.

The framework depicts three linked pathways to the ongoing goal of children's well-being and competence: (1) a direct pathway through provision of individualized services to each child (Figure II.1 left side); (2) a relationship pathway, mediated through staff developing relationships with parents and children that enhance the quality of the parent-child relationship (middle); and (3) a family pathway, that links children and families to needed services and supports families as they work to reach their family development goals (right side). Informed by lessons learned in Baby FACES 2009 and the needs of key stakeholders, future descriptive studies of Early Head Start could assess program implementation, the quality of services, relationships, and contexts that are associated with children's well-being and competence, and examine the relationships among them in a more complex way. This includes continuing with some research questions from Baby FACES 2009 as well as addressing some new research questions.

Figure II.1. Early Head Start Framework for Programs Serving Infants and Toddlers and Their Families



Source: Framework for Programs Serving Infants and Toddlers and Their Families. Head Start Approach to School Readiness. HHS/ACF/OHS. 2012.

2. Research Questions for Future Studies of Early Head Start

As described in Chapter I, the purposes for future descriptive studies of Early Head Start are to document program performance by helping understand what programs are doing and how they are doing it, and to inform policy and T/TA to support program improvement. More specifically, stakeholders are interested in understanding service quality, program T/TA priorities, transitions and continuity of care, and the experiences of key subgroups. Considering these purposes and guided by the conceptual framework, the primary research questions for future Early Head Start studies are similar to those for Baby FACES 2009 but are expanded here as suggested by the members of the TWG. In addition, TWG members suggested that future studies address new cross-cutting questions as well. The researcher questions are grouped first by three broad study objectives (that are similar to the Baby FACES 2009 study objectives): (1) describing Early Head Start program services; (2) describing the population served; and (3) relating program services to child outcomes. The research questions are then grouped by the relevant topical area that falls within each broad objective (e.g., program implementation, program quality). The primary questions include:

Describing Early Head Start Program Services

- Program implementation (Newly developed for future studies of Early Head Start)
  - What are the characteristics of program implementation?
  - Are Early Head Start program strategies implemented with fidelity? What factors (for example, external systems and implementation input) support implementation with high fidelity?
  - What innovations do programs make? What are the mechanisms for program improvement?
- Program quality (Expanded from Baby FACES 2009)
  - What does Early Head Start quality look like (in a representative sample)?
  - What are the factors that shape quality in Early Head Start programs?
- Professional development (Newly developed for future studies of Early Head Start)
  - How is staff professional development delivered in programs, and what are its influences on staff?
  - What are the indicators of effective professional development in Early Head Start?
  - How do both the quality of supervision and staff competencies change over time?
- Service delivery (Expanded from Baby FACES 2009)
  - How do Early Head Start programs deliver comprehensive services to all families? What is the nature of the partnerships they participate in to meet family and child needs across all of the outcome domains?
  - How are services individualized to meet the needs of each child and family?
- Program features (Addressed in Baby FACES 2009)
  - What are the qualifications of staff, and other important program features and characteristics?

Describing the Population Served

- Characteristics of children and families served (Expanded from Baby FACES 2009)
  - What are the characteristics of enrolled children and families (overall and by key subgroups)?
  - What are the needs of families? Are programs meeting those needs?
- Child and family functioning over time (Baby FACES addressed this question, but the sample of children and families were not representative of the Early Head Start population)
  - How do children and families fare over time during Early Head Start program enrollment?

Relating Program Services to Child and Family Outcomes

- Associations of services and outcomes (Expanded from Baby FACES 2009)
  - How are child and family characteristics and outcomes associated? How are they associated with services received over time?

- Are there relationships between program features and outcomes? How do they vary by subgroups?

Cross-Cutting Issues

In addition to the primary research questions listed above, TWG members highlighted some cross-cutting issues that touch on multiple areas of inquiry and could be addressed in future studies.

- How are programs selecting and supporting their chosen curricula and child assessments, and how are assessments used in programs? What curricula and assessment tools are programs using, and do the assessments align with the curricula? How do programs create and individualize plans when they feel they are required to follow a curriculum? What are the features of high quality home visits? What makes a skilled home visitor? How do home visitors work with families with differing risks? What do parents do with their children after the home visit—is there follow-through on what was shared during the visit? What do parents learn during home visits, and what do they expect for subsequent visits? What should home visits include during pregnancy and as children get older?
- What is the level of cooperation and partnership that occurs between Early Head Start programs and other types of child care in the community?
- What are the characteristics of and services for special populations and subgroups? Examples of subgroups include children with identified special needs, highest-risk families, mothers with depression, DLLs, and mothers pregnant at program enrollment.
- For families where languages other than English are spoken, how literate are parents in their first language? How do home visitors and classroom teachers use the home language and what do programs do when there are multiple languages spoken by children in the classroom? What is the quality of language use in programs?
- Does every family and child need the same amount and type of services for 3 years? Can services be more precisely tailored to family needs, with decreases in services as they are no longer needed?
- What are the reasons that families leave Early Head Start before children are 36 months old? Is it because they no longer need services, or do they leave for other reasons? Are the leavers the higher- or lower-risk families?
- What are the combinations of 0 to 5 services that support children's school readiness? What are the experiences of children from 0 to 3 that would help them transition to formal child care at age 3? What helps to sustain boosts from Early Head Start?

CHAPTER III: CONCORDANCE OF MEASUREMENT STRATEGIES WITH RESEARCH QUESTIONS AND STUDY DESIGNS

Future studies of Early Head Start will need to consider updates and improvements to existing measures while balancing needs for keeping constructs fairly constant over time. This will allow comparisons to prior waves while also considering new policy questions and exploring the utility of new measures. Ultimately, the research questions guide the selection of data collection instruments and highlight the key constructs to be assessed. In this chapter, we describe the key measurement domains and constructs that would help answer the research questions identified in Chapter II. The domains include: program features, service inputs, and service quality; child outcomes; and family outcomes.

A. Program Features, Service Inputs, and Service Quality

Understanding how programs are operating requires measures that identify the key aspects of program features, service inputs, and service quality. The measures at the program or service level could include: (1) program implementation; (2) the types of services that programs provide (directly or through referrals) and the frequency of service receipt for individual families, as well as the degree to which services are individualized to meet families' needs; (3) program service quality; and (4) staff characteristics, skills, beliefs/attitudes, and well-being.

1. Program Implementation

Program implementation includes measures that indicate how programs are doing, what is working, and how to improve it. We currently do not have a good measure of program implementation. Baby FACES 2009 used program director self-ratings and survey questions to measure implementation of the Program Performance Standards. However, the measure had low variability and was generally unrelated to other aspects of programs such as staff characteristics and quality or child outcomes (Vogel et al. forthcoming). Future studies of Early Head Start program implementation will need to develop a new approach to measurement that effectively captures the variability in implementation and the features of implementation that relate to other important constructs.

Specific topics for program implementation could include:

- Transitions between service options or from the program to Head Start or other early childhood programs
- Leadership and management
- Use of data for program planning and quality improvement
- Community and child care partnerships
- Staff supervision, professional development, and training/technical assistance
- Implementation of the primary curricula chosen by programs; implementation of core practices and services

2. Program Service Provision and Receipt

Measuring service provision and receipt is difficult, although they are critical to understanding program implementation. In Baby FACES 2009, the provision of center-based and home-based services, as well as family participation in those services, were tracked on a weekly basis by staff reports while children were enrolled in the program. Future studies of Early Head Start should continue to document service intensity and duration at the family level. One option to explore is whether these data are needed from all programs or if a sample of programs would be sufficient for addressing research questions about the association between participation and child outcomes. In Baby FACES 2009, information on service provision and receipt was gathered from all sampled programs. Ideally, these data would be collected by all programs themselves and could be used for research purposes.

Baby FACES 2009 and other previous studies of Early Head Start have collected very little information about the other types of services provided by programs. These services include those intended to promote family development and well-being, such as nutrition and health services, prenatal services for expectant mothers, disability services, parenting supports, and other social services. Opportunities for parent involvement and the level of parent participation in the program are also important. Programs may provide services directly and/or through referral. The frequency with which these services are provided by programs and the levels of receipt among individual families is not well understood. Services include but are not restricted to:

- Developmental screenings and assessments
- Services for children with disabilities
- Nutrition and health care services
- Pediatric dental services
- Prenatal and postpartum health care for expectant mothers
- Child care (other than center-based Early Head Start)
- Group socializations
- Parenting education and supports
- Adult education, job training, and other economic supports for families

3. Program Quality

Program quality measures include: (1) home visit observations that assess aspects of home visits that have been shown to be associated with child and family well-being, (2) environment and process quality in Early Head Start centers, (3) child–adult ratio and group size for the setting, and (4) staff–parent relationship quality. As described in Chapter II, Baby FACES 2009 used the HOVRS-A to assess home visit process quality, but found limited associations with child and family outcomes. Future studies of Early Head Start need to identify or develop a new measure that could better capture associations with child and family outcomes. With regard to classroom quality, Baby FACES 2009 used the ITER-R and CLASS-T for classroom observation. Future studies of Early Head Start could consider the promising new measure of classroom quality that assesses caregiver-child interactions, the Q-CCIIT (Atkins-Burnett et al. 2015), which can be used for a wider age range of children from infants to toddlers. In Baby FACES 2009, Parent-Caregiver Relationship Scale (PCRS; Elicker et al. 1997) was used to measure staff-parent

relationship quality, which has been shown to predict classroom quality (Aikens et al. forthcoming). Future studies could also incorporate the new measure of family and provider/teacher relationship quality measurement funded by ACF (Porter et al. 2012).

4. Staff Characteristics, Skills, Beliefs/Attitudes, and Well-Being

The expertise and educational background of a program's staff members are important contributors to service quality; the performance standards and the conceptual framework both stress the importance of employing qualified staff. The reauthorization of the Head Start Act requires that all center-based Early Head Start teachers must have a minimum of an associate degree in early childhood education (ECE) and have been trained (or have completed equivalent coursework) in early childhood development. By the end of 2013, at least 50 percent of teachers nationwide must have a B.A. or advanced degree with coursework equivalent to a major related to early childhood education. The Act also requires that all Early Head Start teachers have training (or have completed equivalent coursework) in early childhood development with a focus on infant and toddler development by September 30, 2012.

The dimensions in this domain include:

- Early childhood education credentials, ongoing education and training
- Infant/toddler specialization and experience
- Teaching/home visiting skills and competencies
- Cultural/language capacity
- Physical and psychological well-being
- Beliefs and attitudes about work, children's development, their professional development, and persistence in early childhood as a career

Baby FACES 2009 captured many of these dimensions. Future studies of Early Head Start should continue with these dimensions and put more emphasis on staff skills, competencies, beliefs and attitudes.

B. Child Outcomes

Ultimately, the goal of the Early Head Start program is children's well-being and eventual readiness for school. Therefore, assessment of children's developmental progress was an important area for examination in Baby FACES 2009. It will continue to be an area of interest for a future studies because we still know relatively little about what predicts outcomes for children. We recommend that the selection of the child outcome domains for study and the constructs to assess those domains should be guided by the Head Start Approach to School Readiness[5] which defines five essential domains including (1) language and literacy development, (2) cognition and general learning, (3) approaches to learning,

[5] See http://eclkc.ohs.acf.hhs.gov/hslc/hs/sr/approach. Note: In 2015, the Office of Head Start released a newly revised *Head Start Early Learning Outcomes Framework: Ages Birth to Five* designed to represent the continuum of learning for infants, toddlers, and preschoolers and to replace the earlier frameworks. This new framework includes five central domains which align with the five domains discussed here. See http://eclkc.ohs.acf.hhs.gov/hslc/hs/sr/approach/pdf/ohs-framework.pdf.

(4) physical development and health, and (5) social and emotional development. These are the areas of child development that Early Head Start is working to support. Below, we describe the constructs in each of these areas that are likely to be part of future descriptive studies of Early Head Start.

1. Language and Literacy Development

Language skills are foundational to development of other domains, and accordingly, infants and toddlers are primed to rapidly learn to communicate with their parents and caregivers. The huge growth in communication skills across the first three years of life poses measurement challenges and requires identification of a measurement strategy that can capture all types of communication and language skills, from gestures to language comprehension and expression, and assess growth in language skills in the first three years. As noted in Chapter II, Baby FACES 2009 adapted some measures (for example, the MacArthur-Bates Communicative Development Inventories [CDI]) and used IRT methods to create scores that are suitable for examination of growth in the first three years. Future studies of Early Head Start should continue this work. In addition, future measure reviews should also consider recent advances in technology that allow for recording child and adult naturalistic utterances using wearable recording devices.

For DLLs, we need to capture development in both languages. Using conceptual scoring, in which credit is given for correct answers regardless of the language used, is an important complement to standardized scores and provides more information on overall language development. Baby FACES adapted the PLS-4 and conceptually scored the measure. Future studies may take advantage of the new measures available since Baby FACES 2009 (for example, the PLS-5 English and Spanish versions) to better capture language development in DLLs.

Given the programmatic and public policy emphasis on early literacy skills, emergent literacy (such as phonemic awareness, alphabetic principles, and concepts of print) is another area of importance to consider for future studies of Early Head Start. This is an area that is not captured in Baby FACES 2009.

2. Cognitive Development and General Knowledge

Cognitive skills are an essential part of child development and early learning and help children understand the world around them. For infants and toddlers, cognitive development encompasses a range of abilities and skills, including memory, imitation, problem solving, classification, language comprehension and production, and early number and letter concepts. General knowledge is information infants and toddlers gain from their physical and social environment. In Baby FACES 2009, this domain was captured only by the Ages & Stages Questionnaires, Third Edition (ASQ-3) Problem Solving subscale. Future studies should strengthen the measurement in this domain and explore the possibility of new measures available since Baby FACES 2009.

3. Social and Emotional Development

The development of social competence, adaptive behavior, and self-regulation and reduction of behavior problems are seen as critical elements to the development of an integrated sense of self that is necessary for productive functioning in school, work, personal relationships, and participation in family and community life. Children's social-emotional development at very young ages has been linked to later well-being and early elementary school performance (Klute et al. 2007). As children progress through the toddler years and into preschool, important social-emotional milestones are reached—such

as the development of empathy, cooperation, and compliance—making this an appropriate period to explore individual differences in the development of social competence. Social competence refers to positive behaviors that allow children to engage with peers and adults in mutually enjoyable interactions. Most scales that measure social competence include items on pro-social behaviors, such as cooperation and compliance, as well as the absence of negative or problem behaviors, such as anxiety and aggression. Baby FACES 2009 assessed children's social and emotional development through both parent and teacher/home visitor report and found some inconsistencies between reporters. Future studies of Early Head Start will need to examine these inconsistencies, consider alternate assessments, and identify the most appropriate method for assessing social and emotional development of infants and toddlers.

4. Self-Regulation and Approaches to Learning

Self-regulation describes the ability to initiate, maintain, inhibit, or modulate cognitive, emotional, and motor processes to achieve one's goals (Eisenberg & Spinard 2004). The capacity for self-regulation can be demonstrated in behaviors as diverse as delaying gratification, slowing motor responses, and suppressing outbursts of anger. Approaches to learning refer to aspects of children's responses to learning situations and include persistence, emotion regulation, attentiveness, flexibility, and organization (Fantuzzo et al. 2007; McWayne et al. 2004). Research on approaches to learning with infants and toddlers is limited and this area of development is not well understood. Although Baby FACES 2009 included the Orientation/Engagement and Emotion Regulation subscales of the Bayley Scales of Infant Development (Bayley 1993), coverage of self-regulation and approaches to learning was not a major emphasis and should be considered in future studies.

5. Physical Well-Being and Motor Development

Children's physical health is influenced by factors including the quality and timing of prenatal care, health conditions and treatment of the infant/toddler, safety practices, feeding practices, sleep routines, and medical insurance. Most of these could be captured in a parent interview, which is easier and more cost effective than direct assessment. These were covered well in Baby FACES 2009, and should be continued in future studies. Key areas include:

- **Prenatal care and birth outcomes.** Proper prenatal care is important to the health of both mother and child. Routine visits to a medical professional can ensure that the child is developing normally, and that the mother can be screened for certain health problems and counseled about caring for her health and the health of the child after s/he is born.
- **Nutrition and feeding practices.** Parents' feeding practices not only influence child weight gain and growth, but also relate to other outcomes like susceptibility to infections and tooth decay. The benefits of breast feeding have been well documented. Not only is breast milk a complete source of nutrition for infants, but breast feeding also enhances immune system functioning in the infant and reduces risk for chronic diseases. In contrast, other feeding behaviors have been recognized as being unhealthful in early childhood. Prolonged bottle feeding and putting the baby to bed with a bottle can contribute to tooth decay as well as obesity (Gaffney et al. 2004). Excessive consumption of other caloric beverages of low nutritional value, like fruit juices, can also contribute to overweight.
- **Gross motor and fine motor skills.** Children's gross motor and fine motor skills can support their overall health and physical fitness and enhance their progress in other domains. For

example, gross motor skills lead to growing confidence and pride in accomplishments (social and emotional development, self-concept). Children use their fine motor skills to experiment with writing tools and materials (literacy, early writing).

- **Health conditions and treatment.** A standard part of health care for infants and toddlers is the screening, diagnosis, and treatment of health conditions and illnesses, including disabilities.
- **Obesity.** Childhood obesity is increasing nationally, and children from low-income families are especially at risk. Measuring height and weight and combining this information with children's gender and age (i.e., creating BMI percentile) will make it possible to assess the prevalence of overweight and obesity in the study population.
- **Safety practices.** Safety practices are extremely important in this age group, since injuries are the leading cause of death for children ages 14 and under. Examples of safety practices include using gates for the top of the stairs; having covers on all electrical outlets in the home; using a car seat or booster seat in the car with a seatbelt on; and not leaving the child alone with food, near water, or in a hot car.
- **Medical insurance.** Health and dental insurance are relevant to child health, as the lack of insurance could deter parents from seeking prenatal care or medical attention for their child.

C. Family and Home Characteristics, Parent–Child Relationships, and Family Outcomes

Early Head Start aims to involve parents/families in the program by developing strong family–staff relationships that support parents in their role as their child's primary nurturer and in achieving personal and family goals (Forry et al. 2012). Early Head Start family development activities include assessing family needs and either providing relevant services to families directly or linking families to services available in the community. Expected outcomes include enhanced parent–child relationships and stronger families (Kisker et al. 2004). Parent outcomes include demonstrating increased knowledge of child development and awareness of developmental progress; enhanced self-concept, emotional well-being, and decreased parenting stress; progress toward educational, literacy, and employment goals; and stronger relationships among adult family members in support of their children's care and development. In addition, the program also targets enhancing the amount of time parents spend with children in stimulating activities as well as parenting supports, such as providing alternative discipline strategies.

In addition to measures of family background and characteristics, such as family income, parental education and employment, DLLs, and immigrant status, measures of families could cover the areas of parenting, parent well-being (including psychological and physical health and risky behaviors), and the home environment and routines. These constructs were captured in Baby FACES 2009. Future studies of Early Head Start should measure the constructs in a way that is similar to those in Baby FACES 2009.

1. Parenting Outcomes and Home Environment

Parenting outcomes. Parenting outcomes include parenting knowledge—that is, knowledge of developmentally appropriate behaviors and milestones—parenting attitudes, childrearing practices, and family cohesion.

- Parenting knowledge encompasses understanding how to care for children, how children develop, and the diverse roles parents play in children's lives. The general state of knowledge that parents possess in these areas constitutes a vital frame of reference from which parents interpret their children's behaviors. Parenting knowledge affects parents' everyday decisions about their children's care and upbringing, which in turn affects children's development.
- Childrearing practices include whether parents use coercive or more developmentally appropriate forms of child discipline.
- Family cohesion taps how the system of the family works, particularly with respect to childrearing and conflict.

Parent–child relationships and home environment. Parent–child relationships characterized by supportive and engaged interactions are associated with children's development of social, cognitive, and linguistic competencies (Landry, Smith, Miller-Loncar, & Swank 1997; Linver, Brooks-Gunn, & Kohen 2002; Shonkoff & Philips 2000). The quality of stimulation and support available to children in the home environment and in family routines makes a significant contribution to children's development. For example, Hart and Risley's (1995) work reported vast differences in the amount of language spoken in family homes by adults overall, by adults to children under 3 years old, and by young children across families from different socioeconomic groups. They found a 30 million-word gap by age 3 between children of divergent socioeconomic classes, which predicts the achievement gap later in school years. Children with the least exposure to language in the home had the weakest language skills. There is currently a rise in interventions to close this word gap (for example, Providence Talks; the Thirty Million Words Initiative).

Because of the centrality of the parent–child relationship and home environment in early child development, many large scale studies include observational methods for measuring the quality of the parent–child relationship and home environment (Brooks-Gunn & Markman 2005; Caldwell & Bradley 2003) to augment parent reports. There are tradeoffs between using self-reports, live observation ratings, and video-coding of semi-structured interactions for measuring parent–child relationship and home environment.

2. Parent Well-Being

Parent well-being includes parental depression, substance use, parenting stress, family stress, relationships of household members, and social support. These measures are related to parenting and child outcomes, and these data can be collected from self-reports relatively reliably.

D. New and Emerging Measurements Areas

In addition to the measurement domains and constructs described above, there are new and emerging measurement areas for infant/toddler studies that might be promising areas to explore in future studies of Early Head Start. These include executive functioning, toxic stress, and biological and neuro-psychological measurement in infants/toddlers.

Executive functioning underlies self-regulation and involves higher order cognitive processes such as attentional flexibility, planning, and inhibitory control (Carlson 2005; Diamond, Barnett, Thomas, &

Munro 2007; Rothbart, Sheese, & Posner 2007). Executive functioning has grown as an area of interest in recent years because of its predictive power for later development.

Exposure to traumatic, chronic, and/or frequent adverse experiences in childhood can have negative impacts on health and development (Garner et al. 2012; Shonkoff et al. 2012). Exposure to such adverse experiences in the absence of protective factors can result in toxic stress responses which are characterized by strong, frequent, and/or prolonged elevation of the body's physiological stress response system. Prolonged exposure to stress responses can disrupt development and permanently change brain architecture in ways that elevate risk for disease and impairment into adulthood. Children who experience chronic early adversity without the benefit of supportive relationships with adults are at most risk for negative outcomes. As a program that works with disadvantaged families with very young children, Early Head Start is well-positioned to identify children exposed to early adverse experiences that put them at risk for toxic stress responses. It can also help mitigate the effects of this exposure.

There has been an explosion in the use of biological and neuro-psychological measurement in infants/toddlers. For example, researchers measure stress physiology biomarkers when studying children's toxic stress responses—using saliva samples, hair cortisol, and electrocardiogram (ECG) data (ACF 2014). Future studies of Early Head Start might consider incorporating these measurement approaches if they are helpful in answering new research questions, although cost and potential tradeoffs with other measures of children's functioning will be an important consideration.

CHAPTER IV: POTENTIAL OPTIONS FOR FUTURE STUDY DESIGNS

Future descriptive studies of Early Head Start will require a design that is guided by the Early Head Start conceptual framework (see Chapter II). The design will also address gaps in the Baby FACES 2009 design and align with the study's primary research questions. The particular design selected will therefore depend on the research questions prioritized for the study and the levels of data and frequency of reporting required to answer those questions. Questions to address before selecting a design include:

- What has greater priority—tracking characteristics over time or providing a comprehensive snapshot at a single point in time?
- Would ACF and its stakeholders want to provide information across a cross-section of families enrolled in centers/home visits at a given point in time or to be able to represent children and families at key developmental ages?
- Should the study measure children's growth across their time in the program? As children transition to new settings and enter kindergarten?
- What levels of data are important to stakeholders—for example, do stakeholders think having a representative sample at the staff or classroom level is necessary?

The answers to these questions will determine whether a cross-sectional design, a longitudinal design, or some combination of the two is most appropriate.

In this chapter, we describe different design options for future descriptive studies of Early Head Start and discuss the advantages and disadvantages of cross-sectional and longitudinal designs. We also explore combination designs that could be tailored to fit the needs of ACF. Finally, we discuss issues regarding data analysis, reporting and dissemination of findings that should be considered across all design options.

A. Cross-Sectional Design

1. Cross-Sectional Design Overview

Data from a cross-sectional survey can be used to make inferences about a population of interest at a single point in time (Hall 2008). In a cross-sectional survey, data are collected from a broad selection of individuals (for example, children or programs).

A cross-sectional design for future descriptive studies of Early Head Start could provide a comprehensive snapshot at one point in time and provide information that would address a wide range of questions. A cross-sectional design could be used to describe the characteristics of Early Head Start programs, teachers and home visitors, and classrooms and home visits, and the population of Early Head Start children and families (including demographic, household, and family characteristics, family needs, and risk factors). This design could be used to assess the functioning of children and families at a given point in time. With this kind of design, the following research questions are examples of what could be asked:

- What are the characteristics of enrolled children and families (overall and by key subgroups)? What skills and competencies do they demonstrate at different ages?

- What are the characteristics of good program implementation and service quality, including staff characteristics and program take-up?
- How do program and community characteristics relate to implementation and quality?
- Are there changes over time in population served by Early Head Start or in the overall quality of services provided by Early Head Start programs?[6]

2. Sample Design

In a cross-sectional design, individuals can be selected to be representative of a population. The TWG members highlighted the importance of a representative sample at the classroom/home visitor and program level and broader coverage of the entire Early Head Start population for future study designs. Baby FACES 2009 selected a nationally representative sample of Early Head Start programs, and surveyed all children in each of two age ranges within sampled programs. Therefore, it cannot offer descriptions of children and families served by Early Head Start who were outside of those age ranges. Teachers/home visitors and classrooms were studied only when they were linked to the study children and thus are not representative of all Early Head Start staff. This sampling strategy limits the ability to aggregate the data to generate program- or center-level indicators of quality. In addition, small sample sizes limit the child-level subgroups that can be analyzed.

The key element of a cross-sectional design for future descriptive studies of Early Head Start is nationally representative samples of Early Head Start programs, teachers/home visitors, and children. In a cross-sectional design, researchers could select a nationally representative sample of children that is large enough to enable subgroup analysis, as well as a representative sample of teachers/home visitors in Early Head Start that permits aggregating teacher/home visitor level data to the program level.

One way to achieve a sample that addresses the limitations in Baby FACES 2009 is through multi-stage sampling at the program, center, classroom/home visitor, and child levels. Using this approach, the first stage of sampling could involve selecting a representative sample of Early Head Start programs.[7] The second and third stages of sampling would be Early Head Start centers and classrooms or home visitors, respectively. Stage four would involve selecting a representative sample of children/pregnant women enrolled in the sampled programs at a specific time (for example, in the fall of the program year) regardless of entering time and classrooms or association with the sampled home visitors. When child outcomes data are collected, a nationally representative sample of children enrolled in Early Head Start would be selected, spanning all ages that are served (birth through age 3).

Based on the power estimation in the design report for Baby FACES 2009 (Kisker et al. 2003), to achieve the representativeness of the samples for Early Head Start programs, centers and classrooms (or home visitors), and children and families, examine subgroups of interests, and provide a national picture of program performance (for example, program quality) over time, the sample sizes need to increase

---

[6] This question can be addressed with a repeated cross-sectional design discussed in more detail below.

[7] Migrant and Seasonal Head Start (MSHS) programs and programs in Puerto Rico and other U.S. territories would be excluded from the frame prior to sampling to avoid issues that are specific to these program types and for cost and logistical reasons. Whether to include American Indian/Alaska Native (AI/AN) programs or not in the sampling frame depends on OHS's priority, preference, and budget.

substantially compared to Baby FACES 2009. However, the magnitude of the increase needs to be informed by careful power calculation.

There are practical limitations of assessing all children in the programs because the same assessment tools cannot be used for all children due to rapid developmental changes in the first couple of years in life. Researchers will need to consider grouping children by age for assessment purposes.

3. Data Collection Approach

The cross-sectional study could be launched as frequently as every two years, assuming data collection in the first year, and analyses and reporting in the second year. The data collection could occur at one point in time in a program year (e.g., in the fall or spring). In other words, children and families would only participate in the study at one point in time, although the same programs could be sampled over multiple data collection waves if desired. Table IV.1 describes an example of the elements and timing of data collection, including the modes we recommend to employ.

Table IV.1. Summary of Example Data Collection Components and Schedule for Cross-Sectional Design

	Fall
Child/Family Data	
Parent Interview	√
Direct Child Assessment and Assessor Ratings[a]	√
Parent-Child Videotaped Interaction[a]	√
Teacher/Home Visitor Child Rating[a]	√
Home Environment Quality Observation[a]	√
Staff Data	
Teacher/Home Visitor Interview	X
Classroom Quality Observation[a]	√
Home Visit Quality Observation[a]	√
Program Data	
Program Director Interview	X

[a]Data will not be collected for pregnant women.

X = Data collection is not tied to specific families or children in order to have a representative sample of teachers/home visitors and program directors.

4. Advantages and Challenges of the Cross-Sectional Design

The key advantage of a cross-sectional design is that—for a relatively low cost and with relatively low burden on programs and families—it can provide comprehensive snapshots of Early Head Start programs, centers, teachers/home visitors, classrooms, and/or children, depending on the target populations chosen. A cross-sectional study could also provide nationally representative estimates of outcomes for children of different ages, or for key child and family subgroups. It is important to note, however, that ensuring sufficient sample size to support studying such subgroups would add to the costs of a cross-sectional survey. Data from a repeated cross-sectional survey, in which the same population is sampled at different points in time, can be used to measure changes over time in the aggregate. For example, researchers could use data from a repeated cross section to measure the change over time in the proportion of 2-year-olds in Early Head Start who are at or above national norms on an assessment

of interest. If the priority of the study is to provide nationally representative snapshots of Early Head Start, then a cross-sectional design would be sufficient.

Though a cross-sectional design can capture changes over time in aggregate, a limitation is that the study would not follow the *same* children over time. Thus, a cross-sectional design would not be the optimal choice if examining changes in individual children's and families' outcomes over time is a top priority. This design also could not address questions about how staff or program characteristics, program implementation, service quality, or other factors might predict future child and family outcomes, because there are no baseline outcomes to be used as control variables in the analyses. Although there are key policy research questions that can only be captured longitudinally, the value of a longitudinal design may not be sufficient to offset the advantages of a cross-sectional design.

B. Longitudinal Design

1. Longitudinal Design Overview

The defining feature of a longitudinal study is that it follows the same respondents over time. Like the cross-sectional design discussed above, longitudinal studies can survey representative samples of individuals, though maintaining the representativeness of the sample over time can be a challenge due to loss of sample via attrition.

The longitudinal design for future descriptive studies of Early Head Start would track children and families over time and answer questions of child and family functioning and progress over the years in program, in addition to the questions that could be answered in the cross-sectional design, described above. Below are some examples of research questions that a longitudinal design could answer:

- How do children and families grow and change during Early Head Start program enrollment?
- How are child and family needs and outcomes associated with services received over time? Are there relationships between program features and outcomes?

2. Sample Design

As with the sampling strategies in a cross-sectional design, it is important for a longitudinal design to have nationally representative samples of Early Head Start programs, teachers/home visitors, and families/children. The sampling strategies for the cross-sectional design are applicable for the longitudinal design as well. Multi-stage sampling at the program, center and classroom (or home visitor), and child levels could achieve representative samples at different levels. Sample sizes need to take into account attrition rates over time.

There are two primary ways to select the sample of children:(1) children who are enrolled in Early Head Start in the fall of the program year regardless of entering time or (2) those who are newly enrolled during a specified period before the fall. Both approaches have their advantages and disadvantages. Selecting children regardless of their entering time would result in a sample that represents the whole population of Early Head Start children and families. Using this sampling approach, researchers would be able to answer the question of how all children and families enrolled in Early Head Start grow and change in specified time. However, because the sampled children and families will have been in the programs for varying amounts of time, we would not have true baseline data for the sample (that is, no data on children's and families' functioning at the start of their program enrollment). Selecting newly

enrolled children (that is, an enrollment cohort of children), on the other hand, would ensure a true baseline at program entry, but would not be representative of all children and families in Early Head Start. Thus, the design would enable the examination of how the enrollment cohort of children and families rather than all children and families in Early Head Start grow and change during their enrollment in Early Head Start program.

In Baby FACES 2009, the age cohort design limited the representativeness of the sample for all children and families and teachers/home visitors and classrooms. Moreover, Baby FACES 2009 provides no true baseline data at the start of Early Head Start participation. The new longitudinal design could address these issues by having a nationally representative sample of children and families and/or sampling children and families at program entry. Decisions on which approach to use depend on the questions of interest and focus of the study.

3. Data Collection Approach

We provide two options for the data collection schedule for a longitudinal design: (1) short-term longitudinal data collection, and (2) long-term longitudinal data collection. A short-term longitudinal study could be conducted every two years, with the first year for data collection, and the second year for analysis and reporting. The data collection could occur over a single program year, for example, with data collected in the fall and spring.[8] In other words, children and families would participate in the study for one year. Table IV.2 describes an example of the elements and timing of data collection for a short-term longitudinal design, including the modes we recommend to employ.

A long-term longitudinal study could be conducted every five years, with data collection occurring over three years, and final reporting in the last two years. The data on child/family outcomes could be collected at baseline (or program entry) in the fall and then in the spring of each program year (for a maximum of three years). Children would be followed until they are within a pre-defined window around their third birthday. Under either the short- or long-term option, before the data collection of the first cycle of the longitudinal study, there could also be an additional year for planning or piloting work on the measures. Table IV.3 describes an example of the elements and timing of data collection for a long-term longitudinal design, including the recommended modes to employ if this design is selected.

In Baby FACES 2009, about 37 percent of the families left the Early Head Start program before their eligibility ended. The TWG members highlighted the importance of following children and families who leave the program. These follow-ups could occur in subsequent waves even for those who have left the program, to learn more about attrition. For example, parents could be asked about their reasons for leaving Early Head Start and about the services they and their children are currently receiving.

[8] Although Early Head Start is a year-round program, most grantees set a program year start/end date in the summer.

Table IV.2. Summary of Example Data Collection Components and Schedule for Short-Term Longitudinal Design

	Fall	Spring
Child/Family Data		
Parent Interview	√	
Direct Child Assessment and Assessor Ratings[a]	√	√
Parent–Child Videotaped Interaction[a]	√	√
Teacher/Home Visitor Child Rating[a]	√	√
Home Environment Quality Observation[a]		√
Staff Data		
Teacher/Home Visitor Interview	X	X
Classroom Quality Observation[a]		√
Home Visit Quality Observation[a]		√
Program Data		
Program Director Interview	X	X

[a] Data will not be collected for pregnant women in the fall.

X = Data collection is not tied to specific families or children in order to have a representative sample of teachers/home visitors and program directors.

Table IV.3. Summary of Example Data Collection Components and Schedule for Long-Term Longitudinal Design

	Fall Year 1	Spring Year 1	Spring Year 2	Spring Year 3
Child/Family Data				
Parent Interview	√	√	√	√
Direct Child Assessment and Assessor Ratings[a]	√	√	√	√
Parent–child Videotaped Interaction[a]	√	√	√	√
Teacher/Home Visitor Child Rating[a]	√	√	√	√
Home Environment Quality Observation[a]		√	√	√
Staff Data				
Teacher/Home Visitor Interview	X	X	X	X
Classroom Quality Observation[a]		√	√	√
Home Visit Quality Observation[a]		√	√	√
Program Data				
Program Director Interview	X	X	X	X

[a] Data will not be collected for pregnant women at baseline.

X = Data collection is not tied to specific families or children.

4. Advantages and Challenges of the Longitudinal Design

The key advantage of a short-term longitudinal design is that—for a relatively low cost and with relatively low burden on programs and families[9]—it collects "baseline" data in the fall which enable researchers to examine how program features such as program quality are associated with changes in outcomes by controlling for baseline scores. In addition, attrition problems in a short-term longitudinal design will not be as severe as in a long-term longitudinal design. However, the short-term follow-up would not track child progress and family functioning over their years of Early Head Start participation.

By obtaining data on the same children and families throughout their program experiences and transitions out of the program, a long-term longitudinal design confers the ability to track child and family outcomes over time, as well as the program experiences that support children's development and transition to preschool. This includes the quality of early education environments that may help sustain effects of Early Head Start. However, because of program attrition, unlike a repeated cross-sectional design, longitudinal designs do not describe how child and family outcomes in aggregate are changing in Early Head Start nationally over time.

Following individual children over time involves costs to locate and track respondents and burdens families and program staff who respond to repeated survey waves. In addition, if survey items or assessments differ across survey waves, the ability to assess change over time at the level of the individual is lessened. This is a particular concern with the assessment of infants and toddlers because few measures used in assessing the skills of toddlers are also appropriate for use with young infants. Finally, attrition from the study sample complicates analyses of longitudinal data, making the assessment of the experiences of a representative sample of children more difficult. Following Early Head Start children who leave their programs could mitigate attrition and provide useful information but would be costly. Replacing sample members who leave their programs is also possible but threatens comparability across survey waves.

C. Combination Designs and Special Study Options

1. Overview of the Basic Add-On Design

To provide flexibility and timely findings for program improvement and policy, ACF may be interested in considering learning from an approach selected for the FACES redesign project—the Core Plus Design (West et al. 2012). This approach combines ongoing data collection on programs and children to ensure comparability across waves coupled with additional studies that meet potentially changing information needs during the life of the study. Future descriptive studies of Early Head Start could employ a similar design, which we call the "Basic Add-On" design. The Basic Add-On design has two components. The Basic component provides regular data on a key set of program, child, and family indicators—dashboard indicators, described in more detail below—in a representative sample of programs and children. The Add-On elements of the design complement the Basic with information on important topics and may use a range of methods depending on the research questions ACF wants answered.

---

[9] The burden on families is still considerably higher than in a cross-sectional design if we are doing a full battery twice within one year.

TWG members emphasized the importance of collecting data that can address a variety of questions for different purposes and audiences, including policymakers, programs, and researchers. This includes collecting data that can be useful for programs and technical assistance at the national, state, and local levels. With this in mind, the TWG members suggested that future descriptive studies of Early Head Start should have a set of dashboard indicators that are measured with regular periodicity and examined for change and/or improvement. Dashboard indicators refer to data from programs that are collected on a regular basis for policy and national planning. The data could include information that cannot be collected from administrative data (the Head Start Program Information Report, PIR) but that help tell the story of how programs are performing (for example, use of data for program improvement).

The elements of the dashboard that are collected and reported could be re-evaluated periodically to determine if or when new measures should be added. Thus, there is a need for flexibility so the design can address a range of questions about children, families and programs, and easily respond to rapidly evolving research and policy questions.

An issue associated with this design is that a dashboard should have just a few indicators. There is some concern that not enough is known about the relative predictive power of different Early Head Start program implementation inputs to define the key indicators of performance and that the structure of the program (three-plus years of family eligibility depending on the timing of program entry) may not lend itself to this type of design.

Nonetheless, there is a need for data about program performance and practices as they are occurring so that stakeholders can assess current initiatives and make informed program management and policy decisions. There are, however, limitations to relying on simple indicators of performance measurement. Often there is a tension between collecting indicators for the dashboard and the need to contextualize the data. The more context is included, the more the effort becomes like a full study rather than a dashboard. There is also a need for in-depth information on topics of particular interest, for example, program implementation, growth/change in child and family outcomes, and associations among classroom/home visit quality and family and child outcomes. To achieve both goals, the Basic Option provides data on dashboard indicators while the Add-On Option can be used to address in-depth special topics.

The TWG members highlighted two primary purposes of future descriptive studies of Early Head Start: (1) to describe the program, children, and families, including key subgroups; and (2) to measure change/growth. In the section below, we provide examples of how the Basic Add-On designs and options could be used to address these purposes. The Basic Add-On design emphasizes the provision of data on dashboard indicators that are reported frequently and on tracking trends in these indicators. In the meantime, it also allows study of more in-depth special topics.

2. The Basic Options: Repeated Nationally Representative Short-term Longitudinal Implementation and Child/Family Outcome Study

This approach supports reporting on dashboard indicators using a repeated, nationally representative sample of programs, classrooms/home visitors, and children/families that provides sufficient sample sizes to study key child and family subgroups (for example, family risk, immigrant status). The design can help answer research questions such as:

- What are the characteristics of enrolled children and families (overall and by key subgroups)?
- What services are provided and what is the quality of those services?

Sample Design. Either the short-term longitudinal or cross-sectional sampling approaches described earlier could be utilized; we suggest using the same sampling approaches as the short-term longitudinal design for the basic options. The multi-stage sampling strategies would achieve representative samples of programs, staff/classrooms, and/or children.

Data Collection Approach. The Basic study could focus on programs, centers, classrooms, teachers/home visitors, families, children, and/or child outcomes, depending on how we define the indicators of interest.

- **Program Basic Data**. Program data could be collected primarily from program and center directors in the spring with the goal of describing important characteristics of Early Head Start programs, including key policies, practices, and resources. The data collected would go beyond what is available through the PIR (for example, Early Head Start-community partnerships, data use in decision-making, program leadership, program implementation, and program climate and culture). No individual child/family outcomes, parent or family, or individual teacher or classroom data would be collected as part of the Program Basic data.
- **Classroom/Home Visit Basic Data.** Classrooms/home visits could be observed and teachers/home visitors could be surveyed in the spring in order to collect information about Early Head Start classrooms/home visits and children's experiences in those classrooms/home visits. The main focus would be on measuring the quality of the classroom environment, home visits, and the teacher-child interactions. A streamlined program and/or center director survey would be administered to collect a set of limited program/center characteristics that may be used to form subgroups (for example, classrooms that vary in size, percentages of children who are DLLs), providing context for the classroom/home visit data. No individual child outcome, parent or family data would be collected.
- **Child Outcome Basic Data**. Child outcomes data could be collected through parent/staff reports, observations, and the direct administration of a battery of standardized assessments of domains of school readiness in both fall and spring. Parent interviews and teacher surveys could be used to collect information on characteristics of children, their homes, and Head Start classrooms and provide context for understanding children's assessment scores and changes in these scores over time. A streamlined program and/or center director survey could also be administered to collect a set of limited program/center characteristics.

3. The Add-On Options: Longitudinal Studies of Growth/Change and Rapid Cycle Studies of Specific Program Features and Innovations

The Add-On Options, which could supplement the Basic study, offer opportunities to collect information on a wider range of topics and increase the flexibility to respond to new research questions as they emerge. They include (a) more in-depth cross-sectional descriptive studies at less frequent periodicities (rotating or one-time studies, supplements, or topical modules); (b) longitudinal studies that follow children across their time in their Early Head Start programs to investigate growth/change in

child/family outcomes and the associations of program, classroom, family, and child characteristics and child outcomes; and (c) as-needed rapid cycle special studies to explore the relationships between program initiatives, practices, and other topics and child and/or family outcomes.

The Add-On element of the Basic Add-On design could also focus on a range of topics or special studies using a variety of survey methods. There are many possible special study topics. Here we discuss several that may be of interest to ACF and other stakeholders, including topics raised by Baby FACES TWG members.

- **Long-term longitudinal study.** This special study could examine how children and families grow and change during Early Head Start program enrollment. Following Early Head Start children and families throughout their time in the program would allow us to examine relationships between program features, quality, and participation in services and child and family outcomes.
- **Studying leavers and understanding program attrition.** Understanding reasons for leaving Early Head Start before eligibility ends is important, especially given the high and not well understood rate of attrition in Baby FACES 2009. Following children and families who leave their programs could provide important information on reasons for leaving Early Head Start before eligibility ends and what happens when families leave their programs. In Baby FACES 2009, we did not find strong child/family predictors of leaving programs, apart from mobility. Future studies could focus on ways of assessing whether and how program or family characteristics predict children leaving their programs—that is, consider what predicts leaving and not just what predicts outcomes. We would need to think of factors that might be important, but would prefer those that are clearly measurable and that have implications for program intervention.
- **Curriculum, assessment, and professional development systems.** This special study could assess the curriculum, assessment, and professional development systems that are currently in place, how they align with staff professional development needs (or how these needs are identified), and investigate how programs are measuring change in connection with these systems.
- **Home visit quality.** A special study on this topic could involve the use of video-recording technology and other methods to better understand home visits quality. As part of such a study, the possibility of archiving videotapes to use in developing measures and training on their use could be considered. Additionally, home visitor self-reports to learn about skills and beliefs could be developed. As part of this special study, it would be important to learn more about tools and resources used by home visitors, supervision provided to them, how home visitors individualize the services they provide based on children's developmental stage, and how to help parents understand and support development. The special study could follow family-home visitor pairs over time. Researchers would have to make decisions about whether to follow the home visitor or the family when there is a transition (home visitor and family are no longer working together).
- **Program implementation.** Lessons learned from Baby FACES 2009 and other research about Early Head Start implementation point to the need for a new approach to studying implementation in Early Head Start, which could be a special study topic for future studies of Early Head Start. As a mature program, Early Head Start implementation research at this

stage should focus on implementation strategies (structure and processes) at multiple levels to support and sustain high-fidelity implementation and improve child and family outcomes.

- **Additional topics.** Additional topics that may be of interest include
  - Measurement development, especially for DLLs
  - Observational measurement in classrooms and home visits; further validation work with newly developed measures such as the CLASS infant and Q-CCIIT
  - Approaches to family participation and engagement; how they are linked to aspects of Head Start's Parent, Family, and Community Engagement (PFCE) framework
  - Developing and evaluating innovations
  - The formation and maintenance of partnerships; follow up with Early Head Start-Child Care Partnership grantees and/or evaluation of future rounds of Early Head Start-Child Care Partnership grants
  - Assessment of community and family needs and how programs address them
  - Program use of data to create more innovation, quality improvement, and enhancement of child outcomes; data accessibility and usability
  - Program activities to support home language for DLLs
  - Program management, culture, and supervision as dimensions of quality and improvement
  - Transitions from Early Head Start to preschool

In summary, each of the design options described above could provide valuable information about Early Head Start and each has its advantages and challenges. The selection of a particular design would depend on ACF's priorities. Table IV.4 summarizes the key advantages and disadvantages of each option.

Table IV.4. Summary of Advantages and Disadvantages of Different Design Options

Research Objectives	Cross-Sectional Design	Short-Term Longitudinal Design	Long-Term Longitudinal Design	Basic Add-On Design
Describe Early Head Start Services				
Programs	All	All	All	All
Classrooms / Home Visitors	All	All	All	All
Describe Population Served				
Children and Families	All	All if sampled regardless of entering time; enrollment cohort only if sampled among newly entering children and families	All if sampled regardless of entering time; enrollment cohort only if sampled among newly entering children and families	All
How individual children and families change over time	No	Yes	Yes	Yes
How programs and the population of children and families change over time (in aggregate)	Yes, if using a repeated cross-sectional design	No	No	Yes
Relationships between services and outcomes for children and families	No	No	Yes	Yes
Other Considerations				
Flexibility of adding new research questions that would be of interest to ACF and stakeholders	Yes, new questions can be added with each new cohort	Yes, new questions can be added with each new cohort	No	Yes
Quick turn-around of data	Yes	Yes	No	Yes
Cost	Lowest	Middle of range	Highest	Highest

D. Considerations for Analyses, Reporting and Dissemination For All Design Options

1. Analysis Approach

For all design options, carefully planning analyses prior to data collection will ensure not only that each research question will be addressed efficiently, but also that the appropriate data will be collected to generate the answers. Thus, the data analysis plan will need to articulate the fundamental framework for the study and link the research questions with specific constructs that will be measured and the instrument and data collection method to be used.

Analyses will be primarily aimed at identifying trends in the population served, child and family well-being, and program services and quality. The analytic strategies might include descriptive statistics (means, percentages), simple tests of differences across subgroups (*t*-tests, chi-square tests), and multivariate analysis (regression analysis, hierarchical linear modeling). Researchers often analyze data featuring individuals such as children nested within larger units such as Early Head Start programs using hierarchical linear modeling (HLM). This method allows researchers to capture the nested structure of the data; in particular, it enables researchers to determine how much variation occurs at each level of the hierarchical structure (Raudenbush and Bryk 2002). With the longitudinal design option, descriptive, multivariate and HLM (i.e., growth curve analysis) could also be used to analyze longitudinal data and look at relationships between or changes in data points collected over time.

Additionally, the analysis plan will need to include an approach to constructing summary scores for the psychological measures and norm-referenced assessments, such as ratings of children's social behavior and assessments of children's language development. For these types of measures, the scores can be created as suggested by the developers of the specific instrument or according to proven approaches commonly used in the literature. Other scores will be derived from observational and survey data that reflect the quality and characteristics of children's Early Head Start programs and home environments. Psychometric analyses should be conducted to document the internal consistency reliability of the scales. As needed, factor analytic and IRT approaches could be used to explore the properties of scales and subscales. Depending on the research questions of interest, the resulting summary scores could be examined by important subgroups (such as, child race/ethnicity and child gender).

Finally, the analysis plan should describe the weights required to account for nonresponse and ensure that the data provide a nationally representative picture of programs, staff, and children served.

2. Reporting and Dissemination Approach

To maximize the usefulness of the study findings across all design options, they should be disseminated in a highly accessible form to a variety of audiences. Potentially important audiences range from the general public and Early Head Start families and practitioners to the research community and policymakers. Furthermore, to be useful in program management and policy decision-making, the timeliness of data is critical—that is, it is important to share and disseminate findings as quickly as possible and to respond to emerging issues in a timely fashion. Potential report formats might include:

- Comprehensive annual technical reports aimed at documenting the methods used to address each question, analytic results, and conclusions, with technical appendices presenting detailed information about data collection and data quality.

- Short reports and non-technical briefs aimed at communicating findings around specific, high-priority topics that can be easily accessible for programmatic or policy audiences.
- Journal papers, research briefs, and conference presentations aimed at supporting an exchange of information and ideas among researchers that could enhance the quality of study methodology, measurement, and analysis.
- Data tables and key indicator reports aimed at getting information into the hands of Head Start administrators and leaders as quickly as possible (within a few months of the end of each data collection wave) for use in making decisions about the program and its priorities.

REFERENCES

Administration for Children and Families. (1996). *Head Start program performance standards.* Final rule cited in *Federal Register*.

Administration for Children and Families. (2002). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start.* Washington, DC: U.S. Department of Health and Human Services.

Administration for Children and Families. (2003). *Program performance measures for Head Start programs serving infants and toddlers.* Washington, DC: U.S. Department of Health and Human Services.

Administration for Children and Families. (2006). *Research to practice: The program performance measures for Head Start programs serving infants and toddlers.* Washington, DC: U.S. Department of Health and Human Services.

Administration for Children and Families. (2014). *Early Head Start university partnership grants: Buffering children from toxic stress, research grantees.* Retrieved from http://www.acf.hhs.gov/sites/default/files/opre/early_head_start_university_partnership_grants_buffering_children.pdf

Administration on Children, Youth, and Families. (2001). *Building their futures: How Early Head Start programs are enhancing the lives of infants and toddlers in low-income families.* Washington, DC: U.S. Department of Health and Human Services.

Aikens, N., Xue, Y., Bandel, E., Caronongan, P., Vogel, C. A., & Boller, K. (2015). *Early Head Start home visits and classrooms: Stability, predictors, and thresholds of quality.* OPRE Brief 2015-34. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Atkins-Burnett, S., Monahan, S., & Tarullo, L. (2014, July). *Key elements of caregiver interactions with infants and toddlers: Developing the Q-CCIIT observational measure.* Presentation at the National Head Start Research Conference, Washington, DC.

Atkins-Burnett, S., Monahan, S., Tarullo, L., Xue, Y., Cavadel, E., Malone, L., & Akers, L. (2015). *Measuring the Quality of Caregiver-Child Interactions for Infants and Toddlers (Q-CCIIT).* OPRE Report 2015-13. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U. S. Department of Health and Human Services.

Bandel, E., Aikens, N., Vogel, C.A, Boller, K., & Murphy, L. (2014). *Observed quality and psychometric properties of the CLASS-T in the Early Head Start Family and Child Experiences Survey.* OPRE Technical Brief 2014-34. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Bayley, N. (1993). *Bayley Scales of Infant Development, Second Edition, Manual*. New York: The Psychological Corporation, Harcourt Brace & Company.

Berman, J., & Kisker, E. (2013, October). *Negotiating implementation science and evaluation research: Planning for a national teen pregnancy prevention implementation study.* Paper presented at the annual meeting of the American Evaluation Association, Washington, DC.

Boller, K., Tarrant, K., & Schaack, D. D. (2014). *Early care and education quality improvement: A typology of intervention approaches.* OPRE Research Report #2014-36. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Brooks-Gunn, J., & Markman, L. B. (2005). The contribution of parenting to ethnic and racial gaps in school readiness. *The Future of Children*, 15(1), 139–168.

Burchinal, M., Kainz, K., Cai, K., Tout, K., Zaslow, M., Martinez-Beck, I., & Rathgeb, C. (2008, January). *Child care quality and child outcome: Multiple-studies analyses.* Presentation at the working meeting of the Developing the Next Wave of Quality Measures for Early Childhood and School-Age Program.

Burchinal, M., Kainz, K., Cai, K., Tout, K., Zaslow, M., Martinez-Beck, I., & Rathgeb, C. (2009). *Early care and education quality and child outcomes.* Research-to-Policy Research-to-Practice Brief: OPRE Research-to-Policy Brief #1. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, and Child Trends.

Caldwell, B. M., & Bradley, R. H. (2003). *HOME inventory administration manual.* Little Rock, AR: Print Design, Inc.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595–616.

Caronongan, P., Moiduddin, E., West, J., & Vogel, C. (2014). *Children in Early Head Start and Head Start: A profile of early leavers.* Baby FACES and FACES 2009 Research Brief. OPRE Report 2014-54. Washington, DC. Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*(5855), 1387–1388.

Eisenberg, N., & Spinrad, T. L. (2004). Emotion-related regulation: Sharpening the definition. *Child Development, 75*, 334–339.

Elicker, J., I. Noppe, L. Noppe, & Fortner-Wood, C. (1997). The Parent-Caregiver Relationship Scale: Rounding out the relationship system in infant care. *Early Education and Development, 8*, 83–100.

Fantuzzo, J., Bulotsky-Shearer, R., McDermott, P., McWayne, C., Frye, D., & Perlan, S. (2007). Investigation of dimensions of social-emotional classroom behavior and school readiness for low-income urban preschool children. *School Psychology Review, 36*, 44–62.

Filene, J. H., Snell, E. K., Lee, H., Knox, V., Michalopoulos, C., & Duggan, A. (2013). *The Mother and Infant Home Visiting program evaluation-Strong Start: First annual report.* OPRE Report 2013-54. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Forry, N., Bromer, J., Chrisler, A., Rothenberg, L., Simkin, S., & Daneri, P. (2012). *Family-provider relationship quality: Review of conceptual and empirical literature of family-provider relationships*. OPRE Report #2012-46. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Forsetlund, L., Bjørndal, A., Rashidian, A., Jamtvedt, G., O'Brien, M. A., Wolf, F., …Oxman, A. D. (2009). Continuing education meetings and workshops: Effects on professional practice and health care outcomes. *Cochrane Database Systematic Reviews, 2*(2).

Gaffney, K. F., Farrar-Simpson, M. A., Claure, D., & Davilla, G. (2004). Prolonged baby bottle feeding: A health risk factor. *Pediatric Nursing, 30*(3), 242–245.

Garner, A., Shonkoff, J. P., Siegel, B. S., Dobbins, M. I., Earls, M. F., McGuinn, L., …Wood, D. L. (2012). Early childhood adversity, toxic stress, and the role of the pediatrician: Translating developmental science into lifelong health. *Pediatrics, 129*(1), e224–e231.

Hall, J. (2008). Cross-sectional survey design. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp. 172–173). Thousand Oaks, CA: Sage.

Hart, B., & Risley, R. T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.

Kagan, S. L., Moore, E., & Bredekamp, S. (Eds.) (1995). *Reconsidering children's early development and learning: Toward shared beliefs and vocabulary.* Washington, DC: National Education Goals Panel.

Kisker, Ellen Eliason, Boller, Kimberly, & Schochet, Peter (2003). *Design options for a national study of Early Head Start programs*. Princeton, NJ: Mathematica Policy Research, Inc.

Landry, S. H., Smith, K. E., Miller-Loncar, C. L., & Swank, P. R. (1997). Predicting cognitive-language and social growth curves from early maternal behaviors in children at varying degrees of biological risk. *Developmental Psychology, 33*, 1040–1053.

Linver, M., Brooks-Gunn, J., & Kohen, D. (2002). Family processes as pathways from income to young children's development. *Developmental Psychology, 38*, 719–734.

Klute, M. M., Chazan-Cohen, R., Raikes, H. H., & Kisker, E. E. (2007, March). *Age 3 mediators of Early Head Start Impacts and prekindergarten outcomes.* Poster presented at the biennial meeting of the Society for Research in Child Development, Boston, MA.

LoCasale-Crouch, J. (2014, July). *Developing CLASS to observe infant and toddler classroom interactions: What have we learned and how do we improve what we are seeing?* Paper presented at the National Head Start Research Conference, Washington, DC.

McWayne, C. M., Fantuzzo, J. W., & McDermott, P. A. (2004). Preschool competency in context: An investigation of the unique contribution of child competencies to early academic success. *Developmental Psychology, 40*(4), 633–645.

Michalopoulos, C., Duggan, A., Knox, V., Filene, J. H., Lee, H., Snell, E. K., … Ingels, J. B. (2013). *Revised design for the Mother and Infant Home Visiting Program evaluation.* OPRE Report 2013-18. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Miller, A., & Frounfelker, R. (2009) *State health authority yardstick: Report for New Hampshire Bureau of Behavioral Health. Evidence-based supported employment.* Retrieved from http://www.dhhs.nh.gov/dcbcs/bbh/documents/shay.pdf

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2010). *Classroom assessment scoring system toddler manual (CLASS-T)*. Charlottesville, VA: Teachstone, Inc.

Porter, T., Guzman, L., Kuhfeld, M., Caal, S., Rodrigues, K., Moodie, S., Chrisler, A, & Ramos, M. (2012). *Family-Provider Relationship Quality: Review of Existing Measures of Family-Provider Relationships.* OPRE report no. 2012-47. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Roggman, L. (2014, July). *HOVRS-A+ v2: Developing a measure of home visiting quality.* Presentation at the National Head Start Research Conference, Washington, DC.

Roggman, L., Cook, G., Jump, V., Innocenti, M., Christiansen, K., Boyce, L., …Hallgren, K. (2009). *Home Visit Rating Scales-Adapted (HOVRS-A)* (Unpublished document). Princeton, NJ: Mathematica Policy Research.

Rothbart, M. K., Sheese, B. E., & Posner, M. I. (2007). Executive attention and effortful control: Linking temperament, brain networks and genes. *Child Development Perspectives, 1*(1), 2–7.

Shonkoff, J. P., Garner, A., Siegel, B. S., Dobbins, M. I., Earls, M. F., McGuinn, L., … Wood, D. L. (2012). The lifelong effects of early childhood adversity and toxic stress. *Pediatrics, 129*(1), e232–e246.

Shonkoff, J. P., & Philips, D. A. (2000). *From neurons to neighborhoods: The science of early childhood development.* Washington, DC: National Academy Press.

Substance Abuse and Mental Health Services Administration. (2008). *Assertive community treatment: Evaluating your program.* DHHS Pub. No. SMA-08-4344, Rockville, MD: Center for Mental Health Services, Substance Abuse and Mental Health Services Administration, U.S. Department of Health and Human Services.

Texas Christian University. (2005). *Survey of organizational functioning.* Fort Worth, TX: TCU Institute of Behavioral Research.

Vogel, C., Boller, K., Xue, Y., Blair, R., Aikens, N., Burwick, A., & Stein, J. (2011). *Learning as we go: A first snapshot of Early Head Start programs, staff, families, and children.* OPRE Report 2011-7, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Vogel, C., Caronongan, P., Xue, Y. Thomas, J., Bandel, E., Aikens, N., … Murphy, L. (2015). *Toddlers in Early Head Start: A portrait of 3-year-olds, their families and the programs serving them.* OPRE Report 2015-28, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Vogel, C. A., Aikens, N., Burwick, A., Hawkinson, L., Richardson, A., Mendenko, L., & Chazan-Cohen, R. (2006). *Findings from the survey of Early Head Start programs: Communities, programs, and families.* Washington, DC: U.S. Department of Health and Human Services.

Vogel, C. A., Caronongan, P., Thomas, J., Bandel, E., Xue, Y., Henke, J., … Murphy, L. (2015). Administration for Children and Families. *Toddlers in Early Head Start: A portrait of 2-year-olds, their families, and the programs serving them.* OPRE Report 2015-10. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Wasik, B. A., Mattera, S. K., Lloyd, C. M., & Boller, K. (2013). *Intervention dosage in early childhood care and education: It's complicated.* OPRE Research Brief 2013-15. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

West, J., Tarullo, L., Aikens, N., Gleason, P., Malone, L., Moiduddin, E., …Verbitsky-Savitz, N. (2012). *Redesigning FACES: Design options for the Head Start Family and Child Experiences Survey.* Final report submitted to the U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. Washington, DC: Mathematica Policy Research.

Xue, Y., Bandel, E., Boller, K., & Vogel, C. *Psychometric properties of parent- and staff-reported measures and observational measures of infant and toddler development in Early Head Start.* Manuscript submitted to Early Childhood Research Quarterly.

Xue, Y., Bandel, E., Vogel, C. A., & Boller, K. (2015). *Measuring infant/toddler language development: Lessons learned about assessment and screening tools.* OPRE Technical Brief 2015-52. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

APPENDIX A

Table A.1. Key Measures Used in Baby FACES 2009

Program Characteristics and Implementation	
Program Approach: Program Level	Program approach at the program level is based on director responses to questions regarding (a) the types of services their programs offer (center-based, home-based, or combination) and, (b) separately for each service option, the frequency of services offered, using responses from 2009, 2010 and 2011.
Program Option: Family Level	Program approach at the family level is based on information collected during interviews with parents and with information from programs in our Sample Management System (SMS). Parents were asked whether they receive center-based services, home-based services, family child care services, or another type of service (such as a combination of services). Parents also indicated the frequency of center attendance and home visits received.
Program Implementation: Program Level	Program directors completed a self-rating of their implementation in four cornerstone areas in an SAQ in 2009. In 2010 and 2011, they answered questions in the Program Director Survey that were then scored by the analysis team. For this report, we calculate cross-year averages of the 2009-2011 cornerstone ratings. Programs with cross-year ratings equal to or above 3 on each of the cornerstones are designated as "fully implemented."
Population Served	Program directors reported the proportion of families enrolled in their programs who were facing socio-demographic and psychological risks. We use information from 2009 and 2010 to identify programs serving a high proportion (50 percent or more) of families facing these risks.
Staff Characteristics and Program Quality	
The Center for Epidemiologic Studies Depression Scale—Short Form (CESD-SF; Radloff 1977; Ross et al. 1983)	The CESD-SF is the short form of the full-version CESD, which is a self-administered screening tool used to identify symptoms of depression or psychological distress. The tool was used to measure depression symptoms in teachers and home visitors.
Parent-Caregiver Relationship Scale (PCRS; Elicker et al. 1997)	The PCRS was reported by staff and measures the perceived relationship between the parent and the teacher or home visitor of infants and toddlers. Items capture important dimensions of the parent-caregiver relationship, including trust and confidence, communication, respect/ acceptance, caring, competence/ knowledge, partnership/ collaboration, and shared values.
Staff Demographic Characteristics	The teacher and home visitor interviews included sections with items that broadly covered: parent participation in the program, staff training and supervision, staff benefits and morale, languages spoken (by the staff member and by families in the classroom or caseload), racial/ethnic group membership, and education.
Home Visit Rating Scale-Adapted (HOVRS-A; Roggman et al. 2009), modified from the HOVRS (Roggman, Cook, Jump, Boyce and Innocenti 2006b)	Observations of home visits used the HOVRS-A, an adaptation of the HOVRS (Roggman et al. 2006b). The HOVRS-A consists of 7 items measuring the quality of home visitor strategies and effectiveness at involving and engaging the family during home visits.
Home Visit Characteristics and Content (Boller et al. 2009)	During structured observations of home visits, field staff also collected data on the topics covered, activities, and structure of the home visit.

Table A.1. *(continued)*

Classroom Assessment Scoring System, Toddler Version (Pianta, La Paro, and Hamre 2010)	The CLASS-T (Pianta et al. 2010) was used for classroom observation. It is an adaptation of the Pre-K CLASS (Pianta et al. 2008), which focuses on teacher-child interaction quality in toddler child care classrooms. The CLASS-T measures process quality along eight dimensions (Positive Climate, Negative Climate, Teacher Sensitivity, Regard for Child Perspectives, Behavior Guidance, Facilitation of Learning and Development, Quality of Feedback, and Language Modeling) within 2 domains: Emotional and Behavioral Support and Engaged Support for Learning. Dimensions are defined by observable indicators along a 7-point scale, with ratings reflecting scores in the low (1-2), mid (3-5), and high (6-7) ranges.
Child-Adult Ratio	Center-based classroom observations also included child-adult ratios and group sizes.
Services Offered and Received	
Length of program enrollment	Length of enrollment is a continuous variable expressing time of enrollment in Early Head Start in months between initial date of enrollment and graduation or drop out.
Family involvement in past 6 months	Ratings of involvement of the family by staff and collected at each wave of data collection in the Staff-Child Report (SCR). Families are rated on a 4-point scale from not involved to consistently highly involved.
Services offered by programs	This reflects home visits or center days offered to families during their enrollment. The information is derived from weekly Family Services Tracking (FST) data completed by the child's teacher or home visitor on "regularly scheduled" home visits and center days, and adjusted for missed services due to program reasons.
Services received by families	This reflects home visits or center days received during their enrollment. The information is derived from weekly staff-reported FST data.
Family Characteristics, Parenting, and the Home Environment	
Maternal Demographic Risk Index (ACF 2001)	The maternal demographic risk index captures the multiple dimensions of risk of poor developmental outcomes a child may face as a consequence of his or her mother's socioeconomic circumstances. The index comprises three risk groups (lower, medium, and highest). The index was constructed by summing the number of the following risk factors that the mother reported she faced: (1) being a teenage mother, (2) having no high school credential, (3) receiving public assistance, (4) not being employed or in school or training, and (5) being a single mother.
Maternal Psychological Risk (ACF 2001)	This is an index of cumulative risk based on mothers' reports of (1) moderate or severe depressive symptoms, (2) parenting stress one standard deviation or higher than the sample mean on either the Parenting Stress subscale or the Parent-Child Dysfunctional Interaction subscale of the Parenting Stress Index, and (3) substance use problems including parent reports of drug use in the past year or having ever had a drug or drinking problem. Scores are classified as no risk (0 risk factors), medium (1 risk factor), and high (2 or 3 risk factors).
Maternal and Child Characteristics	The parent interview also included sections that broadly covered many different aspects of the family and home environment, including family racial/ethnic membership, languages spoken in the home, program services received, parent and child health, family routines, income and housing, and income and needs.

Table A.1. *(continued)*

Parent-Child Interaction Rating Scales for the Two-Bag Assessment (PCI): Parenting Behaviors (Mathematica Policy Research 2010)	Semi-structured, video-recorded assessments of parent-child play (Two-Bag Task) were coded using the Parent-Child Interaction (PCI) Rating Scales for the Two-Bag Task. The PCI consist of 12 scales that assess a range of child and parent behaviors. Each of eight parent behaviors is rated along a 7-point scale, ranging from a very low incidence of the behavior to a very high incidence of the behavior. A composite parenting score, synchronicity, was derived by averaging scores on parental sensitivity, positive regard, and relationship quality—all of which were highly and significantly correlated.
The Parenting Interactions with Children: Checklist of Observations Linked to Outcomes (PICCOLO; Cook & Roggman 2009; Roggman et al. 2009)	The PICCOLO is an observational instrument designed to measure positive parenting along four domains known to support children's early development: affection, responsiveness, encouragement, and teaching. Twenty-nine behaviors are rated on a 3-point scale, ranging from 0 (absent) to 2 (clearly evident). Behaviors that are infrequently observed are indicated by a score of 1. The domains of affection, responsiveness, and encouragement each consist of seven items; the teaching scale consists of eight items.
Home Observation for Measurement of the Environment (HOME; Caldwell and Bradley 1984)	The HOME measures the quality of stimulation and support available to a child in the home environment. Information needed to score the inventory is obtained through a combination of parent self-report and assessor observation conducted in the home with the child's parent while the child is present. We used selected items from the Infant version of the HOME inventory, the internal environment items from the Early Childhood version of the HOME, and neighborhood rating items from the PHDCN (Sampson 2012). We derived five subscales from this assessment, as well as the total score.
Parent Support for Child Learning Index	This composite index captures the degree of cognitive stimulation provided to children in the home setting by parents or other household members. Component measures include PCI Cognitive Stimulation scores; PICCOLO Teaching scale scores; verbal responsiveness items from the HOME (whether the parent converses with the child at least twice during the visit, answers the child's questions/requests, responds to the child's talk verbally, and uses complex sentence structure in social exchanges); parent-reported frequency of book-reading and storytelling by any household member; and whether any household member helped the child learn shapes/sizes, the alphabet, colors or numbers. Scores were derived by averaging z-scored values for each of the components
Parent Provision of Learning Materials Index	The provision of home learning materials was assessed through a combination of interview items and interviewer observations from the HOME scale. Component items include the number of accessible children's books in the home; the availability of puzzles, toys that teach colors/sizes/shapes, and toys that teach numbers; the accessibility of toys, games, and books appropriate for preschoolers; and whether the child had access to an audio device and at least 5 children's tapes. Scores were derived by averaging z-scored values for each of the components.
Exposure to Violence	Exposure to Violence measures how many violent incidents (out of three) a child has observed in his or her lifetime, according to parent reports. Items come from the Infant-Toddler Social and Emotional Assessment (Carter and Briggs-Gowan 2000), in which parents are asked to respond yes or no to questions such as whether a child has "seen violence in their neighborhood" or "seen someone hit, push or kick a family member."

Table A.1. *(continued)*

External Environment	Using items from the PHDCN, external environment is a measure of the physical and social environment of the face-block (roughly equivalent to the street between two cross streets, or about 10 housing units) where the family lives. Items in this subscale are based entirely on assessor observations of the neighborhood, and include such items as general condition of most of the housing units, garbage in the street or on the sidewalk, volume of traffic, and people arguing or fighting in the street. The items are recoded as 1 (yes) or 0 (no), and then summed. Scores can range from 0 to 8.
Neighborhood Disorder	This construct uses the same items as the External Environment construct above, but scored as a z-score. The scale score is the mean of the item z scores. Higher scores indicate higher levels of disorder.
The Center for Epidemiologic Studies Depression Scale—Short Form (CESD-SF; Radloff 1977; Ross et al. 1983)	The CESD-SF is the short form of the full-version CESD, which is a self-administered screening tool used to identify symptoms of depression or psychological distress. The tool was used in Baby FACES to measure depressive symptoms of mothers (as well as in care providers).
The Parenting Stress Index—Short Form (PSI-SF; Abidin 1995)	The PSI-SF measures the degree of stress in parent-child relationships. Baby FACES includes two subscales: (1) the Parental Distress subscale measures the level of distress the parent is feeling in his or her role as a parent; and (2) the Parent-Child Dysfunctional Interaction subscale measures the parent's perception that the child does not meet expectations and that interactions with the child do not reinforce the parent.
The Family Environment Scale, Family Conflict Subscale (FES) (Moos 2002)	The FES was designed to measure the social and environmental characteristics of families. The Family Conflict subscale measures the extent to which the open expression of anger and aggression and conflict-filled interactions are characteristic of the family.
Child Development	
Preschool Language Scale—Fourth Edition (PLS-4; Zimmerman et al. 2002a, 2002b).	The PLS-4 is a direct child assessment used to evaluate receptive and expressive language skills, as well as understanding and use of grammatical rules for children from birth to 6 years of age. It is composed of two subscales: Auditory Comprehension (AC) and Expressive Communication (EC). We used the AC subscale for both of the English and Spanish editions of PLS-4.
Early Communication Indicator (ECI; Luze et al. 2001; Carta et al. 2010).	The ECI is a semi-structured, play-based assessment designed to measure the expressive communication of infants and toddlers between the ages of 6 and 36 months along four key skill elements: gestures, vocalizations, single-word utterances, and multiple-word utterances. Assessors administered the ECI which was video-recorded for later coding by staff at Mathematica.
Peabody Picture Vocabulary Test-4th Edition (PPVT-4;Dunn and Dunn 2007)	The PPVT-4 is a measure of receptive vocabulary in which children are shown a plate with four pictures and asked to point to the one that indicates the target word that is stated by the assessor "point to [target word]." It is a norm-referenced standardized test and is suitable for a wide range of ages, from 2½ through adulthood.
MacArthur-Bates Communicative Development Inventories—Infant Short Form (CDI; Fenson et al. 2000)	The CDI is designed to assess children's early receptive and expressive language and communication skills through parent report. Two measures were derived from this form: vocabulary comprehension and vocabulary production.
Ages & Stages Questionnaires, Third Edition (ASQ-3; Squires et al. 2009)	The ASQ-3 is a parent-report tool for screening children from 1 month through 5-1/2 years of age for developmental delays in five key developmental areas: (1) communication, (2) gross motor, (3) fine motor, (4) personal-social, and (5) problem solving.

The Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan and Carter 2006)	The BITSEA is the screener version of the longer ITSEA, which is designed to detect delays in the acquisition of social-emotional competencies as well as social-emotional and behavior problems in children 12 to 36 months old.
Parent-Child Interaction Rating Scales for the Two-Bag Assessment: Child Behaviors (Mathematica Policy Research 2010)	Semi-structured, video-recorded assessments of parent-child play (Two-Bag Task) were coded using the Parent-Child Interaction (PCI) Rating Scales for the Two-Bag Task. The PCI includes four child behaviors rated along a 7-point scale, ranging from a very low incidence of the behavior to a very high incidence of the behavior.
Bayley Behavioral Rating Scale (BRS; Bayley 1993)	The BRS measures the child's behavior during child assessment. The BRS is one of the three component scales of the Bayley Scales of Infant Development—Second Edition (Bayley 1993). There are two subscales of the BRS used in Baby FACES: Orientation/Engagement, measuring the child's cooperation with the assessor during the assessment, positive affect, and interest in the test materials; and Emotional Regulation, measuring the child's ability to change tasks and test materials, negative affect, and frustration with tasks during the assessment.
Behavior Problems Index (BPI; Zill and Peterson 1986)	The Behavior Problems Index (BPI) measures child externalizing behavior problems (such as aggression and hyperactivity) and internalizing behavior problems (such as anxiety and depression). The NLSY used the BPI with children 4 years of age or older and the PSID Child Development Supplement used the BPI with children 3 years of age or older. Baby FACES used the BPI in Staff Child Report and Parent Self-Administered Questionnaire, but the items are slightly different than those used in these two studies.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ▪ ANN ARBOR, MI ▪ CAMBRIDGE, MA ▪ CHICAGO, IL ▪ OAKLAND, CA ▪ WASHINGTON, DC

MATHEMATICA
Policy Research

Mathematica® is a registered trademark
of Mathematica Policy Research, Inc.