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.
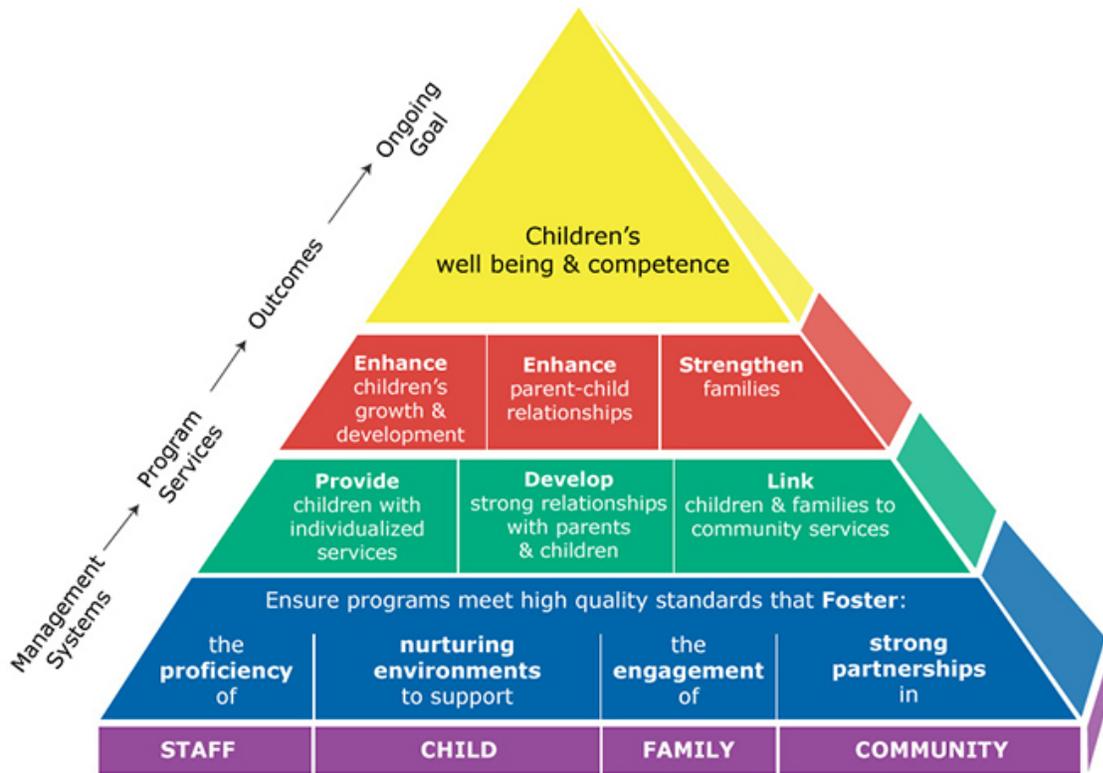
- It is challenging to find measures of infant/toddler development with strong reliability and predictive validity because developmental changes occur quickly and are not strongly predictive of later functioning.

- 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)

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

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

- Program quality (Expanded from Baby FACES 2009)

  - o What does Early Head Start quality look like (in a representative sample)?
  - o What are the factors that shape quality in Early Head Start programs?

- Professional development (Newly developed for future studies of Early Head Start)

  - o How is staff professional development delivered in programs, and what are its influences on staff?
  - o What are the indicators of effective professional development in Early Head Start?
  - o How do both the quality of supervision and staff competencies change over time?

- Service delivery (Expanded from Baby FACES 2009)

  - o 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?
  - o How are services individualized to meet the needs of each child and family?

- Program features (Addressed in Baby FACES 2009)

  - o 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)

  - o What are the characteristics of enrolled children and families (overall and by key subgroups)?
  - o 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)

  - o 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)

  - o How are child and family characteristics and outcomes associated? How are they associated with services received over time?
  - o 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.