

V. DESIGN PLAN FOR ASSESSING FAMILY AND CHILD OUTCOMES

While the primary goal of the EBHV grant program is to build sustainable infrastructure and learn about effective program implementation of the EBHV grantee-selected programs, the goal of home visiting programs is to improve family and child well-being and reduce rates of child maltreatment. To address this, the cross-site evaluation will examine the impact of home visiting programs on families and children by systematically assessing the grantees' evaluations of the effectiveness of their selected home visiting programs.

In this chapter, we provide an overview of the family and child outcomes evaluation design and describe the key research questions. We then describe the cross-site recommended measures, the process for selecting those measures, and the analytic approach (the latter is described in more detail in Chapter VIII).

Overview of Domain and Key Research Questions

The Mathematica-Chapin Hall evaluation approach in the family and child domain is quite different from that in the other domains. The primary evaluation goal for the other domains is to systematically describe what happened within those domains during the grant initiative. For the family and child domain, the primary goal is to assess whether the home visiting programs implemented by the grantees in selected local communities affected the outcomes of families and children. The analyses of the EBHV grantee-selected program impacts will occur once, at the end of the national evaluation.

After the grantees have completed their 17 individual, local impact evaluations of the EBHV grantee-selected programs, the cross-site evaluation team will systematically review the results of these evaluations. The systematic review will build on the local evaluation findings by both appraising and synthesizing their results. In the systematic review, we will determine the level of causal evidence about the effectiveness of the EBHV grantee-selected programs based on the research design and implementation. We will also describe the magnitude and statistical significance of the findings. The systematic review method is a useful way to summarize the results of the 17 evaluations; it is also highly flexible and supports the diversity in grantee program implementation and goals.

A key strength of a systematic review for the cross-site evaluation is that it does not require alignment of the outcome measures across all 17 grantee evaluations. Thus, for the systematic

review, it is not necessary for each evaluation to use the same family and child outcome measures or assess them at the same time (for example, based on child age or time since enrollment). While aligning the measures supports the potential for future cross-grantee analyses, both for this evaluation and for the users of the research data set required under the Mathematica-Chapin Hall contract, mandating one set of measures and data collection points would be prohibitive within the context of the cross-site national evaluation. Grantees are supporting the implementation of multiple home visiting programs, targeting different population groups, and pursuing a variety of evaluation designs. Most relevant for the family and child domain, the EBHV grantee-selected program models target children of different ages, ranging from prenatal through 16 years, and serve families for different lengths of time. As described in Table II.3, 12 of the 17 grantees plan to implement the Nurse-Family Partnership (NFP) model, which targets expectant first-time parents and provides home visits through age 2. Most of the other models target families with children from birth through age 5. One grantee is implementing Triple P, which can include families with children up to 18 years old, though this grantee is planning to enroll families with children up to age 16.

We have strongly encouraged grantees to use a recommended set of constructs, specific measures, and data collection points that resulted from extensive discussions with grantees. However, we recognize that grantees and their local evaluators have different goals and constraints. The systematic review allows the cross-site evaluation to summarize local evaluation results, regardless of how and when the outcomes are measured, affording grantees flexibility in their evaluation designs. At the same time, the systematic review will provide CB/ACF and policymakers with useful information about how effective investments in “real-world” home visiting programs are in producing positive outcomes for parents and children.

Based on the CB’s vision and overall goals for the initiative, reviews of the grantee proposals, additional information obtained during the grantee kickoff meeting, and the PLN domain calls, the cross-site evaluation team identified three overarching family and child outcomes goals for CB’s EBHV grant initiative:

1. Understand how the grantees’ infrastructure investments may affect family and child outcomes—specifically, rates of child maltreatment
2. Understand variations in effects across the different populations (family risk level, age of child, linguistic/cultural family background) that grantees targeted
3. Document how implementation issues may affect progress toward achieving family and child outcomes

Building on these goals, we identified specific research questions for the family and child outcomes domain. Table V.1 presents the family and child outcomes domain research questions. The first bolded question is the primary research question for the domain, as presented in Chapter I. The second bolded question expands on the primary question, examining how effects vary across populations and program models. For each key research question, we present the subquestions we will use to evaluate the research question, as well as an overview of the data collection modes and analytic approach we will use to answer each question. The process domain will also examine how the grantees identified their intended outcomes and/or adjusted their perspective as the initiative matured. Chapters VII and VIII present additional detail on the data collection and analyses for the family and child outcomes domain. Chapter VIII also includes the cross-domain research questions that employ family and child outcomes data.

Table V.1 Family and Child Outcomes Domain Research Questions, Data Collection Mode, and Analytic Approach

Research Questions	Data Collection Mode	Analytic Approach
	Local Evaluation Reports	Systematic Review of Evidence
Do home visiting programs improve parent and child outcomes when programs are implemented in the “real world” and supported by investments in infrastructure?		
Do home visiting programs decrease rates of child maltreatment?	X	X
Are home visiting programs associated with improvements in parent and child health and well-being?	X	X
How do effects vary across different target populations and across program models?		
Do subgroups of the target population experience differential effects of the investments in supports for evidence-based home visiting programs?	X	X
Do effects vary by the program model that grantees implement?	X	X

Family and Child Measures and Analytic Approach

We factored in many considerations when recommending the final measures for the cross-site evaluation. In addition to the preferences and goals of the grantees and the PLN family and child outcomes group, we considered:

- Assessment of constructs potentially influenced by home visiting programs
- Demonstrated sensitivity to similar interventions
- Successful use in other large-scale research
- Appropriateness for families and children from different cultural, racial, ethnic, and linguistic backgrounds (for example, availability in Spanish), as well as across different age groups
- Costs of measures (for purchasing and using copyrighted measures), training required for collecting high-quality data, and time and frequency required for data collection
- Reliability and validity of the measures in general and for Spanish speakers in particular

The cross-site evaluation is designed to assess family and child outcomes in seven measurement domains: (1) parent health, (2) parent mental health, (3) parenting, (4) child physical health/nutrition, (5) overall child development/functioning, (6) child social-emotional development, and (7) child maltreatment/agency action.

Across the grantees, considerable variation exists in the measures they will collect. The PLN family and child outcomes group encouraged grantees to include measures in each of the seven high-priority domains and, if possible, the recommended constructs. Variations across grantees in their measurement approach are driven by the specific model they are implementing, the ages of the children in their study, and the level at which they are working (grantees working only at the systems level focus on aggregate-level outcomes, such as the rates of maltreatment in a county or state).

Eight of the 17 grantees are assessing all the constructs recommended by the cross-site evaluation team; 2 of the 8 are planning to use all the recommended measures and are completely aligned with the cross-site evaluation plan. The remaining nine grantees have excluded at least one of the recommended constructs from their evaluation plans. Of the nine, three are relying on the NFP Clinical Information System database as their primary source for family and child outcome measures. (Chapter VII provides more information about the Clinical Information System.)

Grantees proposed alternative measures for 4 of the 7 recommended family and child outcome constructs, with 11 unique measures proposed as alternatives to the recommended cross-site measures. The cross-site evaluation team reviewed each of the proposed measures and provided feedback to the grantees on which measures met the reliability and validity standards most

psychometricians use to establish minimum thresholds for inclusion in research.¹² Eight of the alternative measures proposed have been approved for use based on their psychometric properties. There are also 21 supplemental measures that met the same reliability and validity standards, which grantees will collect in addition to the recommended cross-site measures. Appendix A contains a list of the approved additional/supplemental outcomes that some of the grantees are collecting.

Below, we list the domains, the recommended constructs within them, and the specific measures recommended for each. Volume II presents the cross-site measures at the item level.

Outcome Domains, Constructs, and Measures

In this section, we describe the overarching measurement domains, constructs, and specific measures, as well as the rationale for including them in the cross-site evaluation. Unless otherwise noted, the recommended outcome measures described below are available in both English and Spanish and meet the reliability and validity standards used in the field of psychology described in the footnote below.¹ The measures have a demonstrated history of success in studies that included parents and children from diverse backgrounds (education, race/ethnicity, age, culture, and economic status). Members of the cross-site evaluation team recently completed measures reviews for two large national evaluations: (1) the Early Head Start Family and Child Experiences Survey (Baby FACES); and (2) the Head Start Family and Child Experiences Survey. The cross-site evaluation recommendations in this section draw on the work conducted for those studies, as well as on reviews of the measures used in other large-scale research/evaluation projects: the National Survey of Child and Adolescent Well-Being (NSCAW II), Building Strong Families (BSF), the Early Head Start Research and Evaluation Project (EHSREP), the Fragile Families and Child Wellbeing Study, and the Early Childhood Longitudinal Surveys, Birth and Kindergarten Cohorts (ECLS-B and ECLS-K).

Parent Health. In this domain, there are a number of constructs on which the home visiting program models selected by the grantees focus. For example, for NFP, a key thrust of the prenatal

¹² For example, internal consistency reliability must be at least .70. Test-retest reliability over a short period of time (less than two weeks) must be at least .85. Kisker et al. (2003) described commonly used heuristics/thresholds for assessing the psychometric properties of outcome measures. In addition, Kisker et al. profiled more than 40 measures, including many of those that we considered for this evaluation and were candidates for sites to select from to increase alignment across the local evaluations.

home visits is to reduce risky health behaviors and increase the use of prenatal health care services that may affect the development of the baby. According to all the models and most of the PLN members participating in the family and child outcomes planning, the construct of substance use (specifically, reducing the incidence of alcohol and drug use during pregnancy and reducing the abuse of substances overall) is an outcome of importance in its own right. In addition, it is a risk factor for child maltreatment (Duggan et al. 2004). One concern about this domain and self-reports by parents is underreporting. Evidence exists that respondents report more instances of substance abuse when questions are administered by more private computer-assisted methods than by another person (Feigelson and Dwight 2000). However, the cost of computer-assisted administration may be prohibitive for grantees. In addition, we would not expect the privacy concerns to affect reporting differently for parents in treatment and control groups in a local evaluation using an experimental design.

The two cross-site measures include short parent self-report instruments that some of the local evaluators have used successfully and that have been used in large-scale research projects (Maisto et al. 1995; Bradley et al. 1998; Teitelbaum and Mullen 2000). The Alcohol Use Disorders Identification Test (AUDIT; Babor et al. 2001) identifies excessive drinking behaviors that can be classified as hazardous or risky. The AUDIT self-report questionnaire includes 10 questions that ask respondents to indicate the frequency of drinking behaviors. The AUDIT has been found to be the most appropriate for use with diverse populations, particularly for identifying early-stage indicators of alcohol misuse (Dawe et al. 2002). The Drug Abuse Screening Test (DAST; Selzer 1971; Gavin et al. 1989) asks respondents to report on 20 items. The DAST assesses drug-related problems in areas that include drug-related impairment affecting daily living and relationships and other problems caused by drug use, such as incarceration. Respondents report whether each item is a “yes” or a “no” for a given behavior or experience. The DAST has sound psychometric and predictive properties (Shields et al. 2007; Yudko et al. 2007).

Parent Mental Health. In this domain, most program models the EBHV grantees will implement focus on three risk factors for child maltreatment: parent depression, stress, and anxiety. Because depression, stress, and anxiety are often highly intercorrelated (Wolfe 2004), we will include only one of these measures as part of the cross-site evaluation. We chose client depression because there is a well-established short-form instrument with good psychometric properties that has been widely used. The Center for Epidemiological Studies Depression 12-item Short Form (CES-D) (Radloff 1977; Ross and Mirowsky 1984; Ross et al. 1983) is a screening tool used to identify

symptoms of depression or psychological distress. The 12-item short form of this self-administered questionnaire takes fewer than 10 minutes to complete. It has been used in many large-scale research projects (the EHSREP as well as FACES, ECLS-K, and ECLS-B). Respondents are asked to rate how often each of the items applied to them in the past week, on a 4-point scale from “Rarely or none of the time” (score of 0) to “Most or all of the time” (score of 3). The scale is best used as an indicator of depressive symptoms rather than a means to diagnose a clinical case. Its strengths include that it is simple to administer and score, and there are no costs associated with using it.

Parenting. Each of the home visiting program models views parenting and the parent-child relationship as the pathway to obtaining improved child outcomes. Parenting encompasses a range of constructs that broadly includes parenting attitudes, parent knowledge of children’s development, approaches to guidance and discipline, supports for children’s learning, and the quality of the parent-child relationship. Parenting is challenging to measure because parent reports about psychological constructs such as the quality of the parent-child relationship are not adequate, and the alternatives (live or videotaped coding of parent-child interactions) are costly to train research staff to collect and to code. Therefore, for the cross-site evaluation, we are focusing on parenting practices. We will include two questions about spanking: (1) whether the parent spanked the child in the past week; and (2) if yes, how often. Parental spanking is predictive of later child well-being and there are demonstrated positive impacts of similar interventions on spanking (Administration for Children and Families 2002).

Child Physical Health/Nutrition. The home visiting program models also address a range of child health and nutrition constructs, from overall physical health, birth outcomes (pre-term births and birth weight) and breastfeeding, to injuries and immunizations. The cross-site evaluation includes injuries and emergency room visits because they are proxies for child abuse. In addition, given the health prevention and promotion focus of many of the program models, immunizations provide a reasonable indicator of these types of behaviors. Parent self-report items on these topics have been used extensively in large-scale research projects and have demonstrated intervention effects and acceptable properties.

Overall Child Development/Functioning. Many of the home visiting program models are designed to support children’s overall development¹³ and prevent developmental delays. Some of them focus on this directly, and others expect indirect effects through referrals for additional assessment and to early intervention service providers. In this domain, the cross-site includes a screening assessment—the Ages and Stages Questionnaires, 3rd Edition (ASQ-3; Bricker et al. 1999; Squires et al. 2009) that addresses a range of child development areas, including communication, motor skills, and social interaction. Grantees were also offered the option of using the Denver Developmental Screening Test II (Frankenburg et al. 1996), but none included it in their implementation plans. It is not discussed further. The cross-site evaluation did not propose a child development measure appropriate for use with children 6 years old or older. This seemed appropriate given the focus of the program models grantees are using with families—the models that can be used with older children tend to focus mainly on the safety of the home environment and on child maltreatment outcomes.

The ASQ-3 are parent-report questionnaires appropriate for parents of children between 1 and 66 months of age. This series of 20 questionnaires with 30 developmental items in each questionnaire is used to screen infants and young children for developmental delays. The questionnaires focus on assessment of five key developmental areas: (1) communication, (2) gross motor, (3) fine motor, (4) personal-social, and (5) problem solving. Items in the problem-solving domain assess attention, memory, reasoning, academic skills, and perception. Parents are asked to respond on a frequency Likert scale to questions such as “If you give your child a bottle, spoon, or pencil upside down, does he turn it right side up so that he can use it properly?” Each questionnaire takes 10 to 15 minutes to complete and approximately 3 minutes to score. Questionnaires are written at a sixth-grade reading level. The ASQ-3 has demonstrated reliability, validity, and accuracy in discriminating children with and without developmental delays. Baby FACES (a study of nationally representative Early Head Start programs) conducted the ASQ-3 by telephone with

¹³ Many developmental psychologists ascribe to a comprehensive view of child development and well-being that includes dimensions of development similar to those viewed as central to school readiness: cognitive development, language development, social-emotional development, approaches to learning, and physical development (Kagan et al. 1995; Love 1999; Love et al. 1994). In identifying the cross-site domains, constructs, and measures, the cross-site evaluation team brought together information about important EBHV outcomes from many sources, including the grantees’ own proposals and the outcomes targeted by the EBHV program models. This led us to group and refer to outcomes in a somewhat different way than is typical in studies of other aspects of child development, such as school readiness.

parents when children were 12 months old, which may also be an option for local evaluators to consider if in-home interviewing is too costly.

Child Social-Emotional Development. The home visiting program models generally focus on supporting parents in their interactions with their children. The goal in this area is to reduce behavior problems by supporting the development of healthy social-emotional behavior and positive interactions between children and adults, and children and their peers. In this area, we will assess children's behavior problems. This is a common area of assessment in the social-emotional domain, given that early behavior problems are often associated with later behavioral issues and poor outcomes for children (Achenbach and Rescorla 2000). In this area, grantees were encouraged to select one of three measures: (1) the Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan and Carter 2005); (2) the Achenbach System of Empirically Based Assessment (ASEBA) Child Behavior Checklists (CBCLs) (Achenbach and Rescorla 2000, 2001); or (3) the Behavior Problems Index (BPI) (Peterson and Zill 1986). For grantees working with children 3 years old and younger, the BITSEA provides an assessment of positive behaviors in addition to problem behaviors. As described below, these measures provide full coverage of the age range of children grantees have targeted to enroll in their home visiting programs. No grantees selected the BPI so we do not discuss that measure further.

The BITSEA is the screener version of the longer ITSEA. Both are designed to detect emerging social and emotional competence, as well as social-emotional and behavior problems and delays in the acquisition of competencies in children 12 to 36 months old. BITSEA scales assess externalizing (activity, aggression), internalizing (inhibition, separation, depression), dysregulation (sleeping, eating), maladaptive habits, fears, and competence (attention, compliance). The BITSEA is a 42-item parent and caregiver report that takes approximately 7 to 10 minutes to complete. The scale focuses on the development of competencies (for example, hugs or feeds dolls or stuffed animals), as well as problem behaviors (for example, avoids physical contact). Respondents are asked to rate each item as "not true/rarely," "somewhat true/sometimes," or "very true/often." Reliability and validity are both in acceptable ranges. The BITSEA was normed on a sample that was not nationally representative: the sample excluded children who, at birth, were expected to have severe developmental delays and excluded parents who could not speak English. Strengths of the BITSEA include that it is available in English and Spanish and that it can be administered both to parents and to primary caregivers.

The CBCL uses information collected from parents to assess the behavioral, emotional, and social functioning (including language development) of children. The Preschool Forms assess children between the ages of 1.5 and 5 years, and the School-Age Forms assess children between the ages of 6 and 18 years. The CBCL consists of a self-administered parent report on a 99-item child behavior checklist. Parents rate their child for how true each item is now or within the past 6 months using the following scale: 0 = not true; 1 = somewhat or sometimes true; 2 = very true or often true. The 99 items in the preschool CBCL are organized into two broad groupings of seven syndromes. The internalizing grouping includes subscales that assess whether the child is emotionally reactive, anxious/depressive, withdrawn, or has somatic complaints. The externalizing grouping includes subscales that assess whether the child has attention problems or exhibits aggressive behavior. A third set of items in the preschool version assess whether the child has sleep problems. The items are also organized into five Diagnostic and Statistical Manual of Mental Disorders (DSM)-oriented scales (American Psychiatric Association 2000). Scales are based on ratings of 1,728 children and are normed on a national sample of 700 children.

For the CBCL School-Age Form checklist (CBCL/6-18), parents provide information for 20 competence areas covering their child's activities, social relations, and school performance. The CBCL/6-18 has 113 items that describe specific behavioral and emotional problems, plus 2 open-ended items for reporting additional problems. The items are also organized into six DSM-oriented scales. The scales are based on factor analyses of parents' ratings of 4,994 clinically referred children, and were normed on 1,753 children ages 6 to 18. The normative sample represented the 48 contiguous states for socioeconomic status, ethnicity, region, and urban-suburban-rural residence. Children were excluded from the normative sample if they had been referred for mental health or special education services in the past year. The main drawback to the CBCL is that it is very long, and the publishers prefer that all the subscales be administered. In addition, some researchers report that parents find the questions too focused on negative behavior.

Child Maltreatment/Agency Action. The specific constructs proposed in this domain include the number of child abuse/neglect reports (both substantiated and unsubstantiated), involvement in the child welfare system, and the number of foster care placements. Many published findings demonstrate that maltreatment rates have been affected by some of the specific home visiting models the EBHV grantees are using and by other similar types of interventions (for example, Appleyard and Berlin 2007; Berlin et al. 2008; Daro 2006; Prinz et al. 2009; Wolfe 2004). The child maltreatment data will be collected from administrative records obtained by local

evaluators using a common set of definitions developed by the cross-site evaluation team in collaboration with the PLN members who have experience working with these data.

The grantee representatives and local evaluators who participated in discussions about the family and child outcomes domain have discussed whether the number of reports or the number of substantiated reports (those verified after an administrative inquiry) is a more appropriate measure. When comparing long-term outcomes and developmental challenges for the children with a report of child abuse during a given period with children who have had a substantiated report of child abuse during the same period, there is little difference (Daro 2009). However, there are advantages and disadvantages associated with each measure. The advantage of focusing on reported cases is that the number of reports is larger than the number substantiated and is considered by many to be a more sensitive predictor of how children in intervention programs are doing. The disadvantage is that the administrative data quality on “reports only” may not be as good as for the substantiated cases. Substantiated cases may constitute a more accurate measure of child maltreatment because a legal standard for maltreatment has been met in the case. However, there are state differences in procedures and criterion for substantiation. For example, some states define abuse incidents less strictly than others, which may lead to different rates of investigation. Similarly, local norms may affect whether members of the public or professionals report suspicions of child abuse and neglect. We will address these limitations by collecting both reported and substantiated cases, and by focusing the systematic review of evidence at the grantee level rather than across all grantees. The local definitions reflect the experiences of families and children in the context in which they live and thus are the outcomes of interest for this grantee cluster as defined by CB/ACF. This will potentially limit the ability to conduct analyses across subgroups for the child maltreatment outcomes.

Recommended Periodicity of Data Collection and Information Submission Requirements to Support the Cross-Site Design

For the systematic review of evidence, it is not necessary, but it is desirable, to align the data collection schedule across grantees as much as possible. Collecting the data on a similar schedule will produce benefits for the study and enhance the utility of the resulting data for subsequent analysis by the local evaluators or other researchers who might access the data through the national data archive (National Data Archive on Child Abuse and Neglect [NDACAN]). Benefits include:

- Enhancing the possibility of collaboration in research efforts among grantees and analyses of subgroups across grantees (for example, coding the quality of parent-child interactions conducted at age 1; analyzing family demographic subgroups of interest across grantees)

- Aligning the timing of the TA/training the cross-site evaluation team provides in support of collecting a given measure

The recommended timing of family and child outcomes assessment will vary by measure, but generally the minimum recommended collection schedule is at baseline and at program exit. Ideally, grantees would also collect outcomes data at the midpoint of the program model implementation (this may not be necessary for the shorter interventions), at program exit, and, if possible, 12 months after families are expected to leave the program. We also recommend that all sites include permission for obtaining child maltreatment data through county/state records beyond the life of the current study as part of their consent process.

Ensuring Data Quality

To strengthen the local evaluations' family and child outcomes results and the utility of the final data file that will be prepared as a restricted-use file for NDACAN (see Chapter IX), the Mathematica-Chapin Hall team has undertaken and planned several technical assistance activities. First, in their draft implementation plans, we asked grantees to provide information about their required sample sizes to detect effects, sample recruitment, response rates, and other information about the major milestones in the project that will affect the soundness of the evaluation design. Both during and after writing the draft plans, grantee liaisons, other Mathematica-Chapin Hall staff, and CB/ACF provided technical assistance to grantees to support them in making sure their designs are as strong as possible. To continue that effort, grantees will be asked to submit a summary of their local evaluation status during key points in their data collection so that Mathematica-Chapin Hall staff can help them review indicators of evaluation quality and adjust recruitment or data collection strategies as necessary. We will ask the sites to provide:

- Study sample and design information (the number of parents and children targeted for recruitment, the number recruited with and without consent, the number assigned to participate, the number of people participating, and the total number of home visitors)
- Information about family and child data (for each cross-site and key local outcome construct, the measure used, the number of parents/children with whom collection of each measure was attempted and completed, the number of measures with substantial missing data, and the oldest and youngest age children on which the measure was used)¹⁴

¹⁴ At the end of the evaluation, for all scales (including the Adapted Working Alliance Inventory), we will also ask grantees to provide the internal consistency reliability, as measured by Cronbach's alpha, for each measure, and separately for English and Spanish versions of the instruments.

- Home visit relationship data (the number of participants and staff who were identified for participation in the study, attempted, and completed the Adapted Working Alliance Inventory at each time point, and the number with substantial missing data)

The cross-site evaluation technical assistance team will work with the local evaluators and grantees to support the initial and ongoing training required to collect the cross-site evaluation data. We will do this using a number of proven strategies learned over many years of experience conducting similar research projects. We will provide detailed training manuals with question-by-question explanations of the items and how to score them. As needed, we will conduct training webinars using a training-of-trainers model. Ideally, local evaluators will only select parent report measures in the recommended cross-site evaluation domains. This will greatly reduce the local evaluation data collection burden. We will work with local evaluators to certify a small number of trainers per site (one is preferred but possibly two if a backup is needed) to collect the parent questionnaire/interview data. After the local trainer is certified, he or she will train the data collectors at the site and report to the cross-site team about whether a data collector met the certification criteria. No data collector should be permitted to collect cross-site evaluation data without being certified. We will work with the local evaluators to ensure that data are reviewed for completeness and quality before they are provided to the cross-site team for inclusion in the data set that will be archived at NDACAN.

Within the constraints of the cross-site evaluation resources, we will also provide additional technical assistance, including on local measures not part of the cross-site evaluation (for example, by providing data collection and coding manuals from large-scale research projects that included measures local evaluators are using). The cross-site evaluation team will also provide technical assistance on family and child outcomes data analysis plans to help grantees conduct analyses and sensitivity tests on topics such as calculating minimum detectable effects and weighting for nonresponse.

Systematic Review of Evidence

To understand whether home visiting programs affect families and children, the cross-site evaluation will undertake a systematic review of evidence of effectiveness of the programs as implemented in grantees' local communities. The systematic review will assess whether home visiting programs supported by the grantees affect the outcomes of families and children. Each grantee will measure family and child outcomes for the common constructs described above, although the number and specific measures may vary by grantee. In addition, grantees will include

other outcomes in their analysis of initiative outcomes, which will also be within the purview of the systematic review of evidence.

At the end of the local evaluations, the cross-site evaluation team will review evaluation designs to determine each local study's level of evidence. Each grantee's evaluation will be categorized into one of three evidence groups: (1) strong evidence about effectiveness, (2) moderate evidence about effectiveness, and (3) exploratory evidence about effectiveness. The level of evidence will be based on the quality of the outcome measures, the rigorousness of the evaluation design, and the implementation of the design. The strong evidence about effectiveness group includes studies with the most rigorous study designs, specifically well-implemented randomized controlled trials. The moderate evidence about effectiveness group includes studies with strong, but somewhat less rigorous, designs, such as quasi-experimental designs with comparable comparison groups. The exploratory evidence of effectiveness group includes studies that do not meet the standards of the strong or moderate evidence groups, such as pre-post studies, which include no comparison group, or studies that use only outcome measures that do not meet the cross-site evaluation's required levels of reliability and validity. Exploratory studies can provide information about whether the results are consistent with the study hypotheses, which provides important information for future research; however, they cannot provide causal evidence about the links between supporting home visiting programs and family and child outcomes. For the specific criteria used to evaluate these methodological considerations, we reviewed the latest methodological developments in this area, surveying the U.S. Department of Education's What Works Clearinghouse, the Campbell Collaboration, and a review of other systematic review methods. The systematic review of evidence is described in more detail in Chapter VIII.