

VIII. DATA ANALYSIS PLAN

The cross-site evaluation will use a mixed-method approach to complement the diversity in home visiting program models, populations, and local evaluation approaches. We will conduct qualitative and quantitative analyses of fidelity, cost, and systems data, both within and across domains. The analyses will focus on systematically and clearly describing these domains and associations among domains. We will use a primarily qualitative process study to examine the overarching context and implementation of the grant initiatives. To determine whether the programs affected family and child outcomes, we will conduct a systematic review of evidence on the impacts of the EBHV grantee-selected programs on family and child outcomes by reviewing the grantees’ local evaluation results. The analyses will also combine measures of the effectiveness and reach of the programs to examine whether supporting EBHV grantees’ systems change is related to improvement in families and children’s outcomes within their communities.

In this chapter, we build on the information provided in previous chapters about the cross-site evaluation’s approach to addressing the study’s research questions. We first provide an overview of the qualitative analyses. Next, we describe the quantitative analyses, including our approach to analyzing measures within each domain, the relationships across domains, and our analysis of the partner networks grantees use to bring about systems change. Finally, we describe the systematic review of evidence we will conduct in the family and child outcomes domain and how we will combine the evidence from that review with measures of intervention reach to examine the overall effects of the EBHV initiative for grantees’ communities. Table VIII.1 provides the analytic approach for each of the cross-site evaluation domains. Chapter IX provides an overview of when we will report on each type of analysis.

Table VIII.1 Cross-Site Evaluation Domains by Analytic Approach

	Qualitative Analysis	Quantitative Analysis	Systematic Review of Evidence	Reach and Effectiveness Analysis
Systems Change	X	X		X
Fidelity to the Evidence-Based Model	X	X		
Costs of Home Visiting Programs	X	X		
Family and Child Outcomes			X	X
Process Study	X			

Qualitative Analysis

The cross-site evaluation will have three types of qualitative data to analyze: (1) site visit data, (2) web-based system data, and (3) partner survey data.²¹ To analyze these data, we plan an iterative process that consists of three steps: “noticing, collecting, and thinking” (Seidel 1998). We will use a case study approach to triangulate data from different sources and identify common themes or categories (Yin 1994). Triangulation will allow us to compare data sources for reliability, as well as identify areas of agreement and disagreement across data sources and interview respondents. Through theme identification, we will reduce the large volumes of qualitative data gathered to a manageable number of topics/themes/categories pertinent to the qualitative aspects of the cross-site evaluation (Coffey and Atkinson 1996).

Overall, we will pursue two main goals for analyzing qualitative data gathered for the cross-site evaluation. First, we will organize the data to develop a detailed understanding of each grantee and its context, the design and implementation of its home visiting program, and its perspective on achieving specified results. Second, we will conduct a cross-grantee analysis to identify themes and patterns in the implementation process of the overall initiative across the grantees.

Summarize Site Visit Interviews

Site visits will provide one source of qualitative data. After each visit, the site visit team will write up interview notes by respondent, as well as a brief site visit summary. For each type of respondent, we will develop report templates that all team members will use for writing up notes from the interviews conducted during their site visits. This will make note-writing easier and will ensure that comparable information and degree of detail in the written notes is captured across grantees. We will provide training on use of these reporting templates during the site visitor training. The reporting template for each respondent will follow the topic areas developed for the site visit protocols that were identified by looking across the evaluation domains to discern what would be gathered during site visits.

²¹ As described in Chapter II, the partner survey includes closed-ended survey questions, network questions, and open-ended questions about partners and how they work together. Partner survey data will be analyzed using a few different techniques: qualitative, quantitative, and traditional network approaches.

The primary purpose of the site visit summary will be to provide feedback to CB/ACF and each grantee on overall themes and issues identified during the visit. The site visit summary will follow a common template that highlights major successes and challenges for the grantee as learned during the visit, and provides recommendations for technical assistance to support implementation and evaluation, as appropriate. To make sure the summaries are useful, we will work with CB/ACF and grantees to develop a reporting template for recording information about the topics of greatest interest to them.

Code Qualitative Data

All qualitative data sources—site visit interviews, web-based system data, and partner survey data—will be systematically coded. We will develop a coding scheme for the site visit interviews early in the data collection period. The coding scheme will align with the research questions for each domain in the cross-site evaluation. We will define codes for each of the key themes and subtopics we anticipate. Once data collection is under way, senior members of the evaluation team will review the codes, along with initial data output, and refine the coding scheme as necessary to better align it with the topics that emerge during data collection. Codes will be reviewed throughout the data collection period to ensure they capture the themes and topics being collected in the data.

Three project team members will be trained to code the qualitative data. To ensure reliability across coders, all three team members will code the initial data for each source and compare codes to identify and resolve discrepancies. In addition, one team member will review a subsample of coded data for all sources to check reliability as coding proceeds.

Mathematica-Chapin Hall will use a qualitative analysis software package, Atlas.ti (Scientific Software Development 1997), to make it easier to organize and synthesize the interview write-ups produced after the site visits. Atlas.ti will be appropriate for coding these qualitative data, as the write-ups will include narratives that require careful coding to be useful for analysis. This software will enable the evaluation team to apply the structured coding system for organizing and categorizing the data, entering them into a database according to the coding scheme, and retrieving data linked to primary research questions. Data can then be retrieved from this system on particular research questions across all grantees, by data source, from individual respondents within sites, or by type of respondent (for example, grantee lead, evaluation partners, or direct service staff). Research team members can also use the system to retrieve all the relevant data on specific topics and assess the consistency and quality of information across respondents and sites.

The evaluation team will code responses to open-ended questions in the web-based reporting system and partner survey into relevant categories by theme. For example, in the web-based system, grantees report on key events that affected project implementation. We will compile responses and then code similar responses together, creating a smaller set of themes. In the partner survey, respondents will report on their three main goals for the EBHV grant initiative. We will compile responses from surveys for each grantee and code similar responses together.

Conduct Within- and Cross-Grantee Analysis

After all qualitative data from site visits have been coded, we will conduct searches to retrieve data on the research questions and subtopics to facilitate theme identification. Data can be retrieved on particular codes across all data sources, from individual informants or categories or data sources, or for specific waves of data collection. We will also be able to retrieve data on particular questions to compare across multiple waves of data collection and data sources. Researchers can use the database to assess the consistency and quality of information across sources and informants.

We will use a within-grantee perspective, followed by a cross-grantee perspective, to identify themes and patterns discernable to an individual grantee, a set of grantees, or all grantees. The within-grantee perspective will be helpful in describing how each grantee implemented its home visiting program and its efforts to develop infrastructure to support the implementation, scale-up, and sustainability of EBHV services. Through the cross-grantee perspective, we will explore relationships across themes—for example, the kinds of implementation challenges grantees faced or similarities in their staffing patterns and partnership arrangements. The Mathematica-Chapin Hall team will use these findings to create grantee-specific case studies and a cross-grantee analysis of the patterns of grantee activities, system attributes, and changes in infrastructure capacity over time.

To facilitate analysis of common themes and patterns across subgroups of grantees, we will also code qualitative site visit data according to selected characteristics for the specific grantee. We will create these codes based on information obtained during the site visit interviews and from other data sources, such as implementation plans or the web-based system. For example, we may want to group sites according to their grantee agency type, selected home visiting model(s), whether they are newly implementing or expanding implementation of their selected home visiting model(s), the length of time that the grantee has been enrolling families for home visiting services, or their overall grantee goals. Likewise, we may want to group sites according to the populations of children and families they serve, such as first-time mothers, risk factors, or age of target child. Creating these

subgroups will enable us to compare, for example, how implementation varied (1) across grantees implementing different home visiting models, (2) among grantees newly implementing a home visiting model, or (3) among grantees not offering direct services.

Quantitative Analysis

The cross-site quantitative data analysis will draw on a variety of data sources and use a number of analytic techniques. The quantitative data will be drawn from the web-based data system, the data provided by the National Service Office of the Nurse-Family Partnership, and surveys of grantees and their partners. To analyze the quantitative data, we will use descriptive statistics and multivariate models to examine the measures of fidelity of implementation, systems, and program costs at a point in time, as well as over time. Before describing our analysis, we address three key data issues that will affect our analytic approach: (1) data collection frequency, (2) the unit of data collection, and (3) the unit of analysis.

Frequency of Data Collection. The systems, fidelity of implementation, and program cost data will be collected at different frequencies, as described in Chapter VII and summarized in Table VIII.2. Much of the fidelity data will be collected monthly, the systems data will be collected biannually, and data covering annual program costs will be collected once. Because these measures will be collected at varying frequencies, we plan to aggregate the fidelity and systems data to a consistent time frame, such as biannually, to facilitate the analyses. Thus, the systems data will provide a “snapshot” of grantees’ systems change activities, infrastructure development goals, and infrastructure capacity over six months. The fidelity of implementation data will provide information about how faithful the delivered services were to the EBHV grantee-selected program model over six months.

Unit of Data Collection. The measures of systems, fidelity of implementation, and program costs will be collected for different units. As described in Chapter I, the 17 grantees are each working with varying numbers of service delivery locations. We will collect systems data from the 17 grantees. We will collect fidelity and cost data from the approximately 60 service delivery locations that are part of the grantees’ initiatives. Approximately 40 of the 60 total service delivery locations will be implemented in the first year of the grant initiative, and we estimate 20 more locations will be added over the course of the grant initiative. Most grantees will provide data for all service delivery locations in their grant initiative because they are working with only a few locations. Illinois and New Jersey, however, are working with multiple service delivery locations; therefore, they will provide fidelity and cost data for a sample of about six to eight locations.

Unit of Analysis. The grantee is the unit of analysis for the systems change data. Corresponding to the data collection units, the cost and the fidelity data will be presented at the service delivery location. For the fidelity data, as described in Chapter III, this will require us to summarize the fidelity data collected at the participant and home visitor levels at the location level.

We will explore the use of participant- and home visitor-level data in our statistical analyses, but for ease of presentation, we discuss fidelity data at the location level below. Because the program fidelity and cost data are nested within systems, we will use hierarchical linear modeling (HLM) for the cross-domain analyses, which is described in more detail below and in Appendix C.

Table VIII.2. Frequency and Unit of Data Collection and Unit of Analysis by Domain

	Systems Change	Fidelity	Program Costs	Family and Child Outcomes
Frequency of data collection	Biannually	Primarily monthly ^a	Once	Once
Unit of data collection	Grantee (17)	Service delivery locations (60)	Service delivery location	Grantee
Unit of analysis	Grantee	Service delivery location ^b	Service delivery location	Grantee reports on local evaluations ^c

^aSome of the fidelity data will be collected for every home visit; however, these data will be submitted monthly to the national cross-site evaluation team. We plan to aggregate the fidelity data to biannual time units; however, we will explore different time units, such as monthly indicators, in the fidelity analyses.

^bThe fidelity data for the service delivery location level will be aggregated from home visitor- and participant-level data. We will explore analyses at the home visitor and participant levels.

^cGrantee local evaluation reports include data collected from parents and children. Local evaluation reports will include one or more separate evaluations of their home visiting program model(s).

Analytic Approach

The quantitative data analysis plan is presented in two sections:

1. The within-domain analysis section describes systems change, fidelity of implementation to selected program model(s), and program costs. The within-domain analysis section includes a description of the network analysis, which will document grantee collaborations to understand the systems in which grantees are working.
2. The cross-domain section describes the relationship between systems change, fidelity, and program costs, as well as the relationship between scale-up and sustainability of the home visiting programs and these three domains.

As discussed in Chapter V, we will assess the impact of the EBHV grantee-selected models on family and child outcomes through a systematic review of evidence, described in more detail below. Finally, we plan to examine the overall effectiveness of the intervention to change systems to support the grantees. We will assess, across the home visiting programs, the combination of the impact of each program and the reach of that program to describe the broader effects of the program on families and children in the grantees' service areas.²²

Within-Domain Analyses

The data collected on systems change, fidelity, and costs will be used to describe grantees' progress within each domain during the grant initiative. As noted in Chapter I, the primary research questions for each domain include:

- ***Systems Change:*** How did grantees build infrastructure capacity to implement with fidelity, scale up, and sustain the home visiting programs?
- ***Fidelity to the Evidence-Based Model:*** Were the home visiting programs implemented and delivered with fidelity?
- ***Costs of Home Visiting Programs:*** How much does the delivery and support of each home visiting program cost?
- ***Family and Child Outcomes:*** Do home visiting programs improve family and child outcomes when programs are implemented in the “real world” and supported by investments in infrastructure?
- ***Process Study:*** How did grantees plan and implement their grant initiative activities?

The quantitative analyses for the three domains will consist of summarizing the data on each domain at points in time. For systems and fidelity data, which will be collected at multiple time points during the initiative, we will also examine change over time within grantees for systems and within locations for fidelity. Finally, we will use multivariate analysis to examine differences in the trajectories of fidelity by grantee and by subgroups, such as home visiting program model, and trajectories of systems by subgroups, such as type of grantee auspice (for example, a state agency or a nonprofit organization). As described in this chapter in the section on qualitative analysis, we plan to code the qualitative data on systems and fidelity and develop quantitative indicators from it. This will allow us to use these indicators in the quantitative analyses.

²² Reach is defined as the proportion of eligible families in the target area who are served by the programs (see Chapter II).

Describe Indicators at Points in Time. We will summarize data from the cost and fidelity measures for each service delivery location and the systems data for each grantee at points in time, such as for the prior six months, using basic descriptive statistics, such as means, proportions, and standard deviations. For fidelity of implementation and program costs, we will examine the indicators at the location level for key subgroups, such as grantee or home visiting program model. We will also examine the systems data, including systems change activities, infrastructure development goals, and infrastructure capacity, for key grantee subgroups, such as by type of grantee or primary home visiting model supported.

In addition to describing each measure, we will use exploratory methods to identify the key indicators that best summarize the findings in each domain. For example, we will use correlational analyses to identify indicators that are highly correlated with multiple other indicators to identify those that are central to assessing each domain. We will also use factor analyses to identify indicators that can be combined to create parsimonious scales with acceptable internal consistency reliability. Reducing the number of indicators to scales and key indicators will enable us to focus our descriptive and multivariate analyses on a subset of the most important outcomes.

Trajectories. Using key indicators and scales, we will examine change over time in fidelity (within each service location) and in systems (within each grantee). We will identify common trajectories across locations, such as improving fidelity, and across grantees, such as expanding infrastructure capacity. We will examine fidelity and systems trajectories by key subgroups, to describe and explain differences across trajectories. For example, we will examine fidelity trajectories for particular home visiting models.²³ To do this, we will analyze the “snapshots” of fidelity, summarized to the location level, at multiple points in time to determine whether fidelity is increasing or decreasing at that location.²⁴ We will group locations by their fidelity trajectories, such

²³ As described in Chapter III, we purposively selected fidelity measures that apply across multiple program models, in order to enable us to examine fidelity trajectories across different models. We plan to measure fidelity as both the fidelity indicator itself (such as, the actual number of home visits), as well as the indicator relative to the program model requirements (such as, whether the clients received the number of home visits recommended by program developers). As described in Chapter VII, we plan to collect the fidelity data in the web-based data system for all home visiting programs, except NFP, making the data collection similar across the programs. We anticipate that because of differences in program requirements, for some programs it will be easier to reach fidelity than for other programs. We plan to analyze and report these differences across program models.

²⁴ As explained in this chapter, for ease of presentation, we discuss fidelity at the location level; however, we will explore the use of fidelity measures at the home visitor or participant level for constructing trajectories. The home visitor

as by improving or declining fidelity, to identify which factors, such as home visiting program model, are associated with trajectories.

For the systems analysis, we will group grantees by trajectories of systems indicators. For example, we will track whether infrastructure capacities expand a great deal or only a little over the course of the grant initiative. Using the same method as employed for fidelity, we will examine similarities and differences among the grantees with similar trajectories. The small number of grantees will be an obstacle to finding statistically significant differences in these analyses; however, we can use these analyses to uncover common patterns of systems change.

Network Analysis. The partner survey contains a set of network questions in which respondents are asked to report on their relationships—for example, frequency and type of communication—with all other respondents for that particular grantee. Completed surveys from partners for each grantee will be analyzed together. In other words, we will conduct 17 sets of network analyses—one per grantee—rather than analyzing the data across grantees.

We will use these data to create a square sociomatrix for each grantee—a tabular representation of relationships among responding organizations where the number of rows and columns equals the number of organizations in the system—for each grantee. The sociomatrix will also be displayed as a diagram (sociogram), where the responding organizations are displayed as nodes, and relationships between organizations are portrayed as lines or arcs between the nodes. (See Appendix B for examples of sociomatrices and sociograms.) The strength of the collaboration can be indicated by the size of the line, with stronger collaborations shown by thicker lines. In addition, placement of circles has significance in the sociogram, as circles that are close together collaborate in similar ways in the network. The sociomatrix and sociogram will be used to describe the size of the given network, and can be used to identify organizations isolated from the collaboration. To explore changes in the pattern and structure of collaboration over time, the cross-site evaluation team will compare the size of the network and the proportion of isolated organizations across the three data collection points.

(continued)

or participant level trajectories would be created in the same way as the trajectories at the location level. We will examine multiple “point-in-time” measures, or “snapshots” of fidelity measured within each service delivery location.

The evaluation team will also calculate the density of the matrix (the proportion of existing collaborative ties relative to all possible collaborative ties) for each grantee. This measure can be examined at multiple time points to establish if, and how, the amount of communication among a grantee's partners changes over time.

We will also analyze the network data to describe the attributes of grantee partners and relationships among grantees' partners within and across infrastructure levels. To do this, we will complete the network analyses described above by subgroups of respondents at each infrastructure level. This analysis will enable us to depict communication patterns both within and across levels and identify potential breaks in communication. For example, perhaps partners at the community level are in frequent communication, but they have little or no communication with partners at the state level. This information about communication patterns within and across levels may be helpful to grantees as they seek to address implementation challenges and improve implementation.

We will also calculate means, standard deviations, and minimum and maximum scores for all continuous variables in the partner survey (such as number of employees, budget, years of experience) to describe the characteristics of the organizations within the grantees' network. Frequencies will be reported for all categorical variables (for example, the infrastructure levels at which the respondent organization works). We will combine selected collaboration items into scales and analyze them as a series of continuous variables for each collaboration construct. By using confirmatory factor analyses, we will establish the multiple dimensions of collaboration. For published measures, we will calculate scores according to the author's specifications (if available), and using confirmatory factor analysis. Finally, we will perform traditional descriptive analyses for the remaining, non-scale-based collaboration survey measures (for example, length of time participating in grantee projects).

Explaining Change Within Domains. The Mathematica-Chapin Hall team will use multivariate methods, such as Ordinary Least Squares (OLS) or logistic regression analysis, to document within-domain changes in grantees and locations with a range of characteristics. We will examine the relationship between subgroup characteristics, such as grantee type or home visiting program model (controlling for contextual factors, such as geographic location), to explore why fidelity might differ across service delivery location and why systems change might differ across grantees. These analyses will be exploratory, and we cannot draw causal conclusions from them. In addition to incorporating qualitative data into the analyses, by coding it into quantitative indicators, we will use the qualitative data to inform this explanatory analysis. Both the quantitative and

qualitative analyses will provide explanations for levels of fidelity of implementation or change in systems. The quantitative results will present a “story” of improvement. For example, the quantitative data might show that locations in states with prior experience using a home visiting model achieve fidelity more quickly than locations without such experience. Similarly, the qualitative interview data will be analyzed to identify themes that explain the facilitators and obstacles to achieving rapid fidelity of implementation, primarily based on the perceptions of staff who deliver the services. In an exploratory analysis, we will contrast the explanations for achieving rapid fidelity of implementation based on the quantitative and qualitative analysis results to identify whether they are similar, complementary, or contrasting. This mixed-method approach will provide a fuller explanation of the change within domains.

Cross-Domain Analyses

The associations among the systems, fidelity of implementation, and program cost domains are a central element of this evaluation. As noted in Chapter I, the cross-domain research questions, which will be analyzed using quantitative data and methods, include:

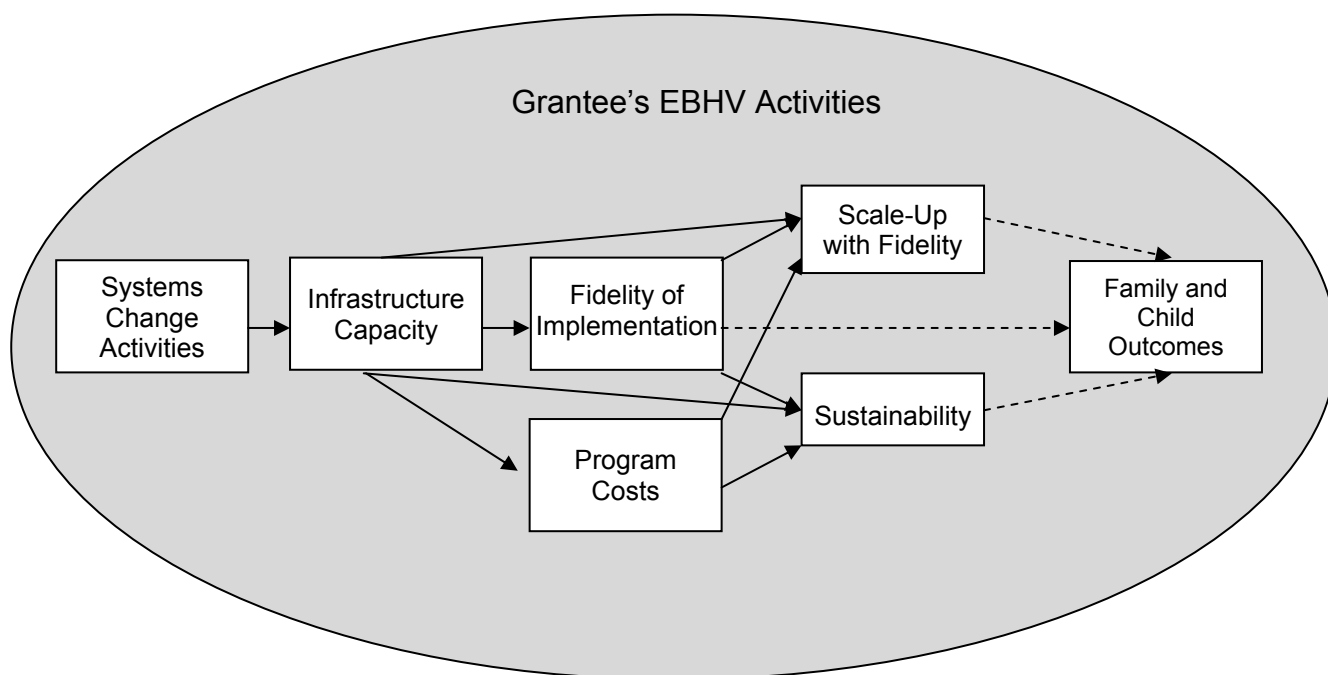
- Are systems and changes in those systems related to the fidelity of implementation? What is the nature of this relationship?
- How are systems, program costs, and fidelity of implementation related to the scale-up and sustainability of home visiting programs?
- Are systems change activities and improvement in infrastructure capacity to support the implementation of home visiting programs with fidelity, scale-up, and sustainability related to positive family and child outcomes?

The cross-domain quantitative analyses will examine the relationship between infrastructure changes resulting from systems activities; program characteristics, including costs; and fidelity of implementation, accounting for differences in relevant grantee and program characteristics. Furthermore, we will examine the relationship between systems change, program costs, and fidelity with two key outcomes: sustainability and scale-up of home visiting programs. Analysis regarding the third cross-domain research questions on improving infrastructure capacity and how that may impact family and child outcomes, is discussed separately at the end of this chapter.

Figure VIII.1 shows the relationships between the domains. The cross-site evaluation will analyze the relationships between the domains, as depicted by the arrows. The solid arrows represent relationships we will analyze directly, while the dashed arrows represent relationships we will assess indirectly. Our analysis of the relationship between systems change and family and child

outcomes (that is, the dashed arrows), is described in the last section of this chapter. To study changes in fidelity, scale-up, and sustainability associated with systems activities and program characteristics, we will use multilevel HLM because the measures of fidelity, at the level of service delivery location, are nested within the measures of systems, at the grantee level. HLM is a powerful tool for analyzing nested data because it can handle nested data in its estimation procedures, and it allows the relationships between fidelity and systems change to differ across grantees. Appendix C provides technical details about the analytic models.

Figure VIII.1 Cross-Domain Relationships Examined in the Quantitative Analyses



To examine how observed improvements in fidelity are associated with the systems change activities, infrastructure capacity, and program cost measures, we will enter the key indicators and scales for each domain into the analytic model, and control for (1) relevant grantee characteristics, such as grantee type and geographic area; and (2) service delivery location characteristics, such as home visiting model or time since implementation. One set of HLM models will include the key fidelity indicators as the dependent variables. The other two sets of HLM models will include fidelity measures as explanatory variables, in addition to systems change activities, infrastructure capacity, and program costs, and will include indicators of scale-up and sustainability as the dependent variables. We will conduct the cross-domain analysis midway through the initiative and at the end of the initiative. We will also perform the cross-domain analyses using subgroups of similar grantees

(for example, those supporting the implementation of the same EBHV home visiting model or those at the same stage of implementation) to determine whether cross-domain relationships differ for key subgroups of grantees.

Systematic Review of Evidence

The goal of the systematic review will be to assess whether the EBHV grantee-selected program models affect the outcomes of families and children. The systematic review of evidence will examine the impacts of the home visiting programs, based on grantees' local evaluations. The cross-site evaluation team will follow three basic steps in the systematic review of evidence: (1) reviewing grantees' evaluation designs, (2) standardizing effects across grantees, and (3) presenting the findings systematically and clearly.

Reviewing Grantees' Evaluation Designs. When the grantees' family and child outcomes evaluations are complete, the cross-site evaluation team will review the evaluation designs to determine the strength of their 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 (RCTs). 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 or outcomes-only studies, which include no comparison groups, and studies with only outcome measures that are not reliable or valid. Exploratory studies can provide information about whether the results are consistent with the study hypotheses, and this provides important information for future research; however, they cannot provide causal evidence about the links between home visiting programs and family and child outcomes. Table VIII.3 shows the minimum requirements necessary to meet the strong and moderate categories of evidence.

Table VIII.3 Overview of Design Elements by Research Design Rigor

Design Elements	Strong	Moderate
Random Assignment, Intent-to-Treat	X	
Random Assignment, Not Intent-to-Treat		X
No Random Assignment, Comparison Group		X
Reliable and Valid Outcome Measures	X	X
Data Collected Comparably Across Treatment and Comparison Groups	X	X

The evaluation designs must be well implemented to meet the requirements of the strong and moderate categories. Study implementation issues that will affect the categorization of the evidence level include sample attrition, particularly differential attrition across the treatment and control groups, which creates bias in the impact estimates, and lack of baseline equivalence of the treatment and comparison groups in quasi-experimental designs. Random assignment designs with high levels of attrition must also demonstrate comparability across the final treatment and comparison samples. As we develop the specific criteria to evaluate these methodological issues, we will look at the latest 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.

Table VIII.4 (first set of columns) shows the designs that the grantees were planning as of September 2009. A challenge to the cross-site evaluation of family and child outcomes is that only nine grantees are planning to use RCTs, which provide the strongest evidence of a causal link between the program and family and child outcomes. Furthermore, for RCTs to be highly rigorous, they must be well implemented. For example, if the treatment and control groups have very different rates of attrition, it is more difficult to draw causal conclusions from an RCT study. Issues, some beyond the control of the grantees, may arise that challenge the execution of well-designed RCTs (for example, funding cuts may reduce the number of families served by grantees). By rating the study design and implementation, the systematic review will draw attention to the results based on the strongest designs, rather than presenting all designs as equally valid in drawing causal conclusions.

A second challenge to the cross-site evaluation is the small sample sizes in the grantees' evaluations, which leads to a lack of power to detect small impacts on family and child outcomes. In particular, the ability to detect impacts on incidence or reports of child maltreatment is likely to be

Table VIII.4 Grantees' Proposed Evaluation Designs and Sample Sizes

State	Grantee	Evaluation Design				Sample Size			
		RCT	QED	Pre-Post	Outcomes	200-299	300-399	400-499	500 or more
CA	County of Solano, Department of Health and Social Services		X			X			
CA	Rady Children's Hospital, San Diego				X				X
CO	Colorado Judicial Department	X				X			
DE	Children & Families First				X	X			
HI	Hawaii Department of Health	X				X			
IL	Illinois Department of Human Services			X				X	
MN	Minnesota Department of Health State Treasurer		X			X			
NJ	New Jersey Department of Children and Families	X					X		
NY	The Society for the Protection and Care of Children, Rochester	X							X
OH	St. Vincent Mercy Medical Center	X							X
OK	University of Oklahoma, Health Sciences Center	X					X		
RI	Rhode Island Kids Count	X				X			
SC	Children's Trust Fund of South Carolina		X				X		
TN	Child and Family Tennessee	X				X			
TN	Le Bonheur Community Outreach				X	X			
TX	DePelchin Children's Center	X							X
UT	Utah Department of Health		X						X
Total		9	4	1	3	8	3	1	5

Source: Grantee plans, October 2009.

QED = Quasi-experimental design; RCT = Randomized controlled trial.

low given the low rates of incidents and reports in the population. Table VIII.4 (second set of columns) shows grantees' proposed sample sizes for their evaluations of impacts on families and children. Table VIII.5 shows the minimum detectable effects (MDEs), which are the smallest impacts that the program must produce in the target population for statistical tests to have a reasonable chance of detecting an effect, for a range of sample sizes. As shown in these tables, for a sample size of 500, the MDE for impacts on child maltreatment reports is a drop of 8 percentage points, assuming a 15 percentage point rate of child maltreatment reports in the grantees' target populations. We assumed this level of child maltreatment reports because the target populations identified by grantees are at higher risk of reports of child maltreatment than the general population (see, for example, Duggan et al. 2007 for rates of substantiated and unsubstantiated reports in the child's first year of life). Many grantees are proposing sample sizes smaller than 500 for their local evaluations. Furthermore, the sample sizes identified in the tables are the number of respondents; thus, if grantees have high levels of attrition in their evaluations, it will be more difficult to attain the power needed to detect effects.

Table VIII.5 Examples of Minimum Detectable Effects (MDEs) for Child Maltreatment and Maternal Depression Measures

Sample Size (at Followup)	MDEs for Child Maltreatment Reports ^a	MDEs for Maternal Depression Using the CES-D—Assumption 1 ^b	MDEs for Maternal Depression Using the CES-D—Assumption 2 ^b
350	9	.24	.21
500	8	.20	.17
1,000	5	.14	.12
2,000	4	.10	.09
3,500	3	.07	.07

^aThe MDEs for child maltreatment reports, as measured by a combination of substantiated and unsubstantiated reports, are calculated based on the following assumptions: $R^2 = .10$ based on controlling for baseline characteristics, $\alpha = .10$ (two-tailed); no clustering in the sample, substantiated and unsubstantiated child maltreatment reports in target population = 15 percent, and power = 0.80.

^bThe MDEs for maternal depression, as measured by the CES-D, are calculated based on the following assumptions: Std Dev = 7, $\alpha = .10$ (two-tailed), no clustering in the sample, power = 0.80, and controlling for baseline characteristics and baseline measure of CES-D with $R^2 = .2$ for assumption 1 and $R^2 = .4$ for assumption 2.

CES-D = Center for Epidemiologic Studies Depression Scale; MDE = Minimum Detectable Effect; QED = Quasi-Experimental Design; RCT = Randomized Controlled Trial; Std Dev = Standard Deviation.

The cross-site evaluation team has addressed the local evaluation sample size issue in several ways. First, as described above, our systematic review includes parental risks for child maltreatment, such as maternal depression, which occur with more frequency than child maltreatment. Examining more prevalent risk factors increases the likelihood of detecting significant effects. Furthermore, we will recommend to grantees that they control for baseline indicators of the parental risk outcomes in their local evaluations. Controlling for baseline measures increases the precision of statistical tests, thereby improving the power of the local evaluations. Parent risk factors, such as maternal depression, can be controlled for baseline levels, while child maltreatment reports cannot. Finally, we encouraged grantees to use evaluation designs that increase the chances of detecting impacts, such as increasing sample sizes if possible and ensuring that the comparisons they plan represent truly different experiences for families (maximizing the possibility of seeing differences), selecting indicators that have the greatest chance of being affected by the program, and focusing on short-term followup to reduce sample attrition. Regardless of steps taken to address this issue, this challenge could pose a problem if sample sizes are low and changes in outcomes are small.

In examining the quality of the family and child outcomes measures, we will be looking for several key indicators. For example, grantees must have used the vetted, age-appropriate indicators of the key cross-site evaluation constructs. Data must have been collected (and scored) by staff with the proper training, and the amount of missing data must be limited. Moreover, data must be collected in the same way for the clients receiving the home visiting services and for those in the comparison group.

Standardizing Effects. For the cross-site evaluation, the Mathematica-Chapin Hall team will work with grantees to calculate their estimates of effects consistently, so that comparable estimates can be presented across grantees. To calculate estimates of effects that are consistent across grantees, the Mathematica-Chapin Hall team will work with grantees to convert their estimates of home visiting impacts to effect sizes to provide consistent measures of effects across differing scales.

Presenting the Findings. After the grantee evaluation results are categorized by level of evidence and their estimates of effects have been calculated in a standardized way, the results will be presented clearly and systematically across the grantees. The goal of the presentation will be to describe the level of evidence of the effects of home visiting programs about family and child outcomes in a way that is straightforward and useful to CB/ACF, the grantees, and other key stakeholders. For example, reports might include tables that provide an overview of (1) the level of

the evidence, (2) the direction (positive or negative) of the effects, and (3) the size of the effects. The reports will also include text that describes how decisions were reached about categorizing the grantees' evaluation designs, as well as providing guidance for readers about how to interpret the level of evidence and the size of the effects.

Systems Change and Family and Child Outcomes

The grant initiative focuses on building infrastructure to support home visiting programs, with the ultimate goal of preventing child maltreatment in the grantees' communities. Thus, the cross-site evaluation of the systems change, fidelity of implementation, and costs will complement the systematic review of evidence, which focuses on the impacts of home visiting programs for families who receive program services. To conduct a global assessment of the potential for the reduction of child maltreatment within grantees' communities, we will examine the results of systems change activities intended to increase scale-up and sustainability of home visiting programs, alongside the program impacts. Based on the work of Abrams et al. (1996), we will examine two key measures from the cross-site evaluation results: (1) measures of the size and significance of the impacts of the EBHV grantee-selected program models on family and child outcomes; and (2) measures of reach of the home visiting program models, defined as the proportion of eligible families who receive services. We will present the measures together and explore methods for combining them.

Effectiveness of EBHV Grantee-Selected Program Models. The measures of effectiveness will include the effect sizes for each family and child outcome grantees include in their local evaluations, which will be calculated for the systematic review of evidence. Whether the evidence is strong, moderate, or exploratory, we plan to include all measures of effectiveness in this analysis; however, we will provide information about the rigor of the evidence when presenting the results of this analysis.

Reach. To calculate the indicator of reach, we will use the information gathered through the scale-up measures in the systems domain, discussed in Chapter II, regarding the number of families referred and enrolled for services and the definitions of the target population. Reach is typically defined as the ratio of (1) clients served to (2) the size of the target population (Abrams et al. 1996). We will work with CB/ACF and grantees to explore different measures of reach to identify the most accurate definition for each grantee, as well as comparable definitions across grantees. For example, the number of clients served could be the capacity of the program in one year or the average number of clients who complete the program each year, over the course of the intervention. The

size of the target population could vary by the geographic boundaries and density of population in the target area.

To calculate the overall effectiveness of the program, we will present the effectiveness and reach indicators together. Furthermore, we will explore methods, such as multiplying them together (Abrams et al. 1996), for combining the two indicators into one scale of overall effectiveness.

Taken together, the analyses will provide a wealth of information designed to answer CB/ACF's research questions, inform grantee program management and decision making, and contribute to policy and research in home visiting. The findings will be summarized in a range of publications and presentations as described in the next chapter.

