

**Maternal and Infant Home Visiting
Program Evaluation (MIHOPE)
0970 - 0402**

**Supporting Statement
Part A: Justification**

April 2012; Updated July 2012

Submitted By:
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

7th Floor, West Aerospace Building
370 L'Enfant Promenade, SW
Washington, D.C. 20447

Project Officer:
Lauren Supplee

A1. Circumstances necessitating data collection

The Administration for Children and Families (ACF) and Health Resources and Services Administration (HRSA) within the U.S. Department of Health and Human Services (HHS) have launched the Mother and Infant Home Visiting Program Evaluation (MIHOPE). This evaluation, mandated by the Affordable Care Act (ACA), will provide information about the effectiveness of the Maternal, Infant, and Early Childhood Home Visiting (MIECHV) program in its first few years of operation, and provide information to help states and others develop and strengthen home visiting programs in the future. It will attempt to fill gaps in research that were identified in recent reviews of home visiting programs funded by the U.S. Department of Health and Human Services through the Home Visiting Evidence of Effectiveness (HomVEE) project. The evaluation is being conducted by MDRC in partnership with Mathematica Policy Research, James Bell Associates, Johns Hopkins University, and the University of Georgia.

The proposed evaluation will be conducted in approximately 85 sites across approximately 12 states. In each site, approximately 60 women will be randomly assigned to either MIECHV-funded home visiting or to a control group, which will be given referrals to other services in the community. Women will be eligible for the study if they are pregnant or have an infant under six months old. The goals of the evaluation are (1) to understand the effects of home visiting programs on parent and child outcomes, both overall and for key subgroups of families, (2) to understand how home visiting programs are implemented and how implementation varies across programs, and (3) to understand which features of local home visiting programs are associated with larger or smaller program impacts.

MIHOPE includes two phases. Phase 1 includes site recruitment, recruitment of women and collection of baseline data on their families, and the collection of data on program implementation at baseline and one year later. Phase 2 is expected to include a survey conducted with parents around the time the child is 15 months old and observations of interactions between parents and children. This document provides support for the data collection efforts of Phase 1. The data collection efforts under Phase 2 will be presented in a subsequent package.

A2. Purpose and use of the information collection: How, by whom, and for what purpose the information is to be used.

Phase 1 of the evaluation will include three broad sets of data collection activities:

1. Collect information from state MIECHV administrators to inform the selection of states and sites for the evaluation.
2. Recruit women into the study and collect baseline information on their families.
3. Collect information on the implementation of home visiting programs.

The remainder of this section provides more detail on the three sets of data collection activities included in Phase 1.

Site Recruitment

The overall goal for site recruitment in MIHOPE is to select approximately 85 sites across approximately 12 states. States and their local program sites will be selected for MIHOPE based on a variety of characteristics including the type of home visiting model, geography, urbanicity, target population, and research feasibility. There is currently limited documentation available to aid site selection. The study team reviewed and analyzed the MIECHV implementation plans each state submitted to the U.S. Department of Health and Human Services. These plans provided a general overview, which allowed the team to prioritize 30 states within the continental United States for further consideration for MIHOPE. The plans did not, however, consistently provide the specific answers needed for site selection. For this reason, the study team is contacting states and their local programs to confirm what was collected from the plans and request some additional information. A Federal Register Notice was published on December 22, 2011 and a Supporting Statement was sent to OMB requesting an emergency clearance for site recruitment activities in order to continue activities to meet a legislatively mandated deadline. The emergency clearance was approved on January 26, 2012 (0970-0402)

The study team's plan for contacting states and local programs and using the information that is collected includes the following steps:

Introduce the evaluation to state administering agencies (January 2012).

Regional project officers from HRSA sent an email (Attachment 1) to state administrators overseeing the MIECHV programming to introduce the study and its goals, introduce the team that will be doing the study on HHS's behalf, and alert state administrators that a study team member may be in contact to explore whether their state would be a good fit for the evaluation.

Telephone contact with state administrators to gather information (January-April 2012).

As approved by the emergency clearance, study team staff have begun calling state administrators to schedule longer telephone appointments to collect the information necessary to allow the study team to understand the universe of states using MIECHV funds and proceed to the next stage of site recruitment. The appointment confirmation includes several documents (Attachment 1): (1) a project description, which explains the study, the process for selection and enrollment, the project timeline; (2) a set of frequently asked questions, which responds to potential questions state administrators may have about the study; (3) a site participation overview, to provide states with an understanding of what participation in the study would entail for their local home visiting programs and the process for their involvement; (4) a list of information to be collected on sites from State administrators, and (5) a protocol for the telephone call with State administrators.

During the call, the study team is answering any questions state administrators have regarding the study and asking for a few key characteristics of each MIECHV-supported program site. This is enabling the team to understand the number of local MIECHV programs, using the study's definition of a local program. At this time, a site is defined as a home visiting program with local administration (separate office and supervision), but the study team will use these conversations to try and understand how the definition may vary across states. The information collected will also help the study team classify these sites according to three main characteristics: geographic region, program model, and urbanicity.

The team will use this information to select approximately 18 high priority states that best suit the evaluation needs.

In-person visits and teleconferences to key states and sites for detailed discussion about the evaluation (March-December 2012).

To recruit and reach agreement with approximately 12 states and 85 local program sites from among the high priority states, the study will visit a state up to three times. Site recruitment staff, working in teams of two, will meet in-person and by phone to discuss the evaluation with state and local program site staff. These visits and telephone calls will be used to collect information to determine which pool of states and sites best meet the criteria for site selection. After each visit, the study team is expected to narrow the pool of eligible states and sites based on the information collected.

A first round of visits will be made to 18 states. An agenda will be used to guide the discussion. The study team will introduce sites to the study using a PowerPoint presentation. Using semi-structured protocols, conversations with state staff will be designed to gain an understanding of the processes for accessing state administrative records and to underscore the state administrators' importance in helping to recruit sites. Important questions concern sites' administrative structures, programmatic experience, when they plan to begin MIECHV services, the community service context, and program size. Initial visits may include groups of sites, but the study team would eventually need to meet with each site individually (although not always in-person) to understand their program flow, respond to questions and concerns, and discuss the terms of an agreement. Materials related to the first round of visits with state administrators are shown in Attachment 2, including the agenda, PowerPoint slides, and protocol.

After the first round of visits, the study team will narrow the pool of eligible states and will schedule follow-up visits to 12-15 states and teleconferences and visits with roughly 120 sites to insure that we will a pool from which to choose approximately 85. Using semi-structured protocols, the goals of the follow-up conversations with state staff are to gain an understanding of the processes for accessing state administrative records, and to underscore the state administrators' importance in helping us to recruit sites. An agenda for the follow-up visits to states and topics for discussion are shown in Attachment 3.

Important questions in our discussions with program sites are about their administrative structures, their programmatic experience, when they plan to begin MIECHV services, the community service context, and their program size. Meetings would be scheduled with each site individually (although not always in-person) to understand their program flow, respond to questions and concerns, and discuss the terms of an agreement. Materials to be used in conversations with local program directors are provided in Attachment 4.

Family Baseline Survey

Baseline information on families will be used by the study team to answer the following research questions within Phase 1:

- What are the characteristics of families that participate in the study?

- How do the families vary by site?
- To what extent does the national evaluation include members of at-risk groups that are mentioned in the ACA as high priority?

Baseline family data will also be used by the study team in the analysis conducted during Phase 2 of the evaluation for two purposes: (1) to define subgroups of families for which home visiting services might have differed and for which the impact of home visiting might differ, and (2) to increase the statistical precision of estimated effects on follow-up outcomes.

Family baseline data will come from three sources: (1) a one-hour family baseline survey conducted by telephone by evaluation staff, (2) state administrative data from birth and child welfare records, and (3) the Home Observation for Measuring the Environment (HOME) to assess home conditions and parenting practices.

The family baseline survey (Attachment 6) will include information on several domains specified in the ACA: newborn health; parental health and well-being; parenting practices, attitudes, and beliefs; domestic violence; history with the criminal justice system; family economic self-sufficiency; and referral and coordination of social services. In addition, the baseline survey will collect information on demographics and household composition to describe the study sample, and will collect contact information for family members or friends who can help locate the family at follow-up if they move. The survey also contains information about the parent's expectations regarding the home visiting program, which will inform research on program implementation.

Table A.1 lists the constructs that will be collected in these various domains, the research questions they will be used to answer, and whether they will be collected during Phase 1 or 2. In general, measures were chosen for one or more of the following reasons:

- Prior research or theory indicates the effects of home visiting vary across subgroups defined by these measures, or that home visiting services are expected to depend on these family characteristics. For example, prior research has found that home visiting programs have different effects for mothers suffering from depression, and that mother's interaction with home visitors will vary with their attachment style.
- They are needed to identify key subgroups of families identified in the ACA, such as low-income pregnant women under age 21, those with a history of substance abuse, families with tobacco users in the home, and families with a parent who serves or has served in the Armed Forces.
- Prior research indicates they will be important predictors of child and family outcomes. Having strong predictors will increase the statistical precision of estimated effects on those outcomes, making it easier to determine whether home visiting programs have had an effect. For families with infants at baseline, questions on birth weight, gestational age, and so on are strong predictors of future health and development as well as family well-being and health system costs. In most cases, the study team chose measures that could be asked of both pregnant women and mothers with infants. For example, questions on parenting beliefs have been found to predict harsh parenting, which can have negative consequences for child development.

Although follow-up data collection is expected to include direct assessments of children (such as the Three/Two Boxes/Bags), the project does not plan to directly assess children at baseline. This decision was made for two reasons. First, a substantial number of women will be enrolled while they are pregnant, limiting the usefulness of baseline child assessments. Second, the study team determined that existing assessments are inappropriate for MIHOPE because of the age of children at baseline, because they are not available in Spanish, or because they require clinicians to administer them. Appendix A summarizes the team's investigation into a number of direct assessments. Although direct assessments are not planned, a number of self-reported measures such as parenting beliefs have been found to predict later child outcomes and have been included in the survey.

Details are not being provided in this statement on administrative data because they would not be collected until Phase 2 of the evaluation (although they would be collected for both the baseline and follow-up period at that time). Details are not being provided for the HOME assessment which does not represent a burden to families since it (1) does not require the family to provide any information, and (2) will be conducted at the same time as the baseline survey. This is consistent with 44 USC, 5 CFR Ch. 11 (1-1-99 Edition), 1320.3, which indicates that "information" does not generally include facts or opinions obtained through direct observation by an employee or agent of the sponsoring agency or through nonstandardized oral communication in connection with such direct observations.

Table A.1.
Baseline Data Collection: Purpose of Family Baseline Data

Domain	Research questions	Constructs	Covariate in impact analysis	Subgroups	Predictor of service delivery and family engagement	Data collection timeline
Child health and development	<p>What characteristics of families participating in MIHOPE are related to child health and development?</p> <p>What are the effects of HV services on child health and development?</p> <p>How does birth weight mediate the effects of HV services on families?</p> <p>How do child health and development characteristics such as birth weight influence the level of service delivery or moderate family engagement?</p>	<p>Birth weight; regular source of care; temperament; NICU visits</p>	X	Birth weight	Birth weight	Phase 1, Phase 2
Parenting	<p>What characteristics of families participating in MIHOPE are related to parenting?</p> <p>What are the effects of HV services on parenting?</p> <p>How do breastfeeding, paternity and father involvement mediate the effects of HV services on families?</p> <p>How do parenting characteristics such as breastfeeding, paternity and father involvement influence the level of service delivery or moderate family engagement?</p>	<p>Nutrition; breastfeeding; parenting attitudes and beliefs; relationship happiness; paternity; father involvement</p>	X	Breastfeeding; paternity; father involvement		Phase 1, Phase 2
Parent health and well-being	<p>What characteristics of families participating in MIHOPE are related to parent health and well-being?</p> <p>What are the effects of HV services on parent health and well-being?</p> <p>How do parents' general health, depression and substance use mediate the effects of HV services on families?</p> <p>How do parenting health and well-being characteristics such as general health; physical functioning; depression; anxiety and attachment style influence the level of service delivery or moderate family engagement?</p>	<p>General health; physical functioning; height and weight; depression; anxiety attachment style; substance, tobacco and alcohol use; desired timing of subsequent births</p>	X	General health; depression; substance use	General health; physical functioning; depression; anxiety; attachment style; substance and alcohol use	Phase 1, Phase 2
Intimate partner violence	<p>What characteristics of families participating in MIHOPE are related to intimate partner violence?</p> <p>What are the effects of services on intimate partner violence?</p> <p>How do intimate partner violence and psychological emotional abuse influence service delivery or moderate family engagement?</p>	<p>Intimate partner violence; psychological and emotional abuse</p>	X		Intimate partner violence; psychological and emotional abuse	Phase 1, Phase 2

Table A.1. (continued)

Domain	Research questions	Constructs	Covariate in impact analysis	Subgroups	Predictor of service delivery and family engagement	Data collection timeline
Crime	What characteristics of families participating in MIHOPE are related to crime? What are the effects of HV services on crime?	Arrests; convictions; incarceration	X			Phase 1, Phase 2
Family self-sufficiency	What characteristics of families participating in MIHOPE are related to family self-sufficiency? What are the effects of HV services on family sufficiency? How do household income, education and employment mediate the effects of HV services on families? How do earnings and education influence the level of services delivery and family engagement?	Income sources; annual household income; employment and employment history; earnings; education; health insurance	X	Household income; education; employment	Earnings; education	Phase 1, Phase 2
Demographics	What characteristics of families participating in MIHOPE are related to demographics? What are the effects of HV services on demographics? How do race, ethnicity, acculturation and language mediate the effects of HV services on families?	Household composition; mobility; language; acculturation; number of children by partner; age; race; ethnicity; sex; language; citizenship	X	Race; ethnicity; acculturation; language		Phase 1, Phase 2
Other services	What characteristics of families participating in MIHOPE are related to other services? What are the effects of HV services on receipt of other services? How does the degree of engagement mediate the effects of HV services on families? How do families' expectations about HV influence the level of service delivery and family engagement?	Maternal health care; enrollment in home visiting; expectations about HV; social services		Degree of engagement in services	Expectations about HV	Phase 1, Phase 2

Data on Program Implementation

Data on program implementation will be used by the study team to address the following research questions:

- What organizations are involved as stakeholders in the local home visiting programs, how are service models defined, who provides services, and how are families referred for home visiting services?
- What are the characteristics of the neighborhoods where families receiving MIECHV services live?
- How do the funded home visiting programs actually operate?
- What is the community context in which home visiting programs operate?
- How are a site's service model and implementation system related to the characteristics of its staff?

The legislation requires examining how impacts vary across programs. Phase 2 of MIHOPE would address this requirement by exploring how features of local home visiting programs are related to impacts of those programs on parent, child, and family outcomes. In particular, variation in estimated effects across the 85 sites will be compared to variation in family characteristics, characteristics of the community (such as the availability of health and social services), features of service models (such as the frequency of home visits), features of the implementation system (such as the training and supervision of home visitors), and the content of home visits. The overarching research question underlying this analysis would be the following:

- What is the relationship between the features of home visiting programs and their effects on family outcomes?

Data related to program implementation will be gathered through six broad categories of activity: 1) structured interviews with mothers at baseline (through the family baseline survey); 2) semi-structured interviews with state MIECHV administrators; 3) surveys of the staff at home visiting program evaluation sites; 4) surveys of administrators of other programs serving the same communities as those served by evaluation sites; 5) logs maintained by supervisors and home visitors; and 6) semi-structured group and individual interviews with home visiting program staff at evaluation sites.

In addition to these data collection activities, research staff will video record the interaction between the home visitor and family on two occasions for 30 percent of families in the study. Because video recording will document the normal activities in a home visit, it will not impose additional burden on families or home visiting staff. Therefore, this supporting statement does not describe this data collection activity in detail.

These data will be used to measure a) characteristics of participating families at the time of enrollment into the study, b) characteristics of program staff, c) organizational factors for service delivery, d) actual service delivery and fidelity, and e) program costs. Table A.2 summarizes the constructs measured and the research questions to be addressed.

Each broad category of data collection is described briefly below.

1) Family baseline survey. The family baseline survey was described in detail earlier in this section. Table A.1 identified the domains with particular relevance for the implementation research. These domains include baseline measures of child health, parental health and well-being, domestic violence, demographics, and expectations regarding home visiting. These domains are relevant for the implementation research because they are likely to be related to actual engagement in services. The baseline interview will take about one hour to complete.

2) Semi-structured interviews with State MIECHV administrators. Semi-structured interviews with State MIECHV administrators will be conducted twice. A baseline interview (Attachment 7) will be carried out when the state enters the evaluation and will include questions about the background of state MIECHV plans, goals, target groups, and service models. Administrators will be asked how they plan to use data that the ACA requires them to collect on MIECHV benchmarks, their current and planned activities for state-level continuous quality improvement, and their current state-level MIS system for MIECHV programs. They will also be asked to describe planned enhancements to the national models being adopted and their rationale. The interview will take about two hours to complete.

The second interview with state administrators will be carried out 12 months later. The content of the second interview (Attachment 8) will parallel that of the baseline interview. The purpose of the second interview will be to elicit the administrators' perception of how well state MIECHV plans have been implemented, reasons for departures from what was planned, and whether/how state MIECHV plans for the near future have changed since what was described at baseline. The second interview will take about two hours to complete.

3) Surveys of staff at participating home visiting program sites. The evaluation will include structured, web-based surveys of the program manager, supervisors, and home visitors at each home visiting program site. Each staff member will be surveyed near the time that family enrollment into the study begins at his or her site (baseline) and again 12 months later. For each survey, an attachment includes the instrument.

Program Manager Surveys

Program manager survey at baseline. The baseline surveys of program managers will be carried out in three parts around the time that family enrollment into the study begins. Part 1 (Attachment 9) will take place four weeks before family enrollment begins and will be done in coordination with data collection by the state's MIHOPE site liaison. Part 2 (Attachment 10) will also take place about four weeks before family enrollment into the study begins at each site and will be carried out via web-based survey. Part 3 (Attachment 11) will take place four weeks after enrollment begins and will also be a web-based survey. Each part of the program manager survey has a unique set of questions.

Program manager survey, baseline, part 1 asks respondents to complete an inventory of the program's policies, procedures, and forms used to guide the program's work. Respondents will be asked whether there is a policy or not, and whether it has been put in

place in the last 12 months. The site liaison will work with each program manager or his/her assistant to gather copies of the policies, procedures and forms. Part 1 will take about 0.50 hours to complete.

Program manager survey, baseline, part 2 asks respondents to describe their service model and implementation system, to identify organizations that have influenced the service model, and to indicate how these organizations have influence the service model. The service model includes the program site's intended goals and outcomes, its target population, its intended services, and intended staffing (roles, responsibilities, and required competencies). The implementation system includes the resources, policies and procedures that the program site uses to implement the service model. These include resources for: staff development (hiring, training, supervising, evaluating); clinical support of staff (availability of consultants, curricula, protocols); administrative supports for staff (management information system, quality improvement activities); and systems interventions (relationships with other community resources for referral and coordination of services). Part 2 will take about one hour to complete.

Program manager survey, baseline, part 3 asks respondents to name community resources for referrals including: prenatal care; early childhood care and education; early intervention services; pediatric primary care; family planning and reproductive health care; substance use and mental health treatment services; domestic violence shelter; domestic violence counseling; and adult education. Respondents will then be asked to rate service availability, accessibility and coordination with each named community resource. Respondents will also be asked to name and provide contact information for other early childhood home visiting and parenting programs for infants in the communities they serve. Part 3 will take about one hour to complete.

Program manager survey at 12 months. The 12-month survey (Attachment 12) will parallel the baseline survey, but will be shorter because it will focus only on changes since the baseline survey was conducted. The 12-month survey will ask respondents to identify significant changes in the organizations that influence how their site defines and implements its service model. It will also ask respondents to describe changes in the service model and the implementation system. The survey at 12 months will be administered as a single survey that will take about 2 hours to complete. It will be designed so that a program manager can complete it in more than one session if that suits his or her schedule.

Supervisor Surveys

Supervisor survey at baseline. The baseline survey of supervisors (Attachment 13) will be carried out about four weeks before family enrollment into the study begins at the site. It will collect information on supervisor characteristics and on supervisor perceptions of organizational factors related to service delivery. The survey will include items about employment, supervision and program outcomes. It will also include questions about supervisors' beliefs about home visitor roles and responsibilities, ratings of her or his own training and skills in supervising staff to carry out activities, and ratings of her or his own ability to secure supervision and professional consultation. Supervisors will be asked to complete standardized measures including the Organizational Social Context scale (OSC), Attachment Style Questionnaire (ASQ), and Center

for Epidemiologic Studies Depression Scale (CES-D). The survey will also include items about demographics and individual background characteristics. The survey will take about 1.25 hours to complete.

While the OSC, ASQ, and CES-D all measure aspects of staff well-being, they measure very different constructs, as detailed below. The OSC is the study's only quantitative measure of organizational functioning, which has been found to influence organizational functioning, staff functioning, and service delivery. The OSC measures features of the organization such as bureaucratic rigidity, organizational expectations for staff competency, and cooperation among staff. The ASQ is a widely used self-report measure of adult attachment style, which has been shown to be an important predictor of communication, satisfaction, trust, and relationship functioning between human service workers and their clients in social service settings, including home visiting. The ASQ measures two primary dimensions of adult attachment security: anxiety and avoidance. The CES-D is a ten-item scale designed to measure depressive symptomology in the general population, and measures an important aspect of home visitor well-being that is distinct from what is measured by the ASQ.

There is theoretical and empirical support for the independent and interactive influence of the OSC, ASQ, and CES-D on service delivery and impact. As a result, there is no way to reduce staff burden while obtaining information on all three constructs.

Supervisor survey at 12 months. The supervisor survey at 12 months (Attachment 14) will parallel the baseline survey. It will measure malleable respondent characteristics and perceptions of organizational factors related to service delivery. The survey will take about 1.25 hours to complete.

Home Visitor Surveys

Home visitor survey at baseline. The baseline survey of home visitors (Attachment 15) will be carried out about four weeks before family enrollment into the study begins at the site. Similar to the supervisor survey, the home visitor survey will include items about employment and supervision, program outcomes, program referrals, roles and responsibilities, and knowledge of child development. Home visitors will be asked to complete standardized measures including the Organizational Social Context scale (OSC), Attachment Style Questionnaire (ASQ), and Center for Epidemiologic Studies Depression Scale (CES-D). The survey will also include items about demographics and individual background characteristics. The survey will take about 1.25 hours to complete.

Home visitor survey at 12 months. The home visitor survey at 12 months (Attachment 16) will parallel the baseline survey. It will measure malleable respondent characteristics and perceptions of organizational factors related to service delivery. The survey will take about 1.25 hours to complete.

4) Surveys of Administrators of Community Resources. Web-based surveys will be conducted with administrators of two types of community resources: a) services to which participating

home visiting programs might make referrals relevant to MIECHV benchmarks and participant outcomes; and b) home visiting programs not participating in the evaluation but serving the same community. These surveys will be carried out with administrators of the organizations identified in part 1 of the baseline program manager survey. Information gathered in these surveys will be incorporated into part 3 of the baseline program manager survey.

Survey of administrators of community services relevant to MIECHV benchmarks. The dual purpose of this survey (Attachment 17) is to assess service availability, accessibility and coordination from the viewpoint of the community resource administrator, and to identify additional community resources beyond those that had been identified by the home visiting program manager. The survey will take about 0.10 hours to complete.

Survey of administrators of other home visiting programs. The purpose of this survey (Attachment 18) is to describe the service model, curriculum, cost, and capacity of home visiting programs not participating in the MIECHV evaluation but serving the same communities as programs that are participating in the evaluation. The survey will take about 0.10 hours to complete.

5) Logs Maintained by Supervisors and Home Visitors. Supervisors and home visitors will maintain a weekly, web-based log of their activities.

Supervisor log. The purpose of the supervisor log (Attachment 19) is to describe implementation system activities that are likely to influence service delivery and program outcomes. Each supervisor will record all individual and group supervision activities in the preceding week for each home visitor in his or her caseload who is participating in the evaluation. The supervisor will provide details on both supervision content and techniques. Supervisor will also report any training activities they complete during the week, noting duration, content, teaching modalities, and evaluation strategies. The log will take about 0.20 hours per week to complete. The length of time will depend on the number of home visitors receiving individual supervision and whether the supervisor participated in any training activities.

Home visitor log. Previous home visiting studies have indicated that the level of service is highly variable both across home visiting sites and across families within site (Filene, Bell and Smith, 2011). Weekly logs (Attachment 20) will be used to allow the evaluation to assess variations and patterns of services provided to families and to describe actual service delivery to families and the training and supervisory activities in which the home visitor participated. Each home visitor will document service delivery for each family on his or her caseload who is participating in the study. The home visitor will record details about actual and attempted contacts with the family. The log will be used to track all activities completed with the family, including assessments, referrals, education, and support. Every 3 months, the log will also elicit the home visitor's rating of the quality of his or her working relationship with each family. The home visitor will also report any training activities completed during the week, noting duration, content, teaching modalities, and evaluation strategies. The log will take about 0.20 hours per week to complete. The length of time will depend on the number of families in the home visitor's caseload who are participating in the evaluation, the nature of the home visitor's contact with them, and whether the home visitor participated in any training or supervisory activities.

6) Semi-structured Group and Individual Interviews with Home Visiting Program Staff. The study team will visit sites in each state about 12 months after the state enters the evaluation. As part of the site visit, the study team will conduct semi-structured in-person group interviews with supervisors and one-third of the home visitors of sites participating in the evaluation. The team will conduct semi-structured, in-person individual interviews with another third of the home visitors. Group and individual interviews will take place at the program site. Additionally, program managers across the state will be interviewed together when possible. If in-person interviews are not possible, group interviews will be conducted via phone. All group and individual interviews will take about 1.5 hours to complete. The purpose of these interviews is to better understand the processes underlying the implementation of home visiting programs from the staff by “unpacking” the quantitative data obtained from baseline web-based surveys and other sources such as home visitor logs. The instruments will differ in nature and type from the web-based surveys in that they will allow for subjective, embedded, and elaborated responses.

One-third of the home visitors will be selected randomly to participate in the group interviews, and one-third will be sampled for the individual interviews. The kinds of information home visitors provide will differ between those who are interviewed in a group vs. those interviewed one-on-one. It is necessary to interview a significant sample of home visitors to accurately reflect their experiences as a home visitor, particularly for experiences that are not captured in the web-based surveys (e.g., perceptions of family strengths and needs, program’s participation in MIECHV, etc.). The responses for the individual and group semi-structured interviews may be combined for some qualitative descriptions of how staff describe their roles and the program, with documentation that both individual and group interviews contributed to the conclusions drawn.

Program manager group interviews (Attachment 21). The interview questions will probe on findings from the quantitative data to elicit program managers’ experiences and insights for interpreting the quantitative findings. Topics will include: the role of other influential organizations on their program site; the program’s intended service model (including intended outcomes and priorities, theory of change, targeted families, intended services, staffing); the home visiting implementation system (including professional development, clinical support, administrative support, organizational culture and climate, and coordination and referral systems); and participation in MIHOPE.

Supervisor group interviews (Attachment 22). The interview questions will probe on findings from the quantitative data to elicit supervisors’ experiences and insights for interpreting the quantitative findings. Topics will include: the role of other influential organizations on their program site; the program’s intended service model (including intended outcomes, theory of change, targeted families, intended services, staffing); and the home visiting implementation system (including professional development, clinical support, administrative support, organizational culture and climate, and coordination and referral systems).

Home visitor group interviews (Attachment 23). The interview questions will probe on findings from the quantitative data to elicit home visitors' experiences and insights for interpreting the quantitative findings. Participants will be asked to provide their perspective on outcome priorities when working with families, their program's theory of change, working with families, approaches to carrying out intended services, and professional development. t.

Home visitor individual interviews (Attachment 25). The purpose and content of the home visitor individual interviews is similar to the group interviews, but with less detail on perceptions of organizational factors for service delivery and a greater focus on psychosocial attributes of families and of the home visitors themselves as factors for service delivery. In particular, home visitors will be asked about situations and interactions with families that make it hard for them to carry out their expected role with families.

Table A.2
Implementation Study Instruments

Research Questions Research Measures	Interviews				Surveys	Web-based Surveys				Staff Logs	
	Baseline & 12 Month	12 Month			Baseline	Baseline & 12 Month				Weekly	
	State Admin	Pgm Mgr	Super	Hver	Mothers	Pgm Mgr	Super	Hver	Comm Prov	Super	Hver
What is the intended local service model in each site and how does it differ from the national program model?											
Intended goals, outcomes, dosage, priorities and approaches of national model ^a											
Intended goals, outcomes, dosage, priorities and approaches of local model		X	X	X		X	X	X			
Processes underlying intended goals, outcomes, dosage, priorities, and approaches of the local model		X	X	X							
What services are actually delivered in each site and how do those differ from the intended local model?											
Intended goals, outcomes, dosage, priorities and approaches of local model		X	X	X		X	X	X			
Actual dosage, priorities and approaches of local model		X	X	X		X	X	X		X	X
Processes underlying actual goals, outcomes, dosage, priorities, and approaches of the local model		X	X	X							
What is the strength of the implementation system in each site?											
Role of state MIECHV agency	X	X									
Staff development supports (e.g. recruitment, hiring, training, supervision, evaluation and feedback)		X	X	X		X	X	X		X	X
Facilitative clinical supports (e.g. screening, curricula, peer support, access to prof. consultation)		X	X	X		X	X	X		X	X
Facilitative admin. supports (e.g. MIS, CQI, org. culture and climate)		X	X	X		X	X	X			
System interventions (MOU for referral/coord., point of contact for referral/coord.)						X			X		
Processes underlying the implementation system		X	X	X							

^a The intended goals, outcomes, dosage, priorities, and approaches of the national model will be collected but is not discussed in this document.

Table A.2. (continued)

Research Questions Research Measures	Interviews				Surveys	Web-based S		
	Baseline & 12 Month	12 Month			Baseline	Baseline and 1		
	State Admin	Pgm Mgr	Super	Hver	Mothers	Pgm Mgr	Super	Hv
What are the characteristics of enrolling families that influence service delivery or moderate family engagement?								
Maternal mental health (CES-D)					X			
Maternal attachment style (modified Attachment Style Questionnaire)					X			
Battery or abuse (WEB; CTS)					X			
Supportiveness of family members for home visiting					X			
Reasons for enrolling in home visiting					X			
Expectations regarding home visiting					X			
What are the characteristics of home visitors that influence service delivery or moderate family engagement?								
Demographics, employment, training and experience, supervision								X
Well-being, attachment style, depression, burnout, morale, sense of competence, perceived adequacy of training								X
Reported priority outcomes for sites								X
Capacity to assess, choose program activities, refer and coordinate home visiting services								X
Processes underlying home visitor behaviors and capacities								
What is the community context within which home visiting services are delivered?								
Neighborhood safety and quality (PHDCN observational scales)								
Availability and accessibility of other services in the community						X		
What types of services do control group members use and what is their service context?								
Use of other home visiting services by control group members					X			
Use of other human services (e.g. health, education, public health) by control group members					X			
By what process did states develop their needs assessments and other key program parameters?	X							

A3. Use of information technology for data collection to reduce respondent burden

This study will use information technology, when possible, to minimize respondent burden and to collect data efficiently.

The burden on state administrators and local program directors from site recruitment efforts is minimal. Information available from the internet will be used to supplement requests for information. Meetings will be centralized or conducted by telephone as much as possible to reduce burden on states and their local program directors.

The baseline family interview will be conducted using computer-assisted telephone interviewing (CATI). CATI allows for the efficient administration of a survey by using skip logic to quickly move to the next appropriate question depending upon a respondent’s previous answer.

Logs maintained by home visitors and supervisors, and surveys of home visitors, supervisors, and program managers will all be collected using web-based applications. These applications will allow for the use of skip patterns to reduce the time needed to complete the various data collection procedures. For example, if a home visitor has not visited a family in a given week, the web-based log would record this information but skip over other questions about the family for that week.

Electronic data collection will also allow the research team to track real-time response rates and to monitor data on a regular basis to ensure data quality. The home visitors and supervisors will receive weekly reports for their own logs, and the research team will also receive weekly reports. These reports will allow the research team to monitor data collection by detailing who has completed the staff survey, whether each home visitor has completed a weekly log for each of their assigned families, and whether each supervisor has completed a weekly log for each of their team members. Electronic data also aid in maintaining and reviewing data quality. Given our real-time access to the web-base data, research staff will be able to regularly review item frequencies and cross-tabulations to guard against inconsistent or incorrect values. In addition, the web-based system is designed such that invalid responses cannot be entered (e.g., a 9,000 minute supervisory session) and will prompt the respondent accordingly.

A4. Efforts to identify duplication and use of similar information

Data being collected for MIHOPE are not available in any other form in a consistent manner across the evaluation's approximately 85 sites. There is currently no comprehensive list of home visiting sites by model within each state that could be used by the study team to select states for participation in MIHOPE. After careful review of the MIECHV plans submitted by each state, the needed information was not found.

Although many home visiting programs assess parents on depression, substance use, smoking, and other information being collected on the family baseline survey, those assessments will differ by local program, and local programs will not collect similar information on control group members. The baseline family survey therefore provides the only opportunity to collect this information in a consistent way for all families in the study.

Likewise, information that is being collected through weekly logs is not expected to be available in any other form. To understand variations in actual services received, the study must collect uniform information across models and program sites. No local program is expected to collect the breadth of information needed by the study team from these logs, and some programs may not collect any of the information in a systematic way. Even if some local programs are collecting some of the information included in the logs, it would be very costly for MIHOPE to align and analyze data from 85 different management information systems (MIS). Program sites vary to the extent that they use an MIS, track service delivery information in the MIS, and track specific service delivery variables. Finally, the study team's experience in conducting analyses in home visiting studies using service delivery data from program site MIS data indicates that the data are often poor quality. Moreover, in past studies it has taken 1–3 months for local programs

to send MIS data, making it impractical to monitor data quality and resolve inconsistencies. Real-time monitoring of data quality will be important for obtaining accurate estimates of service delivery.

A5. Burden on small business

No small businesses are affected by the data collection in this project.

A6. Consequences to collecting information less frequently

Site recruitment. If site recruitment is carried out with fewer conference calls or visits with state and local site administrators, it may take longer to recruit sites for the evaluation, reducing the evaluation's ability to respond to Congress's mandate to provide information about the program by 2015. In addition, local site participation in subsequent data collection activities such as staff surveys and home visitor logs may be of lower quality.

Baseline family data. Baseline family data will be collected only once for each family. Eliminating baseline family data will reduce the ability of the evaluation to answer the proposed research questions. The evaluation would not be able to describe the families that take part in the evaluation. It would not be able to estimate the effects of home visiting for subgroups of families, as required by the authorizing legislation. Eliminating baseline data would reduce the statistical precision of estimated impacts, reducing the ability of the evaluation to answer questions about the overall effectiveness of home visiting programs.

Semi-structured interviews with State MIECHV administrators. State administrators will be interviewed at baseline and 12 months later. Two surveys are needed because state plans, policies and resources related to MIECHV may change over time in ways that influence home visiting program service models, implementation systems, and service delivery. The study team will track these state-level changes in order to understand observed service delivery and impacts on families across states and home visiting program sites.

Surveys of staff at participating home visiting program sites. The program manager, supervisors, and home visitors complete surveys at baseline and 12 months later. Two surveys are needed because home visiting programs change over time. Programs can choose to give specific outcomes a higher or lower priority, change how they target services, redefine staff qualifications, expand or reduce intended services. In the same way, programs can change their implementation system over time. They can develop and refine their system for staff training and supervision, enhance or reduce the availability of clinical supports, expand or contract their use of technologies for service delivery and monitoring, and strengthen or weaken their ties to needed community resources.

For supervisors and home visitors, two rounds of surveys are also needed to assess changes over time in malleable personal attributes. Changes in personal attributes such as psychosocial well-being and competence to carry out one's role are likely to influence service delivery and the resulting program impacts for families. For example, staff skills in carrying out specific activities

might improve over time in response to training and supervision; alternatively, their skills might attenuate due to lack of ongoing training and reinforcement.

Logs maintained by supervisors and home visitors. Compared to less frequent completion of supervisor and home visitor logs, weekly completion of the logs will reduce the time needed to complete each log and will improve the accuracy of recall. Reporting on activities conducted in the previous week will be less time-intensive because it will be easier for respondents to reflect on the previous week rather than a longer period of time. Weekly completion of logs will provide higher quality and more accurate information about program implementation because respondents will have better recall of activities that occurred during the previous week. Although a home visitor is likely to have only about five families ever participate in the study, her full caseload will be much larger. It is common for home visitors to follow 20 or more families at a time. This would make it hard for the home visitor to report client-specific activities on her handful of study families accurately for much longer than a one-week period of recall.

Previous home visiting studies have indicated that the level of service is highly variable both across home visiting sites and across families within site (Filene, Bell and Smith, 2011). Weekly logs will allow the evaluation to assess variations and patterns of services provided to families. For example, the logs will be essential for identifying types of activities completed by home visitors for families who are beginning to disengage from the program. Weekly logs will also allow the evaluation to link child and family functioning outcomes with whether or not specific activities or tasks were completed during a home visit. For example, the logs will allow the evaluation to examine whether impacts on birth outcomes are stronger in local programs where home visitors more frequently referred parents for prenatal care and discussed prenatal health with the parent. More generally, from a cost perspective, it will be important to understand whether frequency of visits or duration of a program have implications for differential outcomes for families. Logs completed weekly by supervisors will provide information on the intensity and methods used in supervising home visitors. Supervisor logs will allow the evaluation to see variation in supervision techniques over time and across home visitors. The amount of supervision has been identified as the most significant predictor of implementation and retention in one home visiting study (McGuigan et al, 2003). An accurate assessment of the amount of supervision provided to each home visitor will be important.

The web-based system for log completion is a MIHOPE research data collection tool; it is not a full MIS system that sites could use to monitor service delivery or prompt staff to conduct programmatic activities. We therefore do not anticipate that MIECHV sites would adopt the MIHOPE log system in place of their MIS systems or integrate the two together. However, we have kept the time spent filling out the MIHOPE web-based log as brief as possible each week. Through the web-based system, home visitors and supervisors will complete a brief log (survey) each week about service delivery and supervision for the duration of a family's services or until the end of the MIHOPE data collection period. Log items are global enough that they apply to all four models included in the study. Although a few variables in the logs might duplicate information collected by study sites, site MIS vary greatly; in fact, some use paper records instead of an MIS.

Semi-structured interviews with home visiting staff. Home visiting staff will be interviewed about 12 months after baseline. The interviews will be the only opportunity for generating a deeper understanding of the processes underlying the provision of home visiting services and to “unpack” initial findings regarding staff goals, intended outcomes, priorities, and service approaches and dosage as measured with the baseline surveys and web-based logs. Though the frequency, once per respondent, is necessary, the plan reduces burden overall by reducing the number of home visitors to 2/3 of all home visitors across both group and individual interviews..

A7. Special Data Collection Circumstances

There are no special circumstances requiring deviation from these guidelines.

A8. Form 5 CFR 1320.8 (d) and consultations prior to OMB Submission

The 60-day Federal Register notice soliciting comments for the MIHOPE phase 1 data collection instruments and requesting subsequent 60 day notices to be waived was posted in the Federal Register, Volume 76, Number 238, pages 77236-77237 on December 12, 2011.

Three comments were received regarding the Notice for phase 1. The comments and responses to comments can be found in Attachment 29.

In addition, a Federal Register Notice was published on December 22, 2011 and a Supporting Statement was sent to OMB requesting an emergency clearance for site recruitment activities in order to continue activities to meet a legislatively mandated deadline. This request was approved on January 26, 2012 (0970-0402). Comments on this notice were addressed in the previous package.

Prior to submitting this package, the evaluation team and staff from OPRE and HRSA sought input from the Secretary’s Advisory Committee on Maternal, Infant and Early Childhood Home Visiting Evaluation. The committee consisted of experts across a range of disciplines and substantive areas. On December 6-7, 2011, the committee met and commented on the various data collection plans described in this package. Their comments resulted in a number of changes to the various instruments.

A9. Justification for Respondent Payments

Incentives are important, especially in a longitudinal study, to gain respondents’ cooperation and ensure a high response rate and their participation throughout the study, both at the baseline and at the follow up interview. (James 1997, Mack et al 1998, Martin et al 2001). Incentives are most appropriately used in Federal statistical surveys with hard-to-find populations or respondents whose failure to participate would jeopardize the quality of the survey data (e.g., in panel surveys experiencing high attrition), or in studies that impose exceptional burden on respondents, such as those asking highly sensitive questions.

Proposed payments for participating women would include the following:

- \$25 for women upon completing the 60-minute baseline family survey, plus an age appropriate book or toy worth \$15 for women with infants. This total amount is

comparable to payments used in FACES (\$35 for 45-60 minute interview), Baby FACES (\$35 for a 120 minute interview) and Building Strong Families (BSF; \$50 for completing two 50-minute parent interviews). The use of toys as incentives has been used in FACES, Baby FACES, and BSF. The incentive amount is sufficient to encourage families to participate in both the study and the survey but is not overly generous. Offering a lower amount could jeopardize the study and actually cost the government more because it could result in a lower uptake of families into the study and more effort expended by the evaluation team to successfully enroll families.

- \$5 for women who respond to a tracking letter by calling a toll-free number to update or confirm their name and address information. This incentive is necessary to track and maintain the study sample over time. The payment is nominal but an important way to encourage participants to respond to the letter by providing their updated contact information. Without an incentive, the study may expend more resources tracking down families that have moved and do not return their information cards.
- \$30 to home visitors and home visiting supervisors for each 75-minute web-based survey they complete (once at baseline and once a year later). The payment is intended to encourage staff to complete both survey waves, and will be a way to thank staff for using time outside of their normal work activities to complete the surveys. This amount is appropriate considering the mean salary for full-time employees over age 25 with a bachelor's degree or higher is \$28.70 per hour. Similar amounts have been used in the Head Start CARES study, where teachers received \$15 on average for completing surveys. In Baby FACES, home visitors received a \$25 gift bag for allowing themselves to be observed during a home visit and \$5 for each child on their caseload for providing information on language and other outcomes. As noted above regarding families, offering a lower amount could jeopardize the study and actually cost the government more by reducing the data that are collected or increasing the resources needed to collect the data.

A10. Confidentiality provided to respondents

The study team is committed to protecting the privacy of participants and maintaining the confidentiality of the data that are entrusted to us; in addition, the study team is experienced in implementing stringent security procedures. Every MDRC and Mathematica employee, including field staff employed for data collection, is required to sign a confidentiality pledge as an assurance of nondisclosure of confidential information. Field staff will also be trained in maintaining respondent privacy and data security.

When participants are recruited into the study, they will provide signed, informed consent. The consent form will include information about study goals, time required and duration, and the nature of questions that will be asked. Parents will be assured that their responses will be shared only with researchers, will be reported only as part of statistical analyses, and will not affect their receipt of services. If an applicant is a minor, it might be necessary to obtain consent from the parent as well, unless the state emancipated minor laws make this unnecessary.

Due to the sensitive nature of this research (for example, questions about substance use, domestic violence, child maltreatment, parental harshness, and depression), the evaluation will obtain a

Certificate of Confidentiality from HRSA. The study team has applied for this Certificate and will provide it to OMB once it is received. The study team has obtained such Certificates for other studies. The Certificate of Confidentiality helps to assure sites and participating mothers that their information will be kept confidential to the fullest extent permitted by law.

Documents shipped from the field and the document transmittal form that accompanies them will contain only identification numbers so that data cannot be attributed to any particular individual. Two exceptions are paper contact sheets used by field staff and signed informed consent forms, although neither type of document will contain data about the family. Completed paper documents will be stored in secured facilities. Security will be maintained on the complete set (and any deliverable backups) of all master survey files and documentation, including sample information, tracking information, and baseline data. Finally, data will be available only to staff associated with the project through password protection and encryption keys.

Staff will be asked to visit the project's website and to indicate whether they consent to participate in research activities. Staff will be informed that their identity will be kept private, that they do not have to answer questions that make them uncomfortable, and that results will only be reported in the aggregate. Staff members' decisions whether to participate in data collection activities and their responses to specific questions will not affect their employment status in any way. Staff will be asked to indicate consent using a check box.

A11. Justification for sensitive questions

Questions in some components of the MIHOPE baseline survey are potentially sensitive for respondents. Parents are asked about personal topics, such as child and parental health, substance abuse, salary and income, intimate partner violence and criminal involvement. To improve understanding of how the home visiting program affected families and children, it will be necessary to ask these types of sensitive questions. For example, maternal substance use is a major risk factor for reduced family well-being and child development and it is important to identify mothers with depression because maternal depression can be associated with poor parenting and has been associated with reduced effects from home visiting. As noted under A4, this information will not be available from other data sources in a consistent manner across the 85 sites and for both program group and control group families.

Parents will also be asked to provide their Social Security number (SSN) to allow the study team to collect state and federal administrative data and to allow the team to track them for purposes of collecting follow-up data. Providing an SSN is first mentioned in the written consent form, where the potential participant is told that providing an SSN is not required to participate in the study or to receive program services. The participant will be asked for their SSN during the baseline family survey. Because participants can refuse to answer any questions on the baseline survey, and because they will be told during the consent process that they are not required to provide an SSN, the study team did not include a separate opt out or consent form. MDRC's Institutional Review Board has approved the current consent and data collection procedures.

To ensure that parents are aware of the sensitive nature of the questions, the family baseline survey will contain instructions that explain questions before they are posed and will remind

participants that they may refuse to answer any question. Also, respondents will be informed by research staff prior to the start of the interviews or surveys that their answers will be kept private, that results will only be reported in the aggregate, and that their responses will not affect any services or benefits they or their family members receive.

Data collected through surveys of home visitors and their supervisors will also be potentially sensitive for respondents. These include questions about depressive symptoms, relationship security, and morale. Such questions are being asked because previous research has found that home visitor psychological well-being influences family engagement, home visit content, and home visitor turnover. Home visitor psychological well-being may thus alter the effects of home visiting services on parent and child outcomes. Likewise, supervisor psychological well-being may influence the effectiveness of their supervisory activities, thus influencing home visitor effectiveness and program impacts. The psychological well-being questions used in the home visitor surveys are from standardized measures or have been used in other studies of home visiting with no evidence of harm.

As part of the consent process, participating home visitors and supervisors will be informed that sensitive questions will be asked. Staff will be informed that their identity will be kept private, that they do not have to answer questions that make them uncomfortable, and that results will not be reported in a way that would identify them or their responses. Staff members' decisions whether to participate in the survey and their responses to specific questions will not affect their employment status in any way. Participating home visitors and supervisors will be asked to provide informed consent acknowledging their understanding of their study procedures and protections and acknowledging that their participation is voluntary.

A12. Estimate of the hour burden of data collection to respondents

Table A.3 shows the annual burden of the activities described in this supporting statement. Explanations for number of respondents on some specific activities are as follows:

Family baseline survey. The team will interview 5,100 respondents at baseline, or 1,700 respondents annually.

Family consent. The team will explain the study and present the consent form to approximately 5,667 women, or 1,889 women annually. This accounts for women who decline to provide consent.

State administrator interview. The team plans to interview two administrators in each of the 12 states in the evaluation, totaling 24 respondents, or 8 annually.

Community service providers survey. Program managers will be asked to provide 18 service providers to which they send participants. One program administrator from each service provider will be contacted, for a total of 1,530 respondents (18 respondents for each of the 85 sites), or 510 annually.

Other home visiting programs survey. Program managers will be asked to list 5 other home visiting service providers in their community. One program administrator from each of the home visiting service providers will be contacted for a total of 425 respondents (5 for each of 85 sites), or 142 annually.

Home visitor group interview. One-third of the home visitors – or about 2 per site – will participate in a group interview, resulting in 170 respondents, or 57 annually.

Home visitor individual interview. One-third of the home visitors will participate in an individual interview, resulting in 170 respondents, or 57 annually.

Table A.3
Annual Burden Estimates

	Number of Respondents	Number of Responses per Respondent	Average Burden Hours per Response	Total Annual Burden Hours	Average Hourly Wage	Total Annual Cost
<u>Site Recruitment</u>						
Telephone contact with state administrators	49	1	1.00	49	\$28.70	\$1,406.30
First round visits with state administrators	18	1	1.50	27	\$28.70	\$774.90
Second round visits with state administrators	15	1	1.50	23	\$28.70	\$660.10
Visits and calls with local program directors	120	1	3.00	360	\$28.70	\$10,332.00
Site Recruitment Total				459		\$13,173.30
<u>Data Collection</u>						
Family screening and consent	1,889	1	0.10	189	\$25.00	\$4,725.00
Family baseline survey	1,700	1	1.00	1700	\$25.00	\$42,500.00
State administrator interview						
Baseline	8	1	2.00	16	\$28.70	\$459.20
12 Month	8	1	2.00	16	\$28.70	\$459.20
Program manager survey						
Part 1, Baseline	29	1	0.50	15	\$28.70	\$430.50
Part 2, Baseline	29	1	1.00	29	\$28.70	\$832.30
Part 3, Baseline	29	1	1.00	29	\$28.70	\$832.30
12 month	29	1	2.00	58	\$28.70	\$1,664.60
Supervisor survey						
Baseline	33	1	1.25	41	\$30.00	\$1,230.00
12 month	33	1	1.25	41	\$30.00	\$1,230.00
Home visitor survey						
Baseline	170	1	1.25	213	\$30.00	\$6,390.00
12 month	170	1	1.25	213	\$30.00	\$6,390.00
Community service providers survey	510	1	0.10	51	\$28.70	\$1,463.70
Other home visiting programs survey	142	1	0.10	14	\$28.70	\$401.80
Supervisor logs	33	60	0.20	396	\$28.70	\$11,365.20
Home visitor logs	170	60	0.20	2040	\$28.70	\$58,548.00
Group interview						
Program manager	29	1	1.50	44	\$28.70	\$1,262.80
Supervisor	33	1	1.50	50	\$28.70	\$1,435.00
Home visitor	57	1	1.50	86	\$28.70	\$2,468.20
Home visitor individual interview	57	1	1.50	86	\$28.70	\$2,468.20
Data Collection Total				5,327		\$146,556.00
ESTIMATED TOTAL				5,786		\$159,729.30

A13. Estimates of Other Total Annual Cost Burden to Respondents and Record Keepers

For efforts involving home visiting program staff other than home visitor and supervisor surveys, an hourly wage of \$28.70 was used (see Table A.3). This is the mean wage for full-time employees over age 25 with a bachelor's degree or higher is \$28.70 per hour according to the Bureau of Labor Statistics' Current Population Survey 2011.

As described in section A9, for the family baseline survey, parents will be provided with a \$25. Home visitors and supervisors will be asked to complete surveys outside of work hours and will be provided \$30.00 per completed survey.

A14. Estimates of costs to federal government

ACF and HRSA are funding these activities. The estimated cost for activities covered in this submission is \$18,168,712. This includes designing data collection instruments, recruiting sites into the study, enrolling families and collecting baseline family information, and collecting all data on program implementation.

A15. Changes in burden

This is an increase in burden approved under the Maternal and Infant Home Visiting Program Evaluation emergency clearance for recruitment (0970-0402). The increase is the result of ongoing information collection for the mandated evaluation.

A16. Tabulation, analysis, and publication plans and schedule

Site recruitment activities began in February 2012 under emergency clearance authorization. States will be contacted between February 2012 and December 2012. Local program sites will be enrolled in the study on a rolling basis from June 2012 through August 2013. Within a site, staff would be surveyed at the time the site enters the study and one year later. In each site, it is expected to take 12-15 months to enroll women, so that sample recruitment and baseline data collection is expected to end in November 2014.

Beginning in February 2012, while awaiting OMB package approval, the MDRC team used iterative pretesting to identify revisions to be made to materials, procedures, and instruments for the baseline parent interview and implementation data collection. The exact timing of the baseline data collection however will depend on receipt of OMB clearance and on progress in site development and program pilots. Any changes to data collection instruments that result from pretesting will be submitted to OMB for review.

Phase 1 will produce a report to Congress in 2015. Regarding family characteristics, this report will include information on the characteristics of families participating in the evaluation. The report will also include information on organizational factors that will be collected from the implementing agencies. The information will include features of the site's intended service model and selected aspects of its implementation system. The service model features include intended goals and outcomes, recipients (such as family eligibility criteria), service delivery

(including dosage, content, and approach), and staffing characteristics. The implementation system refers to staff development (staff training, supervision, and evaluation); facilitative clinical supports (screening and assessment tools, protocols, curricula, peer support and learning, and access to professional consultation and experts); facilitative administrative support (management information systems, technologies for distance and supervision and learning, continuous quality improvement activities); systems characteristics such as formal agreements for referrals and technologies for information sharing; and presence of Memoranda of Understanding with other community resources. Information on organizational factors would come primarily from review of program documents and management interviews; this phase of the project does not include analysis of individual staff surveys and logs other than to extract basic demographics of home visitors.

Phase 2 will include follow-up data collection on family outcomes and a report on program implementation and program impacts. The current plan is to collect follow-up data on families around the time the child is 15 months old. If the first family enters the study in August 2012 with a six-month old child, the first follow-up would occur in May 2013. If the last family enters the study in mid-2014, the last follow-up interview could occur as much as 21-24 months later if the family enters while the mother is pregnant. A new information collection request with relevant instruments will be submitted to OMB for review as part of Phase 2, Data will also be collected through administrative records including, at a minimum, birth records and child welfare records. A report describing the estimated effects of the intervention, program implementation, and the relationship between the two would be published in 2017.

A18. Reasons for not displaying the OMB approval expiration date

All instruments will display the expiration date of OMB approval.

A19. Exceptions to Certification Statement

No exceptions are necessary for this information collection.

Appendix A: Justification for Not Including Direct Child Assessments at Baseline

This memo discusses the potential child assessment measures that could be conducted, and presents our recommendations. The recommendations are informed by consultation with Sally Atkins-Burnett, Jerry West, and members of the Secretary’s Advisory Committee (SAC).

The recommendations are influenced by the three intended uses of the MIHOPE baseline family survey:

1. Describe the characteristics of families that participate in the study
2. Define the analytic subgroups that will be used in the impact analyses
3. Increase the precision of the impact estimates by including measures of key domains at baseline and follow up

The survey will be administered to pregnant women and women with children from birth to 6 months of age, the key groups targeted by the home visiting programs. At this time we do not know the relative proportion of each group, but estimate that approximately one-third to one-half will not be born at the time of the baseline survey. Of those that are born, it is likely that half will be newborns (0 to less than 3 months) and half will be between 3 and 6 months old at the time of the baseline survey. The baseline participant survey will collect data on baseline family characteristics from two data sources: a baseline interview and the observational items from the Home Observation for Measuring the Environment (HOME; Caldwell and Bradley 2003) assessment. The baseline interview will be conducted by computer assisted telephone interview (CATI) to preserve privacy of study participants and to increase the efficiency and security of data collection. The HOME assessment will be conducted by field staff after they obtain informed consent from the family and while the participant is completing the baseline interview on the telephone.

A. CHILD ASSESSMENT MEASURES

There are a number of child assessment measures that could potentially be used for this study at baseline. We list the measures below and some factors to consider in weighing the challenges and benefits of each one as a baseline measure.

ITSEA/BITSEA: This is a parent report measure of child social-emotional well-being and is being used on Baby FACES. However, it is only normed for children aged 12 to 36 months. This was confirmed in an email exchange with the developer, Margaret Briggs-Gowan, who noted that “purposely did not design the BITSEA for less than 12 months due to concern about implying that psychopathology might “exist” at such a young age.” Therefore we cannot use it for the baseline survey. We could potentially use it at follow up.

Bayley Scales of Infant Development (Bayley): This measure can be used as early as 1 month of age. Here we discuss five versions of the assessment, the Bayley-II and the short form based on it developed for the ECLS-B, and the Bayley-II screener, the Bayley-III, and the Bayley-III Screening Test. In addition, we summarize the Social-Emotional Scale included in the

Bayley-III, a parent-completed questionnaire based on the Greenspan Social-Emotional Growth Chart (Greenspan 2004).

Published in 1993, the norming sample from the **Bayley-II** is dated and no longer reflects the population of children in the United States. Short forms of the Bayley-II mental and motor scales for 9-month and 24-month old children (BSF-R, Andreassen and Fletcher 2005) were developed with considerable effort and expense for the ECLS-B to simplify administration and reduce data collection time. The BSF-R was developed in response to a much longer than expected administration time encountered during the 1999 field test of the full BSID-II. Development of the short form was also designed to address difficulties field staff had administering and scoring the items using the standardization rules specified by the test developer. The BSF-R took approximately 36 minutes when the children were 9 months old in the ECLS-B. It would take considerable measurement development and psychometric work to create a short form appropriate for the MIHOPE age range. The Bayley Infant Neurodevelopmental Screener (BINS; Aylward 1995) is based on the Bayley-II and screens infants between the ages of 3 and 24 months for neurological impairments and developmental delays. It takes between 5 and 10 minutes to administer but as a screener, it does not show much variation in typically developing children's development. In addition, it does not extend down to cover the birth through 3 month age range.

The Bayley-III (2006) has not been used in a large-scale national study and was not recommended for the Baby FACES study for that reason and because it has a new, untested approach to separately measuring different outcome domains and computing separate scale scores based on a relatively small number of items appropriate to each age range. The Bayley-III direct child assessment has been organized into three scales and five subtests: (1) the Cognitive Scale is comprised of one subtest, (2) the Language Scale is comprised of the Receptive Communication and Expressive Communication subtests, and (3) the Motor Scale is comprised of the Fine Motor and Gross Motor subtests. In addition, the Social-Emotional Scale and the Adaptive Behavior Scale are two separate parent-report questionnaires. Both questionnaires and any direct assessment items that require the interviewer to speak to the child or parent would have to be translated into Spanish, as neither the Bayley-II nor the Bayley-III are available in Spanish. Other important concerns include (1) the Bayley-III's length, (2) the fact that the test has been normed in English only, (3) the lack of data about how predictive the scales are when used with infants 0-6 months, and (4) the fact that each scale has only a few items in it (which may result in severe floor and ceiling effects). The Bayley-III Screening Test (for 1 to 42 months) maintains the same multi-scale structure of the direct assessments in the full Bayley-III with even fewer items included per subtest (which exacerbates floor and ceiling effects). Given that it is based on the Bayley-III, the same issues described above regarding the norming sample apply.

Information publicly accessible indicates that the National Children's Study (NCS) is piloting a short form of the Bayley-III in four locations across the country using procedures similar to what was done for the ECLS-B that focus on reducing the length of the assessment and increasing the reliability of the administration by field staff. Several consultants suggested that due to these multiple concerns, particularly the lack of predictive validity data and the fact that the NCS version is only expected to extend down to 6 months, which is not far enough for the

MIHOPE baseline (we saw reference to 6-month IRT scores in what was publicly available), the Bayley should not be included at baseline or follow up. The Bayley Infant Neurodevelopmental Screener (BINS; Aylward 1995) is based on the Bayley-II and screens infants between the ages of 3 and 24 months for neurological impairments and developmental delays. It takes between 5 and 10 minutes to administer but as a screener, it does not show much variation in typically developing children's development. In addition, it does not extend down to cover the birth through 3 month age range.

The Greenspan Social-Emotional Growth Chart (Greenspan 2004): This assessment is now part of the Bayley-III and is completed by the child's parent or primary caregiver. It is based on functional emotional milestones that correspond to 8 stages for children from birth to 42 months of age (Bayley 2006). One concern about this measure is the small norming sample and very small sample sizes included in it for ages 0-3, 4-5, and 6-9 months (89, 54, and 51, respectively). In addition, we do not believe it has been used in a large-scale national study of high-risk parents and children.

Mullen (1995): This measure can be used from birth through 68 months. However, the norming sample is out-dated and it is only available in English.

Three/Two Boxes/Bags Task and Coding System: This measure examines parenting constructs such as supportiveness, sensitivity, cognitive stimulation, intrusiveness, and negative regard. It also includes scales that examine child engagement of parent, sustained attention, and negativity toward parent. The semi-structured play task and variations of the original coding scheme by Deborah Vandell and colleagues have been used with children 14, 24, and 36 months old in a number of studies, including the Early Head Start Research and Evaluation project, Fragile Families, ECLS-B, and Baby FACES. It was used with children six months and older in the NICHD Study of Early Child Care and Youth Development and in the Early Head Start Newborn Study. Predictive validity data from use of the task and coding system from 0-6 months is scant. This type of task and coding system are being considered for the follow-up assessment with the full sample.

The Nursing Child Assessment Teaching Scale (NCATS) (1995): This observational measure of the quality of the caregiver-child teaching interaction for children from birth to 3 years of age assesses four parent and two child behaviors. The correlations of the total NCATS scores with the total HOME score among children ages 1 to 36 months, in three age groups, ranged from .41 to .44. Given that the HOME is already planned for MIHOPE, NCATS may not add much additional information given the relative cost of training on the assessment. In addition, the adaptations to shorten the observation period made for administering the NCATS in the EHS-REP revealed internal consistency reliability problems inherent in large-scale live or videotaped coding and administration of the measure. There is scant information available about the predictive validity of the NCATS Teaching Task when conducted with children less than 6 months old.

Brazleton (1973): This measure is used with infants and usually in hospital settings. It is a scale for 0-2 months of age, so its use for this study is limited as our sample at baseline will include children 0 to 6 months of age.

Neonatal Intensive Care Unit Network Neurobehavioral Scale (NNS) (2004). This neurological assessment can be conducted from birth through 48 weeks. The infant should start off in a sleep state that has been maintained for at least 45 minutes. There are 115 items and several position changes are required during which the observer looks for changes in the baby. This assessment requires a highly trained individual, usually a clinician, and does not seem to be suitable for a large-scale study. Although there are a few published articles on the measure, there is little information available on its predictive validity and it has been used primarily for clinical purposes.

Other ECLS-B 9-Month Assessments: The remaining set of measures used at 9 months assesses infant physical development, including weight, length, upper arm circumference, and head circumference. Although a direct assessment may be desirable, we will be getting most of this information from other sources.

B. RECOMMENDATIONS FOR CONDUCTING DIRECT CHILD ASSESSMENT

In developing the MIHOPE baseline survey, we focused on including measures of outcome domains that are most likely to show impacts or that had the potential to mediate or moderate impacts. Direct child assessments of the portion of the sample that includes infants that were born at baseline were considered but rejected because they could only be administered for part of the sample (unborn children would have no data for these measures) and because the developmental experts we consulted with and our SAC recommended against directly assessing children 12 months of age or younger.

For over forty years, the predictive validity of infant assessments, particularly those administered to children less than one year of age, has been an issue for the field of developmental psychology. In the 1970's and 1980's, leading developmentalists debated this issue related to performance of children less than 1 year old on the Bayley and correlations to subsequent cognitive functioning (Lewis and McGurk 1972; Lewis and McGurk 1973; Matheny 1973; McCall 1981; Wilson 1973). Then, as now, researchers have concerns about the predictive validity of assessments conducted with young infants and generally recommend they be used for assessing performance at a given point in time for diagnostic and comparative purposes rather than as predictors of later skills and abilities (for example, Hack et al. 2005). A few measures of information processing for children less than 6 months old have been identified as somewhat more robust predictors to intelligence at 3 years of age (for example, the Fagan Test of Infant Intelligence 2005), but they have not been used in large-scale research projects and are more suitable to laboratory settings than to in-home assessment. The primary arguments against conducting direct child assessments stem from the lack of reliable and valid measures in early infancy; overall, the predictive validity of the measures that are available is either unknown or quite low.

After weighing the information above against the practical issues such as cost, we do not recommend conducting direct child assessments on the MIHOPE study for the following reasons:

1. **Sample Size and Variation.** About one-third to one-half of the sample at baseline will be pregnant women, so we would be able to obtain child assessment data for only part of our sample. The sample of children at baseline will also vary widely, with ages ranging from 1 day old to 6 months. There are few child assessment measures that are suitable for this age group.
2. **Cost.** The cost of conducting child assessments would be high and would require more funds than what have been allocated for the baseline effort. We would need to hire and train a group of staff with experience in complicated direct child assessments. We would need to pay more per hour since they will be doing work that is more difficult. Training will take substantially longer than what we budgeted (4 plus days rather than 2 days). Certification on the measures would be difficult and many staff would not pass, which would require additional hiring, training, and certification.
3. **Logistics.** The logistics of conducting assessments would be more challenging. The baseline visit would be longer, since the field staff would be conducting an assessment. We would need the infant to be awake, which could necessitate going back to the home multiple times to complete the assessment. These logistical considerations would also increase the cost of the baseline data collection.
4. **Low Return on Investment.** There is generally low predictive value of the standard child assessment measures at very young ages (birth to 6 months). We do not believe that the data gathered would provide us with adequate information to make the effort worthwhile. In addition, the measures that could be used at both baseline and follow-up are few and have the limitations described above.

Appendix B: Summary of Changes Made to Family Baseline Survey

Survey Item	Change resulting from pretesting	Rationale
<p>A7 After [CHILD] was born, how long did [he/she] stay in the hospital?</p> <p>A8 After your baby [CHILD] was born, was [he/she] put in an intensive care unit or NICU?</p>	Revised A8 to ask if any of these days were in the NICU, and then if yes, ask for number of days child spent in the NICU.	Will help to clarify how long the baby spent in the NICU.
A13 Do you have a plan to breastfeed?	Revised to “Do you plan to breastfeed?”	Respondents had difficulty with the word “plan.” They often responded with “I hope to” or “I’d like to.” Revised wording will help match respondent’s intent.
A14 How long do you plan to breastfeed?	Revised to “How long would you like to breastfeed?”	
A15 How old was [CHILD] the first time (he/she) ate or drank anything other than (breast milk or) formula?	Replaced with the following item from the ECLS-B 9-month parent interview: “How old was [CHILD] in months when solid food was first introduced? Solid foods include cereal and baby food in jars, but not finger foods.”	Revised wording is more specific to ensure respondent understands the question.
B1 The next questions are about your health. In general, would you say your health is...?	For pregnant women, revised to “The next questions are about your health before your current pregnancy. In general, would you say your health is...?”	To clarify for pregnant women that they should answer about their health before pregnancy, so they do not consider any pregnancy-related ailments when responding.
B5 During (this pregnancy/your pregnancy with [CHILD]), were you told by a doctor, nurse, or other health care worker that you had gestational diabetes (diabetes that started during this pregnancy)?	Add response option, “haven’t been tested yet” for pregnant women.	It is possible that some women may not be far enough along in their pregnancy to have been tested for gestational diabetes.
B8 Is there a place you go for general health care, if you are sick or need advice about your health - that is, any care except prenatal care or family	Added the follow-up item: B8a. What kind of place do you go?	Not all pretest respondents knew that we were asking about a physical location.

Survey Item	Change resulting from pretesting	Rationale
planning?	Clinic Health Center Hospital Doctor's office Some other place	
<p>B9 During the past year, have you ever received family planning or gynecologic services?</p> <p>B9a During the past year, did you ever want or need family planning or gynecologic services?</p> <p>B9b What is the main reason you didn't receive family planning or gynecologic services?</p> <p>B9c Are you currently receiving family planning or gynecologic services?</p>	<p>Replaced with the following items:</p> <p>B9. Is there a place you go, or have gone, for family planning or birth control?</p> <p>B9a. What kind of place do you go/ did you go?</p> <p>The same place I receive general health care Clinic Health Center Hospital Doctor's office Some other place</p>	Some respondents were confused by term "family planning services."
B10 How many more children do you plan to have?	Revised to "How many more children would you like to have?"	Respondents had difficulty with the word "plan." They often responded with "I hope to" or "I'd like to." Revised wording will help match respondent's intent.
C8 What is the highest grade or year of regular school that you have completed?	Removed "regular" from question	Respondents were confused by the term "regular."
Section D items on the woman's spouse or partner	If a woman doesn't have a spouse or partner and doesn't live with the child's biological father, added an item asking if the woman and biological father ever lived together. Added an item asking if the woman is currently in a romantic relationship, and for those who say yes, then ask the intimate partner violence items.	This section didn't flow well during the pretest. This revision will help fill in missing information

Survey Item	Change resulting from pretesting	Rationale
Section E items on household composition and earnings	Revised items about household composition and earnings to accommodate respondents whose household composition is currently different than it was for most of the previous year.	These questions were difficult for respondents to answer if the current household members were not the same as in the prior year (when we ask for total earnings from all household members.) changing these items will make answering them easier for the respondent.
<p>E4 How many months were you employed (did you work for pay) during the past 3 years (including your current job)?</p> <p>RESPONDENT DIDN'T WORK</p> <p>Less than 6 months 7 to 12 MONTHS 13 to 24 MONTHS More than 24 months</p>	Changed format so that interviewer reads the answer choices aloud to respondent, except for “respondent didn’t work.”	Pretest respondents had trouble calculating number of months; providing answer choices helped them respond.
E19 During the past year, have you received Early Head Start or child care services for [CHILD]?	Revised.	Respondents were confused and wondered if we meant EHS only or child care in general.
<p>E20 During the past year, have you ever received Early Intervention services or [INSERT NAME OF PROGRAM FOR STATE] for (CHILD)?</p> <p>E20a Did you ever want or need Early Intervention services for [CHILD]?</p> <p>E20b Are you currently receiving Early Intervention services for [CHILD]?</p>	Deleted from survey.	Since most babies will be too young to have received early intervention services at baseline, we recommend deleting this question and including it in the follow-up survey.
E21/22 a-c Home visiting items	Moved to end of survey, just before contact information.	Moving the home visiting questions to the end eases the transition from the

Survey Item	Change resulting from pretesting	Rationale
		end of the survey to collecting contact information.
E22a-c What do you think will be the three most important benefits of home visiting for you and your family?	Deleted these items.	Responses here were the same as those captured in E21.
F15-F18 questions on receipt of mental health and substance abuse treatment services during past year	Shortened the list by grouping similar items together and using broader categories	The list of items was long, categories were redundant, and the list was cumbersome to administer
<p>G6 Please tell me whether you or any other members of your household received income from the following sources in the past month. This includes anyone who you support and/or supports you and lives in your household.</p> <p>G7 During the past year, have you ever received help in applying for public benefits, including TANF, SNAP, or WIC?</p>	Added “WIC” to the list of sources in G6.	Four respondents said yes to G7 because they received WIC, but didn’t understand that the question was asking if they had received <i>help in applying</i> for services like WIC. Add WIC to G6 to capture this benefit.
	Added questions from the Pearlin mastery scale	Added in response to an NFP comment suggesting the measurement of low psychological resources.
	Added questions from the Wechsler Adult Intelligence Scale Similarities subtest	Added in response to an NFP comment suggesting the measurement of low psychological resources.

Appendix C: Measuring Cognitive Ability

To measure cognitive ability, the MIHOPE baseline survey will contain the Similarities subtest of the Wechsler Adult Intelligence Scales – Third Edition (WAIS-III; Wechsler, 1997). The Similarities subtest is designed to capture abstract reasoning and verbal comprehension abilities, which are two principal dimensions of intellectual abilities (Flanagan and Harrison, 2005; Flanagan, Ortiz and Alfonso, 2007). In the Similarities subtest, respondents are asked a series of questions about how two things are alike. For example, “How are a snake and an alligator alike?” Each item is then scored on a 0 to 2 scale according to general scoring principles and examples that are provided in the testing manual.

This measure is proposed to assess parents’ cognitive and intellectual abilities for a variety of reasons:

- The Wechsler Adult Intelligence Scales are among the most widely used measure of intellectual abilities in the United States and in other countries. The WAIS-III Similarities subtest is also one of the few measures of abstract reasoning and verbal comprehension that is available in both English and Spanish that can be readily administered over the telephone or in person.
- Compared with most other assessments of intellectual abilities, the Similarities subtest is relatively brief – consisting of only 18 items – which places substantially less burden on study participants than most other measures of cognitive and intellectual abilities. Furthermore, study participants need not receive all of the items because the testing includes a discontinuation rule when respondents get three consecutive items incorrect. Thus, the amount of time required to administer the subtest can be quite brief and varies the study participants’ intellectual aptitude thereby reducing the burden of the measure on study participants.
- The English and Spanish versions of the Similarities subtest have been shown to have good psychometric properties. The publishers of the English version of the subtest found that its split-half reliability is 0.87, the test-retest reliability is 0.83, and the inter-rater agreement on scoring the items of the subtest (ICCs) is 0.93 (Tulsky et al., 1997). Elsewhere, Renteria et al. (2008) found the Spanish WAIS-III Similarities subtest had an internal consistency of 0.79 using a sample of primarily Spanish-speaking adults recruited from Chicago neighborhoods.
- The Similarities subtest has been shown to have good validity and demonstrated capabilities for differentiating individuals with qualitatively different levels of intellectual abilities. In numerous studies the Similarities subtest is a strong predictor of the full-scale score of intellectual functioning that can be created when the full battery of subtests from the WAIS-III. Jones et al. (2006), for example, found that the Similarities subtest loads onto the WAIS full-scale score of overall intelligence at 0.81 in a factor analytic model. Moreover, using a sample of adults who are diagnosed with mild intellectual disabilities according to the DSM-IV-TR criteria for intellectual disabilities (e.g., IQs of 40 – 70), the publishers found that this group on the Similarities subtest scored about 2.5 standard

deviations lower than a matched comparison group with average intelligence (Tulsky et al., 1997). Using a sample of adults who meet the DSM-IV-TR criteria for Borderline Intellectual Functioning (e.g., IQs of 71 – 84), the publishers also found that the group scored about 1.4 standard deviations lower on the Similarities subtest than a matched comparison group with average intelligence (Tulsky et al., 1997).

Appendix D: Implementation Study Instruments – Content in Paired Instruments and Revisions per Pretesting and in Response to Public Comments

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
State administrator interview			
Baseline (7)	The baseline survey gathers data on MIECHV- and state-level factors for service delivery, from the perspective of the state's lead agency for MIECHV.	Sections K and L were reformatted to improve clarity.	<i>Comments:</i> None
12 Month (8)	The content of the 12-month interview parallels that of the baseline interview. Items elicit information on changes in factors since the baseline survey.	The 12 month interview was edited to align with the revised baseline instrument.	<i>Comments:</i> None
.....			
Program manager survey			
Part 1, Baseline (9)	The content of each of the three parts of the baseline survey is unique. The three parts are complementary. Together, they gather baseline data on the full set of hypothesized program site factors for service delivery, from the perspective of site	As possible, sections on site policies and procedures were edited to make data collection more efficient by using questions about policies in lieu of requests for copies of the policies. Items on current staff were moved to Part 2 because they fit better with its content.	<i>Comment:</i> It is unclear which survey instruments will be completed by a program manager who is also a supervisor. <i>Response:</i> A program manager who is also a supervisor will complete the program manager survey and sections of the supervisor survey that are not redundant with the program manager survey.
Part 2, Baseline (10)		Items that could be answered more efficiently via other instruments were	<i>Comments:</i> None

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
Part 3, Baseline (11)	leadership.	<p>eliminated.</p> <p>Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.</p> <p>A few items were added to fill identified gaps and eliminate ambiguity in responses.</p> <p>Items on referrals to community resources were moved to Part 3 because they fit better there.</p> <p>Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.</p> <p>A few items were added to fill identified gaps and eliminate ambiguity in responses.</p>	<p><i>Comment:</i> Questions about referral are redundant with questions in the supervisor survey.</p> <p><i>Response:</i> We have eliminated this redundancy by dropping these questions from the supervisor survey. The questions are now a part of only the program manager survey. A site can choose to have a supervisor or other staff member help the program manager answer these questions if the site feels that is more efficient.</p>
12 Month (12)	<p>The content of the 12-month survey parallels that of parts 1 and 2 and a small portion of part 3 of the baseline survey. Thus, comparison of responses from baseline to the 12 month survey allows assessment of change over time.</p>	<p>The 12 month survey was edited to align with the revised baseline instrument.</p>	<p><i>Comments:</i> None</p>
Supervisor survey Baseline (13)	<p>The baseline survey gathers data on hypothesized program</p>	<p>Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.</p>	<p><i>Comment:</i> It is unclear whether a supervisor who is also a home visitor will complete both or only one survey.</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
12 Month (14)	<p>site factors for service delivery from the perspective of supervisors, and on supervisor-specific factors for service delivery.</p> <p>The content of the 12-month survey parallels that of the baseline survey. Thus, comparison of responses from baseline to the 12 month survey allows assessment of change over time.</p>	<p>Items that could be answered more efficiently via other instruments were eliminated.</p> <p>In Sections L-S, items were reorganized, reworded, and some items were eliminated to improve efficiency.</p> <p>Some items were added to fill identified gaps and to eliminate ambiguity in responses.</p> <p>The 12 month survey was edited to align with the revised baseline instrument.</p>	<p><u>Response:</u> A supervisor who is also a home visitor will complete the supervisor survey and portions of the home visitor survey that are not redundant with the supervisor survey.</p> <p><u>Comment:</u> It is unclear which survey instruments will be completed by a replacement supervisor.</p> <p><u>Response:</u> A replacement supervisor will complete a baseline survey upon joining the study. S/he will also complete the 12 month survey if s/he joins the study at least 6 months prior to the 12 month survey.</p> <p><u>Comment:</u> Both the Baseline and the 12 month surveys ask about program expectations, which is unnecessarily repetitious.</p> <p><u>Response:</u> We have deleted a few of the redundant items. Redundancies are by design, to capture expected site-level changes in program models and implementation systems over time. The MIECHV program has already given rise to substantial changes in home visiting at the national, state, local and program site levels. We expect this will continue in the years ahead. Thus, we have designed the 12-month staff surveys to assess changes in both organization- and individual-level factors for service delivery.</p>
Home visitor survey Baseline (15)	<p>The baseline survey gathers data on hypothesized program site factors for service delivery from the perspective of home visitors, and on home visitor-specific factors</p>	<p>Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.</p> <p>Items that could be answered more efficiently via other instruments were eliminated.</p> <p>In Sections L-S, items were reorganized,</p>	<p><u>Comment:</u> It is unclear which survey instruments will be completed by a replacement home visitor.</p> <p><u>Response:</u> A replacement home visitor will complete a baseline survey upon joining the study. S/he will also complete the 12 month survey if s/he joins the study at least 6 months prior to the 12 month survey.</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
	for service delivery.	<p>reworded, and some items were eliminated to improve efficiency.</p> <p>Some items were added to fill identified gaps and eliminate ambiguity in responses.</p>	<p><u>Comment:</u> The home visitor baseline survey remains lengthy.</p> <p><u>Response:</u> Editing as part of pretesting has reduced the number of items by about 20%. Pretesting has established that home visitors can complete the survey within the projected time.</p> <p><u>Comment:</u> 105 items are embedded, not fully shown.</p> <p><u>Response:</u> The source instrument, which is proprietary, was identified by name – the Organizational Social Context (OSC) scales. The commenting organization is familiar with this instrument, having reviewed its items and approved its use in December, 2011 for another home visiting study conducted by MIHOPE team members, in which its sites participate.</p> <p><u>Comment:</u> There was concern that the number of items measuring home visitor psychosocial functioning (n=105 + 39) is burdensome and intrusive.</p> <p><u>Response:</u> This section of the survey includes three instruments: the OSC (105 items), the short form of the CES-D (10 items), and the Attachment Style Questionnaire (29 items). We did not change this section, for several reasons. First, the three instruments in this section measure different constructs, all of which are hypothesized to have independent influences on service delivery and impact. There is theoretical and empirical support for the independence influence of each of these constructs on service delivery and impact. Second, the OSC measures not only individual level factors</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
12 Month (16)	The content of the 12-month survey parallels that of the baseline survey. Thus, comparison of responses from baseline to the 12	The 12 month survey was edited to align with the revised baseline instrument.	<p>(morale and burnout) but is the study primary measure of two key organization-level factors (culture and climate). Third, depressive symptoms and relationship security have been shown to influence service delivery, <i>and</i> to have interactive effects on family engagement. Fourth, leaders of other evidence-based home visiting models expressed their support for assessing staff psychosocial well-being at the ACF/HRSA-sponsored MIHOPE meeting of model developers on October 27, 2011.</p> <p><i>Comment:</i> Questions about home visitors' background as a parent or home visiting recipient seem judgmental.</p> <p><i>Response:</i> These items have been deleted.</p> <p><i>Comment:</i> Questions about referral are redundant with questions in the program manager survey.</p> <p><i>Response:</i> We have kept the referral questions in both instruments. The questions are similar by design, but they serve different purposes. We use answers to referral questions in the program manager survey to assess the site's awareness of and relationship with community resources. We use answers to referral questions in the home visitor survey to measure each home visitor's knowledge of, attitudes toward, and interactions with community resources.</p> <p><i>Comment:</i> Both the Baseline and the 12 month surveys ask about program expectations, which is unnecessarily repetitious.</p> <p><i>Response:</i> We have deleted a few of the redundant items. Redundancies are by design, to capture</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
	month survey allows assessment of change over time.		expected site-level changes in program models and implementation systems over time. The MIECHV program has already given rise to substantial changes in home visiting at the national, state, local and program site levels. We expect this will continue in the years ahead. Thus, we have designed the 12-month staff surveys to assess changes in both organization- and individual-level factors for service delivery.
Community service provider survey (17)	<p>This survey is conducted at baseline only.</p> <p>Its content parallels that of Part 3 of the program manager baseline survey for each type of service provider listed.</p> <p>It elicits the community service provider's perspective on referral and coordination with a specific home visiting site and on service availability, service accessibility and inter-agency agreements as factors for referral and coordination.</p>	<p>We did not pretest this instrument.</p> <p>Two items were added to the survey to address identified gaps (agency address and cost of services).</p> <p>Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.</p>	<u>Comments:</u> None
Other home visiting program survey (18)	This survey is conducted at baseline only. It documents key characteristics of other home visiting or	We did not pretest this instrument.	<u>Comments:</u> None

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
	parenting programs for infants in the community in which control group members might enroll.		
Supervisor logs (19)	These logs are completed weekly to measure supervisor training and actual supervision from the perspective of the supervisor as factors that influence actual service delivery.	Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.	<p><i>Comment:</i> The logs are burdensome because staff are expected to complete them weekly and because they are duplicative of forms that staff complete routinely as part of (NFP) model requirements.</p> <p><i>Response:</i> We reduced the number of items. We expect that each supervisor will complete weekly logs only for home visitors with one or more active families participating in the evaluation. For this reason, repetitiveness is limited.</p> <p>Although the content of the MIHOPE logs overlaps slightly with NFP logs, most items in the MIHOPE logs ask for content different than that in NFP logs.</p> <p><i>Comment:</i> The frequency of log completion should be reconsidered, perhaps to a monthly summative reporting across all home visitors.</p> <p><i>Response:</i> To understand variation in actual services to families and factors that influence service delivery, the study must collect uniform information across all outcome domains for all models and program sites. The logs provide key information about individual-level service delivery and supervision for “black box” analyses as well as for documenting variations in program costs for participant subgroups. No national model requires sites to collect the full set of supervision variables needed for MIHOPE; some sites might not collect any of this information in a systematic way. Our previous research using logs suggests that less</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
Home visitor logs (20)	These logs are completed weekly to measure actual service delivery and home visitor perspectives on actual training and supervision as factors for service delivery.	<p>Items on approaches to service delivery within each content area were dropped to reduce respondent burden.</p> <p>Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.</p>	<p>frequent completion will negatively impact staff recall of events. Our previous research highlights substantial variability in the intensity and content of both home visits and supervision. We need to measure supervision at the home visitor level and service delivery at the client level. These measures will be key variables in analyses factors explaining variations in service delivery and fidelity. Variation in service delivery and fidelity will, in turn, be tested as a moderator of program impacts.</p> <p><i>Comment:</i> The logs are burdensome because staff are expected to complete them weekly and because they are duplicative of forms that staff complete routinely as part of (NFP) model requirements.</p> <p><i>Response:</i> We have reduced the number of items in the logs. Home visitors will complete weekly logs only for active families participating in the evaluation. On average, this will be only about five families, a small portion of the home visitor’s caseload. For this reason, repetitiveness is limited. Although the content of the MIHOPE logs overlaps slightly with NFP logs, most items in the MIHOPE logs ask for content different than that in NFP logs.</p> <p><i>Comment:</i> The frequency of log completion should be reconsidered, perhaps to a monthly summative reporting across all home visitors.</p> <p><i>Response:</i> To understand variation in actual services to families and factors that influence service delivery, the study must collect uniform information across all outcome domains for all models and program sites. The logs provide key information about individual-level service delivery and supervision for “black box” analyses as well as for</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
<p>documenting variations in program costs for participant subgroups. No national model requires sites to collect the full set of supervision variables needed for MIHOPE; some sites might not collect any of this information in a systematic way. Our previous research using logs suggests that less frequent completion will negatively impact staff recall of events. Our previous research highlights substantial variability in the intensity and content of both home visits and supervision. We need to measure supervision at the home visitor level and service delivery at the client level. These measures will be key variables in analyses factors explaining variations in service delivery and fidelity. Variation in service delivery and fidelity will, in turn, be tested as a moderator of program impacts.</p>			
<p>Semi-Structured Interviews</p>			
<p>Group interview – program managers (21)</p>	<p>These group interviews are conducted at 12 months to elicit staff perspectives for interpreting data collected in the surveys and logs, that is to explain the how and why behind quantitative results.</p>	<p>For group interviews with program managers, supervisors and home visitors, we deleted items that were redundant with the staff surveys, added a few questions to fill identified gaps, and edited questions to elicit participants’ perspectives on the reasons and mechanisms for results obtained through the surveys.</p>	<p><i>Comment:</i> There is considerable duplication of questions across the 12 month surveys and interviews.</p>
<p>Group interview – supervisors (22)</p>			<p><i>Response:</i> We have eliminated the Interview participant questionnaire (formally Instrument 24), as it was duplicative of items asked on the baseline surveys.</p>
<p>Group interview – home visitors (23)</p>			<p>We deleted items from the group and individual home visitor interview instruments that were redundant with the baseline and 12 month surveys (Instruments 13-16).</p>
<p>Interview participant questionnaire (24)</p>	<p>This questionnaire elicits basic information to characterize group interview participants</p>	<p>This instrument has been eliminated .</p>	<p>In instruments for both the group and individual interviews, most items are, in fact, either optional or</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
Individual interview – home visitors (25)	<p>(Instruments 21-23)</p> <p>These individual interviews are conducted at 12 months to elicit staff perspectives for interpreting data collected in the surveys and logs, that is to explain the how and why behind quantitative results.</p> <p>The individual interviews seek to elicit views that home visitors are less likely to share candidly in group interviews.</p>	<p>For the individual interviews with home visitors, we deleted items that could be answered adequately in the group interviews, added a few questions to fill identified gaps, and edited questions to elicit participants’ perspectives on the reasons and mechanisms for results obtained through the surveys.</p>	<p>potential probes. We will ask only a subset of questions, with the exact subset to be determined by the specifics of the data collected in the other instruments completed by the participating sites. We’ve edited the instruments to identify optional items and potential probes.</p> <p><i>Comment:</i> It is unclear whether replacement supervisors and home visitors will complete the interviews.</p> <p><i>Response:</i> Replacement home visitors and supervisors will be eligible to participate in the interviews if they joined the study at least 6 months earlier.</p>
Messages to home visiting program staff (28)	<p>These messages thank staff for completing logs and remind staff to do so.</p>	<p>No changes</p>	<p><i>Comments:</i> None</p>

REFERENCES

- Andreassen, C., and P. Fletcher. "Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) Methodology Report for the Nine-Month Data Collection (2001-02), Volume 1: Psychometric Characteristics." Submitted to the U.S. Department of Education. Report No. (NCES 2005-100). Washington, DC: National Center for Education Statistics, 2005.
- Aylward, G. P.. 1995. Bayley Infant Neurodevelopmental Screener. San Antonio, TX: Psychological Corporation.
- Bayley, Nancy. Bayley Scales of Infant and Toddler Development-Third Edition: Administration and Technical Manual. San Antonio, TX: PsychCorp, 2006.
- Bloom, Howard S., Carolyn J. Hill and James A. Riccio. 2003. "Linking Program Implementation and Effectiveness: Lessons from a Pooled Sample of Welfare-to-Work Experiments," *Journal of Policy Analysis and Management*, 22(4): 551 – 575.
- Durlak, J. A., and E. P. DuPre. 2008. "Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation." *American Journal of Community Psychology* 41, 3-4: 327-350.
- Fagan, Joseph F. 2005. "The Fagan Test of Infant Intelligence-Manual." Website: <http://infantest.com/ftii.pdf>.
- Filene, Jill H., James Bell, and Elliott G. Smith. 2011. National Cross-Site Evaluation of the Replication of Family Connections: Final Evaluation Report. Report submitted to the Administration for Children and Families.
- Flanagan, D. P., & Harrison, P. L. (2005). *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. (2nd Edition). New York, NY: The Guilford Press.
- Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of Cross-Battery Assessment*. (2nd Edition). New Jersey: John Wiley & Sons, Inc.
- Greenspan, S.I. *Greenspan Social-Emotional Growth Chart: A Screening Questionnaire for Infants and Young Children*. San Antonio, TX: Harcourt Assessment, 2004.
- James, Tracy. 2001. "Results of the Wave 1 Incentive Experiment in the 1996 Survey of Income and Program Participation." *Proceedings of the Section of Survey Research Methods*, 834-839. Alexandria, VA: American Statistical Association.
- Hack, Maureen, H. Gerry Taylor, Dennis Drotar, Mark Schluchter, Lydia Cartar, Deanne Wilson-Costello, Nancy Klein, Harriet Friedman, Nori Mercuri-Minich and Mary Morrow. 2005. "Poor Predictive Validity of the Bayley Scales of Infant Development for Cognitive Function of Extremely Low Birth Weight Children at School Age." *Pediatrics* 118, 2: 333-341.

Lewis, Michael and Harry McGurk. 1972. "Evaluation of infant intelligence: Infant intelligence scores--true or false?" *Science* 178:1174-1177.

Lewis, Michael and Harry McGurk. 1973. "Testing infant intelligence." *Science* 182:737.

Mack, Stephen, Vicki Huggins, Donald Keathley, and Mahdi Sundukchi. 1998. "Do Monetary Incentives Improve Response Rates in the Survey of Income and Program Participation?" *Proceedings of the Section on Survey Research Methods*, 529-534. Alexandria, VA: American Statistical Association.

Martin, Elizabeth, Denise Abreu, and Franklin Winters. 2001. "Money and Motive: Effects of Incentives on Panel Attrition in the Survey of Income and Program Participation." *Journal of Official Statistics* 17: 267-284.

Matheny, Adam P.. 1973. "Testing Infant Intelligence." *Science* 182: 734.

McCall, Robert B.. 1981. "Early Predictors of Later IQ: The Search Continues." *Intelligence* 5, 2: 141-147.

McGuigan, William M., Aphra R. Katzev, and Clara C. Pratt. 2003. "Multi-Level Determinants of Retention in a Home-Visiting Child Abuse Prevention Program." *Child Abuse & Neglect* 27: 363-380.

Michalopoulos, Charles, Anne Duggan, Virginia Knox, Jill H. Filene, Erika Lundquist, Emily K. Snell, Phaedra S. Corso, Justin B. Ingels, Sue Kim, and Magdalena Mello, 2011. ACF-OPRE Report 2011-16. *Design Options for the Home Visiting Evaluation: Draft Final Report*. Administration for Children and Families, U.S. Department of Health and Human Services, Washington, DC.

Nápoles-Springer AM, Santoyo-Olsson J, O'Brien H, Stewart AL. (2006). Using Cognitive Interviews to Develop Surveys in Diverse Populations. *Med Care*, 44(Suppl 3):S21-S30.

Tulsky, D., Zhu, J. & Ledbetter, M. (Eds.). *WAIS-III WMS-III Technical Manual (Wechsler Adult Intelligence Scale & Wechsler Memory Scale)*. (1997). Harcourt Brace & Company.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition*. San Antonio, TX: The Psychological Corporation.

Willis, Gordon B. 2005. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage.

Wilson, Ronald S.. 1983. "Testing Infant Intelligence." *Science* 182: 734-736.

**Maternal and Infant Home Visiting Program
Evaluation (MIHOPE)
0970 - 0402**

**Supporting Statement
Part B: Statistical Methods**

Updated July 2012

Submitted By:
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

7th Floor, West Aerospace Building
370 L'Enfant Promenade, SW
Washington, D.C. 20447

Project Officer:
Lauren Supplee

B1. Sampling

Exhibit B.1 summarizes the sample sizes for baseline data collection. Approximately twelve states including approximately 85 sites will be included in the evaluation. The average site will include 60 women (30 assigned to the home visiting program and 30 assigned to the control group) and approximately 6 home visitors, for a total of approximately 5,100 women and 510 home visitors. Each site is expected to have one or two home visiting supervisors and one program manager.

States and their local program sites will be selected for MIHOPE in 2012 based on a variety of characteristics including the type of home visiting model, geography, urbanicity, target population, and research feasibility. As described in Part A, the study team will be collecting information from states early in 2012.

From that information, a list of potential local programs will be compiled. Eligible local programs will meet several criteria: (1) having two or more years experience with one of the four evidence-based home visiting service models that were selected by at least 10 states receiving MIECHV funds, (2) excess demand for their services so that they can provide enough families for a control group, (3) the ability to enroll 30 families in their program over a period of about a year, and (4) locations where there are few other home visiting services in order to ensure a strong service differential between the program and control groups.

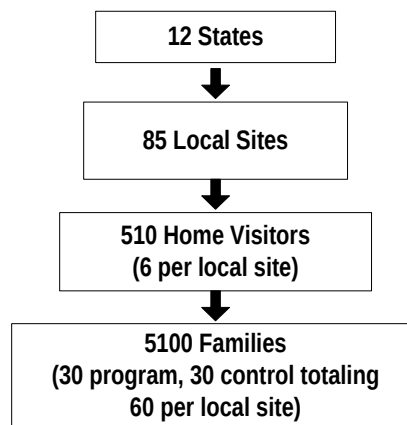
States will be classified in terms of which of four clusters of ACF/HRSA regions the state is in, the number of local sites that appear to be eligible for the evaluation, the urbanicity of the potential program sites, and the national service model of the potential program sites. Once this information is compiled, the study team will choose states so they meet the following criteria: each of four clusters of regions will be represented, the four evidence-based models are represented roughly evenly across the sites, and sites are as representative as possible of the urbanicity of all potential sites. Once 12 states are chosen, 85 sites will be chosen from within those states to meet the same criteria (for example, having the four evidence-based models represented roughly evenly across sites).

Within a site, the evaluation will enroll women who are pregnant or have a child under six months old. Home visiting programs will identify families who appear to be eligible for the study and a field staff person from the research team will go to the family's home to explain the study and obtain informed consent. Families will continue to be recruited until 60 families have been recruited in a site.

Within each state, the evaluation will conduct semi-structured group interviews with program managers and supervisors at one year following the state's recruitment into the study. The group interviews will include the program manager and all supervisors from each site participating in the evaluation. Each site will have only one program manager and most sites will have only one supervisor. Thus, the evaluation needs to include all program managers and supervisors in order to have representation of all sites in eliciting information to explain each site's quantitative results.

Within each state, the evaluation will conduct semi-structured group interviews with one third of the home visitors in each participating site and semi-structured individual interviews with another third of the home visitors in each participating site. The interviews will be carried out at one year following the state’s recruitment into the study. Thus, about 2 home visitors from each site will participate in the group interview and another 2 will participate in individual interviews. This sampling plan will allow for two group interviews with home visitors in each state, with about 7 home visitors in each group interview. This sampling plan will also allow for individual interviews within two different home visitors in each site, to permit examination of personal psychosocial attributes as factors for service delivery while holding site characteristics constant. The responses for the individual and group semi-structured interviews may be combined for some qualitative descriptions of how staff describe their roles and the program, with documentation that both individual and group interviews contributed to the conclusions drawn .

Exhibit B.1
Structure of National Home Visiting Evaluation



Statistical power. Exhibit B.2 shows the “minimum detectable effect” of this sampling plan for the full sample and for differences in impacts across subgroups. A minimum detectable effect is the smallest true effect that is likely to generate statistically significant estimated effects. For purposes of the design, calculations were performed to find the smallest effects that would generate statistically significant findings in 80 percent of studies with a similar design, using two-tailed t-tests with a 10 percent significance level. All results are presented as effect sizes, that is, in terms of the number of standard deviations of the outcome being examined. Results are presented both for administrative data, which would be available for all families, and for data such as surveys, which are assumed to be available for 80 percent of families.

Pooled sample. The minimum detectable effect for the pooled sample would be 0.058 standard deviations for administrative records and 0.065 for survey-based or observational outcomes. For example, if a site had a rate of child abuse and neglect of 20 percent in the control group, this design would have an 80 percent chance of finding a statistically significant impact if the true impact is a reduction of 2.3 percentage points (from 20.0 percent of the control group to 17.7 percent of the program group).

These pooled minimum detectable effects provide reasonable statistical power for the evaluation. For example, HomVEE found average effects of this magnitude or larger for four of the domains

of interest: child health, child development and school readiness, maternal health, and referrals and coordination.

Although families within a site may be more similar to one another than to families in other sites, this would not affect the statistical power of the pooled estimates or the subgroup estimates presented below. That is because individuals will be randomly assigned to the program or control group within a site.

Exhibit B.2
Minimum Detectable Effects of Proposed Sampling Plan

	Administrative data	Survey or observational data
Full sample	0.058	0.065
Subgroups (% in larger subgroup)		
50	0.117	0.130
60	0.119	0.133
70	0.127	0.142
80	0.146	0.163

Results are the smallest true impact that would generate statistically significant impact estimates in 80 percent of studies with a similar design using two-tailed t-tests with a 10 percent significance level. No adjustment for multiple comparisons is assumed. Results are based on fixed effects estimates. Administrative data are assumed available for all families, while survey or observational data would be available for 80 percent of families. Baseline data are assumed to explain 30 percent of variation in outcomes across families.

Subgroup differences. In addition to looking at the average effect across sites, the evaluation would assess whether home visiting had larger effects for some subgroups. For purposes of investigating the statistical power of subgroup estimates, it is assumed that the evaluation would be interested in detecting significant differences across subgroups. Since statistical power depends on the number of families in each subgroup, minimum detectable differences are presented for cases where 50, 60, 70, and 80 percent of the sample is in the larger of two subgroups. For a subgroup that divides the sample in half, for example, the minimum detectable differences are 0.117 standard deviations using administrative data and 0.130 using survey data. If 20 percent of control group families had a substantiated case of child abuse and neglect, the study would have an 80 percent chance of finding significantly larger effects for one subgroup than for another if the difference in true effects was 4.7 percentage points (for example, reducing child abuse and neglect by 4.7 percentage points for one subgroup but having no effect for the other subgroup). These minimum detectable differences increase gradually as the proportion of families in one subgroup increases. They are quite similar if 60 percent of families are in one subgroup, but they increase by 25 percent if 80 percent of families are in one subgroup.

Investigating the effect of program features. The evaluation will include 85 sites to allow it to explore the relationship between program features and program impacts. Program features could include any aspects of the community context, implementation system, service models, organizational influences, or home visitor characteristics. For example, this analysis could explore how program impacts vary with the duration of home visits, the background and training of home visitors, the support provided by supervisors for home visitors, the clarity of the goals of the local program, or the intended targets of the national model being used.

A framework for exploring the links between program features and program impacts is described in Greenberg, Meyer, Michalopoulos, and Wiseman (2003). Within this framework, the precision of the estimated relationships between program features and program impacts depends on a number of factors, including (1) the number of sites in the evaluation, (2) the precision of impact estimates within each site (which will increase with the number of families in the site), (3) the variation in characteristics across sites, (4) the number of program features to be investigated, and (5) how related the various program features are to each other. It is easier to detect differences by program feature if there are more sites, if there are more families in each site, if different sites vary more across the program feature being examined, if fewer program features are being examined at any one time, and if the program features are not closely related to one another. As an example of the last point, it may be very difficult to distinguish the effect of planned duration of home visits from the effect of actual duration, since the two are likely to be closely related in a particular site.

Exhibit B.3 shows the minimum detectable effects of program features for several scenarios. The upper half of the table shows results for a program feature that is binary and takes on one value in half of the sites and a different value in half of the sites. For example, half of the sites might plan to visit families weekly while half would visit only every other week. The lower half of the table shows results for a continuous program feature, such as how many weeks home visits would take place. In each panel, results are presented depending on whether 10, 20, or 30 program features would be examined at one time. As noted above, the ability to detect the effects of program features will worsen as more features are examined. Finally, results for each scenario are presented for three assumptions about how highly correlated various program features are with one another. As noted above, the ability to detect the effects of program features worsens as features become more highly correlated with one another.

Consider the first row of Exhibit B.3, which shows the case where 10 program features are being examined simultaneously and there is a low correlation across them. For outcomes measured using administrative data, the model would be able to detect differences of 0.203 standard deviations between sites of one type and sites of another type. If the overall effect on an outcome were 0.15 standard deviations, for example, the study would have an 80 percent chance of finding a statistically significant relationship between the program feature and impacts if the true impact were 0.252 standard deviations in one set of sites and 0.048 standard deviations in the other set of sites.

The ability to detect an effect of a program feature is only slightly worse if the features are more highly correlated or if 20 program features are being examined. The statistical power gets considerably worse, however, if more features are being examined and the correlation across features is high. For example, the minimum detectable difference is 0.317 standard deviation (for an effect of 0.309 standard deviation in one set of sites compared with -0.009 standard deviation in the second set of sites) if 20 program features are being examined and the correlation across them is high, and the minimum detectable difference is 0.348 standard deviation if 30 features are being examined and the correlation across them is medium.

Mother and Infant Home Visiting Program Evaluation

Exhibit B.3

Minimum Detectable Effects of Program Features

Type of variable	No. of variables representing program features	Correlation across program features	Administrative data	Survey or observational data
Binary, half of sites have the feature				
	10	Low	0.203	0.231
		Medium	0.213	0.243
		High	0.226	0.258
	20	Low	0.231	0.263
		Medium	0.264	0.300
		High	0.317	0.361
	30	Low	0.268	0.305
		Medium	0.348	0.397
		High	0.626	0.713
Continuous				
	10	Low	0.101	0.115
		Medium	0.107	0.122
		High	0.113	0.129
	20	Low	0.115	0.131
		Medium	0.132	0.150
		High	0.158	0.180
	30	Low	0.134	0.153
		Medium	0.174	0.198
		High	0.313	0.356

Results are the smallest true impact that would generate statistically significant impact estimates in 80 percent of studies with a similar design using two-tailed t-tests with a 10 percent significance level. No adjustment for multiple comparisons is assumed. Results are based on fixed effects estimates. Administrative data are assumed available for all families, while survey or observational data would be available for 80 percent of families. The correlation across program features is based on the R^2 statistic when one program feature is regressed on all other program features. For purposes of the calculations, a low correlation means the R^2 increases by .01 with every added feature, by .02 with every added program feature for a medium correlation, and by .03 for a high correlation.

The lower half of Exhibit B.3 shows minimum detectable effects if the program feature is continuous and normalized to have a variance of 1.0 standard deviation across sites. Because there can be greater variability in continuous variables than in binary ones, the design would have a greater ability to detect differences for such measures. For example, for a study examining 10 program features that are not highly correlated, the minimum detectable effect size of the program feature would be 0.101 standard deviation using administrative data and 0.115 standard deviation using survey data. Even for the most extreme case shown in the table — 30 highly correlated program features — the design could detect differences in impacts of 0.313 standard deviations using administrative data and 0.356 standard deviations using survey data.

These minimum detectable differences are well within the range found across previous studies of home visiting. For example, the HomVEE review found that prior studies of home visiting have produced impacts on positive parenting practices with a range of 0.82 standard deviations across studies (Michalopoulos et al. 2011). The range in impacts across prior studies is similar for other domains, including child maltreatment (range of 0.75 standard deviations), child health (0.93), child health and school readiness (.48), maternal health (1.14), and referrals or coordination (1.29). Although some of these differences are due to sampling error, a substantial portion of the differences are likely due to differences in program implementation. For example, a review of over 500 studies of prevention and health promotion programs for children and adolescents found that mean effect sizes were at least two to three times higher when programs were carefully implemented and were free of serious implementation problems (Durlak and Dupre 2008).

The Greenberg et al. framework underlying the calculations shown in Exhibit B.3 assumes that impacts are not correlated across sites. This may not be the case in MIHOPE because sites within a state will be funded through the state MIECHV grantee, which may exercise some control over the operation of local programs. The MIHOPE analysis will take this into account by adjusting the standard errors of estimated effects for such clustering.

It is difficult to say how such clustering will affect the statistical power of the analysis that links program features to program impacts. This is true for two reasons. First, there is little information about how similar sites within a state are likely to be in terms of their program implementation or, just as important, their effects on parent and child outcomes. Second, there is no well-established alternative to the Greenberg et al. framework that would provide an analytical derivation of the standard errors of the analysis linking program features to program impacts. For example, a similar analysis of mandatory welfare-to-work programs assumed program impacts were independent across welfare offices even when welfare offices were in the same city or county and run by the same agency (Bloom, Hill, and Riccio 2003).

One means of assessing the possible effect on statistical power of clustering of sites within a state is to assume the analysis of program features would include state “fixed effects” that would, in essence, base the analysis on variation of programs within a state but not use variation in programs across states. With an intraclass (intrastate) correlation in impacts of 0.01, this would increase the minimum detectable differences by about 10 percent, for example, from 0.20 in the first row of Exhibit B.3 to 0.22. With an intraclass correlation of 0.10, the minimum detectable effects would increase by 15-20 percent. Such differences in effects are still well within the range found in HomVEE.

Although typical levels of intraclass correlation would not affect the statistical power much, the study team will try to minimize the similarity of sites within a state by aiming to include states where local programs vary in features such as the evidence-based model that is being used, the urbanicity of the local site, and the type of local implementing agency.

Implementation data. As shown in Table A.2 above, implementation data will be used to answer several different types of research questions, requiring different types of analyses. They will be used to describe the local program models, their implementation systems, local staff, and service

delivered. Such analyses will rely on primarily on descriptive statistics such as means, medians, and ranges across the 85 sites, for which power calculations are not required. These descriptive analyses will be conducted for all sites combined as well as for the four program models separately. Implementation data will also be used in linear regression analyses or multi-level analyses to understand how program inputs (staff characteristics, organizational characteristics, implementation systems, community characteristics, and family characteristics) are associated with program outputs (service delivery). These will generally be conducted for all sites combined, so that we are analyzing approximately 510 home visitor-level estimates for analyses conducted at the level of individual staff and their service delivery behavior, or 85 site-level estimates for site-level analyses, and so on. Finally, the implementation data for each of the 85 sites will be combined with impact estimates to conduct analyses aimed at “getting inside the black box.” Power estimates for these analyses are described above.

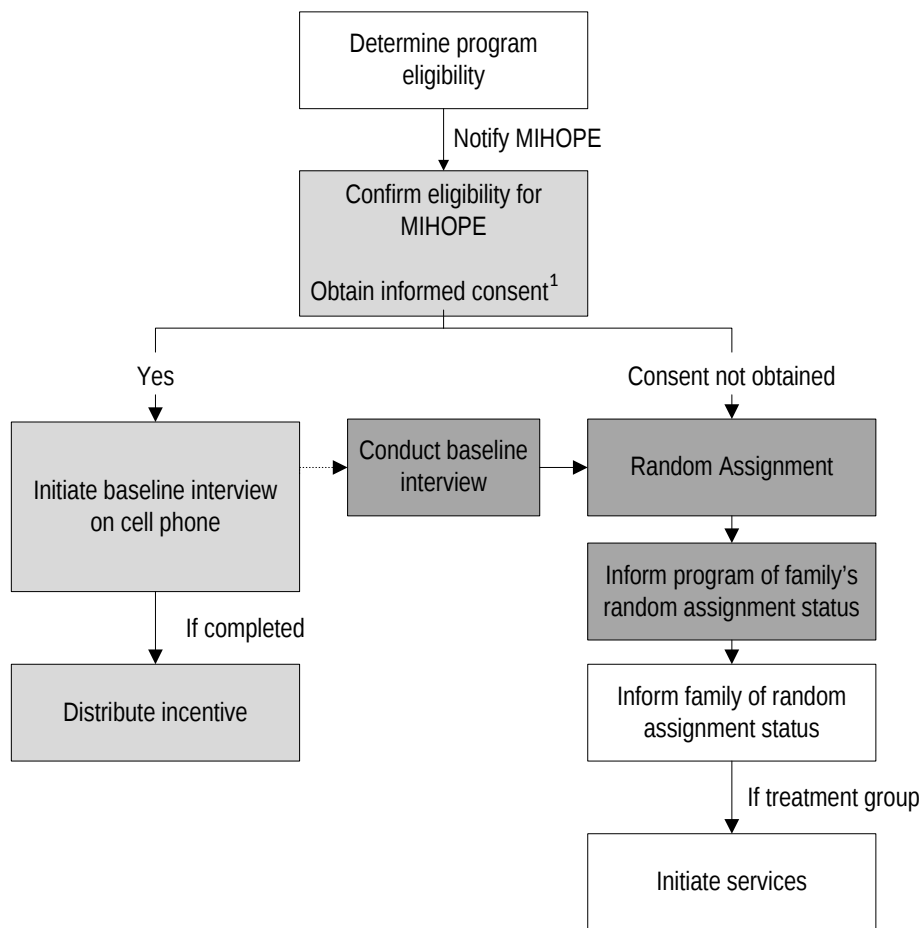
B2. Procedures for collection of information

This section focuses on procedures for quantitative data collection activities: the family baseline survey, the surveys of staff at participating home visiting program sites, and the surveys of administrators of community resources.

Family baseline survey. Exhibit B.4 depicts the process of collecting family baseline data. This process includes determining eligibility for the evaluation, contacting eligible women, and obtaining consent before conducting the family baseline survey. Steps taken to monitor the data quality are also briefly described.

Before recruiting women into the study, following the site’s normal procedures, the home visiting program will collect information to determine whether the woman is eligible for the program’s services. The program will also make an initial determination that the family is eligible for MIHOPE, for example because they have a child under six months old.

**Exhibit B.4
Roles in the MIHOPE Baseline Data Collection Approach**



Home Visiting Staff
 MIHOPE Field Staff
 MIHOPE Survey Operations Center

¹We anticipate that in 15 percent of cases, field staff will be present when eligibility is determined. In the other 85 percent of cases, field staff tasks take place during a separate visit.

For women who appear eligible for MIHOPE, the program will provide survey staff with the woman’s name, address, telephone number, and primary language, as well as the child’s name and date of birth (if the child has already been born).

During the visit, MIHOPE field staff will conduct the following procedures:

- Staff will distribute attractive introductory materials about MIHOPE (Attachment 5)
- Staff will introduce the study, provide a commitment to confidentiality, explain random assignment, and answer questions.
- Staff will attempt to obtain informed consent for the mother to participate in MIHOPE. Informed consent forms (Attachment 5) will reference the baseline and follow-up data collections and will allow the study team to collect state administrative

data on the family. Based on prior studies such as Baby FACES and studies of home visiting programs in Alaska and Hawaii, 90 percent of families are assumed to provide consent to participate in the study. Thus, the evaluation expects to describe the study and attempt to obtain consent from 5,667 families in order to enroll 5,100 families.

- If an applicant is a minor, it might be necessary to obtain consent from the parent as well, unless the state's emancipated minor laws make this unnecessary. This consent will be obtained by telephone if the parent of the minor is not in the home at the time of the consent process. The study team anticipates that 20 percent of program applicants will be minors. The protocol for obtaining consent from parents of minors is included as Attachment 5.
- If the woman consents to participate in random assignment and MIHOPE, she will be given a copy of the consent form. Field staff will then initiate computer-assisted telephone interviewing (CATI) via cell phone and then give the parent a cell phone to complete the 60-minute interview.
- If the mother has two or more children younger than six months old, survey staff will identify one as the focal child and all child-specific activities will be conducted with or about that focal child.
- While in the home, field staff will complete a selection of observational items about the internal and external physical home environment drawn from the HOME.
- At the completion of the interview, field staff will give the parent a \$25 gift card as a token of appreciation and a \$15 gift for the child.
- At the end of the CATI program, the woman will be randomly assigned to the program or control group.
- If the woman does not consent to participate in MIHOPE, random assignment will still be completed to determine if program services will be provided. This ensures that participation in the evaluation does not affect the family's ability to receive MIECHV services.
- The result of random assignment will be uploaded to a secure web-based system and an automated generic email will be sent to the home visiting point of contact to check the web-based system for that participant's random assignment status so that appropriate services can be initiated.
- Program staff will inform women whether they were assigned to receive program services. They will initiate home visiting services for women assigned to the program group and provide referrals to other community services for women assigned to the control group.

The proposed approach has a number of advantages.

- Having survey staff visit the family's home will help build rapport and maximize response rates for follow-up data collection.
- Having survey staff in the home allows the study to obtain written consent, which may be needed to obtain administrative records in some states.
- No burden is placed on home visitors to obtain consent or baseline data.

- Using CATI allows for low cost monitoring of data collection to ensure uniformity and data security (for example because data will not have to be transmitted back to the survey operation center).
- CATI can also accommodate complex instruments and different instruments for different families.
- Phone surveys also provide privacy to the parent in answering sensitive questions. For example, mothers will respond to the CATI survey questions with verbal responses or with a numerical value code, whichever they prefer. Neither the field interviewer nor anyone else in the room will know which survey question the mother is answering or what the question is.
- Field staff will collect information for the HOME while waiting for the survey to be completed. The HOME assessment is expected to take roughly 40 minutes to conduct.
- To facilitate answering of longer questions, the field staff will hand the woman a packet of color-coded show cards. The CATI interviewer will prompt the mother when a show card is needed and which show card to use. For example, the CATI interviewer will say, “For this next question, please take out the yellow show card identified with a G6 in the upper left hand corner. It contains the list of response options that will be used for the next several questions.”
- While the respondent is on the phone the field staff will be available for any questions the respondent may have, and to troubleshoot any technical difficulties with the completion of the CATI interview, for example, if the interview needs to be broken off, or if the call is dropped.
- Field staff can provide a gift card on the spot rather than making participants wait for a gift card to be mailed. This can increase the willingness to respond to the baseline survey.
- Field staff can monitor the infant for any needs while mother is on the phone doing the baseline interview (if there is an infant and if the infant is awake)
- Completing the baseline survey before random assignment ensures a 100% response rate. As noted above, we assume that 10 percent of women who are eligible for the study will decline to participate, so 90 percent of eligible women will complete the baseline survey.

One critique of CATI is that it is more difficult to establish rapport than when conducting the interview in-person. This is addressed by the presence of field staff in the home to introduce the study and obtain consent. Experienced field staff can build rapport to maximize enrollment and support a high response rate in subsequent follow up data collection activities. When possible, the same field staff will collect data during subsequent data collections.

An alternative to CATI is computer assisted personal interview (CAPI) or computer assisted self interview (CASI). CATI is being used for the family baseline survey because it has several advantages over the other methods:

- It provides greater confidentiality of responses for the woman compared with CAPI since others in the home could overhear a CAPI interview. Many items in the baseline survey are of a sensitive nature (domestic violence, drug and alcohol use, cigarette smoking, depression and other mental health issues, attitudes about parenting and attachment) so providing a confidential method of responding is critical. This is an advantage of CATI over CAPI, but both CATI and CASI interview methods provide

a confidential method for participants to answer sensitive questions without being overheard.

- It provides data security since data are collected and stored in a central secure location. CAPI and CASI data are stored on individual laptops and require field staff to upload the data regularly. If field staff do not upload their data, or their laptops are lost or stolen, data will be lost and security may be breached.
- It allows for real time monitoring of interviewers and higher data quality on the study. All telephone interviewers are recorded and ten percent of each interviewer's work is monitored in real time by supervisors. Feedback is given immediately following a monitoring session.
- The cost of developing and implementing a CATI survey is less than a CAPI or CASI survey. This includes the cost of the equipment (laptops) and the labor to program, upload and maintain the laptops.
- Telephone interviewers can halt or stop an interview if a mother needs a break. If needed, the telephone interviewer can schedule an appointment and call back at a more convenient time to complete the interview. This may be especially important for mothers with infants. This is less expensive than sending a field staff person back to the home to complete a CAPI or CASI survey.
- After completing the baseline interview, the CATI program will randomly assign the mother to the treatment or control group. Information on random assignment will be sent in real time to a secure web-based system that home visiting program staff can access, along with a generic email alert to check it so they can start services for treatment families right away. CAPI or CASI systems require that the data from the laptops be uploaded by the interviewer to a secure server and therefore rely on human transmission of the data. There is often a lag of a day or more with this process which would delay receipt of the random assignment status and sending the information to the home visiting program in a prompt fashion.

The study team has used this method on many large scale studies, such as FACES, Baby FACES, and BSF, sending field staff to participants' homes with cellular telephones to complete a survey via CATI. It is more efficient and cost effective than using CAPI or CASI.

Surveys of Staff at Participating Home Visiting Program Sites. Web-based surveys of the program manager, supervisors, and home visitors in each participating program site will be conducted near the time that the state enters the evaluation (baseline) and 12 months later. Site liaisons will notify each site's point of contact about two weeks prior to the targeted date for each survey to discuss the timeline and review survey procedures. Survey completion will be tracked using the management system. If a survey is not completed within one week of the targeted time-frame, the site liaison will follow up with the point of contact at the site to remind the staff member that the survey response is due. These instruments will be designed to preclude backtracking to change responses or printing of the survey.

Surveys of Administrators of Community Resources. Web-based surveys will be conducted with administrators of two types of community resources: a) services to which participating home visiting programs might make referrals relevant to MIECHV benchmarks and participant outcomes; and 2) home visiting programs not participating in the evaluation but serving the same

community. These surveys will be carried out with administrators of the organizations identified in the Program Manager Survey, Baseline, Part 1. For each community service provider and home visiting program identified, program managers will provide their primary contact's name, email address, telephone number, and street address. Web-based surveys of these administrators will be conducted between Parts 1 and 3 of the Program Manager Survey, Baseline. Administrators will be contacted by email with instructions about how to complete the web-based survey. Survey completion will be tracked using the management system. If a survey is not completed within one week of the targeted time-frame, research staff will send a reminder email with follow up by phone if needed. The survey instruments are designed to be completed in a single session of about 0.10 hours.

Logs Maintained by Supervisors and Home Visitors. Data about service delivery, training and supervision will be collected through weekly web-based logs. For sites in which supervisors and home visitors do not have regular access to the internet, paper versions of the logs will be offered. Home visitors and supervisors can complete the paper forms and a support person in the site can enter these data using the site's computers.

Supervisor Logs. Supervisors will use the web-based system to complete logs each week during the period in which home visitors are also completing logs. Supervisors will be prompted each week to complete a log for each of their home visitors that are participating in the study (about 5-8 home visitors). If the supervisor did not have a supervisory session with a particular home visitor, s/he will record the reason (for example, vacation or sick leave, scheduling conflict). The supervisor should enter all data for a given week no later than the end of the first workday of the following week.

Home Visit Logs. Home visit logs will be the major source of standardized data on actual service delivery to families. Home visitors will complete logs each week for the first 15 months of family enrollment. To ensure that log data are completed every week, home visitors will be asked to record information for every family enrolled in the evaluation and assigned to their caseload. If a family did not receive a visit that week, the home visitor will record the reason (for example, no visit was scheduled or a scheduled visit was cancelled). The home visitor should enter all data for a given week by the end of the first workday of the following week.

Group and Individual Interviews with Staff at Participating Home Visiting Program Sites. Within each state, group interviews of program managers and supervisors from participating program sites will be conducted about 12 months after the state's recruitment into the evaluation. All program managers and supervisors will participate in the group interviews.

Within each state, group and individual interviews of home visitors from participating program sites will also be conducted about 12 months after the state's recruitment into the evaluation. Within each site, one-third of the home visitors (about two home visitors per site) will be randomly selected for participation in the group interviews and another third (about two home visitors per site) will be randomly selected for participation in the individual interviews. The interviews will be carried out as part of the evaluation team's 12-month site visit to the state.

Site liaisons will notify each site's point of contact about one month prior to the site visit to discuss the timeline and to review group and individual interview procedures. Interview content will be audiorecorded and notes will be taken to document content.

B3. Maximizing response rates

This section focuses on strategies to maximize response rates for quantitative data collection activities: the family baseline interview, the surveys of staff at participating home visiting program sites, the surveys of administrators of community resources, and the logs maintained by supervisors and home visitors.

Family baseline survey. Minimizing sample attrition is of paramount concern for any longitudinal study. A number of techniques will be used to achieve high response rates:

- Establishing rapport with women at baseline to ensure consent
- Training field staff in respondent cooperation and refusal-avoidance techniques
- Ensuring privacy of participant information
- Providing adequate information about the study at the time participants are recruited
- Conducting random assignment after the baseline interview
- Designing surveys carefully with pretested questions that are easy to answer
- Providing an incentive for participation and to encourage participation in the follow-up survey
- Using MIHOPE's sample management system to track sample recruitment, response rates, and potential sample attrition

Field staff will be trained in how to establish rapport with and gain the trust of women they visit in order to secure their participation in the study. In-person contact at the beginning of the study will provide a solid basis for obtaining participants' cooperation and for tracking participants and ensuring high response rates for the follow-up data collection. Whenever possible, the same field staff will collect data during subsequent data collections, such as the follow-up survey. Field staff will also be trained in refusal-avoidance techniques. Participants will be assured that the information they provide will be secure, treated confidentially, and used only for research purposes. The family will receive a flier (Attachment 5) with information about the study, its importance, an estimated time line of when subsequent visits to the home for data collection will take place, and who the family can call with questions about the study.

Completing the baseline survey before random assignment ensures a 100 percent response rate. After a woman has agreed to participate, completed the baseline interview, and been randomly assigned, field staff will give her \$25 to thank her for completing the interview and a small toy or book for the child of \$15 value (if the child has been born at the time of the interview). Other methods of sample retention we propose are to send a birthday card to the parent on her birthday and to send the child a card when he or she reaches six months of age as a way of maintaining rapport with families and keeping their interest in the study. These mailings also provide additional opportunities outside of the tracking mailings to learn of address changes. Examples of these cards are included in Attachment 27.

The women will also be surveyed when her child is 15 months old. Tracking of study participants for follow-up data collection will begin with the initial visit. First, information will be gathered at the baseline interview to allow the study team to stay in touch with families until the follow-up interview is conducted. This information will include names, dates of birth, Social Security numbers (if possible), addresses, and telephone numbers (home and work) for the parents and detailed contact information for at least two relatives or friends who will know how to reach them in case we have difficulty doing so. Families will also be periodically sent cards that ask them to confirm or update their address and telephone information and return it in the self-addressed, postage-paid return envelope, and they will receive \$5 for completing and returning the card. (An example of this card is included as Attachment 26.) At the second interim contact, the card will also request the sampled child's date of birth for parents who were pregnant at the time of study enrollment. A tracking database will identify when families are due for their tracking letter and generate these materials for mailing. Letters returned as undeliverable will be sent to the survey unit's tracking department for locating and then remailed to the updated address. The study team will call families that do not return a card within three weeks of the mailing in an attempt to verify their contact information by telephone. The study team will contact the secondary contacts for families that we cannot reach by telephone in an attempt to locate them. This script is included as Attachment 26.

Surveys of staff at participating home visiting program sites. When a site enters the study, the research team will explain to program staff the importance of the web-based surveys for advancing the field of home visiting in general and the MIECHV program in particular. Staff will receive a \$30 gift card for each web-based survey they complete. Research staff will closely monitor data completion reports. If a survey is not completed within one week of the targeted time-frame, the site liaison will follow up with the point of contact at the site to remind the staff member that the survey response is due.

Surveys of administrators of community resources. When a site enters the study, the research team will explain the purpose and importance of the community resource survey to the program manager. Our targeting of administrators of community resources nominated by the program manager will maximize response rates as administrators will be more likely to respond to a survey about a program with which they work. Research staff will closely monitor data completion reports. They will send email reminders to administrators who do not complete the survey within a week of the initial contact. They will telephone administrators who do not complete the survey after three weekly email reminders. If research staff reach the administrator by phone, they will offer to complete the survey with the administrator by telephone.

Logs maintained by supervisors and home visitors. Strategies for maximizing response rates are the similar to those described above for the surveys of staff at participating home visiting program sites. When the site enters the study, the research team will explain to program staff the importance of the logs for advancing the field of home visiting in general and the MIECHV program in particular. Research staff will closely monitor weekly log completion reports. They will send program staff two weekly messages (Attachment 28). The first message will remind staff to complete their logs. The second message will document the data that were entered in the previous log by that staff person, thank the staff member for the data provided, and remind those who have not yet completed the previous week's log to do so.

Group and individual interviews with staff at participating home visiting program sites.

When a site enters the study, the research team will explain to program staff the importance of the interviews for advancing the field of home visiting in general and the MIECHV program in particular. The MIHOPE team's in-person presence for the group and individual interviews will also motivate strong staff engagement and participation. In past studies by MDRC and the other MIHOPE partners, program staff have been very willing to participate in in-person interviews (both group and individual) and have attended scheduled interviews at high rates.

B4. Pre-testing

This section focuses on pretesting of quantitative data collection activities: the family baseline interview, the surveys of staff at participating home visiting program sites, the surveys of administrators of community resources, and the logs maintained by supervisors and home visitors. Each type of pretest was conducted with 9 or fewer parents or home visiting staff. Therefore, we have not included pretests in the burden estimates.

Pretest of family baseline survey. The study team is conducting an iterative pretest of the baseline interview. The team will conduct two rounds of pretests. The first pretest occurred in April 2012. The second will occur no later than four weeks before initiation of sample recruitment and after OMB approval.

The first pretest was completed with six participants via telephone by two Mathematica staff experienced with cognitive interviewing techniques. Participants consisted of three pregnant women and three women with young infants ranging from 5 months old to 15 months old. All participants were Maryland residents currently enrolled in home visiting programs, two each from Early Head Start (EHS), Nurse-Family Partnership (NFP), and Parents as Teachers (PAT). Pretest participants were recruited with assistance from EHS, NFP, and PAT program staff. We were unable to make contact with a Healthy Families American staff member, so we did not recruit families from that program.

The pretest interviews began by introducing the survey and informing women that participating in the survey was voluntary and that the data collected would be kept confidential. Five of the six participants consented to audio recording of the interviews. The interviewers asked each survey question exactly as worded and followed-up with specific probes for prescribed questions or if questions appeared confusing to respondents during the interview. Interviews ended with a short debriefing to solicit feedback on the survey experience from participants. Interviews of pregnant women took an average of 45 minutes to administer and interviews of women with infants took an average of 58 minutes to administer.

As a result of the pretest, a number of changes were made to the instrument, as summarized in Appendix B. Most changes were designed to avoid respondent confusion. Several follow-up questions were added, for example, to learn how long a newborn had stayed in a neonatal intensive care unit. One question was eliminated because respondents could not distinguish this question from another one. Two questions about the use of mental health or substance use services were simplified by reducing a long list of options to six options.

Because the average pretest interview took less than one hour to complete, questions were added to respond to a comment from the Nurse Family Partnership that the project identify a subgroup with low psychological resources. Questions from the Pearlin mastery scale and the Wechsler Adult Intelligence Scales Similarities subtest were added for this purpose. A description of the properties of the Wechsler subtest is provided in Appendix C.

Following OMB approval, the instrument will be programmed for CATI format. The second pretest will focus on testing the CATI program, which enables us to test the flow and skip logic of the instrument and to refine our CATI data collection procedures. As with the first round of pretesting, cognitive interviews will be conducted with parents and interviewers will be debriefed. This iterative approach to pretesting helps to ensure that the programmed instrument is almost final, reducing the need for costly changes to programming specifications.

Pretest of Implementation Instruments. Pretesting was carried out in March – May 2012 in preparation for the launch of the study in summer 2012. Appendix D summarizes changes to instruments resulting from pretesting and public comments.

The objectives of pretesting were to:

1. Assess readability and understandability of instructions and questions;
2. Estimate and minimize the time needed for staff completion of each instrument;
3. Confirm that questions and response choices would adequately measure each construct in the study implementation model for each benchmark and participant outcome; and
4. Identify technical problems with web-based administration of the instruments

As a result of pretesting and our own commitment to minimizing respondent burden, nearly all of the instruments have been streamlined, thereby reducing length and eliminating unnecessary repetition across the instruments.

Pretesting focused on the web-based instruments, whose numbers and names are as follows:

- 09 Program Manager Survey Part 1
- 10 Program Manager Survey Part 2
- 11 Program Manager Survey Part 3 / Community Services Inventory
- 13 Supervisor Survey Baseline
- 15 Home Visitor Survey Baseline
- 19 Supervisor Logs
- 20 Home Visitor Logs

Several other implementation study instruments were edited to improve their clarity, minimize the time needed for completion, and assure that questions and response choices would adequately measure each construct. The semi-structured group and individual interviews were edited to delete items that were redundant with the baseline and 12-month web-based surveys and to identify optional items and potential probes. Only a subset of questions will be asked, with the exact subset to be determined by the specifics of the data collected in the other instruments completed by the participating sites. These edits were motivated by public comments and by the need to re-align the instruments with the modified web-based instruments.

1. Overview of Procedures

Each instrument was pretested with at least one staff member from each model included in the MIECHV evaluation: Early Head Start (EHS), Healthy Families America (HFA), Nurse Family Partnership (NFP), and Parents as Teachers (PAT). Each item was tested on nine or fewer people in total. As planned, we used an iterative approach to testing. One round of pretesting was carried out in March, and in April and May additional pretests were conducted to address problems identified in the earlier iterations.

2. Identification and Recruitment of Pretest Sites and Respondents

For pretesting, home visiting program contacts in Florida, Georgia, Maryland, New Jersey, and Washington state were identified. MIHOPE team members had prior relationships with these contacts, and they were thought to be amenable to participating in pre-testing. These contacts were sent introductory emails that described the pretesting opportunity, and they in turn put the team in touch with individual staff members who might be interested in taking part in pretesting. The first round of pretests occurred in Maryland, New Jersey, and Washington states. The subsequent rounds of pretests occurred in Georgia, Maryland, New Jersey, and Washington states.

3. Overview of Pretesting and Cognitive Interview Procedures

A pretest and cognitive interviewing protocol was developed based on best practices from the field (Willis, 2006; Napoles-Springer, Santoyo-Olsson, O'Brien, & Stewart, 2006) and the resources available for pretesting. The cognitive interviews were conducted by MIHOPE implementation study team members over the phone. Team members followed explicit protocols eliciting information to meet the pretesting objectives identified earlier in this section. The content of each cognitive interview was documented in an Excel database for analysis.

[Results from Pretest of Implementation Instruments](#)

The results of pretests are summarized in Appendix D.

B5. Consultants on statistical aspects of the design

There are no consultants on the statistical aspects of Phase 1. We have drawn on the expertise of team members including Charles Michalopoulos and Howard Bloom of MDRC.

Appendix A: Justification for Not Including Direct Child Assessments at Baseline

This memo discusses the potential child assessment measures that could be conducted, and presents our recommendations. The recommendations are informed by consultation with Sally Atkins-Burnett, Jerry West, and members of the Secretary’s Advisory Committee (SAC).

The recommendations are influenced by the three intended uses of the MIHOPE baseline family survey:

5. Describe the characteristics of families that participate in the study
6. Define the analytic subgroups that will be used in the impact analyses
7. Increase the precision of the impact estimates by including measures of key domains at baseline and follow up

The survey will be administered to pregnant women and women with children from birth to 6 months of age, the key groups targeted by the home visiting programs. At this time we do not know the relative proportion of each group, but estimate that approximately one-third to one-half will not be born at the time of the baseline survey. Of those that are born, it is likely that half will be newborns (0 to less than 3 months) and half will be between 3 and 6 months old at the time of the baseline survey. The baseline participant survey will collect data on baseline family characteristics from two data sources: a baseline interview and the observational items from the Home Observation for Measuring the Environment (HOME; Caldwell and Bradley 2003) assessment. The baseline interview will be conducted by computer assisted telephone interview (CATI) to preserve privacy of study participants and to increase the efficiency and security of data collection. The HOME assessment will be conducted by field staff after they obtain informed consent from the family and while the participant is completing the baseline interview on the telephone.

A. CHILD ASSESSMENT MEASURES

There are a number of child assessment measures that could potentially be used for this study at baseline. We list the measures below and some factors to consider in weighing the challenges and benefits of each one as a baseline measure.

ITSEA/BITSEA: This is a parent report measure of child social-emotional well-being and is being used on Baby FACES. However, it is only normed for children aged 12 to 36 months. This was confirmed in an email exchange with the developer, Margaret Briggs-Gowan, who noted that “purposely did not design the BITSEA for less than 12 months due to concern about implying that psychopathology might “exist” at such a young age.” Therefore we cannot use it for the baseline survey. We could potentially use it at follow up.

Bayley Scales of Infant Development (Bayley): This measure can be used as early as 1 month of age. Here we discuss five versions of the assessment, the Bayley-II and the short form based on it developed for the ECLS-B, and the Bayley-II screener, the Bayley-III, and the Bayley-III Screening Test. In addition, we summarize the Social-Emotional Scale included in the

Bayley-III, a parent-completed questionnaire based on the Greenspan Social-Emotional Growth Chart (Greenspan 2004).

Published in 1993, the norming sample from the **Bayley-II** is dated and no longer reflects the population of children in the United States. Short forms of the Bayley-II mental and motor scales for 9-month and 24-month old children (BSF-R, Andreassen and Fletcher 2005) were developed with considerable effort and expense for the ECLS-B to simplify administration and reduce data collection time. The BSF-R was developed in response to a much longer than expected administration time encountered during the 1999 field test of the full BSID-II. Development of the short form was also designed to address difficulties field staff had administering and scoring the items using the standardization rules specified by the test developer. The BSF-R took approximately 36 minutes when the children were 9 months old in the ECLS-B. It would take considerable measurement development and psychometric work to create a short form appropriate for the MIHOPE age range. The Bayley Infant Neurodevelopmental Screener (BINS; Aylward 1995) is based on the Bayley-II and screens infants between the ages of 3 and 24 months for neurological impairments and developmental delays. It takes between 5 and 10 minutes to administer but as a screener, it does not show much variation in typically developing children's development. In addition, it does not extend down to cover the birth through 3 month age range.

The Bayley-III (2006) has not been used in a large-scale national study and was not recommended for the Baby FACES study for that reason and because it has a new, untested approach to separately measuring different outcome domains and computing separate scale scores based on a relatively small number of items appropriate to each age range. The Bayley-III direct child assessment has been organized into three scales and five subtests: (1) the Cognitive Scale is comprised of one subtest, (2) the Language Scale is comprised of the Receptive Communication and Expressive Communication subtests, and (3) the Motor Scale is comprised of the Fine Motor and Gross Motor subtests. In addition, the Social-Emotional Scale and the Adaptive Behavior Scale are two separate parent-report questionnaires. Both questionnaires and any direct assessment items that require the interviewer to speak to the child or parent would have to be translated into Spanish, as neither the Bayley-II nor the Bayley-III are available in Spanish. Other important concerns include (1) the Bayley-III's length, (2) the fact that the test has been normed in English only, (3) the lack of data about how predictive the scales are when used with infants 0-6 months, and (4) the fact that each scale has only a few items in it (which may result in severe floor and ceiling effects). The Bayley-III Screening Test (for 1 to 42 months) maintains the same multi-scale structure of the direct assessments in the full Bayley-III with even fewer items included per subtest (which exacerbates floor and ceiling effects). Given that it is based on the Bayley-III, the same issues described above regarding the norming sample apply.

Information publicly accessible indicates that the National Children's Study (NCS) is piloting a short form of the Bayley-III in four locations across the country using procedures similar to what was done for the ECLS-B that focus on reducing the length of the assessment and increasing the reliability of the administration by field staff. Several consultants suggested that due to these multiple concerns, particularly the lack of predictive validity data and the fact that the NCS version is only expected to extend down to 6 months, which is not far enough for the

MIHOPE baseline (we saw reference to 6-month IRT scores in what was publicly available), the Bayley should not be included at baseline or follow up. The Bayley Infant Neurodevelopmental Screener (BINS; Aylward 1995) is based on the Bayley-II and screens infants between the ages of 3 and 24 months for neurological impairments and developmental delays. It takes between 5 and 10 minutes to administer but as a screener, it does not show much variation in typically developing children's development. In addition, it does not extend down to cover the birth through 3 month age range.

The Greenspan Social-Emotional Growth Chart (Greenspan 2004): This assessment is now part of the Bayley-III and is completed by the child's parent or primary caregiver. It is based on functional emotional milestones that correspond to 8 stages for children from birth to 42 months of age (Bayley 2006). One concern about this measure is the small norming sample and very small sample sizes included in it for ages 0-3, 4-5, and 6-9 months (89, 54, and 51, respectively). In addition, we do not believe it has been used in a large-scale national study of high-risk parents and children.

Mullen (1995): This measure can be used from birth through 68 months. However, the norming sample is out-dated and it is only available in English.

Three/Two Boxes/Bags Task and Coding System: This measure examines parenting constructs such as supportiveness, sensitivity, cognitive stimulation, intrusiveness, and negative regard. It also includes scales that examine child engagement of parent, sustained attention, and negativity toward parent. The semi-structured play task and variations of the original coding scheme by Deborah Vandell and colleagues have been used with children 14, 24, and 36 months old in a number of studies, including the Early Head Start Research and Evaluation project, Fragile Families, ECLS-B, and Baby FACES. It was used with children six months and older in the NICHD Study of Early Child Care and Youth Development and in the Early Head Start Newborn Study. Predictive validity data from use of the task and coding system from 0-6 months is scant. This type of task and coding system are being considered for the follow-up assessment with the full sample.

The Nursing Child Assessment Teaching Scale (NCATS) (1995): This observational measure of the quality of the caregiver-child teaching interaction for children from birth to 3 years of age assesses four parent and two child behaviors. The correlations of the total NCATS scores with the total HOME score among children ages 1 to 36 months, in three age groups, ranged from .41 to .44. Given that the HOME is already planned for MIHOPE, NCATS may not add much additional information given the relative cost of training on the assessment. In addition, the adaptations to shorten the observation period made for administering the NCATS in the EHS-REP revealed internal consistency reliability problems inherent in large-scale live or videotaped coding and administration of the measure. There is scant information available about the predictive validity of the NCATS Teaching Task when conducted with children less than 6 months old.

Brazleton (1973): This measure is used with infants and usually in hospital settings. It is a scale for 0-2 months of age, so its use for this study is limited as our sample at baseline will include children 0 to 6 months of age.

Neonatal Intensive Care Unit Network Neurobehavioral Scale (NNS) (2004). This neurological assessment can be conducted from birth through 48 weeks. The infant should start off in a sleep state that has been maintained for at least 45 minutes. There are 115 items and several position changes are required during which the observer looks for changes in the baby. This assessment requires a highly trained individual, usually a clinician, and does not seem to be suitable for a large-scale study. Although there are a few published articles on the measure, there is little information available on its predictive validity and it has been used primarily for clinical purposes.

Other ECLS-B 9-Month Assessments: The remaining set of measures used at 9 months assesses infant physical development, including weight, length, upper arm circumference, and head circumference. Although a direct assessment may be desirable, we will be getting most of this information from other sources.

B. RECOMMENDATIONS FOR CONDUCTING DIRECT CHILD ASSESSMENT

In developing the MIHOPE baseline survey, we focused on including measures of outcome domains that are most likely to show impacts or that had the potential to mediate or moderate impacts. Direct child assessments of the portion of the sample that includes infants that were born at baseline were considered but rejected because they could only be administered for part of the sample (unborn children would have no data for these measures) and because the developmental experts we consulted with and our SAC recommended against directly assessing children 12 months of age or younger.

For over forty years, the predictive validity of infant assessments, particularly those administered to children less than one year of age, has been an issue for the field of developmental psychology. In the 1970's and 1980's, leading developmentalists debated this issue related to performance of children less than 1 year old on the Bayley and correlations to subsequent cognitive functioning (Lewis and McGurk 1972; Lewis and McGurk 1973; Matheny 1973; McCall 1981; Wilson 1973). Then, as now, researchers have concerns about the predictive validity of assessments conducted with young infants and generally recommend they be used for assessing performance at a given point in time for diagnostic and comparative purposes rather than as predictors of later skills and abilities (for example, Hack et al. 2005). A few measures of information processing for children less than 6 months old have been identified as somewhat more robust predictors to intelligence at 3 years of age (for example, the Fagan Test of Infant Intelligence 2005), but they have not been used in large-scale research projects and are more suitable to laboratory settings than to in-home assessment. The primary arguments against conducting direct child assessments stem from the lack of reliable and valid measures in early infancy; overall, the predictive validity of the measures that are available is either unknown or quite low.

After weighing the information above against the practical issues such as cost, we do not recommend conducting direct child assessments on the MIHOPE study for the following reasons:

8. **Sample Size and Variation.** About one-third to one-half of the sample at baseline will be pregnant women, so we would be able to obtain child assessment data for only part of our sample. The sample of children at baseline will also vary widely, with ages ranging from 1 day old to 6 months. There are few child assessment measures that are suitable for this age group.
9. **Cost.** The cost of conducting child assessments would be high and would require more funds than what have been allocated for the baseline effort. We would need to hire and train a group of staff with experience in complicated direct child assessments. We would need to pay more per hour since they will be doing work that is more difficult. Training will take substantially longer than what we budgeted (4 plus days rather than 2 days). Certification on the measures would be difficult and many staff would not pass, which would require additional hiring, training, and certification.
10. **Logistics.** The logistics of conducting assessments would be more challenging. The baseline visit would be longer, since the field staff would be conducting an assessment. We would need the infant to be awake, which could necessitate going back to the home multiple times to complete the assessment. These logistical considerations would also increase the cost of the baseline data collection.
11. **Low Return on Investment.** There is generally low predictive value of the standard child assessment measures at very young ages (birth to 6 months). We do not believe that the data gathered would provide us with adequate information to make the effort worthwhile. In addition, the measures that could be used at both baseline and follow-up are few and have the limitations described above.

Appendix B: Summary of Changes Made to Family Baseline Survey

Survey Item	Change resulting from pretesting	Rationale
<p>A7 After [CHILD] was born, how long did [he/she] stay in the hospital?</p> <p>A8 After your baby [CHILD] was born, was [he/she] put in an intensive care unit or NICU?</p>	Revised A8 to ask if any of these days were in the NICU, and then if yes, ask for number of days child spent in the NICU.	Will help to clarify how long the baby spent in the NICU.
A13 Do you have a plan to breastfeed?	Revised to “Do you plan to breastfeed?”	Respondents had difficulty with the word “plan.” They often responded with “I hope to” or “I’d like to.” Revised wording will help match respondent’s intent.
A14 How long do you plan to breastfeed?	Revised to “How long would you like to breastfeed?”	
A15 How old was [CHILD] the first time (he/she) ate or drank anything other than (breast milk or) formula?	Replaced with the following item from the ECLS-B 9-month parent interview: “How old was [CHILD] in months when solid food was first introduced? Solid foods include cereal and baby food in jars, but not finger foods.”	Revised wording is more specific to ensure respondent understands the question.
B1 The next questions are about your health. In general, would you say your health is...?	For pregnant women, revised to “The next questions are about your health before your current pregnancy. In general, would you say your health is...?”	To clarify for pregnant women that they should answer about their health before pregnancy, so they do not consider any pregnancy-related ailments when responding.
B5 During (this pregnancy/your pregnancy with [CHILD]), were you told by a doctor, nurse, or other health care worker that you had gestational diabetes (diabetes that started during this pregnancy)?	Add response option, “haven’t been tested yet” for pregnant women.	It is possible that some women may not be far enough along in their pregnancy to have been tested for gestational diabetes.
B8 Is there a place you go for general health care, if you are sick or need advice about your health - that is, any care except prenatal care or family	Added the follow-up item: B8a. What kind of place do you go?	Not all pretest respondents knew that we were asking about a physical location.

Survey Item	Change resulting from pretesting	Rationale
planning?	Clinic Health Center Hospital Doctor's office Some other place	
<p>B9 During the past year, have you ever received family planning or gynecologic services?</p> <p>B9a During the past year, did you ever want or need family planning or gynecologic services?</p> <p>B9b What is the main reason you didn't receive family planning or gynecologic services?</p> <p>B9c Are you currently receiving family planning or gynecologic services?</p>	<p>Replaced with the following items:</p> <p>B9. Is there a place you go, or have gone, for family planning or birth control?</p> <p>B9a. What kind of place do you go/ did you go?</p> <p>The same place I receive general health care Clinic Health Center Hospital Doctor's office Some other place</p>	Some respondents were confused by term "family planning services."
B10 How many more children do you plan to have?	Revised to "How many more children would you like to have?"	Respondents had difficulty with the word "plan." They often responded with "I hope to" or "I'd like to." Revised wording will help match respondent's intent.
C8 What is the highest grade or year of regular school that you have completed?	Removed "regular" from question	Respondents were confused by the term "regular."
Section D items on the woman's spouse or partner	If a woman doesn't have a spouse or partner and doesn't live with the child's biological father, added an item asking if the woman and biological father ever lived together. Added an item asking if the woman is currently in a romantic relationship, and for those who say yes, then ask the intimate partner violence items.	This section didn't flow well during the pretest. This revision will help fill in missing information

Survey Item	Change resulting from pretesting	Rationale
Section E items on household composition and earnings	Revised items about household composition and earnings to accommodate respondents whose household composition is currently different than it was for most of the previous year.	These questions were difficult for respondents to answer if the current household members were not the same as in the prior year (when we ask for total earnings from all household members.) changing these items will make answering them easier for the respondent.
<p>E4 How many months were you employed (did you work for pay) during the past 3 years (including your current job)?</p> <p>RESPONDENT DIDN'T WORK</p> <p>Less than 6 months 7 to 12 MONTHS 13 to 24 MONTHS More than 24 months</p>	Changed format so that interviewer reads the answer choices aloud to respondent, except for “respondent didn’t work.”	Pretest respondents had trouble calculating number of months; providing answer choices helped them respond.
E19 During the past year, have you received Early Head Start or child care services for [CHILD]?	Revised.	Respondents were confused and wondered if we meant EHS only or child care in general.
<p>E20 During the past year, have you ever received Early Intervention services or [INSERT NAME OF PROGRAM FOR STATE] for (CHILD)?</p> <p>E20a Did you ever want or need Early Intervention services for [CHILD]?</p> <p>E20b Are you currently receiving Early Intervention services for [CHILD]?</p>	Deleted from survey.	Since most babies will be too young to have received early intervention services at baseline, we recommend deleting this question and including it in the follow-up survey.
E21/22 a-c Home visiting items	Moved to end of survey, just before contact information.	Moving the home visiting questions to the end eases the transition from the

Survey Item	Change resulting from pretesting	Rationale
		end of the survey to collecting contact information.
E22a-c What do you think will be the three most important benefits of home visiting for you and your family?	Deleted these items.	Responses here were the same as those captured in E21.
F15-F18 questions on receipt of mental health and substance abuse treatment services during past year	Shortened the list by grouping similar items together and using broader categories	The list of items was long, categories were redundant, and the list was cumbersome to administer
<p>G6 Please tell me whether you or any other members of your household received income from the following sources in the past month. This includes anyone who you support and/or supports you and lives in your household.</p> <p>G7 During the past year, have you ever received help in applying for public benefits, including TANF, SNAP, or WIC?</p>	Added “WIC” to the list of sources in G6.	Four respondents said yes to G7 because they received WIC, but didn’t understand that the question was asking if they had received <i>help in applying</i> for services like WIC. Add WIC to G6 to capture this benefit.
	Added questions from the Pearlin mastery scale	Added in response to an NFP comment suggesting the measurement of low psychological resources.
	Added questions from the Wechsler Adult Intelligence Scale Similarities subtest	Added in response to an NFP comment suggesting the measurement of low psychological resources.

Appendix C: Measuring Cognitive Ability

To measure cognitive ability, the MIHOPE baseline survey will contain the Similarities subtest of the Wechsler Adult Intelligence Scales – Third Edition (WAIS-III; Wechsler, 1997). The Similarities subtest is designed to capture abstract reasoning and verbal comprehension abilities, which are two principal dimensions of intellectual abilities (Flanagan and Harrison, 2005; Flanagan, Ortiz and Alfonso, 2007). In the Similarities subtest, respondents are asked a series of questions about how two things are alike. For example, “How are a snake and an alligator alike?” Each item is then scored on a 0 to 2 scale according to general scoring principles and examples that are provided in the testing manual.

This measure is proposed to assess parents’ cognitive and intellectual abilities for a variety of reasons:

- The Wechsler Adult Intelligence Scales are among the most widely used measure of intellectual abilities in the United States and in other countries. The WAIS-III Similarities subtest is also one of the few measures of abstract reasoning and verbal comprehension that is available in both English and Spanish that can be readily administered over the telephone or in person.
- Compared with most other assessments of intellectual abilities, the Similarities subtest is relatively brief – consisting of only 18 items – which places substantially less burden on study participants than most other measures of cognitive and intellectual abilities. Furthermore, study participants need not receive all of the items because the testing includes a discontinuation rule when respondents get three consecutive items incorrect. Thus, the amount of time required to administer the subtest can be quite brief and varies the study participants’ intellectual aptitude thereby reducing the burden of the measure on study participants.
- The English and Spanish versions of the Similarities subtest have been shown to have good psychometric properties. The publishers of the English version of the subtest found that its split-half reliability is 0.87, the test-retest reliability is 0.83, and the inter-rater agreement on scoring the items of the subtest (ICCs) is 0.93 (Tulsky et al., 1997). Elsewhere, Renteria et al. (2008) found the Spanish WAIS-III Similarities subtest had an internal consistency of 0.79 using a sample of primarily Spanish-speaking adults recruited from Chicago neighborhoods.
- The Similarities subtest has been shown to have good validity and demonstrated capabilities for differentiating individuals with qualitatively different levels of intellectual abilities. In numerous studies the Similarities subtest is a strong predictor of the full-scale score of intellectual functioning that can be created when the full battery of subtests from the WAIS-III. Jones et al. (2006), for example, found that the Similarities subtest loads onto the WAIS full-scale score of overall intelligence at 0.81 in a factor analytic model. Moreover, using a sample of adults who are diagnosed with mild intellectual disabilities according to the DSM-IV-TR criteria for intellectual disabilities (e.g., IQs of 40 – 70), the publishers found that this group on the Similarities subtest scored about 2.5 standard

deviations lower than a matched comparison group with average intelligence (Tulsky et al., 1997). Using a sample of adults who meet the DSM-IV-TR criteria for Borderline Intellectual Functioning (e.g., IQs of 71 – 84), the publishers also found that the group scored about 1.4 standard deviations lower on the Similarities subtest than a matched comparison group with average intelligence (Tulsky et al., 1997).

Appendix D: Implementation Study Instruments – Content in Paired Instruments and Revisions per Pretesting and in Response to Public Comments

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
State administrator interview			
Baseline (7)	The baseline survey gathers data on MIECHV- and state-level factors for service delivery, from the perspective of the state’s lead agency for MIECHV.	Sections K and L were reformatted to improve clarity.	<i>Comments:</i> None
12 Month (8)	The content of the 12-month interview parallels that of the baseline interview. Items elicit information on changes in factors since the baseline survey.	The 12 month interview was edited to align with the revised baseline instrument.	<i>Comments:</i> None
.....			
Program manager survey			
Part 1, Baseline (9)	The content of each of the three parts of the baseline survey is unique. The three parts are complementary. Together, they gather baseline data on the full set of hypothesized program site factors for service delivery, from the perspective of site	As possible, sections on site policies and procedures were edited to make data collection more efficient by using questions about policies in lieu of requests for copies of the policies. Items on current staff were moved to Part 2 because they fit better with its content.	<i>Comment:</i> It is unclear which survey instruments will be completed by a program manager who is also a supervisor. <i>Response:</i> A program manager who is also a supervisor will complete the program manager survey and sections of the supervisor survey that are not redundant with the program manager survey.
Part 2, Baseline (10)		Items that could be answered more efficiently via other instruments were	<i>Comments:</i> None

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
Part 3, Baseline (11)	leadership.	<p>eliminated.</p> <p>Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.</p> <p>A few items were added to fill identified gaps and eliminate ambiguity in responses.</p> <p>Items on referrals to community resources were moved to Part 3 because they fit better there.</p> <p>Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.</p> <p>A few items were added to fill identified gaps and eliminate ambiguity in responses.</p>	<p><i>Comment:</i> Questions about referral are redundant with questions in the supervisor survey.</p> <p><i>Response:</i> We have eliminated this redundancy by dropping these questions from the supervisor survey. The questions are now a part of only the program manager survey. A site can choose to have a supervisor or other staff member help the program manager answer these questions if the site feels that is more efficient.</p>
12 Month (12)	<p>The content of the 12-month survey parallels that of parts 1 and 2 and a small portion of part 3 of the baseline survey. Thus, comparison of responses from baseline to the 12 month survey allows assessment of change over time.</p>	<p>The 12 month survey was edited to align with the revised baseline instrument.</p>	<p><i>Comments:</i> None</p>
Supervisor survey Baseline (13)	<p>The baseline survey gathers data on hypothesized program</p>	<p>Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.</p>	<p><i>Comment:</i> It is unclear whether a supervisor who is also a home visitor will complete both or only one survey.</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
12 Month (14)	<p>site factors for service delivery from the perspective of supervisors, and on supervisor-specific factors for service delivery.</p> <p>The content of the 12-month survey parallels that of the baseline survey. Thus, comparison of responses from baseline to the 12 month survey allows assessment of change over time.</p>	<p>Items that could be answered more efficiently via other instruments were eliminated.</p> <p>In Sections L-S, items were reorganized, reworded, and some items were eliminated to improve efficiency.</p> <p>Some items were added to fill identified gaps and to eliminate ambiguity in responses.</p> <p>The 12 month survey was edited to align with the revised baseline instrument.</p>	<p><i>Response:</i> A supervisor who is also a home visitor will complete the supervisor survey and portions of the home visitor survey that are not redundant with the supervisor survey.</p> <p><i>Comment:</i> It is unclear which survey instruments will be completed by a replacement supervisor.</p> <p><i>Response:</i> A replacement supervisor will complete a baseline survey upon joining the study. S/he will also complete the 12 month survey if s/he joins the study at least 6 months prior to the 12 month survey.</p> <p><i>Comment:</i> Both the Baseline and the 12 month surveys ask about program expectations, which is unnecessarily repetitious.</p> <p><i>Response:</i> We have deleted a few of the redundant items. Redundancies are by design, to capture expected site-level changes in program models and implementation systems over time. The MIECHV program has already given rise to substantial changes in home visiting at the national, state, local and program site levels. We expect this will continue in the years ahead. Thus, we have designed the 12-month staff surveys to assess changes in both organization- and individual-level factors for service delivery.</p>
Home visitor survey Baseline (15)	<p>The baseline survey gathers data on hypothesized program site factors for service delivery from the perspective of home visitors, and on home visitor-specific factors</p>	<p>Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.</p> <p>Items that could be answered more efficiently via other instruments were eliminated.</p> <p>In Sections L-S, items were reorganized,</p>	<p><i>Comment:</i> It is unclear which survey instruments will be completed by a replacement home visitor.</p> <p><i>Response:</i> A replacement home visitor will complete a baseline survey upon joining the study. S/he will also complete the 12 month survey if s/he joins the study at least 6 months prior to the 12 month survey.</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
	for service delivery.	<p>reworded, and some items were eliminated to improve efficiency.</p> <p>Some items were added to fill identified gaps and eliminate ambiguity in responses.</p>	<p><i>Comment:</i> The home visitor baseline survey remains lengthy.</p> <p><i>Response:</i> Editing as part of pretesting has reduced the number of items by about 20%. Pretesting has established that home visitors can complete the survey within the projected time.</p> <p><i>Comment:</i> 105 items are embedded, not fully shown.</p> <p><i>Response:</i> The source instrument, which is proprietary, was identified by name – the Organizational Social Context (OSC) scales. The commenting organization is familiar with this instrument, having reviewed its items and approved its use in December, 2011 for another home visiting study conducted by MIHOPE team members, in which its sites participate.</p> <p><i>Comment:</i> There was concern that the number of items measuring home visitor psychosocial functioning (n=105 + 39) is burdensome and intrusive.</p> <p><i>Response:</i> This section of the survey includes three instruments: the OSC (105 items), the short form of the CES-D (10 items), and the Attachment Style Questionnaire (29 items). We did not change this section, for several reasons. First, the three instruments in this section measure different constructs, all of which are hypothesized to have independent influences on service delivery and impact. There is theoretical and empirical support for the independence influence of each of these constructs on service delivery and impact. Second, the OSC measures not only individual level factors</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
12 Month (16)	The content of the 12-month survey parallels that of the baseline survey. Thus, comparison of responses from baseline to the 12	The 12 month survey was edited to align with the revised baseline instrument.	<p>(morale and burnout) but is the study primary measure of two key organization-level factors (culture and climate). Third, depressive symptoms and relationship security have been shown to influence service delivery, <i>and</i> to have interactive effects on family engagement. Fourth, leaders of other evidence-based home visiting models expressed their support for assessing staff psychosocial well-being at the ACF/HRSA-sponsored MIHOPE meeting of model developers on October 27, 2011.</p> <p><i>Comment:</i> Questions about home visitors' background as a parent or home visiting recipient seem judgmental.</p> <p><i>Response:</i> These items have been deleted.</p> <p><i>Comment:</i> Questions about referral are redundant with questions in the program manager survey.</p> <p><i>Response:</i> We have kept the referral questions in both instruments. The questions are similar by design, but they serve different purposes. We use answers to referral questions in the program manager survey to assess the site's awareness of and relationship with community resources. We use answers to referral questions in the home visitor survey to measure each home visitor's knowledge of, attitudes toward, and interactions with community resources.</p> <p><i>Comment:</i> Both the Baseline and the 12 month surveys ask about program expectations, which is unnecessarily repetitious.</p> <p><i>Response:</i> We have deleted a few of the redundant items. Redundancies are by design, to capture</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
	month survey allows assessment of change over time.		expected site-level changes in program models and implementation systems over time. The MIECHV program has already given rise to substantial changes in home visiting at the national, state, local and program site levels. We expect this will continue in the years ahead. Thus, we have designed the 12-month staff surveys to assess changes in both organization- and individual-level factors for service delivery.
Community service provider survey (17)	<p>This survey is conducted at baseline only.</p> <p>Its content parallels that of Part 3 of the program manager baseline survey for each type of service provider listed.</p> <p>It elicits the community service provider's perspective on referral and coordination with a specific home visiting site and on service availability, service accessibility and inter-agency agreements as factors for referral and coordination.</p>	<p>We did not pretest this instrument.</p> <p>Two items were added to the survey to address identified gaps (agency address and cost of services).</p> <p>Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.</p>	<u>Comments:</u> None
Other home visiting program survey (18)	This survey is conducted at baseline only. It documents key characteristics of other home visiting or	We did not pretest this instrument.	<u>Comments:</u> None

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
	parenting programs for infants in the community in which control group members might enroll.		
Supervisor logs (19)	These logs are completed weekly to measure supervisor training and actual supervision from the perspective of the supervisor as factors that influence actual service delivery.	Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.	<p><i>Comment:</i> The logs are burdensome because staff are expected to complete them weekly and because they are duplicative of forms that staff complete routinely as part of (NFP) model requirements.</p> <p><i>Response:</i> We reduced the number of items. We expect that each supervisor will complete weekly logs only for home visitors with one or more active families participating in the evaluation. For this reason, repetitiveness is limited.</p> <p>Although the content of the MIHOPE logs overlaps slightly with NFP logs, most items in the MIHOPE logs ask for content different than that in NFP logs.</p> <p><i>Comment:</i> The frequency of log completion should be reconsidered, perhaps to a monthly summative reporting across all home visitors.</p> <p><i>Response:</i> To understand variation in actual services to families and factors that influence service delivery, the study must collect uniform information across all outcome domains for all models and program sites. The logs provide key information about individual-level service delivery and supervision for “black box” analyses as well as for documenting variations in program costs for participant subgroups. No national model requires sites to collect the full set of supervision variables needed for MIHOPE; some sites might not collect any of this information in a systematic way. Our previous research using logs suggests that less</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
Home visitor logs (20)	These logs are completed weekly to measure actual service delivery and home visitor perspectives on actual training and supervision as factors for service delivery.	<p>Items on approaches to service delivery within each content area were dropped to reduce respondent burden.</p> <p>Items were reworded as needed to improve clarity and to maintain alignment with parallel items in other instruments.</p>	<p>frequent completion will negatively impact staff recall of events. Our previous research highlights substantial variability in the intensity and content of both home visits and supervision. We need to measure supervision at the home visitor level and service delivery at the client level. These measures will be key variables in analyses factors explaining variations in service delivery and fidelity. Variation in service delivery and fidelity will, in turn, be tested as a moderator of program impacts.</p> <p><i>Comment:</i> The logs are burdensome because staff are expected to complete them weekly and because they are duplicative of forms that staff complete routinely as part of (NFP) model requirements.</p> <p><i>Response:</i> We have reduced the number of items in the logs. Home visitors will complete weekly logs only for active families participating in the evaluation. On average, this will be only about five families, a small portion of the home visitor’s caseload. For this reason, repetitiveness is limited. Although the content of the MIHOPE logs overlaps slightly with NFP logs, most items in the MIHOPE logs ask for content different than that in NFP logs.</p> <p><i>Comment:</i> The frequency of log completion should be reconsidered, perhaps to a monthly summative reporting across all home visitors.</p> <p><i>Response:</i> To understand variation in actual services to families and factors that influence service delivery, the study must collect uniform information across all outcome domains for all models and program sites. The logs provide key information about individual-level service delivery and supervision for “black box” analyses as well as for</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
<p>documenting variations in program costs for participant subgroups. No national model requires sites to collect the full set of supervision variables needed for MIHOPE; some sites might not collect any of this information in a systematic way. Our previous research using logs suggests that less frequent completion will negatively impact staff recall of events. Our previous research highlights substantial variability in the intensity and content of both home visits and supervision. We need to measure supervision at the home visitor level and service delivery at the client level. These measures will be key variables in analyses factors explaining variations in service delivery and fidelity. Variation in service delivery and fidelity will, in turn, be tested as a moderator of program impacts.</p>			
<p>Semi-Structured Interviews</p>			
<p>Group interview – program managers (21)</p>	<p>These group interviews are conducted at 12 months to elicit staff perspectives for interpreting data collected in the surveys and logs, that is to explain the how and why behind quantitative results.</p>	<p>For group interviews with program managers, supervisors and home visitors, we deleted items that were redundant with the staff surveys, added a few questions to fill identified gaps, and edited questions to elicit participants’ perspectives on the reasons and mechanisms for results obtained through the surveys.</p>	<p><i>Comment:</i> There is considerable duplication of questions across the 12 month surveys and interviews.</p>
<p>Group interview – supervisors (22)</p>			<p><i>Response:</i> We have eliminated the Interview participant questionnaire (formally Instrument 24), as it was duplicative of items asked on the baseline surveys.</p>
<p>Group interview – home visitors (23)</p>			<p>We deleted items from the group and individual home visitor interview instruments that were redundant with the baseline and 12 month surveys (Instruments 13-16).</p>
<p>Interview participant questionnaire (24)</p>	<p>This questionnaire elicits basic information to characterize group interview participants</p>	<p>This instrument has been eliminated .</p>	<p>In instruments for both the group and individual interviews, most items are, in fact, either optional or</p>

Instrument (Number)	Comparison of Content in Paired Instruments	Revisions Resulting from Pretesting	Response to Public Comments
Individual interview – home visitors (25)	<p>(Instruments 21-23)</p> <p>These individual interviews are conducted at 12 months to elicit staff perspectives for interpreting data collected in the surveys and logs, that is to explain the how and why behind quantitative results.</p> <p>The individual interviews seek to elicit views that home visitors are less likely to share candidly in group interviews.</p>	<p>For the individual interviews with home visitors, we deleted items that could be answered adequately in the group interviews, added a few questions to fill identified gaps, and edited questions to elicit participants’ perspectives on the reasons and mechanisms for results obtained through the surveys.</p>	<p>potential probes. We will ask only a subset of questions, with the exact subset to be determined by the specifics of the data collected in the other instruments completed by the participating sites. We’ve edited the instruments to identify optional items and potential probes.</p> <p><i>Comment:</i> It is unclear whether replacement supervisors and home visitors will complete the interviews.</p> <p><i>Response:</i> Replacement home visitors and supervisors will be eligible to participate in the interviews if they joined the study at least 6 months earlier.</p>
Messages to home visiting program staff (28)	<p>These messages thank staff for completing logs and remind staff to do so.</p>	<p>No changes</p>	<p><i>Comments:</i> None</p>

REFERENCES

- Andreassen, C., and P. Fletcher. "Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) Methodology Report for the Nine-Month Data Collection (2001-02), Volume 1: Psychometric Characteristics." Submitted to the U.S. Department of Education. Report No. (NCES 2005-100). Washington, DC: National Center for Education Statistics, 2005.
- Aylward, G. P.. 1995. Bayley Infant Neurodevelopmental Screener. San Antonio, TX: Psychological Corporation.
- Bayley, Nancy. Bayley Scales of Infant and Toddler Development-Third Edition: Administration and Technical Manual. San Antonio, TX: PsychCorp, 2006.
- Bloom, Howard S., Carolyn J. Hill and James A. Riccio. 2003. "Linking Program Implementation and Effectiveness: Lessons from a Pooled Sample of Welfare-to-Work Experiments," *Journal of Policy Analysis and Management*, 22(4): 551 – 575.
- Durlak, J. A., and E. P. DuPre. 2008. "Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation." *American Journal of Community Psychology* 41, 3-4: 327-350.
- Fagan, Joseph F. 2005. "The Fagan Test of Infant Intelligence-Manual." Website: <http://infantest.com/ftii.pdf>.
- Filene, Jill H., James Bell, and Elliott G. Smith. 2011. National Cross-Site Evaluation of the Replication of Family Connections: Final Evaluation Report. Report submitted to the Administration for Children and Families.
- Flanagan, D. P., & Harrison, P. L. (2005). *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. (2nd Edition). New York, NY: The Guilford Press.
- Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of Cross-Battery Assessment*. (2nd Edition). New Jersey: John Wiley & Sons, Inc.
- Greenspan, S.I. *Greenspan Social-Emotional Growth Chart: A Screening Questionnaire for Infants and Young Children*. San Antonio, TX: Harcourt Assessment, 2004.
- James, Tracy. 2001. "Results of the Wave 1 Incentive Experiment in the 1996 Survey of Income and Program Participation." *Proceedings of the Section of Survey Research Methods*, 834-839. Alexandria, VA: American Statistical Association.
- Hack, Maureen, H. Gerry Taylor, Dennis Drotar, Mark Schluchter, Lydia Cartar, Deanne Wilson-Costello, Nancy Klein, Harriet Friedman, Nori Mercuri-Minich and Mary Morrow. 2005. "Poor Predictive Validity of the Bayley Scales of Infant Development for Cognitive Function of Extremely Low Birth Weight Children at School Age." *Pediatrics* 118, 2: 333-341.

Lewis, Michael and Harry McGurk. 1972. "Evaluation of infant intelligence: Infant intelligence scores--true or false?" *Science* 178:1174-1177.

Lewis, Michael and Harry McGurk. 1973. "Testing infant intelligence." *Science* 182:737.

Mack, Stephen, Vicki Huggins, Donald Keathley, and Mahdi Sundukchi. 1998. "Do Monetary Incentives Improve Response Rates in the Survey of Income and Program Participation?" *Proceedings of the Section on Survey Research Methods*, 529-534. Alexandria, VA: American Statistical Association.

Martin, Elizabeth, Denise Abreu, and Franklin Winters. 2001. "Money and Motive: Effects of Incentives on Panel Attrition in the Survey of Income and Program Participation." *Journal of Official Statistics* 17: 267-284.

Matheny, Adam P.. 1973. "Testing Infant Intelligence." *Science* 182: 734.

McCall, Robert B.. 1981. "Early Predictors of Later IQ: The Search Continues." *Intelligence* 5, 2: 141-147.

McGuigan, William M., Aphra R. Katzev, and Clara C. Pratt. 2003. "Multi-Level Determinants of Retention in a Home-Visiting Child Abuse Prevention Program." *Child Abuse & Neglect* 27: 363-380.

Michalopoulos, Charles, Anne Duggan, Virginia Knox, Jill H. Filene, Erika Lundquist, Emily K. Snell, Phaedra S. Corso, Justin B. Ingels, Sue Kim, and Magdalena Mello, 2011. ACF-OPRE Report 2011-16. *Design Options for the Home Visiting Evaluation: Draft Final Report*. Administration for Children and Families, U.S. Department of Health and Human Services, Washington, DC.

Nápoles-Springer AM, Santoyo-Olsson J, O'Brien H, Stewart AL. (2006). Using Cognitive Interviews to Develop Surveys in Diverse Populations. *Med Care*, 44(Suppl 3):S21-S30.

Tulsky, D., Zhu, J. & Ledbetter, M. (Eds.). *WAIS-III WMS-III Technical Manual (Wechsler Adult Intelligence Scale & Wechsler Memory Scale)*. (1997). Harcourt Brace & Company.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition*. San Antonio, TX: The Psychological Corporation.

Willis, Gordon B. 2005. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage.

Wilson, Ronald S.. 1983. "Testing Infant Intelligence." *Science* 182: 734-736.