

Mother and Infant Home Visiting Program Evaluation (MIHOPE)

Investigating Variation in Program Impacts

September 21, 2015

The legislation authorizing the Maternal, Infant, and Early Childhood Home Visiting program (MIECHV also known as the Federal Home Visiting Program) required MIHOPE to study the effectiveness of early-childhood home visiting programs on a range of parent and child outcomes. In addition to fulfilling this requirement through the study's random assignment impact analysis, in-depth information on implementation including program features, home visitor characteristics, service delivery, and local conditions is being collected in order to link information on communities, organizations, and families to program impacts. The analysis of impact variation is intended to deepen understanding of the features of local programs that are associated with greater benefits for families and thus to strengthen the future implementation of home visiting programs.

The plan for the core impact analysis is presented in a separate memo. This memo provides our current thinking on the analysis linking program features to impacts.¹ Specifically, it proposes analyses to address the following broad research questions:

- How much variation is there in impacts across local home visiting programs?
- What is the relationship between the features of local home visiting programs and their effects on family outcomes?
- What is the relationship between the actual home visiting services that families receive and family outcomes?

Rather than producing one final judgment about which aspects of local programs are associated with greater effectiveness, the set of proposed analyses will examine whether and why impacts vary from several different perspectives. The analyses would start with an examination of how much impacts appear to vary across local programs. It would then attempt to “explain” variation in those impacts by looking at various aspects of local program implementation. Because we are concerned that some features of program implementation may not satisfy assumptions needed to credibly interpret them as “causing” program impacts — particularly the requirement that measured features be unrelated to any unmeasured characteristics of the family, location, or local program — the analysis would proceed in stages, beginning with those features (such as national model) that are most likely to satisfy the needed assumptions, and sequentially adding those that may be least likely to satisfy the assumptions. The last type of analysis to be conducted would investigate how aspects of the dosage of home visiting services — such as number, intensity, duration, and content of visits, referrals by home visitors to community services, and use of community services — are associated with program effects. For the latter set of analyses, we are considering a number of methods such as instrumental variables,

¹Throughout the memo, the term “program” is used to refer to the local implementation of home visiting while “model” is used to refer to the four national models that are being studied. Each local program would be implementing a national model, but implementation may vary at the local level.

causal mediation analysis, and principal stratification, each of which presents a way of linking services to impacts and requires statistical assumptions in order to interpret any correlations between services and impacts as causal.

A range of possible methods is presented in the memo so that nothing is ruled out at this stage, and so that we may seek the Committee's advice on where to best put our resources. Our current plan is to examine how much variation there is across local programs and across the four national models and to use a regression framework to link program features to program impacts. We also currently plan to use instrumental variables analyses to estimate the relationship between home visiting services received (such as number of visits) and program impacts. We emphasize these approaches because we think that the assumptions underlying the methods are most likely to be met and therefore will provide the most credible results. We are investigating other methods described in the memo and will likely include them in the analysis to some extent if time and resources permit.

After providing some additional background motivating this phase of the analysis, the memo describes the information being collected for the impact and implementation analyses. It then describes some approaches to addressing the three broad research questions mentioned above.

Questions for the Committee

We are especially interested in the Committee's thoughts on the following questions:

- We propose making final decisions about which outcomes to analyze in this phase of the study by looking at where the implementation analysis finds substantial variation in program features, community characteristics, or service delivery that are likely to be tied to impacts. Does the committee agree with that approach? Keeping in mind that we do not have the resources to do this analysis across all possible outcomes, does the Committee have recommendations on which outcomes we should include in the analysis?
- As with all nonexperimental methods, each of the methods described for “explaining” variation in impacts across local programs is based on statistical assumptions that are difficult to test and may be implausible in the context of home visiting. For that reason, we will make clear to readers of the study's findings that the results should be interpreted with caution. Do you agree that we should proceed despite concerns about some of the methods we discuss?
- In linking program features to program impacts, we are concerned that some aspects of program features may violate the standard assumptions underlying regression analyses. In particular, they may be correlated with improved family outcomes but not actually the cause of those improved outcomes. Which types of program features do you think are least likely and most likely to suffer from such a problem? In the parlance of econometrics, which program features do you think are most likely to be “exogenous” predictors of home visiting services and impacts?

- Are there ways of investigating the link between program features and impacts and between dosage of services and impacts that are not discussed in this memo but that you would recommend we investigate?
- In linking service receipt to program impacts, service receipt can be defined in many different ways, such as number of visits (both overall and in certain time periods such as while the mother is pregnant), length and duration of visits, number of visits during which certain topics are discussed, parent responsiveness, number of referrals made by home visitors for services related to the study's primary outcomes, and whether other types of services available in the community are received. Does the Committee have recommendations on which aspects of service receipt should be the focus of this part of the analysis?

Background and Context

The research design and scale of MIHOPE offer unique and powerful opportunities for learning about what works, for whom, under what conditions. “What works” refers to the aspects of service delivery that are associated with greater impacts for families. “For whom” refers to an awareness that some aspects of service delivery and support structures may result in relatively greater improvements for some types of families than for others. “Under what conditions” refers to an appreciation that particularly successful service delivery strategies for particular types of participants might be enabled or constrained depending on conditions in the local program office, the service delivery system, or the broader community.

Interest in these types of questions has arisen as policymakers, local program managers, and others seek to make the most efficient and effective use of public expenditures and private philanthropy. In the past, learning about what works best for whom has often relied on sharing “best practices”, on correlational analyses (simple or more complex), and occasionally on random assignment studies. Because random assignment study findings that could shed light on “what works best for whom under what conditions” are not available to inform every possible decision that policymakers and program managers must make, other ways of building this knowledge are sought.

A promising approach, which we pursue in this part of MIHOPE, is to build on the research design that allows estimation of causal impacts across local programs and also to incorporate rich information from the implementation study. As described in this memo's introduction and as illustrated in Figure 1 in the “Overview” memo for this meeting, the impact analysis will estimate program impacts on a range of family outcomes. The implementation analysis will identify and measure the outputs of the local programs (service delivery for program group members and, for some outputs, impacts on service delivery compared with the control group) as well as the inputs (service models, implementation system, community context, families) that together lead to those outputs. The analysis linking program features to impacts — the subject of this memo — will first examine associations between inputs and impacts and then between outputs and impacts.

An interest in better understanding the links between inputs and impacts and between outputs and impacts is not unique to home visiting. These questions are of broad interest to many social service programs, schools, or indeed any application or setting where policies or programs are not “self-implementing” and where local decision makers and program managers may have some discretion over the design or implementation of programs. While the questions are of broad interest, the strategy being pursued in this paper is not widely used, likely due to the data and effort it requires. (This type of research has been conducted in studies of welfare-to-work programs and in Head Start.)²

Building Blocks

Our analysis of variation in program impacts builds on the impact and implementation analyses. In this memo we provide a brief summary of the key outcomes in the impact analysis and the information that will be available through the study’s implementation research. For more detail on data sources and plans for analyzing impacts, studying how programs are implemented, and investigating differences across subgroups of families, please see the Impact and Implementation memos.

Outcome measures for the impact analysis

The Impact memo recommends the following 13 primary outcomes from five domains:

- Maternal and child health: (1) new pregnancy after study entry, (2) mother has health insurance coverage, (3) number of well-child visits, (4) child has health insurance coverage, (5) any child emergency department use;
- Child development: (6) behavior problems total score and (7) language skills in normal range;
- Child maltreatment: (8) frequency of minor physical assault, (9) frequency of psychological aggression, (10) any health care encounter for injury or ingestion;
- Parenting: (11) quality of home environment and (12) parental supportiveness;
- Economic self-sufficiency: (13) whether the mother is receiving education or training.

We intend to conduct the analyses described in this memo on these 13 outcomes. In addition, we may conduct the analyses on some of the study’s secondary outcomes if there is reason to believe there is substantial variation in local program implementation that could contribute to substantial variation in local impacts.

Explanatory variables (measures of program implementation and service delivery)

The Implementation memo provides details on implementation research questions, data sources, key constructs, and preliminary analysis plans. The following paragraphs summarize the major constructs to be measured for the implementation study. These will also be used in the analysis described in the current memo.

²Bloom, Hill, and Riccio (2003); Feller, Grindal, Miratrix, and Page (2014); Greenberg, Michalopoulos, and Robins (2003); Peck and Bell (2014).

- Community Context. This includes information on neighborhood characteristics measured from the 2010 Census and from field staff observations of families' neighborhoods. It also includes information on service availability and accessibility provided by home visiting program staff.
- Intended services. This includes (1) goals and outcomes, (2) recipients, (3) service delivery and linkages with other services, and (4) staffing. Model developers have been asked to indicate the high priority goals of their program as well as intended service dosage, including visit frequency and length and the duration of family enrollment. Intended service content is being assessed by examining program developers' curricular materials and by interviewing developers. Intended staffing includes information on qualifications for hiring staff, staff roles and responsibilities, required competencies, and caseload limits. Parallel measures of each of these items will be created for each local program since local programs may deviate from the national models (and three of the four national models explicitly allow local discretion in some aspects of intended services).
- Implementation System. The implementation system includes the policies, procedures, and resources needed to implement the service model. The defining features of the implementation system can be categorized as (1) policies and procedures for staff selection, training, supervision, and evaluation; (2) facilitative clinical supports; (3) facilitative administrative supports; (4) systems interventions; and (5) organizational culture and climate.
- Staff Characteristics. In addition to organizational influences, home visiting services are affected by the individuals who participate in programs, including home visitors and supervisors. Through surveys of home visitors and supervisors, the study is collecting information on a number of staff characteristics, including demographics; educational and employment background; psychological well-being; and beliefs, attitudes, knowledge, and skills.
- Home Visiting Services Delivered. MIHOPE is examining five aspects of service delivery: dosage (frequency, intensity, and duration measured through logs completed by home visitors after each visit), content (curricula, referrals, topics discussed, also available through logs), techniques, family responsiveness, and quality of visits. Information on parent responsiveness is limited to two ratings that the home visitor provides weekly (of level of engagement during the visit and follow-up on activities between visits). Information on the quality of actual services delivered and detailed observation of family responsiveness will be available only for a subset of sites and families, so it cannot be included in the full black box analysis. For investigating the link between dosage and impacts, it is best to use an estimate of the difference between the program and control groups on services received, although some measures of service delivery will be available only for program group members (such as those derived from MIHOPE logs).

Variation in Impacts

This section and the two that follow describe the proposed analyses. The starting point is estimating the amount of variation in impacts across local home visiting programs. The analysis would then use a regression framework to explore which program features — “inputs” in the MIHOPE conceptual framework (see Figure 1 in the overview memo to the Advisory Committee) — are associated with larger effects for those programs. Finally, the study would explore the link between services that families receive — “outputs” in the MIHOPE conceptual framework — and the programs’ effects on their outcomes.

Variation in impacts across local programs

In considering whether there is variation in impacts, it is important to distinguish between variation in impact estimates and variation in “true” impacts. In particular, estimates will vary from site to site (or across other units) because of sampling error as well as true variation in impacts. That is, estimated effects of one local program will differ from those of another local program not only because of true differences in those effects but also because both will be measured using information on 60 sample members in most sites (and follow-up data on perhaps 45 where data are collected directly from families). For this reason, the first step of the analysis will be to estimate whether the observed variation in impact estimates is more than would be predicted through sampling error alone.

The first step would be to estimate the variation in impacts across local programs. The framework underlying this part of the analysis is captured by the following equation:

$$Y_{ij} = \alpha_j + \beta X_i + \delta_j T_{ij} + u_{ij}.$$

That is, an outcome Y for family i in site j will be influenced by the family’s characteristics (X_i) and whether they are assigned to the study’s program or control group (T_{ij}). In addition, outcome levels for control group families may vary across sites (α_j) as may impacts (δ_j). The term u_{ij} captures variation in family outcomes not captured through the other terms.

This phase of the analysis would use a “random effects” framework. In particular, impacts of the local program — δ_j — would be assumed to be normally distributed, and the variance of that normal distribution would provide a measure of how much true impacts vary across local programs.³

Variation in impacts across home visitors

The study also has the possibility of exploring variation in impacts across home visitors. Such an analysis could be important because there is likely to be substantial variation in service delivery across home visitors, even within specific local programs. Because of concerns about data quality and statistical power discussed below, we envision this analysis as more exploratory

³By comparison, the main impact analysis would include a fixed-effect intent-to-treat estimate of the average effect across all sites and for subgroups of families.

than the analysis of variation by local program. It would likely be presented in a technical appendix to the report with a summary of the results described briefly in the body of the report.

This analysis would take advantage of following feature of MIHOPE. Before families were enrolled in the study, local home visiting programs entered information about them in the study's intake system. One piece of that information was the home visitor the local program planned to assign to that family if the family were randomly assigned to the program group. Because this information was gathered before random assignment, it is available equally for program and control group members. In essence, MIHOPE randomly assigned each home visitor's potential caseload to the program and control groups, providing a means of estimating impacts for each home visitor.

This analysis would be similar to the analysis of variation in impacts across local programs, except that families would be nested in home visitors who are nested in local programs. Because each local program has several home visitors, however, there are fewer families per home visitor than per local program. This may limit the statistical power of extending the analysis to the home visitor level. We will make a final determination of the statistical power of the analysis once sample enrollment is completed in September 2015.

This analysis may also be complicated by two limitations to the data. First, not all programs provided information on the projected home visitor. As a result, the information is available for about 65 percent of families in the study. In addition, information on the projected home visitor is not always accurate. In about 18 percent of cases, the home visitor indicated by the program before random assignment was not the home visitor who actually provided services to the family. Thus, there is a fair amount of measurement error in identifying the home visitor for control group families.

One option the team is considering is to use matching of home visitors to control group families to supplement or replace the preliminary home visitor field. For example, in sites where the preliminary home visitor was not indicated, we know which home visitor completed weekly logs on each program group family, and hence which home visitor provided services to that family. Using baseline characteristics, we could match each program group family to a similar control group family and estimate the impacts for those home visitors using the matched comparison. Although such matching may introduce bias into the estimates, we could conduct sensitivity checks to ensure that using this method does not change the results substantially. Because there is substantial measurement error in the preliminary home visitor field, we are considering the possibility of using this matching method more broadly.

How Impacts Are Associated with Program Features

The next stage of the analysis would explore how various inputs into home visiting programs (program features) are related to impacts of those programs. Because local programs are not randomized to have different program features, a finding that local programs with certain features had larger effects would not necessarily mean that those features were responsible for the larger effects. Instead, those program features might be related to aspects of the program that were not measured or not included in the analysis. Unbiased estimates generated through random

assignment of the effects of home visiting at each site would be linked to program features of that site, but the associations uncovered through the analysis might not be causal.

Linking features of local programs to their impacts

The idea behind this analysis is a natural extension of the random effects model presented earlier:

$$Y_{ij} = \alpha_j + \beta X_i + \gamma C_j + T_{ij}(\delta + \lambda X_i + \mu C_i + \kappa LP_j + \epsilon_j) + u_{ij}$$

The essential change to the model is that impacts of local programs would be allowed to vary with family characteristics (X_i), characteristics of the local community (C_j), and implementation features of the local programs (LP_j). Program features could include any of the implementation factors summarized earlier in the memo and described in more detail in the memo on the MIHOPE implementation analysis, such as intended services and characteristics of the implementation systems.

One concern about this framework is that some program features may violate the standard regression assumption of no unexplained variation in outcomes that is correlated with the explanatory variables. Because of that concern, the analysis would proceed in steps, starting with explanatory factors that are more likely to meet the regression assumptions (that is, to be exogenous) and adding in others that may be less likely to be exogenous.

A logical first step is to examine how much impacts vary by national model. There is certainly great interest in how the national models compare with one another on key outcomes, and the study's design document indicates that this would be one of the study's research questions.⁴ Because the national models differ to some extent in which types of families they aim to serve, we would next add information on the characteristics of families. This would provide us with a comparison of the four national models, adjusting for differences in family characteristics.⁵

The next step would be to add community characteristics and features of planned services that may vary across local programs within a national model to investigate whether local variation within sites operating the same national model is associated with variation in impacts. An example of a feature that might vary across local programs within a national model is whether the local program places a high priority on improving a particular outcome. Both community characteristics and planned services can generally be considered exogenous because they are either present or determined by the local program before the family has entered the study.

⁴Michalopoulos et al. (2013).

⁵Because home visiting programs are intended to tailor their services to family needs, it is plausible that some program features are more beneficial for some types of families than for others. Likewise, the actual home visits may be more important for some types of families than for others. For that reason, in this analysis and the analysis of mediators discussed later in the memo, we would explore how the link between program features and mediators varies with family characteristics.

The regression model could then be expanded by adding features of the implementation system, such as ratings of the training used for home visitors, the quality of the supervision of home visitors, what supports are available for facilitating program administration, and so on. These implementation system features are likely independent of service model features (described in the previous paragraph), so that both groups of features could be included in one regression. The results of this step would need to be interpreted with caution, because these implementation system features can theoretically be influenced by characteristics of the home visitors and families in the study site and by their responses to the program as it is implemented.

Absent from this framework is what actually happens during home visits. This omission is intentional because the content and frequency of home visits will depend on a family's needs and engagement with the program. These aspects are likely to be correlated with unobservable family characteristics that are also correlated with outcomes; therefore, the regression framework assumptions would be violated by their inclusion in the model. We discuss this type of analysis in the next section of the memo.

We will need to be parsimonious in selecting features to include in the model in order to preserve statistical power. As described in the Implementation memo, both theory and prior evidence would be used to choose which features to include. For example, theory and prior evidence might suggest that one set of features is important for studying how maternal health is improved but that a different set of features is important for studying how child development is improved. Likewise, theory might suggest that the duration and intensity of home visits may affect the full range of outcomes.

Because the statistical power of this analysis depends on how closely related program features are to one another, final decisions about the analysis would not be made until after data are collected. If the data suggest that many program features are unrelated to one another, a more expansive analysis could be conducted. If, as is more likely, program features are highly related within a site, we would try to combine related measures into indexes or factors in order to reduce the number of explanatory dimensions.

In addition, we would give priority to program features that show substantial variation across local programs since it would be difficult to link variation in program features to variation in impacts if there is little variation in the features.

Because of concerns about selection or omitted variable bias, we would interpret the results with caution. For example, we state findings using language such as, "Impacts on [Outcome *Y*] are larger in local programs that place a high priority on child health." We would not, however, be able to definitively conclude that impacts on [Outcome *Y*] are larger *because* local programs place a high priority on child health. Local programs that place a high priority on child health may be recruiting mothers whose choices can be influenced by home visitors, and other programs might not achieve the same effect merely by emphasizing child health without choosing families in the same way.

Linking variation in home visitors to their impacts

Using the preliminary home visitor assignment provided by programs at study intake, the analysis may be extended to estimate the relationship between home visitor characteristics and impacts. Because of concerns about data and statistical power discussed earlier, this extension would be considered exploratory and would be conducted only if time and project resources permitted.

The Role of Mediators

The previous section discussed an approach for relating program features to program impacts, using either variation across local programs or across home visitors. Excluded from the list of possible explanatory variables in the models was anything that occurs during specific home visits. These measures were excluded because they are very likely to violate one of the key assumptions underlying regression analysis, namely that explanatory variables are not correlated with the unexplained portion of the outcome. Mothers have to agree to schedule visits, have to let the home visitor in the door, and have to spend time with the home visitor. It is likely that mothers who benefit the most from home visits will be those who are most engaged in the program and who consequently remain enrolled in the program over a longer period of time. Larger impacts for such mothers would not necessarily mean that increasing the number or length of home visits for other mothers would lead to similar improvements in their family's outcomes. Alternatively, it is possible that mothers who can schedule and keep multiple appointments with a home visitor may have better parenting skills, be better able to navigate the health care system, and be more likely to delay having their next child than other parents. Such parents might not actually benefit much from the program, and impacts might be smaller for them than for other mothers. In this case, a simple correlation between dosage and impacts might suggest that increasing dosage is not associated with improved family outcomes; such a conclusion would not necessarily be valid.

We refer to the services that families receive because of home visiting as “mediators.” In a random assignment study such as MIHOPE, a mediator is usually considered anything that is affected by assignment to the program that may in turn affect subsequent outcomes. For example, being assigned to the program would presumably increase the number of home visits that a family receives, which may improve a range of outcomes. The next step of the analysis is thus to examine the relationship between services received and impacts.

For the primary MIHOPE analysis, the analysis of mediators will be limited to aspects of home visiting services — such as the number, length, content, and quality of home visits — and community services.⁶ Mediators related to home visiting services may be general — such as the total number of visits a family receives — or they may be specific to particular outcomes — such as the number of visits in which the home visitor and the mother discussed a particular outcome

⁶Although outcomes could also be influenced by improvements in intermediate outcomes such as maternal depression, such mediators will not be considered here, although we expect them to be the subject of later research (either by the team or by other researchers who gain access to the public use file that the study will eventually produce).

area or in which the home visitor provides a referral to community services relevant to that outcome area.

Table 1

Comparison of Methods to Explore Link Between Program Services and Program Impacts

Method and what is estimated	Basic estimation method	Key assumptions	Comment
<u>Instrumental variables</u> Effect of the mediator on the outcome.	Use random assignment to obtain exogenous variation in the value of the mediator. Intuitively, the results reflect how differences in impacts on the mediators across local programs are associated with differences in their effects on outcomes.	All intervention effects flow through the mediators. The effect of local programs on a mediator is unrelated to the effect of the mediator on the outcome.	Relies on variation across sites or home visitors in impacts on mediators or outcomes.
<u>Causal mediation analysis</u> Effect of the mediator on the outcome ("indirect effect"), allowing for treatment assignment to independently affect the outcome ("direct effect").	Include value of the mediator as an explanatory variable.	Enough other covariates are included (such as family characteristics) that mediator levels are not associated with omitted variables that are related to the outcome.	May have greater statistical power than instrumental variables because it uses variation within sites, but key assumption is likely to be violated, producing biased estimates of the effect of the mediators.
<u>Endogenous subgroups</u> (principal stratification) Effects for different subgroups of families defined based on event or outcomes that occur after random assignment.	Predict subgroup membership and compare estimated impacts across subgroups.	Differences in impacts across subgroups are due to treatment differences, not differences in family characteristics. Families will not receive fewer home visiting or other services if they are assigned to the program group.	In general, limited to one categorical mediator.

Table 1 presents some features of three approaches that are discussed below: instrumental variables, causal mediation analysis, and endogenous subgroups (including principal stratification). The three methods all address the same basic question, namely how mediators such as home visiting services are related to impacts, but they do so in somewhat different ways, using somewhat different statistical assumptions. We are considering all three precisely because they rely on different assumptions. Our confidence in the results would be strengthened if all three methods tell a similar story, but it would also be important to know if they tell very different stories, in which case we would not be confident in our recommendations to the field.

Although we think that all three approaches are worth pursuing, based on our reading of the literature and discussions with others at MDRC who are involved in a project that is making advances in the use of all three methods for studying impact variation, the team has the most confidence in instrumental variables for investigating the role of mediators. In its most basic form, however, causal mediation analysis is straightforward to estimate, and we would probably include some version of that method in the study’s findings. Principal stratification, because of its focus on strata or subgroups of families, addresses somewhat narrower questions than either instrumental variables or causal mediation analysis, and the team would explore it if a key set of strata could be identified that are likely to meet the method’s assumptions.

Instrumental variables. Instrumental variable analysis was developed to estimate relationships in which an explanatory variable (such as home visits received) is correlated with the outcome outside the causal relationship between the two variables (for example, if more motivated mothers receive more home visits). The classic example in economics is estimating a demand curve when a series of price and quantity pairs are observed. Since supply affects price and quantity, it is unclear to what extent a simple regression of quantity on price is capturing demand versus supply. To disentangle the two, economists look for supply shocks, which would make the supply curve move up or down but have no effect on the demand curve. By identifying a series of such shocks, one could trace out the path of the demand curve. The key requirement — called the “exclusion restriction” — is that the instrument be associated with the outcome only through its association with the mediator.

In MIHOPE, a simple instrumental variable model might be the following:

$$\begin{aligned} M(T) &= M_0 + BT \\ Y(T) &= Y_0 + \Delta T \\ Y(T) &= Y_0^* + \Gamma M(t) \end{aligned}$$

The first equation indicates that the treatment increases the value of the mediator by an amount B . The second equation indicates that the treatment increases the value of the outcome by an amount Δ . The third equation indicates that increasing the value of the mediator by one unit increases the value of the outcome by Γ units.

Estimates of B and Δ can be easily obtained from intent-to-treat estimates of program impacts on the mediator and outcome, respectively. The difficulty lies in determining the relationship between the mediator and the outcome. Thus, as noted in Table 1, estimating this relationship is the focus of this method. The basic idea behind the method is that, because of random

assignment, a family's assignment to the MIHOPE treatment group is not related to the effect home visiting would have on that family's receipt of services or on that family's outcomes. Variation in the effects on family outcomes across local programs would be related to variation in their effects on service receipt. If impacts are larger in locations where service receipt is larger, this would provide evidence that service receipt matters and provide a means of estimating the relationship between service receipt and outcomes.

For a multi-site study with multiple mediators such as MIHOPE, Reardon and Raudenbush (2013) discuss the assumptions needed to estimate the relationship between mediators and outcomes. Two assumptions, in particular, may not be credible in MIHOPE. First, the treatment assignment is assumed to affect family outcome only through the mediators. In principle, this should hold in MIHOPE since home visiting should produce effects only to the extent that home visits are made, that various types of content are discussed, that referrals are made and kept, and so on. In practice, however, the assumption would hold only if all relevant aspects of those services are included as mediators. Even with the rich measurement that MIHOPE includes, some aspects of service delivery will not be measured well or at all, but we hope that we can include enough aspects of service delivery that this assumption will be close to holding.

A second assumption discussed by Reardon and Raudenbush (2013) is that variation in the impact on the mediators across sites is not related to variation in the effect of the mediators on outcomes across the sites. This assumption would rule out the possibility that some sites have mothers who are more motivated to receive home visiting (for example) and who therefore receive a greater quantity of services and see their outcomes improve by a greater amount. If their greater motivation is part of the reason their outcomes improve, results for those mothers could not be generalized to mothers with less motivation.⁷

Causal mediation analysis. In causal mediation analyses,⁸ the effect of the treatment on the outcome is divided into a direct effect and an indirect effect (see Table 1). The direct effect represents the change in family outcomes from being assigned to the program group but having none of the mediators change. In MIHOPE, this would mean that being assigned to a MIECHV program improves outcomes even if it does not increase the amount of home visiting received by the family (in whatever way we are able to measure it). The indirect effect represents the effect of changing the mediator but not changing the treatment status (which is likely to be more of a thought experiment than a real outcome in a study such as MIHOPE, where the quantity and quality of home visiting services (the mediators) are closely tied to which group the family is assigned to).⁹

⁷A third issue arises when the relationship between the outcome measure and the mediators is not linear. For example, with a binary outcome, it may be appropriate to estimate a model such as a logistic regression. Terza, Bradford, and Dismuke (2008) show that the most common method of obtaining instrumental variable estimates is biased unless the residual from a regression of the mediator on treatment assignment is included as a "control function." If we apply instrumental variable analysis to such nonlinear situations, we will explore the benefits of using the approach suggested by Terza, Bradford, and Dismuke (2008).

⁸For example, Imai, Keele, Tingley, and Yamamoto (2011).

⁹Hong, Deutsch, and Hill (2015) describe a number of other direct and indirect effects that might be of interest and could potentially be estimated using a causal mediation framework. However, because their approach still relies on the assumption of sequential ignorability, their extensions are not discussed here.

In its simplest form, this approach can be represented by the following three-equation model:

$$\begin{aligned} Y_i &= \alpha_1 + \beta_3 T_i + \xi_1 X_i + \varepsilon_{1i} \\ M_i &= \alpha_2 + \beta_2 T_i + \xi_2 X_i + \varepsilon_{2i} \\ Y_i &= \alpha_3 + \beta_1 T_i + \gamma M_i + \xi_3 X_i + \varepsilon_{3i} \end{aligned}$$

Randomization provides a way of obtaining unbiased estimates of β_3 and β_2 . However, obtaining unbiased estimates of β_1 and γ is unlikely for the reasons mentioned earlier: The quantity and content of home visiting services are likely to be related to characteristics of the mother that are not captured in the regression and therefore are correlated with ε_{3i} .

The key assumption used to get around this problem is called “sequential ignorability.” Sequential ignorability contains two parts. First, treatment assignment is not correlated with ε_{3i} after conditioning on pretreatment covariates. This holds trivially in random assignment studies. Second, sequential ignorability assumes there are no unmeasured covariates that confound the relationship between the mediator and the outcome, conditional on treatment assignment and family characteristics. This will be true only if the analysis includes all family and other characteristics that are correlated with both the mediator and the outcome.

If sequential ignorability is incorrect, the error term in the equation relating treatment to the mediator value will be correlated with the error term in the equation relating the mediator and treatment status to the outcome. This provides a sensitivity check on the results. One can estimate the amount of bias that would result from different levels of that correlation. More confidence can be placed in the results if plausibly high levels of correlation result in little bias in the estimated relationships. Even if sequential ignorability is incorrect, it is a natural place to start the investigation of services received since it amounts to adding measures of services received to the right-hand side of the regression framework presented in the previous section.

Endogenous subgroups (including principal stratification). This approach refers to methods to estimate impacts on groups of families that are placed into subgroups based on post-random assignment outcomes or services received.¹⁰ For example, Peck (2003) used this approach in a random assignment study to compare the impacts for program group members who received program services with those who were no-shows. In brief, her approach used baseline characteristics to predict whether individuals were likely to have received program services. Because the predictions were based on baseline characteristics, which are similar in aggregate between randomly assigned program and control groups, this provided a means of forming subgroups of program and control group members that should have been the same except for their assignment. It thus provided a means of estimating program impacts for individuals in the predicted subgroups.

Peck and Bell (2014) showed that this framework could be extended to other types of endogenous subgroups. In particular, the more recent paper used data from the National Head Start Impact Study to compare the effects for program group members who were in a high

¹⁰By comparison, the MIHOPE impact analysis would compare impacts across subgroups of families defined by baseline (that is, pre-random assignment) characteristics.

quality Head Start program with those who were in a low quality Head Start center or not in Head Start at all. The approach Peck and Bell applied to the Head Start Impact Study has an obvious analog for MIHOPE, namely whether impacts are larger for families with a high dosage of home visiting than for those with a low dosage or those who never receive a home visit.

Although this approach is intriguing, it requires several assumptions. First, Peck's approach provides estimates of impacts based on predicted subgroup membership. Additional assumptions are needed to infer estimated effects for the true subgroups (for example, those who actually participated rather than those who were predicted to participate). As outlined in Peck and Bell (2014), these include assuming that the average impact for people who are correctly predicted to be in a subgroup are the same as for those who are incorrectly predicted to be in other subgroups. For example, the average impact for people who actually receive high dosage is the same for those who are predicted to receive high dosage as for those who are predicted to receive low dosage and for those who are predicted to receive no home visiting. Since those who are predicted to receive no home visiting act very differently than their prediction, it seems likely that the average impact for that group would be different than for those who are correctly predicted to receive a high dosage. An alternative assumption is that covariates perfectly predict subgroup membership, which also seems implausible. Finally, like principal stratification, the Peck approach may confound the effects of the family characteristics that predict program participation (or high or low dosage) with the actual effects of that participation.

Principal stratification is a different approach to estimating the effects for subgroups of families defined based on post-random assignment outcomes. One difference from Peck's approach is that it provides a means of estimating effects for families who change their behavior — for example by moving from not receiving home visiting to receiving a high dosage of home visiting — because they are assigned to the program group. In analyzing data from the Head Start Impact Study, for example, Feller, Grindal, Miratrix, and Page (2014) applied principal stratification to estimate the effects for families who *changed* their behavior from using home-based child care or center-based care to using Head Start.

As in the Peck approach, an insight underlying principal stratification modeling in random assignment studies is that each subgroup of families in the program group has an equivalent in the control group. Thus, the approach seeks to place families in both the program and control groups into the endogenously defined subgroups and then estimates the effects within each subgroup by comparing outcomes for program and control group members in the subgroup.

Principal stratification seeks to use information about outcomes in making the assignment. For example, Feller, Grindal, Miratrix, and Page (2014) assumed that control group families who use Head Start would also have used Head Start if they had been in the program group, and that being assigned to the program group would consequently have had no effect on outcomes for those families. The distribution of outcomes for control group members who used Head Start centers thus provides information on the distribution of outcomes for their program group counterparts. Thus, program group members would be more likely to be placed into this endogenous subgroup if they had outcomes that were similar to that for control group members in this subgroup. A similar argument and set of assumptions allows them to define an endogenous subgroup for program group families that used home-based or center care rather

than Head Start, and to identify control group members who were likely to be their counterparts. A combination of predictive modeling (linking family characteristics to strata) and other assumptions (such as assuming that outcomes are normally distributed) provides a means of estimating the effects for those who changed their behavior because they were assigned to the program group.

In MIHOPE, this approach could be used to investigate the effects of home visiting for families who move from no home visiting in the control group to low dosage or high dosage in the program group, or for those who move from low dosage home visiting in the control group to high dosage in the program group.

One of the assumptions underlying the work by Feller and colleagues is that being assigned to the program group had no effect on outcomes for families that did not change their behavior. For example, those who would have used Head Start if they had been assigned to the control group would also have used Head Start if they had been assigned to the program group, and being assigned to the program group would not have altered their outcomes. When applied to home visiting, this assumption may be difficult to justify. Families who would have used home visiting services if assigned to the control group may use very different types of services if they are assigned to a MIECHV program. In many MIHOPE locations, for example, control group members had access to public health home visiting programs that might have provided a very small number of visits, rather than the frequent visits provided over multiple years by the evidence-based programs. Thus, being assigned to the program group might have made a substantial difference for these families.

Another concern is that the approach works best if baseline characteristics are good predictors of stratum membership. That is, families with some set of characteristics will be predicted to be those who would have received home visiting even if they had been assigned to the control group, or may predict which program group families would receive a high dosage of home visiting. Larger estimated impacts for those with high dosage than for those with low dosage may consequently be due to differences in their characteristics rather than differences in their dosage. Helping families of the low-dosage type to achieve a higher dosage would not necessarily result in larger effects for them, or effects that match those for the high-dosage group.

Next Steps

Assuming that the Committee is comfortable with the methods proposed in this memo, the next steps in this analysis include the following:

- Choose the program features to be considered for this analysis based on information from the implementation analysis, including the amount of local variation in program features, the association between program features and service receipt at the local level, and measurement work that is used to create summary measures such as indexes.

- Once the final sample member is enrolled in September 2015, estimate the power to detect variation in impacts across home visitors and to explore the relationship between home visitor characteristics and impacts on family outcomes.
- Finalize a set of secondary outcomes using information on variation in program features across local programs.
- Finalize a set of mediators to be investigated in the instrumental variables and causal mediation analyses.
- Define the strata that would be investigated using principal stratification, if that method is pursued.
- Provide final specifications for all statistical tests and regression analyses.

References

- 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.
- Feller, Avi, Todd Grindal, Luke W. Miratrix, and Lindsay Page. 2014. *Compared to What? Variation in the Impact of Early Childhood Education by Alternative Care-Type Settings*. Rochester, NY: Social Science Research Network .
- Greenberg, David H., Charles Michalopoulos, and Philip K. Robins. 2003. "A Meta-Analysis of Government-Sponsored Training Programs." *Industrial & Labor Relations Review* 57, 1: 31-53.
- Hong, Guanglei, Jonah Deutsch, and Heather D. Hill. 2015. "Ratio-of-Mediator-Probability Weighting for Causal Mediation Analysis in the Presence of Treatment-by-Mediator Interaction." *Journal of Educational and Behavioral Statistics* 40, 3: 307-340.
- Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning About Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review* 105, 4: 765-789.
- Michalopoulos, Charles, Anne Duggan, Virginia Knox, Jill H. Filene, Helen Lee, Emily E. Snell, Sarah Crowne, Erika Lundquist, Phaedra S. Corso, and Justin B. Ingels. 2013. *Revised Design for the Mother and Infant Home Visiting Program Evaluation*. OPRE Report 2013-18. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.
- Michalopoulos, Charles, Helen Lee, Anne Duggan, Erika Lundquist, Ada Tso, Sarah Crowne, Lori Burrell, Jennifer Somers, Jill H. Filene, and Virginia Knox. 2015. *The Mother and Infant Home Visiting Program Evaluation: Early Findings on the Maternal, Infant, and Early Childhood Home Visiting Program*. OPRE Report 2015-11. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Peck, Laura R. 2003. "Subgroup Analysis in Social Experiments: Measuring Program Impacts Based on Post-Treatment Choice." *American Journal of Evaluation* 24, 2: 157-187.
- Peck, Laura R., and Stephen H. Bell. 2014. *The Role of Program Quality in Determining Head Start's Impact on Child Development*. OPRE Report 2014-10. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Reardon, Sean F., and Stephen W. Raudenbush. 2013. "Under What Assumptions Do Site-by-Treatment Instruments Identify Average Causal Effects?" *Sociological Methods & Research* 42, 2: 143-163.

Terza, Joseph V., W. David Bradford, and Clara E. Dismuke. 2008. "The Use of Linear Instrumental Variables Methods in Health Services Research and Health Economics: A Cautionary Note." *Health Services Research* 43, 3: 1102-1120.