Descriptive Study of Early Head (Early Head Start Family and Child Experiences Study; Baby FACES)

Supporting Statement for Request for OMB Approval of Program Recruitment and Feasibility Study: Section B

April 16, 2008

U.S. Department of Health and Human Services
Administration for Children, Youth and Families
4th Fl. West, 370 L'Enfant Promenade, SW
Washington, DC 20047

Project Officer:
Rachel Chazan Cohen

CONTENTS

Section **Page**

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

B.1 Sample Universe, Sampling Method, and Expected Response Rate

Baby FACES will use a stratified clustered sample design. We will select a probability sample of Early Head Start programs using the Head Start Program Information Report (PIR) as the sample frame. As specified in the federal RFP, we will exclude from the sample any programs in Alaska, Hawaii, Puerto Rico, and U.S. territories, as well as migrant programs, and American Indian programs. We also will exclude from the frame any programs not directly providing Early Head Start services, and any programs under the management of the national interim grantee contractor.

When sampling programs, we will form eight explicit strata, first stratifying the frame by the program's total enrollment size (four strata) and then by whether the majority of children served by the program are likely to be English language learners (For sampling purposes, ELL will be based on primary home language). We plan to implicitly stratify (sort the frame by) program service approach (center-based, home-based, or mixed) within explicit strata. After sorting by program service approach, we will also implicitly stratify by census region and urbanicity (MSA versus non-MSA). Before selecting the sample, we will use an optimal allocation approach (balancing cost and variance) to determining the number of programs to allocate to each size stratum. Selecting more programs from the larger strata will help ensure that we end up with enough study-eligible children in later stages of selection. We will proportionally allocate the program sample between the ELL and non-ELL substrata within each program size stratum. Within each explicit stratum, we will select a sequential, equal probability sample. The sequential sampling technique, based on a procedure developed by Chromy,[3] offers all the advantages of the systematic sampling approach but eliminates the risk of bias associated with it.

[3] The procedure makes independent selections within each of the sampling intervals while controlling the selection opportunities for units crossing interval boundaries. Chromy, J.R. "Sequential Sample Selection Methods." Proceedings of the Survey Research Methods Section of the American Statistical Association, 1979, pp. 401-406.

We will initially select 180 programs, and then pair up adjacent selected programs within strata. (These paired programs would be similar to one another with respect to the implicit stratification variables.) We will then randomly select one from each pair to be released as part of the main sample of programs. After initial 90 programs are selected, we will ask the Office of Head Start to call the regional ACF offices to confirm that the 90 selected programs are in good standing. If confirmed, each program will be called and recruited to participate in the study. If the program is not in good standing, or is in good standing but refuses to participate, we will release into the sample the other member of the program's pair and go through the same process of confirmation and recruitment with that program. The goal is 90 participating programs.

B.2 Statistical Methods for Sample Selection and Degree of Accuracy Needed

We now describe in greater detail the statistical methodology for the sample selection and the data collection methodology.

a. Statistical Methodology for Sample Selection

(1) Sample Selection

Sample of Early Head Start Programs. Each year, every Head Start grantee and delegate agency submits a report that provides a broad profile of the program's operations. For the Baby FACES study, the most recent available PIR information (probably from the 2006-2007 Head Start year) will be employed in the selection of a stratified sample (based on enrollment size and majority ELL) of 180 Early Head Start programs that provide direct services and are not under transitional management. From these 180 programs, we will randomly select 90 to be the main release, with the other 90 serving as replacements if needed. Early Head Start programs in Alaska, Hawaii, Puerto Rico, the U.S. territories, Migrant programs, and American Indian programs are excluded from the sampling frame.

After programs are selected, we plan to include in the sample all children and pregnant women who are receiving services at the sampled programs at the time of our spring 2009 visit, as long as they meet the inclusion criteria based on their date of birth (or gestational age for pregnant women). While this will be discussed more fully in a separate OMB submission, we plan to include all pregnant women in their last trimester, any children up to 2 months of age, and any children between 10 and 14 months of age. These sample members will form two cohorts to be followed over time: a perinatal cohort and an age 1 cohort.

(2) Estimation Procedures

We will create weights to account for variations in the probabilities of selection and variations in the eligibility and cooperation rates among those selected. For the program sample, within each of the eight explicit sampling strata, we will calculate the probability of selection. The inverse of the probability of selection at each stage is called the sampling or base weight. The sampling weight takes into account the stratified sampling approach, including the over sampling of programs in the larger-size strata and the release of any "replacement" program within a pair. Within stratum, we will then multiply the sampling weight by the inverse of the weighted response rate so that the responding programs' weights account for both themselves and nonresponding programs. Thus, the program-level weight adjusts for the probability of selection of the program and response at the program level. The weighting steps are represented schematically in the formula below, where P represents the probability of selection within stratum (for the sample of 180), the second term represents the probability of release within the program pair, and RR represents the response rate among eligible programs within the program pair.[4]

[4] If only one program is released in a pair, that means the initially released program is a respondent, and the response rate is equal to 1. If both programs are released within a pair because first program is ineligible, and the second program responds, then the response rate is equal to 1. If both programs are released because the first program is an eligible nonrespondent, and the second program responds, then the response rate within the pair is

$$W_{pgm} = \frac{1}{P_{pgm}} \cdot \frac{2}{1} \cdot \frac{1}{RR_{pgm}}$$ for responding programs in pairs with only one program released

$$W_{pgm} = \frac{1}{P_{pgm}} \cdot \frac{1}{1} \cdot \frac{1}{RR_{pgm}}$$ for responding programs in pairs with both programs released

The program-level weight will be used for program-level analysis, but it is also a component of weights in the subsequent stage of sampling (not included as part of this statement).

(3) Variance Estimation

The sampling plan, which includes stratification and unequal probabilities of selection, requires the use of specialized procedures to calculate the variance of estimates. Standard statistical software assumes independent and identically distributed samples, which would indeed be the case with a simple random sample. A complex sample, however, generally has higher variances than would be calculated with such standard software. Two approaches for estimating variances under complex sampling, Taylor Series and replication methods,[5] can be estimated by using SUDAAN. When doing program-level analysis, one would specify the sampling strata used for selecting the programs. These strata will be based on program size and proportion of English language learners. There is no clustering to be specified for program-level analysis.

(4) Adequacy of Sample Size

Most of the analysis will be at the child level, and the adequacy of that sample size will be discussed in a separate OMB submission. Suffice it to say that this first stage of sampling 90 programs will allow for the detection of meaningful differences between subgroups of children,

---

equal to 1/2. If both programs in the pair are nonrespondents, then a nonresponse weighting adjustment will be made at the program stratum level.

[5] There are technical advantages to the Taylor Series approach, along with its ease of use; however, the replication methods provide more flexibility in terms of the types of estimates available. We will discuss the pros and cons of each approach with ACF and calculate replicate weights for each of the constructed weights if that is the direction that is taken.

given various assumptions about the sample design and its impact on the variance of estimates. However, some analysis will take place at the program level, and the 90 programs in the sample will allow for estimates of program-level characteristics. Table 3 shows the minimum detectable differences and effect sizes for comparing various program subgroup sizes with 80 percent power. We assume a small unequal weighting effect (1.155) on the variance.

TABLE 3

MINIMUM DETECTABLE DIFFERENCES AND EFFECT SIZES FOR PROGRAM-LEVEL MEASURES COMPARING TWO SUBGROUPS AT A POINT–IN-TIME

Subgroups	Subgroup 1 (Nominal Sample Size)	Subgroup 2 (Nominal Sample Size)	Minimum Detectable Differences Between Subgroups for a Proportion p = .50	Minimum Detectable Effects (Presented As a Proportion of a Standard Deviation)
1/2, 1/2	45.0	45.0	.321	.634
1/3, 1/3	29.7	29.7	.397	.781
1/3, 2/3	29.7	60.3	.342	.675
1/4, 3/4	22.5	67.5	.373	.733
1/5, 4/5	18.0	72.0	.406	.793
4/10, 6/10	36.0	54.0	.328	.647

Note: Two-sided $\alpha = .05$. Power = .80.

If we wanted to compare two equaled-sized subgroups of programs, we would be able to detect a difference of about two-thirds of a standard deviation with a sample size of 90 programs. Note that these detectable differences and effect sizes are fairly large, which is to be expected for a nominal sample size of 90 programs.

While child-level analysis will be dealt with more fully in a separate OMB submission, we present here some power calculations for child-level analysis, because the main reason we are sampling programs is in order to sample children served by Early Head Start and their families.

The sample size of about 2,000 children and pregnant women should be large enough to detect meaningful differences, given various assumptions about the sample design and its impact on the variance of estimates. Tables 4 and 5 show the minimum detectable differences (MDDs) with 80 percent power and various sample and subgroup sizes, and with different assumptions about the impact of weighting and clustering on the variance of estimates from the child assessments. We assume an intracluster correlation of .05 and, for the pre-post change estimates,

TABLE 4

MINIMUM DETECTABLE DIFFERENCES FOR CHILD ASSESSMENTS
COMPARING TWO SUBGROUPS AT A POINT-IN-TIME

			Effective Sample Sizes		Minimum Detectable Differences Between Subgroups	
Cohort			Subgroup 1	Subgroup 2	Proportion p = .50 Std. Dev. = 0.50	Normalized Variable Mean = 100 Std. Dev. = 15
Perinatal	Age 1	1/2, 1/2	273.0	273.0	.190	3.595
		1/3, 2/3	182.0	264.0	.180	3.813
	Age 3	1/2, 1/2	184.0	184.0	.231	4.379
		1/3, 2/3	122.7	245.3	.220	4.644
Age 1	Age 1	1/2, 1/2	273.5	273.5	.190	3.592

	1/3, 2/3	182.3	364.7	.180	3.809
Age 3	1/2, 1/2	184.0	184.0	.231	4.379
	1/3, 2/3	122.7	245.3	.220	4.644

Note: Two-sided α = .05. Power = .80.

TABLE 5

MINIMUM DETECTABLE DIFFERENCES FOR CHILD ASSESSMENTS
PRE-POST COMPARISONS (ASSUME PRE-POST CORRELATION = 0.5)

	Effective Sample Size		Minimum Detectable Pre-Post Differences	
Cohort	Time 1 (Age 1)	Time 2 (Age 3)	Proportion p = .50 Std. Dev. = 0.50	Normalized Variable Mean = 100 Std. Dev. = 15
Perinatal	546	368	.077	2.307
Age 1	547	368	.077	2.307

Note: Two-sided α = .05. Power = .80.

an average correlation between measures at baseline and age 3 of 0.5. We also adjust the nominal sample size for design effects due to clustering and unequal weighting according to the over sampling design described above, to yield the effective sample sizes in the table.

At the child level, if we compared normalized assessment scores (mean of 100, standard deviation of 15) of perinatal cohort children at age 3 for two approximately equal-sized subgroups (that is, each having about half the total sample, or about 184 children), this design would allow us to detect a minimum difference of 4.4 points with 80 percent power. If we did a pre-post comparison (age 1 to age 3) for the same assessment measure for children in the age 1 cohort (that is, an *n* of about 368 at age 3), we would be able to detect a minimum difference of 2.3 points. The latter is more than enough power to detect developmentally meaningful change; the former is adequate for most developmental measures. Differences of one-quarter to one-third of a standard deviation were found on some outcomes in the Early Head Start Research and Evaluation Project, but are considered to be large differences. In this study of Early Head Start participants, smaller differences between subgroups are more likely to exist and such differences will have a small probability of being detected under the 90-site design.

(5) Unusual Problems Requiring Specialized Sampling Procedures

We do not anticipate any unusual problems that require specialized sampling procedures.

B.3 Methods for Maximizing Response Rates

MPR has a successful record of high response rates in research studies of Early Head Start and preschool programs, centers, teachers, children, and families and will make every effort to maximize program response rates on Baby FACES 2009. To do this, MPR will continue the procedures that have worked well for us on other projects to recruit programs for Baby FACES.

Early Head Start programs will be motivated to participate because they are vested in the success of the Early Head Start program. ACF will send a letter signed by Dr. Rachel Chazan Cohen, the federal Project Officer as well as a member of the senior staff at the Office of Head Start, to the programs encouraging their participation in the feasibility study prior to any contact by MPR. A brochure promoting participation in the study will be available for review by the Early Head Start program directors.

Recruitment: After the Early Head Start program is sampled for the study, materials explaining Baby FACES will be sent to the program directors for their review, including a brochure, a brief study description, parental consent form, and a list of frequently asked questions.

Each recruited Early Head Start program will have an On Site Coordinator and will be supported by an MPR Baby FACES coordinator. Through biweekly calls, the MPR coordinator will guide the On Site Coordinator through the process of recruiting program families to participate, explain all recruitment materials, and respond to questions that arise.

B.4 Tests of Procedures to Minimize Burden

The assessments proposed for use in the Baby FACES study have all been used in other large-scale studies such as FACES and the National Head Start Impact Study. New procedures using electronic enhancements were used for the FACES 2006 study, and we are confident that these procedures will minimize burden on Early Head Start staff. Iterative pretesting of the Web-based Teacher Child Report Forms to ensure that they are user-friendly is scheduled for the spring of 2008.

B.5 Identity of Person Consulted on Statistical Aspects of Design, and Identity of Contractors

The contractors for this project are Mathematica Policy Research under contract number HHSP23320072914YC and its subcontractors Twin Peaks Partners, LLC; Branch Associates, Inc.; Shugoll Research; Zero to Three; Brenda Jones Harden; and Alphabet Soup Bookstore.

Staff at Mathematica Policy Research

Daniel Kasprzyck, Director, Statistical Services	(202) 264-3482
John Hall, Senior Statistician,	(609) 275-2357
Barbara Carlson, Senior Statistician	(609) 275-2374
John Deke, Senior Researcher	(609) 275-2230