# Head Start Family and Child Experiences Survey (FACES 2019): Recruitment of Programs and Selecting Centers

## OMB Information Collection Request
## 0970 - 0151

# Supporting Statement

# Part B

**February 2018**

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

4th Floor, Mary E. Switzer Building
330 C Street, SW
Washington, D.C. 20201

Project Officers:
Mary Mueggenborg and Meryl Barofsky

# CONTENTS

# APPENDICES

## ATTACHMENTS

**TABLES**

**FIGURES**

**B.1.     Respondent Universe and Sampling Methods**

The Head Start Family and Child Experiences Survey (FACES) will provide data on a set of key indicators in Head Start Regions I–X (FACES 2019) and Region XI (AI/AN FACES 2019). There is a slightly different sample design and recruitment strategy for FACES 2019 and AI/AN FACES 2019 (whose Head Start grants are awarded to tribal governments or consortiums of tribes). ACF proposes to contact 230 Head Start programs in Regions I-X and 30 Head Start programs in Region XI[1] that will be selected to participate in FACES 2019 or AI/AN FACES 2019 for the purpose of gathering information that will be used (1) to develop a sampling frame of Head Start centers in each program, and (2) to facilitate the selection of the center and classroom samples and, (3) for a subsample of programs, the selection of child samples. In this package, we present the sampling plans for all the information collection activities needed to recruit Head Start programs and centers into FACES 2019 and AI/AN FACES 2019 and contacting those in Regions I-X again in 2022. We will submit a separate information collection request for the FACES 2019 and AI/AN FACES 2019 data collection, including selecting classrooms and children for the study and gathering consent for children, data collection instruments and procedures, data analyses, and the reporting of study findings.

**Target population.** The target population for FACES 2019 and AI/AN FACES 2019 is all Region I through XI Head Start programs in the United States (in all 50 states plus the District of Columbia), their classrooms, and the children and families they serve. The sample design is similar to the one used for FACES 2014 and the first AI/AN FACES in 2015. FACES 2019 and AI/AN FACES 2019 will use a stratified multistage sample design with four stages of sample selection: (1) Head Start programs, with programs defined as grantees or delegate agencies providing direct services; (2) centers within programs; (3) classes within centers; and (4) for a random subsample of programs, children within classes. To minimize the burden on parents/guardians who have more than one child selected for the sample, we will also randomly subsample one selected child per parent/guardian, a step that was introduced in FACES 2009.

**Sampling frame and coverage of target population.** The frame that will be used to sample programs is the latest available Head Start Program Information Report (PIR), which has been used as the frame for previous rounds of FACES. Because we will begin recruiting programs for AI/AN FACES 2019 from Region XI in fall 2018, we will use the most current PIR (most likely the 2016-2017 PIR) to select those programs. FACES 2019 recruitment will begin spring 2019, so we will likely use the 2017-2018 PIR for the programs from Regions I-X. We will exclude from the sampling frame: Early Head Start programs, programs in Puerto Rico and other U.S. territories, Region XII migrant and seasonal worker programs, programs that do not directly provide services to children, programs under transitional management, and programs that are (or will soon be) defunded.[2] The center frame will be developed through contacts with the sampled programs. Similarly, the classroom and child frames will be constructed after center and classroom samples are drawn, respectively. All centers, classrooms, and children in study-

---

[1] This package request is to contact 230 Head Start programs in Regions I through X and 30 AI/AN Region XI Head Start programs with the goal of obtaining participation from 180 Head Start programs and 22 AI/AN Region XI Head start programs.

[2] We will work with the Office of Head Start (OHS) to update the list of programs before finalizing the sampling frame. Grantees and programs that were known by OHS to have lost their funding or otherwise closed between the latest PIR and the time of sampling will be removed from the frame, and programs associated with new grants awarded since then will be added to the frame.

eligible, sampled programs will be included in the center, classroom, and child frames, respectively, with three exceptions. Centers that are providing services to Head Start children through child care partnership arrangements are excluded. Classrooms that receive no Head Start funding (such as prekindergarten classrooms in a public school setting that also has Head Start-funded classrooms) are ineligible. Also, sampled children who leave Head Start between fall and spring of the 2019-2020 program year become ineligible for spring 2020. Sampling of centers is included in this information collection; sampling of classrooms and children are described below, but will be included in a future information collection request.

### *Sample design*

**FACES 2019.** The sample design for the new round of FACES is based on the one used for FACES 2014, which in turn was based on the designs of the five previous rounds. Like FACES 2014, the sample design for FACES 2019 will involve sampling for two study components: Classroom + Child Outcomes Core and Classroom Core. The Classroom + Child Outcomes Core study will involve sampling at all four stages (programs, centers, classrooms, and children) and the Classroom Core study will involve sampling at the first three stages only (excluding sampling of children within classes). Under this design, the collective sample size across the two studies will be large enough at the program, center, and classroom levels to allow for sufficiently powerful analyses of program quality, especially at the classroom level. We investigated whether the sample design, including its nested structure, supports answering the study's important questions, and found that it allowed for point estimates with adequate precision at the classroom and child levels, and allowed for the detection of expected differences between subgroups, even after accounting for sample design complexities. (See section on "Degree of Accuracy" below.) We also found that adding more classrooms per center than in the current design would not provide much in the way of gains in the precision for detecting the impact of program-level characteristics on classroom-level estimates. We found that doubling the number of classrooms would only reduce the minimum detectable differences for classroom quality estimates by about 0.03 standard deviations, whether in the context of statistically testing the difference between classroom means of two program-defined subgroups or in the hierarchical linear model (HLM) framework.

As was the case for FACES 2014, the child-level sample for FACES 2019 will represent children enrolled in Head Start for the first time and those who are attending a second year of Head Start. This allows for a direct comparison of first- and second-year program participants and analysis of child gains during the second year. Prior to 2014, FACES followed newly enrolled children through one or two years of Head Start and then through spring of kindergarten. FACES 2019 will follow all children, but only through the fall and spring of one program year.

To minimize the effects of unequal weighting on the variance of estimates, we propose sampling with probability proportional to size (PPS) in the first two stages. At the third stage we will select an equal probability of classrooms within each sampled center and, in centers where children are to be sampled, an equal probability sample of children within each sampled classroom. The measure of size for PPS sampling in each of the first two stages will be the number of classrooms. This sampling approach maximizes the precision of classroom-level estimates and allows for easier in-field sampling of classrooms and children within classrooms. We will select a total of 180 programs across both Core study components for Regions I-X. Sixty

of the 180 programs sampled for the Core study will be randomly subsampled with equal probability within strata to be included in the Classroom + Child Outcomes study. Within these 60 programs, we will select, if possible, two centers per program, two classes per center, and a sufficient number of children to yield 10 consented children per class, for a total of about 2,400 children at baseline.

Based on our experience with earlier rounds of FACES, we estimate that 70 percent of the 2,400 baseline children (about 1,680) will be new to Head Start. We expect a program and study retention rate of 90 percent from fall to spring, for a sample of 2,160 study children in both fall 2019 and spring 2020, of which about 1,512 (70 percent) are estimated to have completed their first Head Start year.

The FACES 2019 Classroom Core study component will include the 60 programs where children are sampled plus the remaining 120 programs from the sample of 180 in Regions I-X. We will select, from the additional 120 programs, two centers per program, and two classrooms per center. Across both study components, we will have a total of 360 centers and 720 classrooms for spring 2020 data collection. For follow-up data collection in spring 2022, we will select a refresher sample[3] of programs and their centers so that the new sample will be representative of all programs and centers in Regions I-X at the time of follow-up data collection, and we will select a new sample of classrooms in all centers. Figure B.1 is a diagram of the sample selection and data collection procedures. At each sampling stage, we will use a sequential sampling technique based on a procedure developed by Chromy.[4]
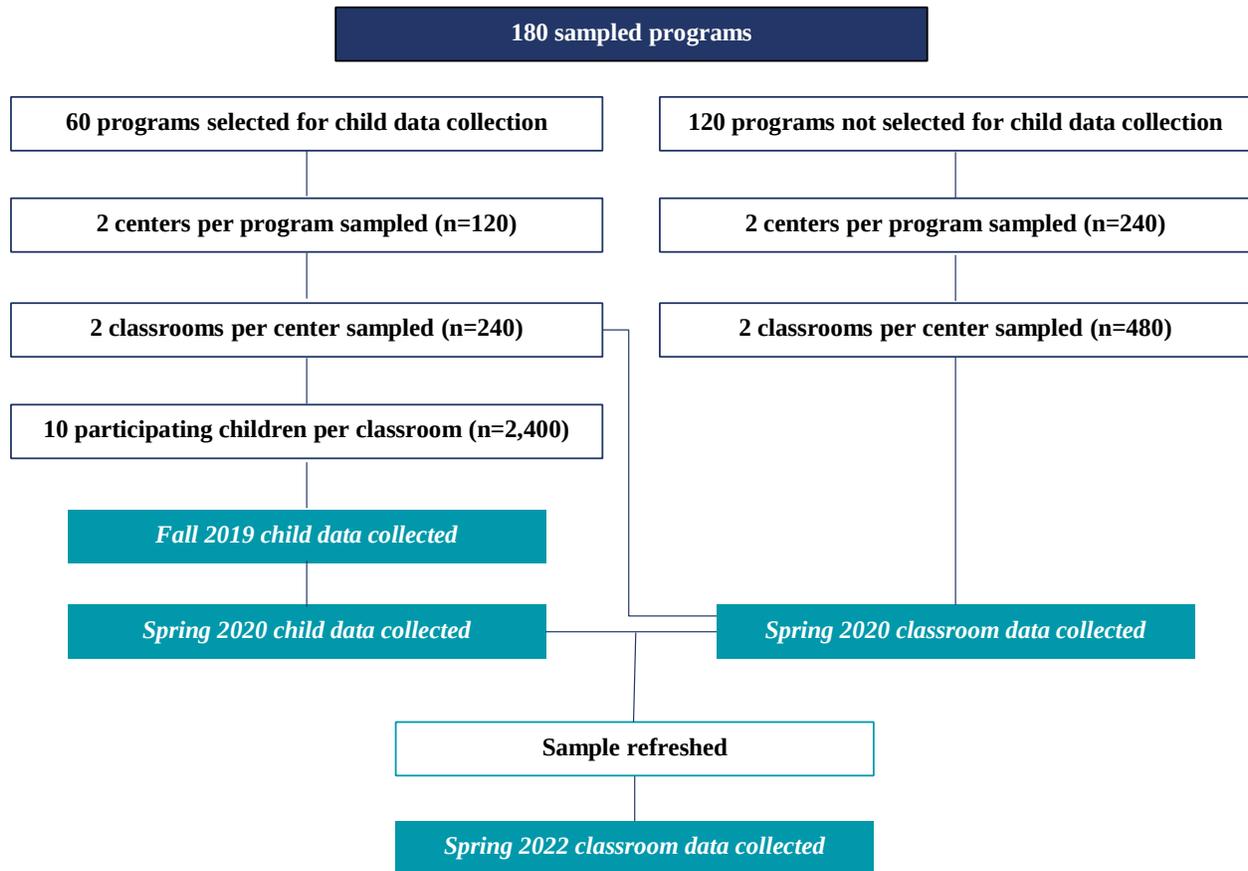
We will initially select double the target number of programs, and pair adjacent selected programs within strata. (These paired programs would be similar to one another with respect to the implicit stratification variables.) We will then randomly select one from each pair to be released as part of the main sample of programs. After the initially released programs are selected, we will ask the Office of Head Start (OHS) to confirm that these programs are in good standing. If confirmed, each program will be contacted and recruited to participate in the study: the 60 programs subsampled for the Classroom + Child Outcomes Core will be recruited in spring 2019 (to begin participation in fall 2019); the remaining 120 programs will be recruited in fall 2019 (to begin participation in spring 2020). If the program is not in good standing or 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. All released programs will be accounted for as part of the sample for purposes of calculating response rates and weighting adjustments. At subsequent stages of sampling, we will release all sampled cases, expecting full participation among the selected centers and classes. At the child level, we estimate that out of 12 selected children per class, we will end up with 10 eligible children with parental consent, which is our target. We expect to lose, on average, two children per class because they are no longer enrolled, because parental consent was not granted, or because of the subsampling of selected siblings.

---

[3] The freshening of the program sample for FACES 2019 will use well-established methods that ensure that the refreshed sample can be treated as a valid probability sample.

[4] The procedure offers all the advantages of the systematic sampling approach but eliminates the risk of bias associated with that approach. 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,* Alexandria, VA: American Statistical Association, 1979, pp. 401–406.

**Figure B.1. Flow of sample selection procedures for Regions I-X**

| 180 sampled programs |
|---|

| 60 programs selected for child data collection | 120 programs not selected for child data collection |
|---|---|
| 2 centers per program sampled (n=120) | 2 centers per program sampled (n=240) |
| 2 classrooms per center sampled (n=240) | 2 classrooms per center sampled (n=480) |
| 10 participating children per classroom (n=2,400) | |
| *Fall 2019 child data collected* | |
| *Spring 2020 child data collected* | *Spring 2020 classroom data collected* |

| Sample refreshed |
|---|

| *Spring 2022 classroom data collected* |
|---|

We will select centers PPS within each sampled program using the number of classrooms as the measure of size, again using the Chromy procedure. For the Classroom + Child Outcomes Core, we will randomly select classrooms and home visitors within center with equal probability. Classrooms with very few children will be grouped with other classrooms in the same center for sampling purposes to ensure a sufficient sample yield.[5] Once classrooms are selected, we will select an equal probability sample of 12 children per class, with the expectation that 10 will be eligible and will receive parental consent. For the additional 120 programs in the classroom core, we will randomly select classrooms within center. Home visitors will not be sampled in this study component, and classroom grouping will not be necessary, as child sample sizes are not relevant for this set of programs.

**AI/AN FACES 2019**. In this section, we will highlight any differences in the sample design relative to FACES 2019 described above. We will select a total of 22 programs for Region XI. Within these 22 programs, we will select, if possible, two centers per program, two classes per center, and a sufficient number of children to yield 10 consented children per class, for a total of about 800 children from Region XI. Based on our experience with AI/AN FACES 2015, we expect a program and study retention rate of 90 percent from fall to spring, for a sample of 720 study children in both fall 2019 and spring 2020. We will have a total of 37 centers and 80 classrooms for spring 2020 data collection. Figure B.2 is a diagram of the sample selection and

---

[5] If the number of children per class is not available at the time of classroom sampling, we will randomly sample three classrooms, and then randomly subsample two for initial release. If these two classrooms are not likely to yield 20 children, we will release the third classroom as well.

data collection procedures. The primary stratification of the program sample is based on program structure (number of centers per program, and number of classrooms per single-center programs, as described in the next section in more detail). At the child level, we estimate that out of 13 selected children per class, we will end up with 10 eligible children with parental consent, which is our target. We expect to lose, on average, three children per class because they are no longer enrolled, because parental consent was not granted, or because of the subsampling of selected siblings.

**Figure B.2. Flow of sample selection procedures for Region XI**
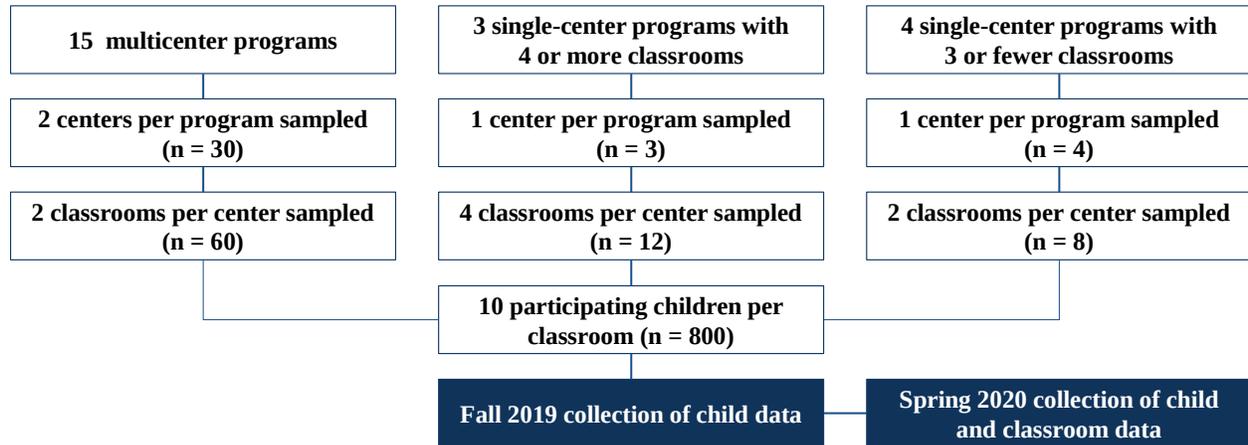


*Sampling and estimation procedures*

**Statistical methodology for stratification and sample selection.** The sampling methodology is described under section B1 above. When sampling programs from Regions I-X, we will form explicit strata using census region, metro/nonmetro status, and percentage of racial/ethnic minority enrollment. Sample allocation will be proportional to the estimated fraction of eligible classrooms represented by the programs in each stratum.[6] We will implicitly stratify (sort) the sample frame by other characteristics, such as percentage of dual language learner (DLL) children (categorized), whether the program is a public school district grantee, and the percentage of children with disabilities. No explicit stratification will be used for selecting centers within programs, classes within centers, or children within classes, although some implicit stratification (such as the percentage of children who are dual language learners) may be used for center selection. For sampling programs from Region XI, we will form three explicit strata using program structure (number of centers and classrooms), and further divide the largest of these three strata multicenter programs) into five geographic regions to form a total of seven strata. The AI/AN FACES Workgroup provided guidance prior to the AI/AN FACES 2015 data collection on how to combine into five groups the states in which Region XI Head Start programs exist. We will implicitly stratify the frame by the percentage of children in the program who are AI/AN (from PIR information) to ensure we get a range of programs with respect to the percentage of children who are AI/AN.

**Estimation procedure.** For both FACES 2019 and AI/AN FACES 2019, we will create analysis weights to account for variations in the probabilities of selection and variations in the eligibility and cooperation rates among those selected. For each stage of sampling (program,

---

[6] We will stochastically round the stratum sizes as needed.

center, class, and child) and within each explicit sampling stratum, we will calculate the probability of selection. The inverse of the probability of selection within stratum at each stage is the sampling or base weight. The sampling weight takes into account the PPS sampling approach, the presence of any certainty selections, and the actual number of cases released. We treat the eligibility status of each sampled unit as known at each stage. Then, at each stage, we will multiply the sampling weight by the inverse of the weighted response rate within weighting cells (defined by sampling stratum) to obtain the analysis weight, so that the respondents' analysis weights account for both the respondents and nonrespondents.

Thus, the program-level weight adjusts for the probability of selection of the program and response at the program level; the center-level weight adjusts for the probability of center selection and center-level response; and the class-level weight adjusts for the probability of selection of the class and class-level response. For FACES 2019 only, the child-level weights adjust for the subsampling probability of programs for the Classroom + Child Outcomes Core. For FACES 2019 and AI/AN FACES 2019, we then adjust for the probability of selection of the child within classroom, whether parental consent was obtained, and whether various child-level instruments (for example, direct child assessments and parent surveys) were obtained. The formulas below represent the various weighting steps for the cumulative weights through prior stages of selection, where P represents the probability of selection and RR the response rate at that stage of selection.

$$W_{pgm} = \frac{1}{P_{pgm}} \cdot \frac{1}{RR_{pgm}}$$

$$W_{center} = W_{pgm} \cdot \frac{1}{P_{center}} \cdot \frac{1}{RR_{center}}$$

$$W_{class} = W_{center} \cdot \frac{1}{P_{class}} \cdot \frac{1}{RR_{class}}$$

$$W_{child} = W_{class} \cdot \frac{1}{P_{pgm-subsample}} \cdot \frac{1}{P_{child}} \cdot \frac{1}{RR_{child}}$$

**Degree of accuracy needed for the purpose described in the justification.** The complex sampling plan, which includes several stages, stratification, clustering, and unequal probabilities of selection, requires using 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 larger variances than would be calculated with standard software. Two approaches for estimating variances under complex sampling, Taylor Series and replication methods, can be estimated by using SUDAAN and special procedures in SAS, Stata, and other packages.

Most of the analyses will be at the child and classroom levels. Given various assumptions about the sample design and its impact of estimates, the sample size should be sufficiently large to detect meaningful differences. In Tables B.1 and B.2 (for Regions I-X and Region XI,

respectively), we show the minimum detectable differences with 80 percent power (and alpha=0.05) and various sample and subgroup sizes, assuming different intraclass correlation coefficients for classroom- and child-level estimates at the various stages of clustering (see table footnote).

For point-in-time estimates, we are making the conservative assumption that there is no covariance between estimates for two subgroups, even though the observations may be in the same classes, centers, and/or programs. By conservative, we mean that smaller differences than those shown will likely be detectable. For pre-post estimates, we do assume covariance between the estimates at two points in time. Evidence from another survey shows expected correlations between fall and spring estimates of about 0.5. Using this information, we applied another design effect component to the variance of estimates of pre-post differences to reflect the fact that it is efficient to have many of the same children or classes at both time points.

The top section of Tables B.1 and B.2 (labeled "Point in Time Subgroup Comparisons") shows the minimum differences that would be detectable for point-in-time (cross-sectional) estimates at the class and child levels. We have incorporated the design effect attributable to clustering. We show minimum detectable differences between point-in-time child subgroups defined two different ways: (1) assuming the subgroup is defined by program-level characteristics, and (2) assuming the subgroup is defined by child-level characteristics (which reduces the clustering effect in each subgroup). The bottom section (labeled "Estimates of Program Year Gains") shows detectable pre-post difference estimates at the child level. Examples are given below.

The columns farthest to the left ("Subgroups" and "Time Points") show several sample subgroup proportions (for example, a comparison of male children to female children would be represented by "50, 50"). The child-level estimates represent two scenarios: (1) all consented children in fall 2019 (n = 2,400) and (2) all children in spring 2020 who remained in Head Start (n = 2,160). For example, the n = 2,400 row within the "33, 67" section represents a subgroup comparison involving children at the beginning of data collection for two subgroups, one representing one-third of that sample (for example, children in bilingual homes), the other representing the remaining two-thirds (for example, children from English-only homes).

The last few columns ("MDD") show various types of variables from which an estimate might be made; the first two are estimates in the form of proportions, the next is an estimate for a normalized variable (such as an assessment score) with a mean of 100 and standard deviation of 15 (for child-level estimates only), and the last shows the minimum detectable effect size—the MDD in standard deviation-sized units. The numbers for a given row and column show the minimum underlying differences between the two subgroups that would be detectable for a given type of variable with the given sample size and design assumptions.

**Table B.1. FACES 2019 minimum detectable differences – Regions I-X**

| POINT IN TIME SUBGROUP COMPARISONS FOR CLASSROOMS AND CHILDREN | | | | | | | |
|---|---|---|---|---|---|---|---|
| **CLASSROOM SUBGROUPS** | | | | **Minimum detectable difference** | | | |
| Time point | Percentage in group 1 | Percentage in group 2 | Classes in group 1 | Classes in group 2 | Proportion of 0.1 or 0.9 | Proportion of 0.5 | | Minimum detectable effect size |
| Spring 2020 | 50 | 50 | 360 | 360 | .084 | .140 | | .280 |
| | 33 | 67 | 238 | 482 | .090 | .149 | | .298 |
| | 15 | 85 | 108 | 612 | .119 | .198 | | .392 |
| **CHILD SUBGROUPS** | | | | **Minimum detectable difference** **(Program-defined subgroups / Child-defined subgroups)** | | | |
| Time point | Percentage in group 1 | Percentage in group 2 | Children in group 1 | Children in group 2 | Proportion of 0.1 or 0.9 | Proportion of 0.5 | Normalized variable (mean = 100, s.d.= 15) | Minimum detectable effect size |
| Fall 2019 | 50 | 50 | 1,200 | 1,200 | 0.091/0.067 | 0.151/0.112 | 4.5/3.4 | 0.301/0.224 |
| | 33 | 67 | 792 | 1,608 | 0.096/0.068 | 0.161/0.113 | 4.8/3.4 | 0.320/0.227 |
| | 40 | 30 | 960 | 720 | 0.110/0.070 | 0.183/0.117 | 5.5/3.5 | 0.364/0.233 |
| Spring 2020 | 50 | 50 | 1,080 | 1,080 | 0.091/0.068 | 0.152/0.113 | 4.5/3.4 | 0.303/0.226 |
| **ESTIMATES OF PROGRAM YEAR GAINS FOR CHILDREN** | | | | | | | |
| **TIME POINTS** | | | | **Minimum detectable difference** | | | |
| Time 1 | Time 2 | Percent Subgroup at both times | Children at time 1 | Children at time 2 | Proportion of 0.1 or 0.9 | Proportion of 0.5 | Normalized variable (mean = 100, s.d.= 15) | Minimum detectable effect size |
| Fall 2019 | Spring 2020 | 100 | 2,400 | 2,160 | 0.048 | 0.079 | 2.375 | 0.158 |
| | | 70 | 1,680 | 1,512 | 0.057 | 0.095 | 2.839 | 0.189 |
| | | 40 | 960 | 864 | 0.075 | 0.126 | 3.756 | 0.250 |

Note: Conservative assumption of no covariance for point-in-time subgroup comparisons. Covariance adjustment made for pre-post difference (Kish, p. 462, Table 12.4.II, Difference with Partial Overlap). Assumes =.05 (two-sided), .80 power. For classroom-level estimates, assumes 180 programs, 360 centers, between-program ICC=.2, between-center ICC = .2. For child-level estimates, assumes 60 programs, 120 centers, between-program ICC = .10, between-center ICC = .05, between-classroom ICC = .12.

Note: The minimum detectable effect size is the minimum detectable difference in standard deviation-sized units.

**Table B.2. AI/AN FACES 2019 minimum detectable differences – Region XI**

| POINT IN TIME SUBGROUP COMPARISONS FOR CHILDREN | | | | | | | |
|---|---|---|---|---|---|---|---|
| | CHILD SUBGROUPS | | | | Minimum detectable difference (Program-defined subgroups / Child-defined subgroups) | | |
| Time point | Percentage in group 1 | Percentage in group 2 | Children in group 1 | Children in group 2 | Proportion of 0.1 or 0.9 | Proportion of 0.5 | Normalized variable (mean = 100, s.d.= 15) | Minimum detectable effect size |
| | 50 | 50 | 400 | 400 | 0.155/0.115 | 0.258/0.191 | 7.7/5.7 | 0.511/0.380 |
| | 33 | 67 | 264 | 536 | 0.165/0.116 | 0.274/0.194 | 8.1/5.8 | 0.543/0.385 |
| Fall 2019 | 40 | 30 | 320 | 240 | 0.187/0.120 | 0.312/0.200 | 9.3/6.0 | 0.617/0.397 |
| Spring 2020 | 50 | 50 | 360 | 360 | 0.155/0.116 | 0.259/0.193 | 7.7/5.8 | 0.514/0.385 |
| ESTIMATES OF PROGRAM YEAR GAINS FOR CHILDREN | | | | | | | |
| TIME POINTS | | | | | Minimum detectable difference | | |
| Time 1 | Time 2 | Percent subgroup at both times | Children at time 1 | Children at time 2 | Proportion of 0.1 or 0.9 | Proportion of 0.5 | Normalized variable (mean = 100, s.d.= 15) | Minimum detectable effect size |
| | | 100 | 800 | 720 | 0.081 | 0.135 | 4.0 | 0.269 |
| Fall 2019 | Spring 2020 | 70 | 560 | 504 | 0.097 | 0.162 | 4.8 | 0.321 |
| | | 40 | 320 | 288 | 0.129 | 0.215 | 6.4 | 0.425 |

Note:     Conservative assumption of no covariance for point-in-time subgroup comparisons. Covariance adjustment made for pre-post difference (Kish, p. 462, Table 12.4.II, Difference with Partial Overlap). Assumes =.05 (two-sided), .80 power. Assumes 22 programs, 37 centers, 80 classrooms, between-program ICC = .10, between-center ICC = .05, between-classroom ICC = .12.

Note:     The minimum detectable effect size is the minimum detectable difference in standard deviation-sized units.

If we were to compare two equal-sized subgroups of the 720 classrooms (Regions I-X) in spring 2020, our design would allow us to detect a minimum difference of .280 standard deviations with 80 percent power. At the child level, if we were to compare normalized assessment scores with a sample size of 2,400 children in fall 2019, and two approximately equal-sized child-defined subgroups (such as boys and girls), our design would allow us to detect a minimum difference of 3.4 points with 80 percent power. If we were to compare these two subgroups again in the spring of 2020, our design would allow us to detect a minimum difference of 3.4 points.

If we were to perform a pre-post comparison (fall 2019 to spring 2020) for the same normalized assessment measure, we would be able to detect a minimum difference of 2.4 points. If we were to perform the same pre-post comparison for a subgroup representing 40 percent of the entire sample (n = 960 in fall 2019; n = 864 in spring 2020), we would be able to detect a minimum difference of 3.8 points.

The main purpose of the AI/AN FACES 2019 is to provide descriptive statistics for this population of children. Comparisons between child subgroups are a secondary purpose, given the smaller sample size. If we were to compare normalized assessment scores for two equal-sized subgroups (say, boys and girls) of the 800 children in Region XI (AI/AN) in fall 2019, our design would allow us to detect a minimum difference of 5.7 points with 80 percent power. If we were to compare these two subgroups again in the spring of 2020, our design would allow us to detect a minimum difference of 5.8 points. If we were to perform a pre-post comparison (fall 2019 to spring 2020) for the same normalized assessment measure, we would be able to detect a minimum difference of 4.0 points.

## B.2.    Procedures for Collection of Information

Upon OMB approval of this information collection request, a memo signed by the Acting Director of the Office of Head Start will be sent to each sampled program's director. The memo will describe the study goals and the importance of the study and introduce the Mathematica team who will be doing the study on ACF's behalf (Part A, Appendix A-1.a for FACES 2019 and A-2.a for AI/AN FACES 2019). A memo and study fact sheet will be sent by Mathematica along with the introductory letter (Part A, Appendix A-1.b and A-1.c for FACES 2019 and Appendix A-2.b and A-2.c for AI/AN FACES 2019).

Program directors will then receive a phone call from a member of the study team to answer any questions about the study. Using a prepared script (Attachment 1 for FACES 2019 and Attachment 2 for AI/AN FACES 2019), the study team will review our request for information and ask for information about centers (names, addresses, and estimated enrollment), how services are organized (center-based, home-based, combination, or locally designed), and scheduling specifics (hours of operation, program year start and end dates, and whether they are a full- or part-day program). This information will be recorded for use in preparing the data collection plans for the study programs. Directors will also be asked to identify a staff member to serve as an on-site coordinator who will work with the study team to recruit participants, develop a data collection plan, and help schedule site visits.

As another element of this information collection, a member of the study team will call on-site coordinators (after they have received a letter describing the study; see Part A, Appendix A-1.d for FACES 2019 and A-2.d for AI/AN FACES 2019) and, using a prepared script

(Attachment 3 for FACES 2019 and 4 for AI/AN FACES 2019), ask the coordinator to provide or confirm information about the centers in their program. Finally, once centers are selected, center directors will receive a letter (Part A, Appendix A-1.e for FACES 2019 and A-2.e for AI/AN FACES 2019) along with the study fact sheet.

Recruitment procedures for AI/AN FACES 2019 will mirror those described above for FACES 2019, with a few important differences. Members of the Mathematica study team will be paired with an AI/AN FACES Workgroup member to recruit programs and obtain tribal approval for the study. During program recruitment calls, the study team will discuss how AI/AN FACES is both scientifically rigorous and culturally responsive. As part of the recruitment packets we send to programs, we will include an Agreement of Collaboration and Participation (Appendix E, refined from AI/AN FACES 2015) that explicitly details the expectations for both Mathematica and the tribal programs so we remain transparent about the study's intent and processes. As part of our scripts, we will determine the tribal approval process for a given program and community. Based on our experience in AI/AN FACES 2015, we will conduct in-person presentations to tribal communities as requested (see Appendix F for a presentation template).

### B.3. Methods to Maximize Response Rates and Deal with Nonresponse

*Expected Response Rates*

Program response rates in FACES 2006, 2009 and 2014 exceeded 90 percent. For AI/AN FACES, our prior experience in 2015-16 shows a somewhat lower participation rate at the program level (just under 80 percent weighted). We will use the same approach and procedures that were used successfully in prior rounds, so we expect to achieve similar response rates.

*Dealing with Nonresponse*

For any response rates lower than 80 percent, we will conduct an analysis to examine the risk of nonresponse bias at the program level before and after applying weighting adjustments. The results of this analysis at the program level will be used to infer potential nonresponse bias at the child level, which cannot be estimated for programs not participating in the study.

*Maximizing Response Rates*

We do not anticipate problems contacting and gaining the cooperation of Head Start programs, and in gathering information from program directors and on-site coordinators. The study team will conduct calls with program directors and on-site coordinators during business hours, at times that coincide best with their schedules.

### B.4. Test of Procedures or Methods to be Undertaken

The proposed procedures were used successfully in FACES 2014 and AI/AN FACES 2015, and there are no plans to test the procedures.

### B.5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

The team is led by Mary Mueggenborg and Dr. Meryl Barofsky, Federal Contracting Officer's Representatives (CORs); Dr. Nina Philipsen Hetzner, Contract Project Specialist; Dr. Laura Hoard, Federal Project Officer; Dr. Lizabeth Malone, project director; Drs. Louisa Tarullo and Nikki Aikens, co-principal investigators; and Annalee Kelly, survey director. Additional staff consulted on statistical issues include Barbara Carlson, a senior statistician at Mathematica, and Drs. Margaret Burchinal and Marty Zaslow, consultants to Mathematica on statistical and analytic issues.