# The Early Head Start Family and Child Experiences Survey (Baby FACES)—2020/2022

## OMB Information Collection Request
## 0970 - 0354

# Supporting Statement

# Part B

**MARCH 2022**

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 Officer
**Amy Madigan, Ph.D.**

# Part B

**B1.    Objectives**

*Study Objectives*

The Administration for Children and Families (ACF) at the U.S. Department of Health and Human Services (HHS) seeks approval to collect descriptive information for the Early Head Start Family and Child Experiences Survey 2020/2022 (Baby FACES 2020/2022). The goal of this information collection is to provide updated nationally representative data on Early Head Start (EHS) programs, staff, and families to guide program planning, technical assistance, and research. In addition to collecting data for comparison to Baby FACES 2018, Baby FACES 2020/2022 will take an in depth look at home visits.

*Generalizability of Results*

Data will be collected from a nationally representative sample of EHS programs, their associated staff, and those they serve. The results are intended to be generalizable to the EHS program as a whole, with a few restrictions. EHS programs in Alaska, Hawaii, and U.S. territories are excluded, as are those under the direction of ACF Region XI (American Indian and Alaska Native Head Start), Region XII (Migrant and Seasonal Worker Head Start), and programs under transitional management.  This limitation will be clearly stated in published results.

*Appropriateness of Study Design and Methods for Planned Uses*

Baby FACES was primarily designed to address questions of importance for technical assistance, program planning, and to the research community. The study's conceptual framework (briefly described in Section A and attached in depth as Appendix A) guided the study's overall design, sampling approach, and the information to be collected from each category of respondent. The study's sample is designed so that the resulting weighted estimates are unbiased, sufficiently precise, and with adequate power to detect relevant differences at the national level.

The topical focus on home visiting for Baby FACES 2020/2022 is reflected in previously approved observational data collection on home visit quality and parent child interactions. The sub-framework for home-based services (Appendix A, Figure 3) illustrates how program processes and activities are hypothesized to be associated with high quality service delivery and enhanced family and infant/toddler outcomes.  Restricted-used data from Baby FACES 2020/2022 will be archived for secondary analysis by researchers interested in exploring nonexperimental associations between program processes and intermediate and longer-term outcomes, and the published frameworks will be included in documentation to support responsible secondary data use.

**B2.    Methods and Design**

*Target Population*

The target population for Baby FACES 2020/2022 is a nationally representative sample of EHS programs and their associated centers, home visitors, classrooms, teachers, and the families and children they

serve.[1] Based on administrative data, there are approximately 1,000 EHS programs eligible for inclusion in the study population.  These programs directly provide services to children; are located in ACF Regions 1 through 10; are based in one of the 48 contiguous states; and not under transitional management. These programs serve about 140,000 children and pregnant women, in about 11,000 center-based classrooms or through home visits from about 5,500 home visitors.

*Sampling and Site Selection*

*Sampling Overview*

To produce nationally representative estimates and track changes over time, Baby FACES 2020/2022 will build upon the Baby FACES 2018 sampling strategy.  To balance the desire for precise and unbiased estimates with the logistical realities of data collection, Baby FACES 2018 collected data from staff, classrooms, and families within randomly selected EHS programs.  The first stage of sample selection was EHS programs. Within each program, the team selected a sample of centers and/or home visitors, depending on the type(s) of services the program provided. Within each center, we selected a sample of classrooms (and their associated teachers), and a sample of children within classrooms. For each sampled home visitor, we selected a sample of pregnant women and children from their caseloads. These data were used to produce estimates of representative EHS programs; their associated centers and home visitors; classrooms and teachers; and the families they serve.

*Selection of EHS Programs*

Baby FACES 2018 selected a probability proportional to size (PPS) sample of EHS programs from a sample frame derived from the 2016 Head Start Program Information Report (PIR)[2].  The 2016 PIR included administrative information from all Head Start and EHS programs (grantees and delegate agencies) from program year 2015–2016.[3] PIR data were used PIR as explicit and implicit stratification variables in the selection of the EHS sample.

In explicit stratification, the sample is allocated and selected separately within each stratum, allowing for more control over how the sample is distributed across important characteristics. The explicit stratification variables were whether the program provided center-based services only, home-based services only, or both. In implicit stratification, the sampling frame is sorted by one or more additional important characteristics within explicit strata before selecting the sample, as a way of enhancing the sample's representativeness. The implicit stratification variables for the Baby FACES program sample are whether the program has a majority of Spanish-speaking enrollees, whether the program is located in a

---

[1] Although Baby FACES 2018 included pregnant women in the sample, we subsequently dropped them from analysis due to small sample sizes.  We therefore will omit pregnant women in the 2022 round of Baby FACES.

[2] The PIR is an administrative data system for the Head Start program as a whole that includes data collected annually from all programs. Head Start programs collect the information as approved under OMB control number 0970-0427.

[3] Before selecting the sample, we excluded all Head Start programs (i.e., programs serving only preschool-aged children) as well as any EHS programs that are overseen by ACF regional offices XI (American Indian and Alaska Native) and XII (Migrant and Seasonal), any programs that are under transitional management, any programs outside the 50 states and the District of Columbia, and any programs that do not directly provide services to children and families. About 1,000 EHS programs remained after these exclusions. We also combined programs that had different grant numbers but had the same program director into a single unit for sampling, data collection, and analysis.

metropolitan or non-metropolitan area,[4] and the program's ACF region. Explicit stratification is not used to oversample any type of program, but it is employed to ensure that the sample of programs represents the most policy-relevant characteristics proportional to their distribution in the population. The program sample is allocated across explicit strata to maximize precision at the end of the multistage sampling process. We select a PPS sample of programs using a sequential sampling procedure in SAS developed by Chromy (1979). As with any PPS selection, we appropriately account for programs that are so large relative to others in their stratum that they are selected with certainty. We used funded enrollment as the measure of size to allocate and select the program sample.

Baby FACES 2018 aimed to collect data on 140 EHS programs. To achieve this goal, we first selected an augmented PPS sample of more than 300 programs. Second, we formed pairs of adjacent selections within strata (those with similar implicit stratification characteristics). Finally, we randomly selected one program from each pair to release initially, with the other member of the pair becoming its backup. We released the backup only if the main release turned out to be ineligible (for example, closed or in imminent danger of losing its funding) or refused to participate. We also selected one or more extra pairs within each stratum in case any main pair yielded no eligible and participating sampled programs. We properly accounted for programs released into the sample initially or as a backup in the weights and in the response rates.

For **Baby FACES 2020/2022,** we will begin by retaining as many of the programs sampled in 2018 as possible. However, to produce representative estimates for 2022, we need to "freshen" the sample by including programs that came into being since the original sample was selected.  The total sample will also be reduced by 18 programs for resource reasons. To address resource constraints while allowing for the inclusion of new programs, we will randomly select out some of the programs sampled and participating in 2018, after accounting for natural attrition from programs that closed or lost their funding since 2018.  After adjusting for this subsampling, the program weight for Baby FACES 2018 program i in stratum h that is retained for Baby FACES 2020/2022 is:

$$W\,2022_{hi} = W\,2018_{hi} \cdot \left(n_h/m_h\right) \cdot NRADJ_h$$

Where $W2018_{hi}$ is the final nonresponse-adjusted weight for program i from Baby FACES 2018, $m_h$ is the number of programs in stratum h retained in the random subsample for Baby FACES 2020/2022, $n_h$ is the total number of programs in stratum h that are still eligible in 2022, and $NRADJh$ is the nonresponse adjustment factor to account for 2018 programs that are still eligible in 2022 but opt not to participate.

Using the 2018 PIR, and with assistance from the Office of Head Start, we will first identify any 2018 sampled and participating programs that closed or lost funding in the interim, which will simply be dropped with no further weighting adjustments, and all programs that came into being since the previous sample was drawn. The new programs will form three new sampling strata, from which we will select a modest number using PPS sequential selection (with the size measure being funded enrollment), again explicitly stratifying by types of services offered (center-only, home-visiting only, or mixed). As was done for the original Baby FACES 2018

---

[4] Metropolitan area status was merged onto the file using the program zip code in the PIR.

sample, we will select twice the number needed,[5] and form pairs of similar programs. Within each main pair, one program will be randomly selected to be released. The other member of the program pair will be released if the first program refuses or turns out to be ineligible. The probability of selection for new program i in new stratum h' is calculated as:

$$P_{h'i} = n_{h'} \left( PMOS_{h'i} / \sum_{j=1}^{N_{h'}} PMOS_{h'j} \right) \cdot (1/2) \cdot ¿$$

Where $PMOS_{h'i}$ is the measure of size (enrollment) for program i in stratum h', $n_{h'}$ is the number of programs selected in stratum h' for the augmented sample (including backups), $N_{h'}$ is the total number of programs in stratum h', and the last two terms are preliminary adjustments (before backup releases) for within-pair and main ($a_{h'}$) vs. extra ($b_{h'}$) pair selections.

We expect to randomly subsample about 15 programs from the Baby FACES 2018 sample to exclude from the Baby FACES 2020/2022 sample, stratifying by service type. (The exact number will depend on the number of programs that closed or lost funding at the time of the 2020 sampling and the number of new programs available for sampling.)  Among the 2018 programs that are retained, some will likely refuse to participate again in 2022. We will release programs from the extra pairs in one of the corresponding new strata to account for the 2018 refusers, and account for all nonresponse among selected and eligible programs, both those from the 2018 sample and those new to the 202 sample.

By March 2020, when the COVID-19 pandemic interrupted data collection, we had nearly finished recruitment (117 of 123 programs) and had begun data collection (19 programs).  Our plan is to restart data collection in spring 2022, reaching out in fall 2021 to our previously recruited and not-yet-recruited programs. Should any of our currently sampled and released programs (recruited or not-yet-recruited) decline to participate in 2022, we plan to continue releasing backup programs not yet released from our spring 2020 sample as needed.  If this does not yield 123 participating programs, we can sample additional programs from the "new program" sample frame.

*Selection of Nested Samples*

Once programs have been successfully recruited or retained for 2022 data collection, we will select fresh samples of associated home visitors, centers, classrooms, and children in all participating programs. Because of a stronger emphasis on home visiting services in Baby FACES 2020/2022 relative to 2018, we plan to select more home visitors and fewer centers per program in this data collection.

As each sampled program is recruited into the study, we will ask the program to provide a list of all its centers and home visitors, along with characteristics such as number of classrooms (for centers) and size of caseload (for home visitors). Based on our experience in 2018, we expect 88 percent of programs to provide center-based services and 80 percent to provide home-based care (with 68 percent providing both). This will result in a sample of 123 programs, including 108 programs offering center-based and 98 programs offering home-based services. Not all programs will provide center-based services, and not all will provide home visiting services, but we expect about 83 will provide both. Following the protocol established in Baby FACES 2018, we will sample **centers** and **home visitors** on a rolling basis, as each sampled program is recruited or retained.

---

[5] We will also select one or two extra pairs in each sampling stratum, to be released should any of the initial pairs not yield a participating program, and to allow for additional attrition of 2018 programs due to refusal.

**Alternative Supporting Statement for Information Collections Designed for
Research, Public Health Surveillance, and Program Evaluation Purposes**

In Baby FACES 2018, we sampled an average of 4.2 centers per program with center-based services for EHS children, 1.9 classrooms per participating center, and 2.9 children per sampled classroom. For Baby FACES 2020/2022, we will sample an average of 3 centers per program. We expect that roughly half of programs will have fewer than 3 such centers; for those programs, we will include all centers in the sample. In all other programs, we will select 3 or more centers with PPS, selecting centers with certainty if their size (number of enrolled children per center) is large enough relative to the other centers in the program. Center i is classified as a certainty selection if $CMOS_{hi} > \sum_{j=1}^{N_h} CMOS_{hj}/n_h$ , where $CMOS_{hi}$ is the measure of size (number of enrolled children) for center i in stratum (program) h, $N_h$ is the total number of centers in stratum h, and $n_h$ is the number of centers to sample in stratum h.

By the time data collection stopped in March 2020 due to the COVID-19 pandemic, we had sampled centers in 99 recruited programs. For those programs that agree to participate when we restart data collection in spring 2022, we will ask if the list of centers we used as a sampling frame for 2020 selection is unchanged, we will keep the originally sampled centers for spring 2022 data collection. If the only change to the list is that one or more of the nonsampled centers has since closed, we can also keep the originally sampled centers for spring 2022 data collection but will perform a ratio adjustment to the center sampling weight to account for the reduced number of centers. For all other changes, including new centers coming into existence, or one of the sampled centers closing, we will need to select a new sample of centers using a new sampling frame.

Within participating centers, we will obtain a list of all classrooms and sample 2 that serve EHS-funded children. (We will select all new samples of classrooms in spring 2022.) Within the sample of centers, we will randomly subsample half from which to select a random sample of 3 EHS-funded children per classroom, of which we expect 2.6 to participate.  (In Baby FACES 2018, all sampled centers were used to select samples of children.)

In Baby FACES 2018, we sampled an average of 6.2 home visitors per program with home-based services for EHS children. We subsampled half of these home visitors from which to sample 2.9 children from their caseloads. In Baby FACES 2020/2022, we will sample an average of 8 home visitors per program. For those programs with fewer than 8 EHS-funded home visitors, we will include all of their home visitors. For all other programs, we will select 8 or more home visitors with PPS, selecting them with certainty if their size (home visitor caseload) is large enough relative to the caseloads of other home visitors in the program. Home visitor i is classified as a certainty selection if

$HMOS_{hi} > \sum_{j=1}^{N_h} HMOS_{hj}/n_h$, where $HMOS_{hi}$ is the measure of size (number of children in caseload) for home visitor i in stratum (program) h, $N_h$ is the total number of home visitors in stratum h, and $n_h$ is the number of home visitors to sample in stratum h.

Within each participating home visitor's caseload (not a random half as in Baby FACES 2018), we will randomly select 3 EHS-funded children, of which we expect 2.3 to participate. (We will select all new samples of home visitors in spring 2022.)

The specific procedures for sampling at levels below the center level (for center-based services) and below the program level (for home visiting services) are described next. A few weeks before the first data collection visit, the Baby FACES liaisons will contact each sampled center to obtain a list of all EHS-funded classrooms and the age range of EHS-funded children in those classrooms, using the

classroom/home visitor sampling form (Attachment 1). The Baby FACES liaison will enter this information into a sampling program. If a center has only 1 or 2 classrooms, the sampling program will include all classrooms in the sample; otherwise, it will select a systematic sample of 2 classrooms, implicitly stratified (sorted) by whether the room is predominantly an infant or toddler classroom. We expect this process will yield 618 center-based classrooms in the sample. Within the randomly subsampled half of centers, using the child roster form (Attachment 2), the Baby FACES liaison will then obtain classroom rosters for each of the two sampled classrooms and enter that information into the sampling program. The sampling program will then select a systematic sample of 3 children per classroom (implicitly stratifying by date of birth); we expect that about 2.3 children per classroom will have parental consent and complete the data collection instruments. If we happen to select for the sample more than one child from the same household in a given center, the program will randomly subsample one child to remain in the study sample to minimize burden on the family. (We will select all new samples of center-based children in spring 2022.)

For sampled programs that provide EHS-funded home visiting services, the Baby FACES liaison will contact each program and ask for a list of these home visitors gathering information listed in the classroom/home visitor sampling form (Attachment 1). After sampling home visitors, the Baby FACES liaison will obtain from the program a list of EHS-funded children on each sampled home visitor's caseload, along with their date of enrollment and their date of birth. The information we will collect is listed in the child roster form (Attachment 2). After the liaison enters this information into the sampling program, the program will select a systematic sample of 3 children per home visitor in the sample (implicitly stratifying by age within the child category). We expect that 1.9 children per home visitor will have study consent and complete all or most data collection instruments. If more than one child from the same household happens to end up in the sample, the program will randomly subsample one to remain in the sample to minimize burden on the family.[6] (We will select all new samples of home-based children in spring 2022.)

**Size of the sample and precision needed for key estimates**

**Sample size.** In Baby FACES 2018, we collected data from 137 sampled programs: 120 with center-based services and 110 with home-based services. Table B.1 shows sample sizes and response rates for Baby FACES 2018 instruments. In Baby FACES 2020/2022, we plan to collect data from 123 programs: 108 with center-based services and 98 with home-based services. Table B.2 shows the approved sample sizes for Baby FACES 2020/2022 and Table B.3 shows the expected number of participating sample members and instrument completions. We expect to select 325 centers, half of which will be subsampled for child-level data collection. We expect 309 of these 325 centers to participate in the study. We expect to select 787 home visitors, with 737 participating in the study and 706 of these completing the staff survey. We expect to select 618 classrooms, with 609 of the associated teachers completing the staff survey and 613 having their classrooms observed.

In Baby FACES 2020/2022, we expect to sample 927 center-based children (805 for whom we expect to get parental consent) and 2,210 children receiving home-based services (1,690 for whom we expect to get parental consent), for a total of 2,495 study participants. Of these, we expect to have 2,083 completed parent interviews, 2,229 completed staff reports, and (for children only) 2,008 completed parent-child reports.

---

[6] In Baby FACES 2018, we had a few situations in which one family member was selected in the center-based sample and another in the home-based sample. A similar subsampling procedure was carried out manually for these cases, and we plan to do the same for Baby FACES 2020/2022.

**Alternative Supporting Statement for Information Collections Designed for Research, Public Health Surveillance, and Program Evaluation Purposes**

Table B.1 Baby FACES 2018 sample sizes and response rates

| Sampling level | Study participation or instrument response | Number of participants or respondents | Unweighted marginal response rate (percent) | Unweighted cumulative response rate (percent) | Weighted marginal response rate (percent) | Weighted cumulative response rate (percent) |
|---|---|---|---|---|---|---|
| Program | Participation | 137 | 83.5 | 83.5 | 92.1 | 92.1 |
| | Program director survey | 134 | 97.8 | 81.7 | 98.7 | 90.8 |
| Center | Participation | 468 | 96.3 | 80.4 | 95.5 | 88.0 |
| | Center director survey | 446 | 95.3 | 76.7 | 95.7 | 84.2 |
| Home visitor | Participation | 611 | 98.7 | 82.5 | 99.0 | 91.2 |
| | Home visitor interview | 586 | 95.9 | 79.1 | 96.1 | 87.6 |
| Classroom | Participation | 871 | 100.0 | 80.4 | 100.0 | 88.0 |
| | Teacher interview | 859 | 98.6 | 79.3 | 98.3 | 86.4 |
| | Classroom observation | 864 | 99.2 | 79.8 | 99.3 | 87.3 |
| Child or pregnant woman | Parental consent | 2,868 | 85.7 | 69.9 | 84.8 | 76.1 |
| | Parent interview | 2,350 | 81.9 | 57.3 | 81.8 | 62.2 |
| | Parent child report (children only) | 2,495 | 88.0 | 62.0 | 86.8 | 66.6 |
| | Staff child or pregnant woman report | 2,708 | 94.4 | 66.0 | 93.4 | 71.0 |

Note 1. Study eligibility status is assumed to be known for all sampling units, including nonrespondents.

Note 2. The marginal response rate is the response rate only among those attempted for that study component or instrument and does not account for nonparticipation at higher level sampling stages. For example, the unweighted marginal response rate for the center director survey is 95.3 percent, which represents the 446 center directors in the 468 participating centers who completed the survey. The unweighted cumulative response rate of 76.7 percent for the center director survey incorporates the 83.5 percent program participation rate and the 96.3 center participation rate as well as the center director instrument response rate.

Note 3. Weighted response rates use the inverse of marginal selection probabilities

Note 4. Child-level includes both center-based and home-visited children.

Table B.2. Expected selected sample counts for Baby FACES 2020/2022

| Sampling stage | | All programs | Programs with only center-based services | Programs with both center- and home-based services | Programs with only home-based services |
|---|---|---|---|---|---|
| Programs | Total participating | 123 | 25 | 83 | 15 |
| Home visitors | Mean per program | n.a. | 0 | 8 | 8 |
| | Total | 787 | 0 | 667 | 120 |
| Centers | Mean per program | n.a. | 3 | 3 | 0 |
| | Total | 325 | 75 | 250 | 0 |
| Subsampled centers for child sampling | Mean per program | n.a. | 1.5 | 1.5 | 0 |
| | Total | 162 | 37 | 125 | 0 |
| Classrooms | Mean per participating center | n.a. | 2 | 2 | 0 |
| | Total | 618 | 143 | 475 | 0 |
| Classrooms in participating subsampled centers for child sampling | Mean per participating center | n.a. | 2 | 2 | 0 |
| | Total | 309 | 71 | 238 | 0 |
| Home-based children | Mean per participating home visitor | n.a. | 0 | 3 | 3 |
| | Total | 2,210 | 0 | 1,872 | 338 |
| Center-based children | Mean per classroom in participating subsampled center | n.a. | 3 | 3 | 0 |
| | Total | 927 | 214 | 713 | 0 |

n.a. = not applicable.

Note: This table shows the number of selected units at each stage. The expected number of participating and responding units can be found in Table B.3.

Table B.3. Expected response rates and number of responses by instrument

| Data source | Number of consented sample members | Expected response rate (percentage) | Expected number of responses |
|---|---|---|---|
| 1. Parent survey | 2,495 | 83.5 | 2,084 |
| 2. Parent Child Report | 2,411 | 83.3 | 2,008 |
| 3. Staff survey (Teacher survey and Home Visitor survey) | 1,354 (618 classroom teachers, and 737 home visitors) | 97.1 | 1,317[7] |
| 4. Staff Child Report | 2,495 | 89.4 | 2,230 (each of the 1,046 participating staff with sampled children will report on 2.1 children on average) |
| 5. Program director survey | 123 | 97.8 | 120 |
| 6. Center director survey | 309 | 95.3 | 294 |

Note:    We have assumed that 20 percent of the programs have centers only, 12 percent have home visiting only, and 68 percent have both centers and home visitors. We will be selecting an average of 3 centers per program and 2 classrooms per participating center. This yields a total of 618 classrooms. But we will subsample 1.5 centers per program from which to sample children. For home visitors, we will select an average of 8 home visitors per program for a total of 787.

**Precision needed for key estimates.** Baby FACES 2020/2022 uses a complex, multistage clustered sample design. Such a design has many advantages, but there is a cost in terms of the precision of estimates. Clustering and unequal weighting increase the variance of estimates, and this can be quantified in terms of the design effect.[8] Table B.4 shows the precision of estimates for each point estimate after accounting for expected design effects; Table B.5 shows the minimum detectable effect sizes for comparing subgroups (with approximated subgroup sizes).

In the tables, we make the following assumptions. We assume a Type I error rate of 0.05 (two-sided) and power of 0.80. For estimates shown in the tables, we assume a design effect due to unequal weighting of 1.2, mostly due to nonresponse adjustment, and assuming that our multistage PPS sample results in fairly even cumulative sampling weights for these estimates. Based on findings from similarly designed studies of Head Start (Aikens et al. 2012), we assume the following intraclass correlation coefficients (ICC) to estimate the design effect due to clustering:

---

[7] This is equal to the expected number of completed staff surveys from teachers (579) and home visitors (738) combined.

[8] The design effect is the ratio of the variance of the estimate (properly accounting for the impact of the sample design on the variance) divided by the variance of the estimate one would have obtained from a simple random sample of the same size. For example, a design effect of 1.5 means that the complex sample design inflated the variance of a particular estimate by 50 percent, effectively reducing the sample size by one-third.

- For estimates of classroom quality
    - ICC = 0.20 for between-program variation
    - ICC = 0.20 for between-center, within-program variation
- For estimates of home visitors (or home visitors combined with classroom teachers)
    - ICC = 0.20 for between-program variation
- For estimates of home-based children (or home- and center-based children combined)
    - ICC = 0.05 for between-program variation
    - ICC = 0.05 for between-home visitor (or center), within-program variation
- For estimates of center-based children
    - ICC = 0.05 for between-program variation
    - ICC = 0.05 for between-center, within-program variation
    - ICC = 0.05 for between-classroom, within-center variation

Table B.4. Precision of estimates and minimum detectable correlations

| | Sampled | Responding sample | Effective sample size | 95 percent confidence intervals (half widths) for outcome proportion of 0.50 | Minimum detectable correlations |
|---|---|---|---|---|---|
| Home visitors | 787 | 706 | 263 | .061 | .173 |
| Teachers/classrooms | 618 | 611[9] | 240 | .063 | .181 |
| Teachers and home visitors | 1,405 | 1,317 | 373 | .051 | .145 |
| All children | 3,136 | 2,107 | 937 | .032 | .092 |
| Home-based children | 2,210 | 1,387 | 680 | .038 | .107 |
| Center-based children | 927 | 720 | 392 | .050 | .142 |

In Table B.4, we can see that, under the Baby FACES 2020/2022 sample design, we will be able to make percentage-based estimates of home visitors within plus or minus 6.1 percentage points with 95 percent certainty. For estimates of teachers plus home visitors, we will be able to make percentage-based estimates within plus or minus 5.1 percentage points. For all children, we will be able to make estimates within plus or minus 3.2 percentage points.

---

[9] This number is an average between the expected number of staff interviews (609) and the expected number of classrooms observations (613).

Table B.5. Minimum detectable effect sizes (between subgroups)

| | Subgroup 1 | | Subgroup 2 | | Minimum detectable effect |
|---|---|---|---|---|---|
| | Description | Proportion | Description | Proportion | |
| Home visitors | Speaks only English | .50 | Speaks another language | .50 | .285 |
| Teachers and home visitors | More than 5 years of experience | .70 | 5 or fewer years of experience | .30 | .241 |
| All children | Lower risk | .75 | High risk | .25 | .174 |
| | Not DLL | .60 | DLL | .40 | .160 |
| Home-based children | Lower risk | .75 | High risk | .25 | .208 |

Note:  Effect sizes are in standard deviation-sized units.

DLL = dual language learner.

In Table B.5, we provide examples of the estimated precision of various subgroup comparisons. For example, if response rate goals are met, we will be able to detect underlying differences of .241 standard deviations between teachers and home visitors with more than five years of experience and teachers and home visitors with five or fewer years of experience, with 80 percent power. For all children, we will be able to detect underlying differences of .174 standard deviations between lower risk and higher risk children and underlying differences of .160 standard deviations between dual language learner (DLL) and non-DLL children.

**Weighting.** The purpose of analytic weights is to enable the computation of unbiased population estimates based on sample survey responses. Weights will take into account both the probability of selection into the sample and differential response patterns in the completed data collection. After data collection, we will construct weights at the program, center, home visitor, classroom/teacher, and child levels. We will know the selection probabilities for each stage of sampling from the original sample selection and adjust selection probability to account for any backup sample releases. The inverse of the selection probability is the sampling weight. Nonresponse (nonparticipation) adjustments at each stage will mitigate the risk of nonresponse bias on observable factors using weighting class adjustments. In this technique, we will be essentially using the inverse of the response rate (or response propensity) to inflate the respondents' sampling weights to account for non-responding sample members with similar characteristics. Although this method is used to reduce bias, it will also increase the design effect due to unequal weighting over and above the design effect from the complex sample design itself.

We will use the program weights as components of center- and home visitor-level weights, and the center weights as components of classroom-level weights. Similarly, the classroom and home visitor weights will be used as components of child-level weights.

## B3.   Design of Data Collection Instruments

*Development of Data Collection Instruments*

The Baby FACES data collection instruments are based upon a conceptual framework (presented in Appendix A) that was developed through expert consultation and with ACF involvement (as described in Supplemental Statement A) to ensure the data's relevance to policy and the research field.

**Alternative Supporting Statement for Information Collections Designed for
Research, Public Health Surveillance, and Program Evaluation Purposes**

The data collection protocol includes surveys of program directors, center directors, teachers, home visitors, and parents. It also includes observations of the quality of EHS center-based classrooms, observations of the quality of home visits, observations of parent-child interactions, and parent and staff reports of child development. Information on staff members' perspectives on children's development and about staff relationships with the family allow the data to address research questions about the associations between the staff–family relationship, family engagement with the program, and outcomes.

Wherever possible, surveys use established scales with known validity and reliability. When there were few instruments to measure the constructs of interest at the program or center level, expert consultation supported the identification of potential measures or the development of new items tapping these constructs. This effort fills a gap in the knowledge base about EHS program processes and will answer questions about relationships between program characteristics and other levels of the conceptual framework.

Ahead of the Baby FACES 2018 data collection, we conducted pre-tests[10] with parents, teachers, home visitors, program directors, and center directors using a variety of modes: in person, telephone, and self-administered. For the 2020 data collection, we made modifications to specific questions and measures based on lessons learned in Baby FACES 2018 and conducted another round of pre-tests (with fewer than 10 people). In addition to wording adjustments to specific items, we also included options for "don't know" responses and options to report ranges or instructions for the respondents' best estimate when numeric values are requested. Although item level nonresponse is quite low overall (see section B.5) we hope to even further reduce item level non response by including these response options to relevant items. To reflect the mode that was originally intended to be used in the data collection for each instrument, program and center directors completed electronic versions, teachers and home visitors completed telephone versions, and parents completed hard copy versions of the Parent Child Report. We only pretested new questions from the parent survey by reading them aloud to parents who also completed the Parent Child Report.

The instruments and forms (Instruments 1-10) are annotated to identify sources of questions from prior studies, as well as new questions developed for Baby FACES 2020/2022 (Appendix A). We are including previously approved new items related to COVID-19 to allow us to interpret 2022 data relative to 2018 data. Since the last approval we made non-substantive changes to put instruments online, changed paper invitations to emails with embedded links, and created a paper reminder with QR codes to allow respondents easier access instruments. This was due to the ongoing restrictions resulting from the COVID-19 pandemic.

**B4.    Collection of Data and Quality Control**

All data will be collected by the Contractor. Modes for all instruments are detailed in table B.6.

Table B.6. Data collection activities

| Component | Administration characteristics | Spring 2020 |
|---|---|---|
| Parent survey | | |
| | Mode | Telephone survey (CATI) |

---

[10] All pretests were done with 9 or fewer people.

| Component | Administration characteristics | Spring 2020 |
|---|---|---|
| | Location | Calls initiated from Mathematica's SOC |
| | Time | 32 minutes |
| | Token of appreciation | $20 |
| Parent child report | | |
| | Mode | Paper SAQ distributed one week prior to the scheduled on-site visit week and collected during on-site visit or Web version on similar schedule if no in-person visit is possible |
| | Location | EHS program |
| | Time | 20 minutes |
| | Token of appreciation | $5 |
| Staff child report | | |
| | Mode | Paper/Web SAQ distributed and collected during on-site visit |
| | Location | EHS program |
| | Time | 15 minutes per child |
| | Token of appreciation | $5 per child |
| Staff/home visitor survey | | |
| | Mode | Web SAQ which may be facilitated using a study tablet during onsite visit or via CATI if no in-person visit possible |
| | Location | EHS program |
| | Time | 30 minutes |
| | Token of appreciation | Not applicable |
| Classroom observation | | |
| | Mode | In-person observations (CADE). Each classroom will be observed using Q-CCIIT. |
| | Location | EHS classrooms |
| | Time | 4 hours |
| | Token of appreciation | Two books worth up to $10 each |
| Program director survey | | |
| | Mode | Web with in-person and/or phone follow-up |
| | Location | Web via EHS program, other location |
| | Time | 30 minutes |
| | Honorarium | $250 to be used at discretion of program director and shared with centers |
| Center director survey | | |
| | Mode | Web with in-person and/or phone follow-up |

**Alternative Supporting Statement for Information Collections Designed for
Research, Public Health Surveillance, and Program Evaluation Purposes**

| Component | Administration characteristics | Spring 2020 |
|---|---|---|
| | Location | Web via EHS program, other location |
| | Time | 30 minutes |
| | Honorarium | Not applicable (may come from program) |
| Home visitor observation | | |
| | Mode | In-person or virtual observations (CADE) |
| | Location | Families' homes |
| | Time | 90 minutes |
| | Token of appreciation | Not applicable |
| Observations of parent-child interactions | Mode | In-person observations (CADE) |
| | Location | Families' homes |
| | Time | 10 minutes |
| | Token of appreciation | $35 for parent (for allowing us to come into the home to complete the parent-child interaction and home visit observation) and a book or toy worth $7 for the child. Also give parents $35 for virtual observations that they allow us to join. |

Baby FACES 2020/2022 will deploy monitoring and quality control protocols developed during Baby FACES 2018.

**Recruitment protocol.** Starting in fall 2021, we will send previously sampled and recruited programs from spring 2020 an update about the plans for spring 2022, and for those that already agreed to participate, requesting their cooperation again. It will include an official request from the Office of Head Start, along with letters of support from ACF, and Mathematica, fully informing program directors about the Baby FACES 2020/2022 study, the planned data collection, the assistance we will need to recruit and sample families, and our planned visits to their programs (Appendices E and F). The mailing will include a brochure and fact sheet about the study. Should any sampled program be ineligible or decline to participate, we will release its replacement program and repeat the program recruitment process.

**Telephone interview monitoring.** For the parent telephone interview and staff interviews professional Mathematica Survey Operation Center (SOC) monitors will monitor the telephone interviewers and observe all aspects of the interviewers' administration—from dialing through completion. Each interviewer will have his or her first interview monitored and will receive feedback. For ongoing quality assurance, over the course of data collection we will monitor 10 percent of the telephone interviews.  Monitors will also listen to interviews conducted by interviewers who have had problems during a previous monitoring session.

**Web instrument monitoring.** For the program and center director, and teacher and home visitor web surveys, we will review completed surveys for missing responses and review partial surveys for

follow-up with respondents. We will conduct a preliminary data review after the first 10–20 completions to confirm that the web program is working as expected and to check for inconsistencies in the data. We will build soft checks into the web surveys to alert respondents to potential inconsistencies while they are responding.

**Field data collection monitoring.** Our approach to monitoring field data collection is multi-faceted. First, Baby FACES liaisons will hold regular calls with on-site coordinators and team leaders to discuss each site visit and any challenges that arose. Second, trained SOC staff will review all materials returned from the field for completeness and follow up with field staff, if needed. Third, gold standard observers[11] will accompany each field staff conducting Q-CCIIT classroom observations. Those conducting HOVRS-3 home visit observations will be asked to score home visit videos to conduct a Quality Assurance (QA) observation. We will compare the field staff's classroom and home visit observation scores with the gold standard observers' scores to determine the field staff's inter-rater reliability. We will address issues with field staff whose inter-rater reliability scores are lower than required, including providing feedback on errors, the opportunity for refresher trainings with the study's gold standard observer who conducted the QA observation, and the opportunity to engage in a second independent observation or code additional home visit videos in which the inter-rater reliability scores are checked again. Field staff who are unable to achieve the desired level of inter-rater reliability on either the classroom or home visit observations will not be allowed to continue conducting those observations. In these instances, we will bring in another team member to conduct the observations. We will similarly monitor reliability in coders of the video-recorded parent-child interaction task. Once coders are certified to reliability with gold standards (before they begin coding), we will compare coder and gold standard scores on a random selection of videos each week.

**Response rate monitoring.** We will use reports generated from the sample management system and web instruments to actively monitor response rates for each instrument and EHS program/center. Using these reports, we will provide on-site coordinators and team leaders with progress reports on response rates and work with them to identify challenges and solutions to obtaining expected response rates.

**Addressing lower than expected response rates.** We plan to be flexible with data collection modes when possible or necessary. For example, we can offer to complete the home visit observation virtually if needed due to ongoing COVID restrictions. We can also offer to complete program and center director surveys in person or over the phone as an alternative to the web survey. Likewise, we will offer our tablets to teachers and home visitors to complete their web surveys and will offer to complete it by phone after the visit week as needed. Similarly, we can offer to administer the Parent Child Report on the web, by paper, or in person as options  to make completing it more convenient for parents. While on site, the field team can provide reminders to program staff and parents throughout the data collection week. We are also prepared to offer make-up visits to programs, with different respondents, if necessary.

## B5.     Response Rates and Potential Nonresponse Bias

*Response Rates*

---

[11] Training team members who are reliable with the observation trainers.

**Alternative Supporting Statement for Information Collections Designed for
Research, Public Health Surveillance, and Program Evaluation Purposes**

As with Baby FACES 2018, we will use AAPOR response rate formula 3 (RR3)[12] to calculate response rates. We expect to know the eligibility status for all sampled units, and so will not have to estimate the eligibility rate of nonrespondents with unknown eligibility status. We will assign a final status of complete to partial completes with sufficiently completed instruments, and exclude all other partial completes from the response rate numerator (to coordinate with the rules used for weighting). We will calculate response rates both unweighted and weighted by the sampling weight. And we will incorporate participation in prior stages of sampling in the response rates for subsequent stages. For example, the center-level response rate will incorporate the response rate from the higher program level sample.

Our expected response rates for Baby FACES 2020/2022 are based on actual response rates from Baby FACES 2018. We expect response rates of 98 and 95 percent, respectively, for the program and center director surveys in Baby FACES 2020/2022. Among those children for whom we expect to obtain parental consent, we anticipate we will complete the parent survey for 84 percent, the Parent Child Report for 83 percent, and the Staff Child Report for 89 percent. Table B.3 provides expected response rates and expected number of responses for each study instrument.

*Nonresponse*

The parent survey is our main concern for potential non-response bias affecting the quality or precision of resulting estimates.  For those parent/guardian characteristics we are able to identify from the consent form, we plan to monitor parent response rates throughout the field period to proactively address any emerging non-response: center vs home-based care, relationship to child, and preferred language. We may experience overall lower consent rates if we are not able to have an in-person visit. We will monitor and follow up via phone and email and ask staff to help us gain consent from sampled families.

Following completion of data collection, we will construct weights that adjust for nonresponse. These weights will be used for creating survey estimates and will minimize the risk of nonresponse bias. These weights will build on sampling weights that account for differential selection probabilities as well as nonresponse at each stage of sampling, recruitment, and data collection. When marginal response rates are low (below 80 percent) we plan to conduct a nonresponse bias analysis to compare distributions of characteristics for respondents to those of nonrespondents using any information available for both types of sample members, and then compare the distributions for respondents when weighted using the nonresponse-adjusted weights to see if any observed differences appear to have been mitigated by the weights. For program-level nonresponse, we have a fair amount of data available from the PIR. For centers, home visitors, classrooms, and children, we only have the information is collected on the lists or rosters used for sampling as well as program level characteristics.

The data we plan to collect do not include answers to any especially critical questions that would require follow-up if they were missing. Furthermore, based on our experience with previous rounds of Baby FACES, we expect a minimal item nonresponse rate (5 percent or less) in general. As noted earlier, we created "don't know" and categorical ranges for items that require a count or percentage. These items had higher levels of missingness, potentially because they required looking up records to give an exact response. Although some of the more sensitive questions garnered higher item nonresponse in

---

[12] The American Association for Public Opinion Research. 2016. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition. AAPOR.

2018, the level of missingness on these items was below 2 percent, with the exception of household income, which had an item nonresponse rate of 14 percent).

## B6. Production of Estimates and Projections

All analyses will be run using the final analysis weights, so that the estimates can be generalized to the target population. Documentation for the restricted use analytic files will include instructions, descriptive tables, and coding examples to support the proper use of weights and variance estimation by secondary analysts.

## B7. Data Handling and Analysis

*Data Handling*

Once the electronic instruments are programmed, Mathematica uses a random data generator (RDG) to check questionnaire skip logic, validations, and question properties. The RDG produces a test data set of randomly generated survey responses. The process runs all programmed script code and follows all skip logic included in the questionnaire, simulating real interviews. This process allows any coding errors to be addressed prior to data collection.

During and after data collection, Mathematica staff responsible for each instrument will make edits to the data when necessary. The survey team will develop a document for data editing to identify when survey staff select a variable, noting the current value, the new value, and the reason why the value is being edited. A programmer will read the specifications from these documents and update the data file. All data edits will be documented and saved in a designated file. We anticipate that most data edits will correct interviewer coding errors identified during frequency review (for example, filling in missing data with "M" or clearing out "other specify" verbatim data when the response has been backcoded). This process will continue until all of the data are clean for each instrument.

*Data Analysis*

The instruments included in this OMB package will yield data to be analyzed using quantitative methods. We will carefully link the study's research questions with the data we collect, constructs we measure, and our analyses. Baby FACES 2020/2022 includes three categories of research questions:

1. **Descriptive.** We will address descriptive questions about relationship quality in EHS, classroom features and practices, home visit processes, program processes and functioning that support responsive relationships, and the outcomes of infants and toddlers and families that EHS serves.

2. **Associations with relationship quality.** We will examine associations of relationship quality in EHS with classroom features and practices, home visit processes, and program processes and functioning, along with associations of teacher–child and parent–child relationships with infant/toddler outcomes.

3. **Mediators.** We will study mechanisms for hypothesized associations by examining elements that may mediate associations.

Many research questions will be answered by calculating the means and percentages of classrooms, teachers and home visitors, programs, or children and families grouped into various categories, and comparing these averages across subgroups. We can perform hierarchical linear modeling for more complex analyses of associations between relationship quality and program, classroom, and home visit processes as well as program, teacher, and home visitor characteristics. We

can conduct similar analyses to examine the associations of relationship quality and classroom and home visit processes with children's outcomes. We will conduct mediation analyses to examine the mechanisms for the associations through structural equation modeling.

To properly incorporate the final analytic weights and complex sample design, including unequal weighting and clustering design effects, we will use specialized statistical software or procedures to calculate estimates and their variance. For example, we plan to use SAS Survey procedures that use the Taylor Series Linearization methodology for variance estimation.

*Data Use*

For each round of the study, we release a data user's guide to inform and assist researchers who might be interested in using the data for future analyses. The manual includes (1) background information about the study, including its conceptual framework; (2) information about the Baby FACES sample design on the number of study participants, response rates, and weighting procedures; (3) an overview of the data collection procedures, data collection instruments, and measures; and (4) data preparation and the structure of Baby FACES data files, including data entry, frequency review, data edits, and creation of data files.

Mathematica will produce several other publications based on analysis of data from Baby FACES 2020/2022:

- Descriptive tables of findings from all surveys. The intention is to quickly produce findings that Federal agencies can use.

- A final report including information from the descriptive tables and additional narrative explanation of the findings. This report will be accessible to a broad audience, using graphics and figures to communicate key findings.

- Specific topics briefs of interest to the government. These briefs will be focused and accessible to a broad audience.

- Restricted-use data files and documentation that will be available for secondary analysis.

**B8. Contact Persons**

Mathematica Policy Research and consultants Dr. Margaret Burchinal of the Frank Porter Graham Child Development Center at the University of North Carolina-Chapel Hill, Dr. Jon Korfmacher of the Erikson Institute, and Dr. Virginia Marchman of Stanford University are conducting this project under contract number HHSP233201500035I. Mathematica developed the plans for statistical analyses for this study. To complement the study team's knowledge and experience, we also consulted with a technical working group of outside experts, as described in Section A8 of Supporting Statement Part A.

The following individuals at ACF and Mathematica are leading the study team:

| | |
|---|---|
| Amy Madigan, Ph.D.<br>Project Officer<br>Office of Planning, Research, and Evaluation<br>Amy.Madigan@acf.hhs.gov | Jenessa Malin, Ph.D.<br>Senior Social Science Research Analyst<br>Office of Planning, Research, and Evaluation<br>Jenessa.Malin@acf.hhs.gov |
| Cheri Vogel, Ph.D. | Sally Atkins-Burnett |

**Alternative Supporting Statement for Information Collections Designed for Research, Public Health Surveillance, and Program Evaluation Purposes**

| | |
|---|---|
| Project Director<br>Mathematica Policy Research<br>CVogel@mathematica-mpr.com | Co-Principal Investigator<br>Mathematica Policy Research<br>SAtkins-Burnett@Mathematica-Mpr.com |
| Yange Xue, Ph.D.<br>Co-Principal Investigator<br>Mathematica Policy Research<br>YXue@mathematica-mpr.com | Laura Kalb, B.A.<br>Survey Director<br>Mathematica Policy Research<br>LKalb@mathematica-mpr.com |
| Harshini Shah, Ph.D.<br>Deputy Survey Director<br>Mathematica Policy Research<br>HShah@mathematica-mpr.com | Eileen Bandel, Ph.D.<br>Measurement Task Lead<br>Mathematica Policy Research<br>EBandel@mathematica-mpr.com |
| Barbara Carlson, M.A.<br>Senior Statistician<br>Mathematica Policy Research<br>BCarlson@mathematica-mpr.com | Kimberly Boller, Ph.D.*<br>Mathematica Policy Research |

*Kimberly Boller, who is no longer at Mathematica, served as the Co-Principal Investigator for Baby FACES 2018.

**Alternative Supporting Statement for Information Collections Designed for Research, Public Health Surveillance, and Program Evaluation Purposes**

Appendices

Appendix A.        Conceptual Frameworks and Research Questions

Appendix B.        NIH Certificate of Confidentiality

Appendix C.        Advance Materials

Appendix D.        Brochure

**Instruments**

Instrument 1.    Classroom/home visitor sampling form from Early Head Start staff

Instrument 2.    Child roster form from Early Head Start staff

Instrument 3.    Parent consent form

Instrument 4.    Parent survey

Instrument 5.    Parent Child Report

Instrument 6a.   Staff survey (Teacher survey)

Instrument 6b.   Staff survey (Home Visitor survey)

Instrument 7a.   Staff Child Report (Teacher)

Instrument 7b.   Staff Child Report (Home Visitor)

Instrument 8.    Program director survey

Instrument 9.    Center director survey

Instrument 10.  Parent–child interaction