Head Start Family and Child Experiences Survey (FACES)

OMB Control Number: 0970-0151

Nonresponse Bias Analysis for AI/AN FACES Program Participation

The third Plus study for the 2014 Head Start Family and Child Experiences Survey (FACES) is a descriptive study of children and families who attend Head Start tribal programs in Head Start Region XI. The study is referred to as the American Indian and Alaska Native Head Start Family and Child Experiences Survey (AI/AN FACES) and was conducted in fall and spring of the 2015–2016 program year. It answers questions about demographic characteristics, home environments, levels of school readiness, and needs of children and families served by Region XI.[1]

The study aimed to select and recruit a representative sample of 22 Region XI Head Start programs from across the United States. Because the study had a Head Start program participation rate of less than 80 percent, we conducted an analysis of the potential for nonresponse bias for programs participating in the study. Our goal was to assess whether the programs that participated in the study exhibited a potential for nonresponse bias compared to programs that did not participate in the study. This was the only stage of sampling and data collection for which response rates were not high. Although estimates are not planned at the program level, the program-level participation rate factors into the child-level response rate, because we are missing from the study the children who would have been sampled in those nonparticipating programs. We will extrapolate our findings from the program level and apply them to child-level estimates for children sampled within these participating programs. This memorandum describes:

- Our approach to selecting the sample of programs for AI/AN FACES
- Our approach to conducting the nonresponse bias analysis
- The results of this analysis
- Implications for researchers using AI/AN FACES data

Sampling approach and participation rate. Selection of AI/AN FACES programs was the first of four stages of selection. We selected an augmented sample of 70 programs, stratified by geography[2] and program structure[3] (number of centers and classrooms) and selected with probability proportional to size (PPS), where the size measure used was the number of

[1] The Office of Management and Budget (OMB) submission containing a description of the study design is available at http://www.reginfo.gov/public/do/PRAViewICR?ref_nbr=201511-0970-002.

[2] Based on input from the members of the AI/AN FACES workgroup, we formed five geographic strata, based on state.

[3] To ensure enough available centers, classrooms, and children in the sampled programs, we stratified the programs by whether they had two or more centers and, among those with only one center, whether they had four or more classrooms. Only those with two or more centers were further stratified by the geographic state groups described in footnote 2.

classrooms in the program. The initially released sample was 22 programs; however, 9 additional sampled programs were released later so there would be enough participating programs. Of the 31 programs ultimately released, 20 participated in fall 2015, yielding an unweighted participation rate of 64.5 percent. The weighted participation rate, which accounts for the PPS sampling, is 77.1 percent. This means that smaller sampled programs (with lower chances of selection and therefore higher sampling weights) were more likely to participate than larger sampled programs. One additional program participated in the study in spring 2016 once tribal approval was obtained, bringing the total number of participating programs to 21.

Nonresponse bias analysis approach. As participation or response rates decrease, the risk for nonresponse bias for an estimate increases if nonrespondents would have responded differently from respondents. Nonresponse bias cannot usually be directly measured; however, we can look for indications of the potential for nonresponse bias on key outcomes and examine whether the nonresponse-adjusted weights developed at the program level (a building block for child-level analysis weights) appear to have mitigated the risk for bias. We focused our analysis on the 31 programs sampled and released (all of which were study eligible) for which we had Head Start Program Information Report (PIR) information.

Our analysis involved two steps:

1. Identifying which of these associated variables had significantly different participation profiles
2. Examining whether these differences were diminished after applying the nonresponse-adjusted program weights

We analyzed the following PIR variables for indication of potential nonresponse bias:

- Whether the program's zip code is in a metropolitan statistical area[4]
- Geographic strata[5]
- Size of the program by total enrollment
- Ratio of children's ages (number of children age 3 or younger: number of children age 4 or older)
- Percentage of enrolled children who are American Indian or Alaska Native
- Percentage of children with a disability

[4] http://www.census.gov/population/metro/. "Metropolitan … statistical areas (metro … areas) are geographic entities delineated by the Office of Management and Budget (OMB) for use by Federal statistical agencies in collecting, tabulating, and publishing Federal statistics. … A metro area contains a core urban area of 50,000 or more population…. Each metro … area consists of one or more counties and includes the counties containing the core urban area, as well as any adjacent counties that have a high degree of social and economic integration (as measured by commuting to work) with the urban core."

[5] All programs sampled from geographic state group 2 (see footnote 2) participated, so we could not compare participating to nonparticipating programs for that group for this nonresponse bias analysis. For this analysis only, we combined geographic strata 2 and 4.

We examined whether the distributions of these variables differed between participating and nonparticipating programs.[6] None of the variables were found to be significantly associated with the probability of participating in the study at $\alpha = 0.05$ (using a Rao-Scott Chi-square test in SAS SurveyFreq procedure), although, given the small effective sample size, we likely did not have sufficient power to reject any of the null hypotheses. The analysis accounted for stratification and unequal weights. We used the program-level sampling weight (before and after nonresponse adjustment) for these analyses.

Analysis results. We did not find any statistically significant differences in participation patterns across the six variables examined. In Table 1, however, we present the participation rates and observed weighted distributions for all six variables. Related to the point made above about power, the very small effective sample size for this analysis should be kept in mind when comparing point estimates; that is, any differences seen can be due to sampling error and not reflect differences in the underlying populations. While we are conducting this analysis at the program level, the study is not designed for program-level analysis; still, this analysis can provides some indication of how well the sample maps to the population of Region XI. We show five columns. Column 1 shows the program participation rate for each level of the variable, column 2 shows the variable's distribution among all sampled and eligible programs, column 3 shows the distribution among participating programs only, and column 4 shows the distribution among participating programs after the nonresponse adjustment has been applied. Column 5 shows the relative bias of the estimate after nonresponse adjustment. This is calculated as the bias (absolute difference between columns 2 and 4), relative to column 2, where the best scenario would be a value of 0. This helps put the size of the difference in perspective.

Table 1 shows that differences in participation rates across levels of a variable (column 1) may lead to differences between the variable's distribution for all eligible cases (column 2) and for participating programs only (column 3), thereby indicating a need for nonresponse-adjusted weights. Ideally, these weights would correct for differential participation behavior and close the gap between the variable's distribution for all cases (2) and participating programs only (3). We hope that applying the nonresponse-adjusted weight (4) would result in a distribution that looks more like that of all cases (2).

[6] We have no information beyond the program level (for example, centers, classrooms, children) for nonresponding programs.

Table 1. Program-level variables examined before and after nonresponse adjustment

Variable	Response categories	(1) Weighted program participation rate (percentage)	Unadjusted weighted distributions (percentage)		Nonresponse-adjusted weighted distributions (percentage)	(5) Relative bias after nonresponse weighting adjustments
			(2) All sampled and eligible programs (n = 31)	(3) Participating programs only (n = 20)	(4) Participating programs (n = 20)	
Program is in a metropolitan area	Yes	84.10	27.07	29.54	26.62	0.017
	No	74.46	72.93	70.46	73.38	0.006
Geographic region	State group 1	23.67	9.22	2.83	6.77	0.266
	State group 3	85.92	34.25	38.19	31.80	0.072
	State groups 2 and 4	83.60	45.89	49.78	49.39	0.076
	State group 5	66.67	10.64	9.20	12.04	0.132
Size of program by total enrollment	125 or fewer	80.45	65.42	68.29	64.79	0.010
	More than 125	70.66	34.58	31.71	35.21	0.018
Ratio of children age ≤ 3 to children age ≥ 4	More 3s than 4s	83.02	29.89	32.20	31.71	0.061
	More 4s than 3s	74.53	70.11	67.80	68.29	0.026
Percentage of AI/AN children	83 or less	83.52	54.65	59.23	57.15	0.046
	More than 83	69.29	45.35	40.77	42.85	0.055
Percentage of children with disability	13 or less	79.31	45.62	46.95	54.01	0.184
	More than 13	75.19	54.38	53.05	45.99	0.154

Source: 2013 Head Start PIR.

Note: None of these variables had statistically significantly different participation rates.

Note: The weighted estimates in columns 2 and 3 have not been adjusted for program nonparticipation.

For all variables examined, weighted distributions are, on the whole, similar for participating programs (column 3) as for all sampled cases (column 2). Except for one initial deviation of 6 percentage points (first region listed), differences between columns 2 and 3 are fewer than 5 percentage points. Application of nonresponse weights usually improved the fit of the participating programs to the distribution of the entire sampled group (that is, difference between columns 2 and 4), with two exceptions—the percentage of disabled children in the program, which grew from a difference of 1.3 percentage points to a difference of 8.4 percentage points after weighting, and the percentages in the various geographic categories. Deviations in all other weighted distribution estimates after the application of nonresponse adjustments were reduced to no larger than 3.5 percentage points. As stated previously, these point estimates should be considered along with the fact that the effective sample size is quite small in this analysis, and any differences seen can be due to sampling error alone.

Implications. None of the variables we examined had statistically significantly different distributions between participating programs and nonparticipating programs before nonresponse adjustments were made to the sampling weights. That is, we were unable to reject the null

hypotheses that participating programs did not differ from nonparticipating programs on each of the six program-level variables we examined; however, some estimated percentages did appear to differ between participating and nonparticipating programs. Nonresponse adjustments to the weights mostly improved these distributions, although in one variable they resulted in greater deviations than initially observed.

Because of the small sample size used for this nonresponse bias analysis, researchers should be cautious in interpreting its findings. For program size, urbanicity, and the percentage of children who are AI/AN, these three variables, we saw small differences before nonresponse adjustments and even smaller differences after those adjustments. In turn, this likely means that the program-level nonparticipation will have minimal impact on ***child-level*** estimates that will result from these participating programs, because child-level weights are built upon the final adjusted program weights. Furthermore, the study was designed to produce child-level, not program-level estimates, with a primary focus on point estimates, rather than comparisons between child subgroups. Our child sample size exceeded those laid out in the study design. Therefore, we believe researchers should feel comfortable using the AI/AN child-level data, along with the appropriate weights.