

Analysis of Nonresponse Bias in the Early Head Start Research and Evaluation Project

Barbara Lepidus Carlson

Mathematica Policy Research, P.O. Box 2393, Princeton, NJ 08543-2393

Abstract

This paper describes the nonresponse patterns among the participants in the Early Head Start Research and Evaluation Project, a Congressionally mandated random assignment study that began in 1995. Across the 17 study sites, 3,001 infants and their families were randomized at the time of enrollment to be in the program or control group. The program group children were allowed to participate in the Early Head Start program in their site; the control group children were not. Information was collected by program staff at baseline, and the children were followed up by research staff when children were 14, 24, and 36 months of age, and in pre-kindergarten. To better understand response patterns across waves, the paper looks at how the retained sample differs from the baseline sample, and the patterning of loss; that is, how attrition differed (if at all) between the program and control groups.

Key Words: Nonresponse, Bias, Attrition, Early Head Start

1. Background

It is a well-known fact that survey response rates have been declining across the board, and longitudinal surveys of highly mobile populations have the additional problem of tracking respondents over time, which leads to increased nonresponse. However, a low response rate does not necessarily mean that nonresponse bias is present, and a high response rate does not guarantee that you are free from nonresponse bias. Accordingly, the focus of survey research has been moving towards evaluating nonresponse bias whenever possible. This paper looks at nonresponse bias in the Early Head Start Research and Evaluation Study.

Early Head Start is a federal program that serves low-income pregnant women, children from birth to age 3, and their families. Early Head Start offers services at home, in centers, or both. In 1996, when Early Head Start was beginning, a randomized study was begun in 17 of the original sites. These sites were not randomly selected, but were chosen to represent a diverse set of programs in terms of geographical region, urban versus rural setting, and the language, race, and ethnicity of the families served. These first programs were rather small, and would have undoubtedly had to turn away a number of interested families, so participating in a randomized study where interested families would be randomly assigned to either the program group (to participate in Early Head Start) or the control group for purposes of this study posed fewer ethical issues for these programs.

In this study, 3,001 families were enrolled and randomized, either during pregnancy or up until the age of 1. Among these baseline families, 1,513 were randomly assigned to the program group and 1,488 to the control group. Children were followed up at about 14 months, 24 months, 36 months, pre-kindergarten (age 5), and grade 5 (age 11)--waves 1 through 5, respectively. A fair amount of data were collected on all 3,001 families at baseline: information about the child, the child's mother, the child's family, and the program at which the child was randomized. This provides a wealth of information for

comparing respondents to nonrespondents over time, a luxury not afforded many surveys. In fact, having so many variables poses its own problem in terms of deciding what to look at. We have the ability to look at response patterns over time, by site, and by randomization group for dozens of baseline variables. We can compare respondents to the full sample, or compare respondents to nonrespondents.

In this paper, we decided to focus on differential patterns that may indicate nonresponse bias. By definition, we cannot measure bias in the key outcome measures directly because we have no data for the nonrespondents. But we can examine baseline variables available for everyone that we believe could be correlated with outcome measures such as a child's cognitive, socioemotional, and physical development. We compared the distribution of baseline characteristics for the full sample to the distribution of these characteristics among those responding to the grade 5 followup. We analyzed the data for the full sample (combining program and control groups) and separately by randomization group.

2. Response Patterns

2.1 Response at Grade 5 Follow Up

For purposes of this paper, we considered a case a respondent in grade 5 if we completed either a child assessment (CA) or a parent interview (PI).¹ Table 1 shows the response distribution at the grade 5 follow up.

Table 1: Response at Grade 5 Follow Up

	Final Status	Study Sample	Percentage
Not worked	No response since baseline	300	10.0
	Moved out of country	89	3.0
	Harsh refusal from prior wave	47	1.6
Worked but not completed	Unlocatable	459	15.3
	Located but no data collected	469	15.6
	Completed only PCI or HO	4	0.1
Completed	Completed CA or PI	1,633	54.4
Total baseline sample		3,001	100.0

We decided before fielding the grade 5 follow up that we would not spend valuable resources going after the 300 families for whom no data had been collected since baseline, although they were still part of the target population. In addition, we did not attempt to complete data collection on 89 families known to have moved out of the country, nor 47 whom the site thought were adamant refusers from the prior round of data collection (in the pre-kindergarten year). Among the remaining 2,565 families, we were able to locate 2,106 and obtain data on 1,633. These completed cases represent 54.4 percent of the full sample, 63.7 percent of the attempted cases, and 77.5 percent of the located cases. While we have grade 5 data on only the 1,633, we have baseline

¹ We also collected data from a parent-child interaction (PCI) and a home observation (HO). There were four cases in which we obtained both of these but not a child assessment or a parent interview, and these cases were treated as nonrespondents.

information on all 3,001 families, and so can use these characteristics to compare the grade 5 respondents to the full sample.

2.2 Response Patterns Over Time

Table 2 shows response patterns over time, using the same definition of response in each wave; that is, a completed child assessment or parent interview. We see that almost two-thirds of the sample participated in four or more of the five waves of data collection.

Table 2: Participation in Study Across Waves

Total Number of Waves Participated	Study Sample	Percentage
0	300	10.0
1	194	6.5
2	258	8.6
3	392	13.1
4	762	25.4
5	1,095	36.5
Total Sample	3,001	100.0

It should be noted that, of those missing one or more waves of data collection, 55 percent are attrition cases (that is, once they missed a wave they did not return in a subsequent wave) and 45 percent returned for at least one wave after missing one.

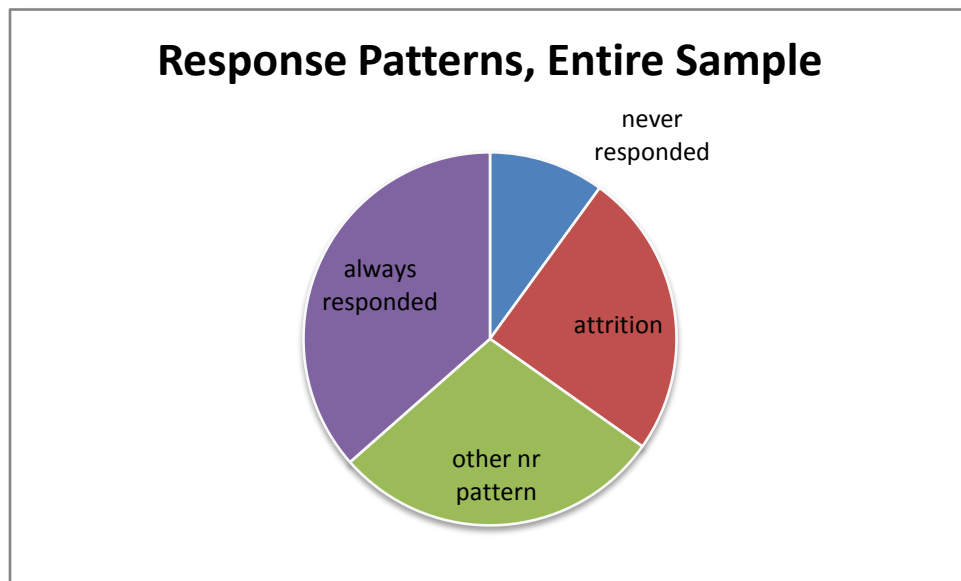


Figure 1: Distribution of Response Patterns

Figure 1 shows the distribution of response patterns for the entire sample. The distribution by type of nonresponse *pattern* (no response since baseline, attrition, or out-and-back-in) was not significantly different for the program and control groups. However, we did see statistically significant differences in these distributions by a number of characteristics we looked at, including site, sociodemographic variables (gender, birth order, race-ethnicity, primary language, mother's education, living

arrangement, maternal risk level, low birthweight, work requirements) and study variables (randomization date, child’s age at randomization).

For example, among those whose primary language is English and who did not respond in all five waves: 14 percent had no response since baseline; 38 percent attrited; and 47 percent returned after missing a wave. Among those nonrespondents whose primary language is *not* English: 18 percent had no response since baseline; 43 percent attrited; and 38 percent returned after missing a wave. (Chi-square p-value = .0038.) Among those nonrespondents in a state or county requiring AFDC recipients who are mothers of infants to work: 13 percent had no response since baseline; 45 percent attrited; and 42 percent returned after missing a wave. Among those nonrespondents in areas *without* that requirement: 17 percent had no response since baseline; 35 percent attrited; and 47 percent returned after missing a wave. (Chi-square p-value < .0001.)

While Table 2 shows the number of waves that sampled families completed, Table 3 and Figure 2 show the trend over time—how nonresponse rates changed from wave to wave, by randomization group and combined. As expected, nonresponse increased over time. We also see that the control group has a consistently higher nonresponse rate than the program group, by a few percentage points.

Table 3: Percent Nonresponse Over Time

Wave	Approx. Child Age	Control Group	Program Group	Combined
1	14 months	22.2	19.5	20.8
2	24 months	28.6	24.9	26.7
3	36 months	32.1	26.2	29.1
4	5 years	33.0	28.8	30.9
5	11 years	46.8	44.3	45.6

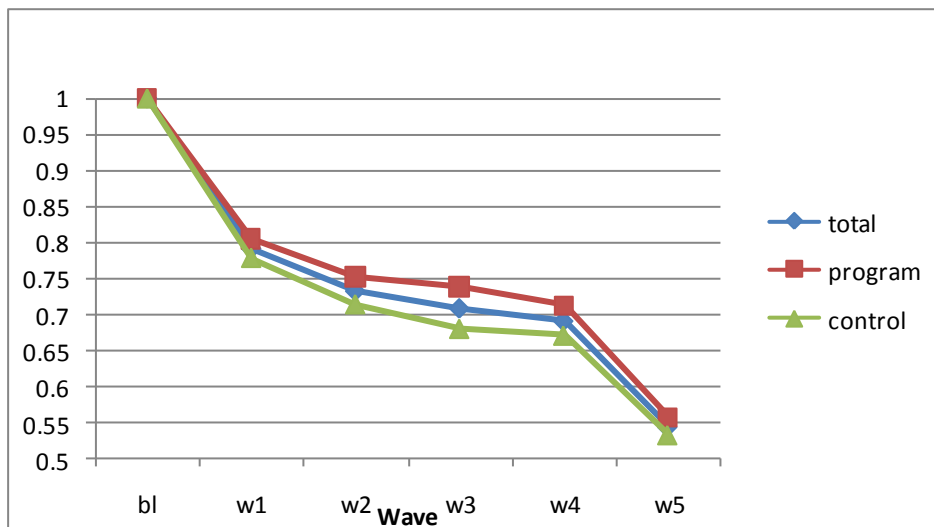


Figure 2: Percent Responding Over Time

“Nonresponse” includes those cases that were not attempted, not located, refused, or not completed for any other reason. The next two tables and Figure 3 show the nonresponse

rates in waves 1, 3, and 5 by site (Table 4a), and by site and randomization group (Table 4b). They also show the percent of families that were nonrespondents in any of the five waves and the percent that were nonrespondents in all five waves.

Table 4a: Percent Coded as Nonresponse by Wave and by Site*

Site	Wave 1	Wave 3	Wave 5	Nonresponse Any Wave	Nonresponse All 5 Waves
A	7.9	18.5	32.5	48.3	6.6
B	31.8	35.8	63.1	78.8	12.8
C	14.2	22.8	52.5	64.2	4.3
D	15.6	22.4	44.9	59.2	6.8
E	12.0	27.3	51.3	60.0	6.7
F	27.6	33.8	55.6	72.0	12.4
G	34.9	45.3	54.3	84.4	18.4
H	17.3	23.0	41.8	55.6	7.7
I	31.1	48.7	42.5	83.9	16.1
J	21.3	37.6	57.4	71.6	14.2
K	24.5	28.0	46.0	67.5	9.0
L	15.8	28.9	40.1	65.1	3.9
M	17.3	22.8	33.7	49.5	7.4
N	14.4	27.4	34.9	60.3	6.8
O	25.3	24.2	42.9	55.5	12.1
P	21.1	23.2	39.0	50.5	15.3
Q	9.8	19.1	39.9	45.7	4.0
Total	20.8	29.1	45.6	63.5	10.0

*Denominator is total baseline sample for that site.

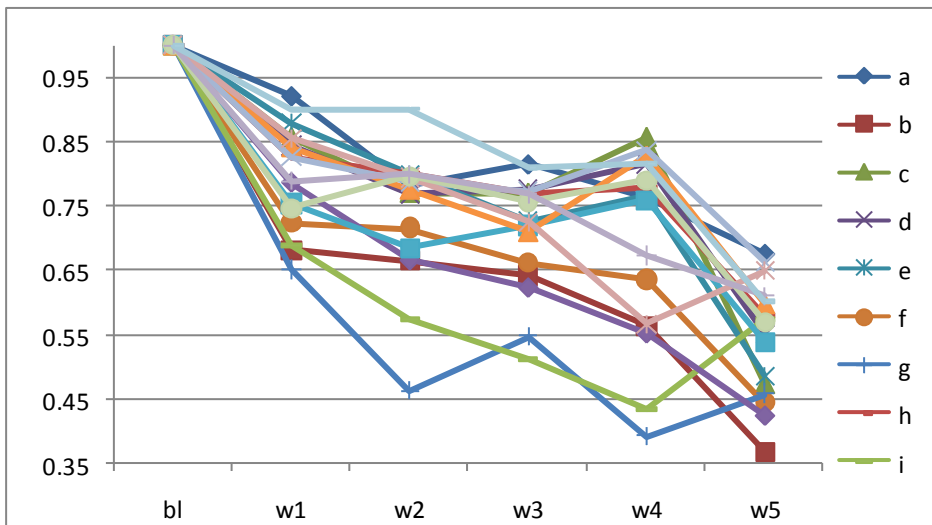


Figure 3: Percent Responding Over Time By Site

Table 4b: Percent Coded as Nonresponse by Wave, Site, and Randomization Group*

Site	Wave 1		Wave 3		Wave 5		Nonresponse Any Wave		Nonresponse All 5 Waves	
	Ctrl	Pgm	Ctrl	Pgm	Ctrl	Pgm	Ctrl	Pgm	Ctrl	Pgm
A	11.7	4.1	23.4	13.5	36.4	28.4	53.3	43.2	10.4	2.7
B	38.4	25.8	43.0	29.0	67.4	59.1	83.7	74.2	14.0	11.8
C	11.5	16.7	23.1	22.6	46.2	58.3	56.4	71.4	2.6	6.0
D	20.8	10.7	33.3	12.0	50.0	40.0	70.8	48.0	11.1	2.7
E	10.5	13.5	35.5	18.9	51.3	51.4	61.8	58.1	5.3	8.1
F	24.5	30.4	32.7	34.8	56.4	54.8	72.7	71.3	10.0	14.8
G	40.7	28.8	46.3	44.2	53.7	54.8	88.0	80.8	20.4	16.3
H	19.4	15.3	27.6	18.4	40.8	42.9	55.1	56.1	9.2	6.1
I	33.7	28.6	50.5	46.9	41.1	43.9	84.2	83.7	17.9	14.3
J	22.9	19.7	35.7	39.4	62.9	52.1	74.3	69.0	17.1	11.3
K	22.9	26.0	27.1	28.8	55.2	37.5	70.8	64.4	7.3	10.6
L	21.5	9.6	32.9	24.7	44.3	35.6	72.2	57.5	6.3	1.4
M	14.3	20.2	27.6	18.3	36.7	30.8	54.1	45.2	7.1	7.7
N	15.5	13.3	21.1	33.3	33.8	36.0	53.5	66.7	8.5	5.3
O	23.9	26.7	23.9	24.4	40.2	45.6	53.3	57.8	10.9	13.3
P	23.2	18.9	24.2	22.1	40.0	37.9	49.5	51.6	17.9	12.6
Q	11.5	8.1	32.2	5.8	39.1	40.7	49.4	41.9	3.4	4.7
Total	22.2	19.5	31.1	26.2	46.8	44.3	65.3	61.8	10.8	9.3

*Denominator is total baseline sample for that site and randomization group.

From these tables, we see that there is quite a bit of variation in nonresponse rates from site to site, as well as in the difference between the program and control group rates. The control group generally has higher nonresponse rates than the program group. This strongly suggests that nonresponse weighting adjustments control for both site and randomization group. Tables 5a and 5b focus on the nonresponse that was due to unlocatability, which is a significant problem particularly for longitudinal studies such as these. Because we did not attempt to contact families with no completes in prior follow-up waves, there are no cases that were unlocatable in all five waves, by definition.

Table 5a: Percent Coded as Unlocatable by Wave and by Site*

Site	Wave 1	Wave 3	Wave 5	Unlocatable Any Wave
A	2.0	2.6	16.6	24.5
B	19.6	19.0	12.9	48.6
C	6.8	8.6	25.9	36.4
D	5.4	4.8	25.2	38.1
E	5.3	12.0	22.7	35.3
F	5.3	7.1	21.3	39.1
G	6.1	14.6	9.4	40.1
H	6.1	12.8	19.4	38.3
I	11.9	36.8	7.3	48.7
J	12.8	18.4	14.9	41.1
K	7.0	10.0	13.0	35.0
L	3.9	12.5	18.4	29.6
M	5.0	5.0	5.9	19.3
N	5.5	7.5	10.3	23.3
O	7.1	11.0	18.7	37.4
P	8.4	7.4	6.3	25.8
Q	4.6	2.9	17.3	24.3
Total	7.3	11.5	15.3	34.6

*Denominator is total baseline sample for that site. By design, no families were unlocatable in all five waves because we did not attempt wave 5 for the 300 families without any completed follow-up waves.

Table 5b: Percent Coded as Unlocatable by Wave, Site, and Randomization Group*

Site	Wave 1		Wave 3		Wave 5		Unlocatable Any Wave	
	Ctrl	Pgm	Ctrl	Pgm	Ctrl	Pgm	Ctrl	Pgm
A	3.9	0.0	3.9	1.4	14.3	18.9	26.0	23.0
B	26.7	12.9	24.4	14.0	14.0	11.8	57.0	40.9
C	5.1	8.3	7.7	9.5	28.2	23.8	37.2	35.7
D	6.9	4.0	6.9	2.7	29.2	21.3	48.6	28.0
E	3.9	6.8	17.1	6.8	19.7	25.7	35.5	35.1
F	4.6	6.1	10.0	4.4	25.5	17.4	41.8	36.5
G	9.3	2.9	15.7	13.5	5.6	13.5	39.8	40.4
H	7.1	5.1	15.3	10.2	15.3	23.5	39.8	36.7
I	11.9	12.2	38.9	34.7	8.4	6.1	51.6	45.9
J	14.3	11.3	14.3	22.5	8.6	21.1	35.7	46.5
K	8.3	5.8	13.5	6.7	12.5	13.5	37.5	32.7
L	5.1	2.7	10.1	15.1	15.2	21.9	25.3	34.3
M	6.1	3.9	8.2	1.9	8.2	3.9	22.5	16.5
N	7.0	4.0	4.2	10.7	4.2	16.0	15.5	30.7
O	5.4	8.9	7.6	14.4	21.7	15.6	37.0	37.8
P	9.5	7.4	10.5	4.2	7.4	5.3	28.4	23.2
Q	4.6	4.7	4.6	1.2	14.9	19.8	21.8	26.7
Total	8.2	6.3	12.8	10.2	14.7	15.9	35.7	33.6

*Denominator is total baseline sample for that site and randomization group. By design, no families were unlocatable in all five waves because we did not attempt wave 5 for the 300 families without any completed follow-up waves.

Tables 6a and 6b look at nonresponse rates by three important baseline characteristics potentially related to child development or other outcomes of interest: (1) maternal risk level, (2) race/ethnicity, and (3) program service type (center-based, home-based, or mixed approach). The maternal risk level looks at five risk factors for the mother of the randomized child at baseline: teen mother, single mother, no high school diploma or GED, recipient of AFDC, and unemployed. A mother who had 0 to 2 of these factors was considered to be “low risk.” A mother who had 3 factors was considered to be “medium risk.” And a mother who had 4 or 5 of these factors was considered to be “high risk.”

Table 6a: Percent Coded as Nonresponse by Wave and Baseline Characteristic*

Characteristic	Wave 1	Wave 3	Wave 5	Nonresponse Any Wave	Nonresponse All 5 Waves
Low Risk	17.2	24.7	40.1	57.4	7.6
Medium Risk	20.4	30.5	46.7	66.4	9.0
High Risk	26.6	34.1	52.8	70.7	13.3
White Non-Hispanic	19.5	24.3	43.2	57.9	9.1
Black Non-Hispanic	20.4	32.1	44.5	67.5	8.5
Hispanic	23.2	31.0	49.1	65.9	11.7
Other Race/Missing	21.7	34.0	52.2	65.5	16.8
Center-Based	13.4	24.0	43.3	55.6	7.7
Home-Based	24.3	30.7	45.9	65.3	11.5
Mixed Approach	20.5	30.1	46.5	65.8	9.4

*Denominator is total baseline sample for that baseline characteristic. Risk level percents exclude those with missing values.

Table 6b: Percent Coded as Nonresponse by Wave, Baseline Characteristic, and Randomization Group*

Characteristic	Wave 1		Wave 3		Wave 5		Nonresponse Any Wave		Nonresponse All 5 Waves	
	Ctrl	Pgm	Ctrl	Pgm	Ctrl	Pgm	Ctrl	Pgm	Ctrl	Pgm
Low Risk	19.2	15.2	28.8	20.5	41.9	38.2	59.3	55.4	8.8	6.3
Medium Risk	20.6	20.2	32.8	28.2	47.5	45.9	68.4	64.5	8.4	9.6
High Risk	26.8	26.3	35.2	33.2	52.8	52.9	71.0	70.5	14.5	12.1
White Non-Hispanic	19.4	19.6	26.5	22.2	45.5	40.9	59.7	56.2	9.1	9.0
Black Non-Hispanic	23.2	17.7	34.3	29.9	45.0	44.0	69.5	65.4	9.1	7.9
Hispanic	25.4	21.1	36.7	25.6	50.9	47.3	67.8	64.2	13.9	9.6
Other Race/Missing	21.1	22.3	34.9	33.0	49.5	55.3	65.1	66.0	16.5	17.0
Center-Based	16.3	10.5	30.1	17.0	46.4	40.2	61.1	50.0	10.1	5.2
Home-Based	24.9	23.8	33.0	28.4	48.1	43.9	67.0	63.8	11.8	11.2
Mixed Approach	22.0	19.0	31.3	28.8	45.4	47.6	65.5	66.2	9.7	9.0

*Denominator is total baseline sample for that baseline characteristic and randomization group. Risk level percents exclude those with missing values.

We see from these tables that higher risk mothers are less likely to respond than lower risk mothers, and this pattern persists across waves. We also see that racial-ethnic minorities are less likely to respond than non-Hispanic whites, and that Hispanics are less

likely to respond than blacks (except for wave 2). When we look at the type of service provided by the program at which families were randomized (and at which the program group received services), the pattern changes across waves. In wave 1, families from the center-based programs had the lowest level of nonresponse and those from the home-based programs had the highest. In wave 3, the home-based and mixed-approach programs had very similar nonresponse rates. And by wave 5, there was not much difference across the three types of programs. We also see differences in the disparity of nonresponse levels between program and control groups by baseline characteristic. For example, we see little difference in nonresponse levels between the two randomization groups among high-risk mothers, but fairly large differences among the low-risk mothers.

3. Characteristics of Respondents

The preceding tables show that nonresponse varies by site, randomization group, a few important baseline characteristics, and wave, which are not unexpected findings. More pertinent to the examination of nonresponse bias is a comparison of the characteristics of respondents compared with those of the full sample. In other words, while we know nonresponse has increased over time, has the nature of the sample changed over time as a result? We looked at 21 baseline characteristics of the randomized child and his/her family and 13 characteristics of the program at which the child was randomized and the area in which it is situated.

Child/Family Variables (21)	Program/Area Variables (13)
<ul style="list-style-type: none"> ➤ Gender of focus child ➤ Randomization date (before 10/06, 10/96 – 6/97, after 6/97) ➤ Age group of focus child at random assignment ➤ Whether focus child was firstborn ➤ Whether primary caregiver pregnant or gave birth to first or later child ➤ Race/ethnicity of primary caregiver ➤ Whether primary caregiver's primary language is English ➤ Whether the mother is possibly depressed based on CESD short form (≥ 10) ➤ Whether the mother is at risk for depression based on CESD long form (≥ 16) ➤ Whether the mother is probably depressed based on CESD long form (≥ 23) ➤ Whether age of mother at birth of focus child less than 20 	<ul style="list-style-type: none"> ➤ Program type (center-based, home-based, mixed) ➤ Whether program was fully implemented ➤ Percent of children in formal care (grouped) ➤ Program part of Head Start Agency ➤ Percent of children in Head Start (grouped) ➤ Most common placement for 3-year-olds (Head Start and child care, centers, Head Start) ➤ Most common placement for 4-year-olds (Head Start and child care, centers, Head Start) ➤ Capacity to care for 3-year-olds

Child/Family Variables (21)	Program/Area Variables (13)
<ul style="list-style-type: none"> ➤ Whether family received AFDC/TANF ➤ Percent of poverty level (<33, 33-66, 67-99, 100+) ➤ Primary caregiver's highest grade completed (<12, diploma/GED, >12) ➤ Lives with husband, other adults, no other adults ➤ Employed, in school/training, neither ➤ Number of risk factors out of 5 (teen mother, single parent, low education, welfare, not working) ➤ Trimester began receiving prenatal care ➤ Birth weight less than 2500 grams ➤ Born more than three weeks early ➤ Newborn stayed in hospital due to medical problem 	<ul style="list-style-type: none"> (low/medium/high) ➤ Capacity to care for 4-year-olds (low/medium/high) ➤ ECE availability for 3-year-olds ➤ Program provided transition services (low/medium/high) ➤ Whether or not program is in urban setting ➤ Whether state or county requires welfare mothers of infants to work

Figure 4: Baseline Characteristics Examined

For each characteristic, we compared the distribution at baseline to the distribution among respondents at wave 3 and wave 5. We also compared the distribution of each characteristic between the program and control group at each wave. For each comparison, we did a chi-square test, combining the baseline cases and the follow-up respondents (a subset of the former) to check the expected against the observed distributions of the characteristic for each timepoint. (We did this including and excluding a missing value category for the child/family-level variables. Results shown here exclude the missing value category.) For example, for mother's educational attainment (shown below), the chi-square p-value (with 2 d.f.) was 0.0292:

Table 7: Comparison of Mother's Education between Baseline and Grade 5 Respondents

Variable	Value	Baseline Sample		Wave 5 (Grade 5) Respondents	
		Number	Percent	Number	Percent
Highest grade completed	<12	1,375	47.8	690	43.9
	12 or GED	822	28.6	464	29.5
	>12	682	23.7	418	26.6
	Missing	122		61	
	Total	3,001	100.0	1,633	100.0

While this example shows a statistically significant difference in distribution between the baseline and wave 5 respondents, this is not typical of our findings. Rather, only 2

child/family characteristics and 7 program/area characteristics had distributions that differed enough to be statistically significant at $\alpha=.05$. This does not account for multiple comparisons. We would have expected to see one or two significant results out of 34 chi-square tests just by chance if we use .05 as the Type I error rate. If we had done a Bonferroni correction, the p-value would have had to be less than .0015 (.05/34) to be significant, and none met that criterion. However, the significant results without such a correction are certainly suggestive of a trend.

Table 8: Comparison of Baseline, Wave 3, and Wave 5 Respondents (Column Percents)

Baseline Variable	Baseline Value	Full Baseline Sample	Wave 3 (36 mos) Respondents	Wave 5 (grade 5) Respondents
		n = 3,001	n = 2,127	n = 1,633
<i>Child/Family Variables</i>				
Focus Child Gender	Female	49.0	49.5	48.8
	Male	51.0	50.5	51.2
Randomization Date	Before 10/96	36.3	35.6	35.7
	10/96 – 6/97	30.5	29.9	30.7
	>6/97	33.2	34.6	33.6
Age of Focus Child at Random Assmt	Unborn	25.4	24.9	26.3
	0-4 months	35.4	35.1	34.5
	5+ months	39.2	40.1	39.3
Focus Child Birth Order	Not first	37.4	38.8	36.7
	First	62.6	61.2	63.3
Primary Caregiver's Pregnancy	Pregnant with 1st	16.7	16.1	17.4
	First child born	45.9	45.1	45.9
	Pregnant with later	8.5	8.9	9.1
	Later child born	29.0	29.9	27.6
Race-Ethnicity	White non-Hisp.	37.2	39.7	38.7
	Black non-Hisp.	34.6	33.1	35.1
	Hispanic	23.6	22.9	22.0
	Other	4.6	4.3	4.2
Primary Language	Not English	21.4	20.4	19.2
	English	78.6	79.6	80.8
Mother Possibly Depressed (Sh Form)	No	42.3	41.3	42.0
	Yes	57.7	58.7	58.0
Mother At Risk for Depressn (Lg Form)	No	48.0	47.7	48.2
	Yes	52.0	52.3	51.8
Mother Probably Depressed (Lg Form)	No/Mild	72.2	73.4	73.9
	Yes	27.8	26.6	26.1
Mother <20 When Focus Child Born	No	61.7	62.7	62.3
	Yes	38.3	37.3	37.7
Got AFDC	No	64.9	67.0	67.9
	Yes	35.1	33.0	32.1
Percent of Poverty Level	<33	30.1	28.7	29.5
	33-66	30.9	30.8	29.7
	67-99	25.3	26.3	25.7
	100+	13.8	14.2	15.1

Baseline Variable	Baseline Value	Full Baseline Sample	Wave 3 (36 mos) Respondents	Wave 5 (grade 5) Respondents
		n = 3,001	n = 2,127	n = 1,633
Highest Grade Comp	<12	47.8	45.6	*43.9
	12 or GED	28.6	29.1	29.5
	>12	23.7	25.3	26.6
Living Arrangement	Husband	25.2	26.5	25.8
	Other adults	38.7	38.7	39.7
	Alone	36.1	34.8	34.5
Primary Occupation	Employed	23.4	24.7	26.3
	School/Training	21.7	21.3	22.0
	Other	54.9	54.0	51.7
Maternal Risk Level (teen, single, no dipl, AFDC, unemployed)	0-2 risks	42.9	45.5	*47.2
	3 risks	30.8	30.1	30.1
	4-5 risks	26.3	24.4	22.8
Trimester Began Prenatal Care	1st	80.7	81.8	81.9
	2nd	16.3	15.5	15.4
	3rd	3.0	2.7	2.7
Low Birthweight (<2500 g)	No	91.0	92.0	91.7
	Yes	9.0	8.0	8.3
Baby Born More Than 3 Weeks Early	No	86.4	87.2	86.7
	Yes	13.6	12.8	13.3
Newborn Stayed in Hospital – Med Prob	No	83.0	83.2	83.5
	Yes	17.0	16.8	16.5
Program/Area Variables				
Program Type	Center	20.4	21.9	21.2
	Home	46.2	45.1	45.9
	Mixed	33.5	33.0	32.9
Overall program fully implemented		69.7	*72.3	72.3
Percent of EHS Children in Formal Care	< 33%	12.7	12.7	12.2
	33 – 84%	22.8	23.5	23.0
	85% +	64.5	63.8	64.7
Program Part of HS Agency		59.3	60.3	62.0
Percent of EHS Children in HS	< 33%	43.8	41.5	42.9
	33 – 84%	22.6	23.0	22.4
	85% +	33.6	35.5	34.7
Most Common Placement for 3 yr olds	HS and child care	21.9	*23.1	22.8
	Centers	31.0	27.3	29.2
	HS	47.2	49.6	47.9
Most Common Placement for 4 yr olds	HS and child care	33.5	*35.3	35.4
	Centers	31.0	27.3	29.2
	HS	35.6	37.4	35.4
Program Capacity to Care for 3 yr olds	Low	37.4	35.9	35.1
	Medium	47.3	48.4	49.1
	High	15.3	15.7	15.9
Program Capacity to Care for 4 yr olds	Low	21.2	19.2	*18.7
	Medium	40.8	43.8	44.9
	High	38.0	37.0	36.4

Baseline Variable	Baseline Value	Full Baseline Sample	Wave 3 (36 mos) Respondents	Wave 5 (grade 5) Respondents
		n = 3,001	n = 2,127	n = 1,633
Program ECE Availability for 3 yr olds	Difficult	11.0	10.5	*8.5
	Somewhat difficult	72.0	71.7	74.5
	Not difficult	17.0	17.8	17.0
Program Provided Transition Services	Low	12.6	12.2	10.7
	Medium	33.7	32.8	34.4
	High	53.7	55.0	55.0
Program Located in Urban Area		58.7	56.4	*54.6
St/Co Reqs AFDC Moms of Infants Work		41.9	*45.0	43.6

The percents that are in **boldface type with an asterisk** indicate significantly different² distributions among that wave's respondents than those among the full baseline population. (Percents that are *italicized* indicate a significantly different distribution only when the missing value category—not shown—was included.)

We can see that, among the child/family-level baseline characteristics, the two that showed significantly different distributions for the wave 5 respondents were mother's educational attainment and maternal risk level. Wave 5 responding mothers were less likely than the full sample to have been high school dropouts and less likely to have been considered high risk at baseline. There were a fair number of program characteristics that had significantly different distributions, but these are difficult to interpret. We do see that wave 5 respondents were less likely to have been associated with urban programs, compared to the full baseline sample.

When we compared the program and control group distributions at baseline, wave 3, and wave 5 (not shown), only one pair of distributions was significantly different (again, not accounting for multiple comparisons), and that was at baseline, for the premature birth variable. No program-control differences were significant among wave 3 respondents or among wave 5 respondents for any of the baseline variables examined.

4. Conclusion

When examining the response patterns and characteristics of respondents over five waves of follow up, overall and by randomization group, we found few things to suggest the possibility of nonresponse bias by the time of the fifth grade follow up, 11 years after randomization. Differences between grade 5 respondents and the full baseline sample would cause concern for generalizing the grade 5 findings to the original randomized population (which was not itself a random sample of sites or children). Different response patterns by randomization group would cause concern for impact analyses that compare the paths of the program and control group children over time. Fortunately, we found similar patterns of nonresponse for the program and control groups over time, with relatively constant differences in nonresponse rates between the two groups at each wave.

² Significant ($p < .05$) without a Bonferroni correction for multiple comparisons.

While a major cause of nonresponse for a longitudinal study such as this would be locatability—that is, tracking sample members over time—we found patterns for overall nonresponse similar to those for unlocatability. The control group consistently had nonresponse rates a few percentage points higher than the program group, although for some sites, for some waves, and for some population subgroups, this program-control pattern was reversed. It is important to note that, at each wave of data collection, the characteristics of the program and control group members were statistically similar, with only one exception (rate of premature birth for the full baseline sample).

There were differences in response patterns by site, and by randomization group by site. To the extent that site influenced key outcome variables, then the differential response by site—and by randomization group within site—can be problematic if not appropriately accounted for in the nonresponse-adjusted weights. We did in fact control for site and randomization group when constructing the analysis weights, with the intention of reducing bias for any analyses that combine across sites, as well as those that estimate impacts.

As one would expect for a high-risk and highly mobile population like the one studied here, there was a fair amount of attrition over 11 years of follow-up. Nonetheless, characteristics of respondents at grade 5 are remarkably similar to those of the full baseline sample. We looked at a number of baseline characteristics, comparing the distribution for the full sample to that of the grade 5 respondents, and only a handful appeared to be different. We were more likely to lose lower-educated and higher-risk mothers by wave 5, as well as those in urban programs. This pattern (retaining the more advantaged sample over time) is similar to the pattern one sees in many other longitudinal studies. But the vast majority of characteristics we looked at showed no differential distributions between the full sample and the grade 5 respondents and that, combined with the lack of differences in program-control characteristics at grade 5, is reassuring in terms of the likelihood of nonresponse bias—particularly if the analysis weights are used. A decision will be made by the Administration for Children and Families as to whether to do another followup at grade 9, and this nonresponse bias analysis is one factor in that decision.

Acknowledgements

The author would like to acknowledge the input and suggestions made by Rachel Chazan Cohen of the Administration for Children and Families, Richard Faldowski from the University of North Carolina at Greensboro, Tom Cook of Northwestern University, and the other members of the Early Head Start Nonresponse Bias Work Group convened in March of 2008, as well as Jerry West, John Hall, and others from Mathematica Policy Research.