SSM - Population Health 18 (2022) 101108

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: www.elsevier.com/locate/ssmph

Distributional effects on children's cognitive and social-emotional outcomes in the Head Start Impact Study: A quantile regression approach

Sun Yeop Lee [a], Justin Rodgers [b], Rockli Kim [c,d], S.V. Subramanian [b,e,*]

[a] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[b] Harvard Center for Population & Development Studies, Cambridge, MA, USA
[c] Division of Health Policy and Management, College of Health Science, Korea University, Seoul, South Korea
[d] Interdisciplinary Program in Precision Public Health, Department of Public Health Sciences, Graduate School of Korea University, Seoul, South Korea
[e] Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA

ARTICLE INFO

Keywords:
Distributional effect
Treatment effect heterogeneity
Quantile regression
Head Start Impact Study
Child development

ABSTRACT

Heterogeneity in treatment effects of the Head Start, a federally funded early childhood development program in the United States, has previously been found in the Head Start Impact Study (HSIS), a nationally representative randomized controlled trial. While individual characteristics have been extensively examined as sources of effect heterogeneity, treatment effects may vary as a function of outcome distribution (i.e., distributional effect). Using quantile regressions, we investigated distributional effects of the Head Start on eight child developmental outcomes for first year, second year, third year, and the 3rd grade year follow-up in the HSIS data. For PPVT and Applied Problems, the effects varied substantially across quantiles in the first follow-up, but they were positive overall. The effects at the lower quantiles were larger and were sustained beyond the first follow-up (PPVT [95% CI] at 10th and 90th quantiles: 8.74 [6.22, 11.27], 3.32 [0.82, 5.81]) in the first follow-up and 5.72 [2.66, 8.77], −1.66 [-3.69, 0.37] in the second follow-up). For Behavior Problems, the effects were only positive for the lower quantiles in the first follow-up, but they became null in the latter follow-ups. For Letter-Word Identification, Spelling, and Pre-Academic, the effects were positive in the first follow-up with moderate variation across quantiles. In the second follow-up, only the effects at the lower quantiles were statistically significant, although they faded in the latter follow-ups. For Oral Comprehension and Social Skills, effects were null for all follow-ups. The Head Start had meaningful distributional effects for a range of child developmental outcomes, and distributional effects should be routinely assessed for better understanding of child developmental programs.

1. Background

The Head Start program, administered by the Office of Head Start within the U.S. Department of Health and Human Services and delivered through hundreds of local community agencies such as centers and schools, is a federally funded early childhood development program in the United States. The Head Start provides a variety of educational, health, and social services to children aged five and below and their families with the aim of promoting school readiness by supporting the development of the whole child. In 2002, after 37 years of operation, the Office of Head Start launched the Head Start Impact Study (HSIS) to evaluate the effectiveness of Head Start with respect to children's cognitive, social-emotional, and health outcomes in a randomized controlled trial (RCT) setting. The HSIS recruited a nationally representative sample of Head Start applicant children and followed them from age three or four to third grade (Puma et al., 2010).

The HSIS official report found that the Head Start positively affected children in the short-term (i.e., after the first year of the Head Start) and that the effects mostly faded away in a few years (Puma et al., 2010, 2012). However, a number of follow-up studies using the same data have revealed substantial variation in the effects of Head Start, with one study reporting a pattern of the Head Start having larger positive effects for systematically more excluded subgroups of children such as Hispanics, those with low maternal education level, and those who had low cognitive measures at baseline (S. Y. Lee et al., 2021).

While previous attempts at uncovering heterogeneous effects of the Head Start have focused on individual or environmental characteristics as sources of effect heterogeneity, another source is an outcome

* Corresponding author. Harvard Center for Population & Development Studies, 9 Bow Street, Cambridge, MA 02138, USA.
E-mail address: svsubram@hsph.harvard.edu (S.V. Subramanian).

https://doi.org/10.1016/j.ssmph.2022.101108
Received 6 September 2021; Received in revised form 21 April 2022; Accepted 22 April 2022
Available online 27 April 2022

distribution itself; the Head Start effect may vary as a function of a distribution of a targeted outcome (i.e., distributional effect). The distributional effect describes how the treatment affected the outcome distribution, whereas an ordinary least squares linear regression only describes the effect on average. This approach may be especially relevant for child development outcomes, such as cognitive test scores or desired behaviors, as many of them have skewed distributions due, in part, to floor and ceiling effects (Petscher & Logan, 2014). Indeed, a couple of studies using quantile regressions have found that those who had low scores on a cognitive outcome at baseline, Peabody Picture Vocabulary Test (PPVT), experienced a larger improvement in PPVT after a year of Head Start (Bitler et al., 2014; Feller et al., 2016). However, both studies were limited to evaluating the Head Start effect on a single cognitive outcome at a single short-term (i.e., one-year) follow-up. Other cognitive and social-emotional outcomes at later follow-ups remained unexplored in terms of effect heterogeneity along the outcome distributions (S. Y. Lee et al., 2021).

In this study, we estimated distributional effects of the Head Start on a range of child developmental outcomes. Using the HSIS data, we extended previous attempts at using quantile regressions to investigate heterogeneity in the Head Start effects along the outcome distributions. Specifically, we analyzed the quantile effects for six cognitive and two social-emotional outcomes at four time points (first year, second year, third year, and the 3rd grade year follow-ups).

2. Methods

2.1. Sample

The HSIS utilized a multi-stage sampling procedure in order to select Head Start programs, centers, and children into the study (Puma et al., 2010, 2012). The sampling procedure, initiated in 1998, first categorized the initial 1,715 programs into 161 geographic clusters and 25 strata (based on region, state-level childcare policy, race/ethnicity, and urbanicity). Next, one cluster was randomly selected from each stratum, excluding programs that were closed, merged, or saturated and grouping those with small sample sizes. These programs were then stratified by type and local contextual characteristics. From there, three programs per stratum were randomly chosen from each stratum. Lastly, centers were randomly selected from the final set of programs. This process produced a final sample of 4,442 children including of age three or four associated with 378 centers within 84 programs. More details are available in the HSIS official reports (Puma et al., 2010, 2012).

The HSIS data are hosted by Inter-university Consortium for Political and Social Research. Restrictions apply to the availability of these datasets. All methods were carried out in accordance with relevant guidelines and regulations. The Harvard Longwood Campus Institutional Review Board (IRB) allows researchers to self-determine whether their research meets the requirements of IRB oversight using the IRB Decision tool. The HSIS data were not collected specifically for this study, and no one on the study team has access to identifiers linked to the data. These activities do not meet the regulatory definition of human subject research. As such, our study was determined to be exempt from a full institutional review.

2.2. Treatment

In the baseline year (i.e., 2002), children were randomly assigned to treatment and control groups. The Head Start provides educational (e.g., language activities, math activities), health (e.g., health education, medical/dental services, nutritional support), and social services (e.g., family needs assessments, crisis intervention) for participating children. Therefore, the treatment of interest is having access to a mixture of multiple services that are helpful for school readiness and child development. While each Head Start program can adjust its services to the specific needs of their children, all Head Start programs are federally regulated to adhere to the Head Start Performance Standards (Puma et al., 2010).

2.3. Outcomes

Children were assessed on a multitude of developmental outcomes, commencing during children's preschool years (age 3 or 4) in the baseline year of 2002. For the present study, we only included outcomes that are reliable and compatible for our analytic approach, resulting in a total of eight outcomes: six cognitive outcomes (PPVT, Letter-Word Identification, Applied Problems, Oral Comprehension, Spelling, and Pre-Academic) and two social-emotional outcomes (Behavior Problems, Social Skills) (Table A. 1 for more details). Outcomes were excluded if 1) they did not have evidence on reliability of the measure, 2) the HSIS official report raised concerns in scoring and interpretation of their results, 3) they were subjective academic performance measures, 4) they were not available for both 3- and 4-year-old cohorts at a given follow-up year, and 5) they were categorical variables. For all outcomes except Behavior Problems, an increase means improvement, and a lower quantile means a worse-performing part of the outcome distribution. For Behavior Problems, a decrease means improvement, and a higher quantile means a worse-performing part of the outcome distribution.

2.4. Covariates

We followed recommendations of the HSIS official reports for covariate adjustment, which suggested adjusting for 1) strong predictors of the outcome, such as sociodemographic variables and baseline outcomes, to enhance statistical precision and 2) baseline outcomes to account for any systematic bias at baseline (Puma et al., 2010, 2011, 2012). The selected sociodemographic covariates were gender (male, female), race/ethnicity (White/other, Black, Hispanic), primary language at baseline (English, Spanish), special needs (yes, no), primary caregiver's age (continuous), teen mom at birth (yes, no), living with a single parent (yes, no), recent immigrant parents (yes, no), parents' marital status (not married, married, separated/divorced/widowed), parental education level (less than high school, high school graduates, beyond high school), urbanicity (urban, rural), household risk (low, moderate, high), age cohort (age 3, age 4), and baseline outcomes (PPVT, Pre-Academic, Behavior Problems, and Social Skills).

2.5. Statistical analysis

To examine distributional effects of the Head Start, we utilized conditional quantile regressions to estimate treatment effects across quantiles of each continuous outcome. Conditional quantile regression can estimate treatment effects on quantiles of an outcome distribution conditional on covariates, rather than just the mean, and is appropriate for continuous outcomes (Koenker & Hallock, 2001).

Quantile regression models were specified for each continuous outcome variable according to the following structure: $Q_\tau(y_i) = \beta_{0\tau} + \beta'_{1\tau} T_i + \beta_\tau x'_i$, where $\beta_{0\tau}$ is an intercept, T_i is a treatment assignment, $\beta_{1\tau}{}'$ is a vector of treatment effect parameters, x'_i is a set of covariates, and τ is a corresponding quantile. Pre-specified quantiles to be estimated were 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th percentiles. This quantile regression models the relationship between the Head Start and outcomes for each specified quantile conditional on covariates. Such estimation procedure provides quantile estimates with higher precision and lower bias compared to analogous subgroup analyses provided by stratified ordinary least squares linear regression. Further, quantile regression estimates are robust to non-normal errors and outliers as the model makes no assumptions about the error distribution. However, quantile regression estimates will be more precise for quantiles nearer the center of the distribution (maximized at the median) as compared to more extreme quantiles (e.g., 10th and 90th quantiles).

All models were run for the first year, second year, third year, and the 3rd grade year follow-ups. List-wise deletion was applied for missing data because the amount of missingness was negligible. All quantile regressions were fitted using *quantreg* package (version 5.88) in R (version 4.1.1) (R Core Team, 2020).

3. Results

At baseline, there was a total sample size of 4,442 participating children, of which 2,646 were assigned to the treatment group and 1,796 were assigned to the control group (Table 1). The total sample was made up of 33.7% White/other, 30.3% Black, and 36.0% Hispanics. Approximately a quarter (25.7%) used Spanish at the baseline assessments. About half (50.4%) of children lived with a single biological parent, 38.0% had mothers who did not graduate from high school, and about one-fifth (19.2%) were recent immigrants. The average age for children's primary caregivers was 29, and 16.9% gave birth to the child as a teenager. Baseline estimates were comparable between treatment and control groups for all covariates and outcome variables except PPVT, where the estimates for the control group were slightly higher (250.03 vs. 246.97). The response rates varied across the outcomes, ranging from 80.2 to 81.8%. As with any RCT, there were noncompliance to the random assignment of the treatment. Twelve percent of the control group enrolled in Head Start, and 19% of the treatment group did not actually enroll in Head Start. The percentage of missing data for each variable in the analyses ranged from 0 to 1.5%.

Table 1
Sample characteristics at baseline by the treatment and control groups.

		Overall	Control	Head Start	Missing
N		4442	1796	2646	0
Age cohort (%)	3	2449 (55.1)	985 (54.8)	1464 (55.3)	
	4	1993 (44.9)	811 (45.2)	1182 (44.7)	
Gender (%)	Male	2239 (50.4)	912 (50.8)	1327 (50.2)	0
Race/ ethnicity (%)	White	1496 (33.7)	623 (34.7)	873 (33.0)	0
	Black	1348 (30.3)	536 (29.8)	812 (30.7)	
	Hispanic & others	1598 (36.0)	637 (35.5)	961 (36.3)	
Primary language (%)	English	3301 (74.3)	1345 (74.9)	1956 (73.9)	0
	Spanish	1141 (25.7)	451 (25.1)	690 (26.1)	
Parental education (%)	More	1274 (28.7)	505 (28.1)	769 (29.1)	0
	High school	1481 (33.3)	592 (33.0)	889 (33.6)	
	Less	1687 (38.0)	699 (38.9)	988 (37.3)	
Single parent (%)		2239 (50.4)	907 (50.5)	1332 (50.3)	0
Recent immigrant (%)		855 (19.2)	337 (18.8)	518 (19.6)	0
Marital status (%)	Married	1972 (44.4)	806 (44.9)	1166 (44.1)	0.1
	Separated & divorced & widowed	724 (16.3)	290 (16.1)	434 (16.4)	
	Never	1742 (39.2)	699 (38.9)	1043 (39.4)	
Special needs (%)		570 (12.8)	204 (11.4)	366 (13.8)	0
Teen mom (%)		752 (16.9)	330 (18.4)	422 (15.9)	0
Urban (%)		3746 (84.3)	1513 (84.2)	2233 (84.4)	0
Household risk (%)	Low	3383 (76.2)	1399 (77.9)	1984 (75.0)	0
	Moderate	741 (16.7)	277 (15.4)	464 (17.5)	
	High	318 (7.2)	120 (6.7)	198 (7.5)	
Caregiver's age (mean (SD))		28.91 (7.34)	28.65 (7.06)	29.08 (7.52)	0
PPVT (mean (SD))		248.21 (42.64)	250.03 (42.76)	246.97 (42.53)	1.5
Pre-Academic (mean (SD))		347.27 (22.99)	346.75 (22.82)	347.61 (23.11)	1.5
Behavior Problems (mean (SD))		6.15 (3.65)	6.21 (3.68)	6.11 (3.62)	0
Social Skills (mean (SD))		12.25 (1.79)	12.25 (1.77)	12.25 (1.80)	0

For PPVT, Applied Problems, and Behavior Problems, the Head Start effects were positive in the first follow-up, and larger effects were observed in the worse-performing quantiles from the first follow-up (Table 2). For PPVT and Applied Problems, the quantile effect estimates were statistically significant for most quantiles in the first follow-up, but they ranged from 3.32 to 8.74 for PPVT and from 0.35 to 5.60 for Applied Problems, suggesting substantial variation in the effects along the outcome distribution. The effects were especially larger at the lower part of the outcome distribution and were sustained beyond the first follow-up. For Behavior Problems, the quantile effect estimates were only statistically significant at the worse-performing quantiles in the first follow-up. However, the effect became mostly null in the later follow-ups.

For PPVT, the effect estimates [95% CI] for the 10th and 90th percentiles were 8.74 [6.22, 11.27], 3.32 [0.82, 5.81] in the first follow-up and 5.72 [2.66, 8.77], −1.66 [−3.69, 0.37] in the second follow-up. The positive effects for the 10th and 20th percentiles continued until the third follow-up, but the effects faded away in the 3rd grade follow-up. For Applied Problems, the effect estimates [95% CI] for the 10th and 90th percentiles were 5.15 [2.37, 7.92], 0.35 [−0.96, 1.65] in the first follow-up and 3.28 [0.53, 6.03], 0.33 [−0.96, 1.34] in the second follow-up. The positive effects for the 10th, 20th, and 30th percentiles continued until the second follow-up, but the effects disappeared in the third follow-up and the 3rd grade follow-up. For Behavior Problems, the effect estimates [95% CI] for the 60th, 70th, and 80th percentiles were −0.25 [−0.43, −0.06], −0.32 [−0.58, −0.05], and −0.27 [−0.58, 0.00], respectively, in the first follow-up year. In the second follow-up year, the effect estimate [95% CI] was −0.41 [−0.73, −0.09] for the 90th percentile.

For Letter-Word Identification, Spelling, and Pre-Academic, the Head Start had positive effects for all quantiles in the first follow-up, and the variation in the effect estimates across the nine quantiles was moderate with no clear pattern, ranging from 4.97 to 6.74 for Letter-Word Identification, from 1.81 to 3.90 for Spelling, and from 1.60 to 4.79 for Pre-Academic (Table 3). In the second follow-up, only the effects at the worse-performing part of the outcome distribution were statistically significant, while the effects at other quantiles disappeared. In the second follow-up year, the effect estimates [95% CI] were statistically significant at the 10th and 20th percentiles (3.84 [2.03, 5.66], 3.38 [0.97, 5.79]) for Letter-Word Identification, at the 20th and 30th percentiles (2.89 [0.48, 5.29], 1.99 [0.08, 3.90]) for Spelling, and at the 10th, 20th, and 40th percentiles (3.32 [1.21, 5.44], 2.14 [0.46, 3.82], 1.52 [0.00, 3.05]) for Pre-Academic. The effects for all quantiles faded in the third follow-up and the 3rd grade follow-up.

There was no clear evidence of the Head Start effects for Oral Comprehension and Social Skills for all four follow-ups (Table 4). For both outcomes, the effect estimates did not vary substantially across the nine quantiles.

4. Discussion

Using the HSIS data, we estimated distributional effects of the Head Start on a range of cognitive and social-emotional outcomes. With the quantile regression approach, we found that the Head Start had larger

Table 2
Quantile treatment effect estimates (95% confidence intervals) for PPVT, Applied Problems, and Behavior Problems.

Quantile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
PPVT									
1st year	**8.74**	**6.92**	**6.30**	**5.08**	**4.69**	**5.34**	**4.08**	**3.57**	**3.32**
	(6.22, 11.27)	**(4.64, 9.21)**	**(4.04, 8.56)**	**(3.27, 6.90)**	**(3.01, 6.37)**	**(3.44, 7.24)**	**(2.17, 5.99)**	**(1.72, 5.42)**	**(0.82, 5.81)**
2nd year	**5.72**	**2.92**	**2.33**	**3.03**	2.04	2.06	1.47	−0.85	−1.66
	(2.66, 8.77)	**(0.13, 5.71)**	**(0.31, 4.36)**	**(0.83, 5.23)**	(-0.26, 4.34)	(-0.07, 4.20)	(-0.63, 3.58)	(-3.40, 1.71)	(-3.69, 0.37)
3rd year	**3.43**	**2.32**	1.64	1.71	**1.96**	1.36	1.10	1.19	**3.34**
	(1.85, 5.00)	**(0.40, 4.24)**	(-0.12, 3.41)	(-0.10, 3.52)	**(0.23, 3.70)**	(-0.40, 3.11)	(-0.71, 2.91)	(-0.62, 3.00)	**(1.15, 5.53)**
3rd grade	1.86	1.13	1.20	0.76	1.10	1.46	0.08	1.71	**2.50**
	(-0.46, 4.17)	(-1.03, 3.29)	(-0.65, 3.05)	(-0.95, 2.46)	(-0.75, 2.95)	(-0.40, 3.32)	(-1.65, 1.82)	(-0.41, 3.84)	**(0.82, 4.18)**
Applied Problems									
1st year	**5.15**	**5.60**	**4.56**	**3.25**	**1.72**	0.90	**1.17**	0.74	0.35
	(2.37, 7.92)	**(3.65, 7.55)**	**(2.44, 6.67)**	**(1.28, 5.22)**	**(0.37, 3.07)**	(-0.39, 2.18)	**(0.01, 2.34)**	(-0.26, 1.75)	(-0.96, 1.65)
2nd year	**3.28**	**1.51**	**1.81**	1.50	1.32	**1.74**	0.79	0.83	0.33
	(0.53, 6.03)	**(0.08, 2.93)**	**(0.34, 3.28)**	(-0.03, 3.02)	(-0.02, 2.66)	**(0.55, 2.93)**	(-0.32, 1.89)	(-0.28, 1.94)	(-0.69, 1.34)
3rd year	−1.35	−0.66	−0.28	0.53	−0.23	0.06	0.54	−0.21	−0.32
	(-3.60, 0.90)	(-2.59, 1.26)	(-1.91, 1.34)	(-0.82, 1.88)	(-1.65, 1.19)	(-1.34, 1.45)	(-0.98, 2.06)	(-1.40, 0.97)	(-1.50, 0.85)
3rd grade	**−2.00**	−0.53	−1.03	−0.32	−0.46	−0.47	−0.98	−0.79	0.13
	(-3.77, -0.23)	(-2.47, 1.41)	(-2.62, 0.56)	(-1.71, 1.07)	(-1.91, 0.98)	(-1.90, 0.97)	(-2.46, 0.50)	(-2.24, 0.67)	(-1.18, 1.43)
Behavior Problems									
1st year	**−0.32**	−0.17	−0.16	**−0.23**	**−0.28**	**−0.25**	**−0.32**	**−0.27**	−0.20
	(-0.55, -0.10)	(-0.39, 0.04)	(-0.38, 0.06)	**(-0.45, -0.01)**	**(-0.51, -0.05)**	**(-0.43, -0.06)**	**(-0.58, -0.05)**	**(-0.57, 0.00)**	(-0.49, 0.09)
2nd year	−0.04	−0.05	−0.16	−0.10	0.00	0.00	0.06	−0.14	**−0.41**
	(-0.22, 0.13)	(-0.26, 0.16)	(-0.36, 0.04)	(-0.32, 0.12)	(-0.24, 0.23)	(-0.26, 0.27)	(-0.23, 0.34)	(-0.46, 0.19)	**(-0.73, -0.09)**
3rd year	0.07	−0.03	0.05	0.12	0.04	0.06	−0.02	−0.28	−0.33
	(-0.07, 0.20)	(-0.19, 0.14)	(-0.14, 0.24)	(-0.12, 0.35)	(-0.24, 0.31)	(-0.21, 0.34)	(-0.35, 0.31)	(-0.61, 0.05)	(-0.80, 0.14)
3rd grade	0.07	0.11	0.17	0.03	0.03	−0.11	−0.24	−0.20	−0.42
	(-0.10, 0.25)	(-0.11, 0.33)	(-0.11, 0.45)	(-0.26, 0.31)	(-0.26, 0.33)	(-0.48, 0.27)	(-0.64, 0.17)	(-0.62, 0.21)	(-0.84, 0.01)

Notes: Bolded estimates are statistically significant at 0.05 level.

Table 3
Quantile treatment effect estimates (95% confidence intervals) for Letter-Word Identification, Spelling, and Pre-Academic.

Quantile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
Letter-Word Identification									
1st year	**5.13**	**5.67**	**5.78**	**6.16**	**6.13**	**6.74**	**6.28**	**5.46**	**4.97**
	(3.69, 6.56)	**(3.92, 7.42)**	**(4.03, 7.53)**	**(4.41, 7.90)**	**(44.6, 7.79)**	**(4.98, 8.49)**	**(4.39, 8.18)**	**(3.41, 7.50)**	**(2.68, 7.27)**
2nd year	**3.84**	**3.38**	2.10	1.59	1.68	1.27	−0.14	−0.83	−0.65
	(2.03, 5.66)	**(0.97, 5.79)**	(-0.22, 4.43)	(-0.51, 3.69)	(-0.53, 3.89)	(-1.03, 3.57)	(-2.66, 2.37)	(-2.89, 1.24)	(-2.64, 1.33)
3rd year	1.30	0.80	0.85	−1.84	−0.98	−0.93	−0.38	−0.75	1.58
	(-1.40, 4.00)	(-2.06, 3.67)	(-1.96, 3.65)	(-4.38, 0.69)	(-3.19, 1.23)	(-3.08, 1.23)	(-2.54, 1.79)	(-3.14, 1.64)	(-1.18, 4.34)
3rd grade	0.92	−0.65	−0.24	−0.12	0.87	0.89	0.47	1.44	2.16
	(-1.99, 3.82)	(-3.19, 1.90)	(-2.62, 2.15)	(-1.97, 1.73)	(-1.15, 2.90)	(-1.10, 2.89)	(-1.48, 2.42)	(-0.47, 3.35)	(-0.20, 4.52)
Spelling									
1st year	**3.72**	**3.57**	**3.90**	**3.55**	**2.73**	**2.05**	**2.23**	**1.81**	**2.09**
	(1.81, 5.64)	**(2.01, 5.14)**	**(2.48, 5.32)**	**(2.07, 5.02)**	**(1.24, 4.22)**	**(0.61, 3.49)**	**(0.66, 3.80)**	**(0.16, 3.46)**	**(0.23, 3.94)**
2nd year	1.53	**2.89**	**1.99**	1.37	0.12	−0.33	−0.50	−0.24	0.67
	(-1.00, 4.07)	**(0.48, 5.29)**	**(0.08, 3.90)**	(-0.55, 3.30)	(-1.85, 2.09)	(-2.22, 1.56)	(-2.33, 1.33)	(-1.96, 1.47)	(-1.02, 2.36)
3rd year	−0.17	−1.22	0.56	0.14	0.58	1.27	1.47	0.56	0.53
	(-2.45, 2.10)	(-3.36, 0.92)	(-1.39, 2.51)	(-1.47, 1.74)	(-1.03, 2.18)	(-0.30, 2.84)	(-0.10, 3.04)	(-1.18, 2.29)	(-1.08, 2.14)
3rd grade	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Pre-Academic									
1st year	**4.79**	**4.70**	**4.37**	**4.02**	**4.13**	**3.43**	**3.70**	**2.85**	**1.60**
	(3.03, 6.55)	**(3.47, 5.94)**	**(3.04, 5.70)**	**(2.84, 5.20)**	**(2.99, 5.27)**	**(2.27, 4.59)**	**(2.46, 4.94)**	**(1.66, 4.05)**	**(0.18, 3.43)**
2nd year	**3.32**	**2.14**	1.52	**1.52**	0.85	−0.12	0.48	−0.14	−0.24
	(1.21, 5.44)	**(0.46, 3.82)**	(-0.01, 3.06)	**(0.00, 3.05)**	(-0.69, 2.38)	(-1.59, 1.35)	(-0.90, 1.87)	(-1.61, 1.32)	(-1.90, 1.43)
3rd year	−0.39	−1.06	−0.19	−0.38	0.28	0.26	0.47	0.29	0.54 (−1.31, 2.38)
	(-3.48, 2.71)	(-2.78, 0.68)	(-2.08, 1.71)	(-1.85, 1.08)	(-1.32, 1.8)	(-1.09, 1.61)	(-1.10, 2.03)	(-1.01, 1.60)	
3rd grade	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Notes: Bolded estimates are statistically significant at 0.05 level.

beneficial effects at the worse-performing part of the outcome distribution (i.e., lower quantiles) for five out of six cognitive outcomes, which were sustained until two to three years after the intervention depending on the outcomes. For one out of two social-emotional outcomes, the Head Start also had larger beneficial effects at the worse-performing part of the outcome distribution (i.e., higher quantiles) in the first year after Head Start. Overall, a substantial variation in treatment effects along the outcome distribution was observed for multiple child developmental outcomes, highlighting the importance of distributional effect estimation.

Since the conclusion of the HSIS, studies have examined the potential heterogeneity of treatment effects of the Head Start, and there have been

Table 4
Quantile treatment effect estimates (95% confidence intervals) for Oral Comprehension and Social Skills.

Quantile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
Oral Comprehension									
1st year	**0.92**	0.56	0.24	0.25	0.08	−0.20	0.01	−0.72	0.06
	(0.19, 1.65)	(-0.46, 1.58)	(-0.66, 1.14)	(-0.52, 1.03)	(-0.69, 0.85)	(-0.97, 0.57)	(-0.98, 1.00)	(-1.65, 0.21)	(-1.00, 1.11)
2nd year	0.75	0.66	0.66	0.33	0.05	−0.12	0.52	0.40	0.74
	(-0.45, 1.96)	(-0.46, 1.77)	(-0.36, 1.68)	(-0.62, 1.27)	(-0.90, 1.00)	(-1.07, 0.82)	(-0.52, 1.56)	(-0.44, 1.25)	(-0.42, 1.90)
3rd year	0.25	0.66	0.80	0.36	0.77	0.54	0.89	**1.03**	0.61
	(-1.24, 1.75)	(-0.70, 2.02)	(-0.43, 2.03)	(-0.73, 1.44)	(-0.24, 1.79)	(-0.39, 1.61)	(-0.20, 1.98)	**(0.07, 2.00)**	(-0.83, 2.05)
3rd grade	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Social Skills									
1st year	0.12	0.00	0.09	0.09	0.03	0.02	0.00	0.00	0.00
	(-0.08, 0.31)	(-0.18, 0.18)	(-0.04, 0.22)	(-0.03, 0.20)	(-0.09, 0.16)	(-0.08, 0.12)	(-0.02, 0.02)	(-0.00, 0.00)	(-0.00, 0.00)
2nd year	0.11	−0.02	−0.06	−0.07	−0.04	0.03	0.01	0.00	0.00
	(-0.14, 0.35)	(-0.21, 0.18)	(-0.20, 0.08)	(-0.19 0.05)	(-0.14, 0.06)	(-0.06, 0.11)	(-0.01, 0.02)	(-0.00, 0.00)	(-0.00, 0.00)
3rd year	0.26	0.14	**0.17**	0.13	0.07	0.09	0.03	0.00	0.00
	(-0.01, 0.53)	(-0.04, 0.32)	**(0.02, 0.03)**	(0.00, 0.26)	(-0.04, 0.18)	(-0.02, 0.19)	(0.00, 0.06)	(-0.00, 0.00)	(-0.00, 0.00)
3rd grade	0.09	−0.02	0.05	0.08	0.02	0.02	0.03	0.00	0.00
	(-0.18, 0.37)	(-0.26, 0.21)	(-0.13, 0.23)	(-0.08, 0.24)	(-0.12, 0.17)	(-0.11, 0.16)	(-0.05, 0.11)	(-0.04, 0.04)	(-0.00, 0.00)

Notes: Bolded estimates are statistically significant at 0.05 level.

reviews summarizing the findings (S. Y. Lee et al., 2021; Morris et al., 2018). Most attempts focused on subgroup or interaction analyses, investigating children's individual or environmental characteristics as sources of effect heterogeneity. Distributional effects, on the other hand, were rarely examined. In this study, we found substantial amounts of variation in treatment effects for a number of child development outcomes. Distributional effect estimation is especially useful in an RCT setting as in the HSIS because, in addition to the mean, the outcome distribution at baseline is expected to be comparable between the treatment and control groups. Our findings on distributional effects further the understanding of previous findings in multiple ways.

Our findings on cognitive outcomes are aligned with previous studies utilizing quantile regressions for the first follow-up, where larger effects for PPVT were found at the lower quantiles (Bitler et al., 2014; Feller et al., 2016). We extended these analyses to five additional cognitive outcomes and four additional follow-ups, among which four of them were found to have a similar trend as PPVT (i.e., larger effects at the lower quantiles). In other words, the Head Start was able to shift up the cognitive outcome distribution at the lower tail, or those who performed worse at baseline. These findings add support to the previous observations that the Head Start had compensatory effects on cognitive outcomes, having larger positive effects for multiple systematically excluded subgroups such as Spanish-speaking children (Bitler et al., 2014; Bloom & Weiland, 2015), children with non-parent care at baseline (Lipscomb et al., 2013) or special needs (K. Lee & Rispoli, 2016). By analyzing additional follow-ups, we found that larger benefits in multiple cognitive outcomes for those at the lower quantiles were sustained until two to three years after the intervention. It was previously reported that the effects were present for PPVT and Applied Problems until two to three years after the Head Start, and we showed that these effects were only present at the lower quantiles. Besides, the average effects on Letter-Word Identification and Spelling have previously been reported to be null at the second follow-up, but we found that positive effects exist at lower quantiles. As such, the average treatment effect estimation alone may be insufficient in evaluating the Head Start effect for certain outcomes.

For the social-emotional outcomes, this is the first study using the HSIS data to explore the quantile effects of the Head Start beyond the first follow-up. For Behavior Problems, the Head Start also had larger positive effects at the worse-performing part of the outcome distribution (i.e., higher quantiles, or those with more behavioral problems) in the first follow-up. Previously, the positive effects for Behavior Problems have been reported for the first follow-up, but our finding provides a new insight into the compensatory effect on social-emotional outcomes. Although some studies did not find larger effects for the social-emotional outcomes among high-risk subgroups such as children with foster care at baseline (K. Lee & Lee, 2016) or who had experienced violence (K. Lee & Ludington, 2016), our observation on quantiles suggest that the Head Start indeed benefitted children with more behavior problems at baseline, and a mechanism underlying this phenomenon should be uncovered in future research.

A pattern of larger effects at the worse-performing part of the outcome distribution reflects a reduced variability, or inequality, in the outcome across children, which is aligned with a previous finding that variance of the outcome has reduced after the Head Start intervention (S. Y. Lee et al., 2022). With the assumption that the effect was monotonic along the outcome distribution, or that the rank of children did not change, we may conclude that the Head Start can shrink the outcome distribution by pulling up those who performed worse and ultimately promote leveling the playing field in child development. This, on one hand, is aligned with the goal of the program, which is to assist development of children who may be left behind if the additional support were not given. On the other hand, the children included in the HSIS were already disadvantaged compared to an average child according to the eligibility criteria for the program. This means that some disadvantaged children are not being benefitted from the Head Start. We need to further our understanding on why the Head Start does not work well on some children and why it is more effective for the more disadvantaged even within the already disadvantaged group of children.

The present study is not without limitations. First, measurement error in Head Start outcomes may inflate estimates of variance. Second, crossover in treatment assignment may have effects on the precision and magnitude of the effect estimates. One study suggests that the crossover may underestimate the Head Start effects (Feller et al., 2016). Third, there were missing observations due to lost to follow-up over the study period. While this could also lead to biased estimates, the HSIS official report found minimal impacts of loss to follow-up. Fourth, the conditional quantile regressions make inference on the distribution itself, not to the individuals.

The present study extended the previous attempt at estimating a short-term distributional effect of the Head Start on a single cognitive outcome by adding more outcomes and further analyzing additional longer-term follow-ups. Our findings confirmed that compensatory effects of the Head Start were at play for multiple developmental outcomes and even for two to three years after the intervention. Some effects were known to be present on average, but we found that they were present only for worse-performing quantiles. Furthermore, some outcomes with null effect on average were found to have statistically significant effects on some quantiles. Taken together, variation is an important component when evaluating the effectiveness of an intervention, as treatment effects may vary by individual characteristics, study design and implementation, and the outcome distributions themselves. Distributional effect estimation should be a routine practice in child development program evaluations.

Funding

This work was supported by the Robert Wood Johnson Foundation (grant ID: 75602).

Author contributions

RK and SVS conceptualized and designed the study. JR contributed to the initial analysis of the data as well as to writing the first draft of the manuscript. SYL contributed to the conceptualization of the study and interpretation of the data. SYL led the revision. All authors approved of the final draft.

Ethical statement

The HSIS data are not collected specifically for this study and no one on the study team has access to identifiers linked to the data. These activities do not meet the regulatory definition of human subject research. As such, an Institutional Review Board (IRB) review was not required. The Harvard Longwood Campus IRB allows researchers to self-determine when their research does not meet the requirements for IRB oversight via guidance online regarding when an IRB application is required using an IRB Decision Tool.

Declaration of competing interest

The authors have no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssmph.2022.101108.

References

Bitler, M. P., Hoynes, H. W., & Domina, T. (2014). *Experimental evidence on distributional effects of Head Start (No. 20434*. National Bureau of Economic Research, Inc. NBER Working Paper).

Bloom, H. S., & Weiland, C. (2015). *Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact Study*. MDRC.

Feller, A., Grindal, T., Miratrix, L., & Page, L. (2016). Compared to what? Variation in the impacts of early childhood education by alternative care type. *Annals of Applied Statistics, 10*(3), 1245–1285. https://doi.org/10.1214/16-aoas910

Lee, S. Y., Kim, R., Rodgers, J., & Subramanian, S. V. (2021). Treatment effect heterogeneity in the head start impact study: A systematic review of study characteristics and findings. *SSM - Population Health, 16*, Article 100916. https://doi.org/10.1016/j.ssmph.2021.100916

Lee, S. Y., Kim, R., Rodgers, J., & Subramanian, S. V. (2022). Assessment of heterogeneous Head Start treatment effects on cognitive and social-emotional outcomes. *Scientific Reports, 12*(1), 1–11. https://doi.org/10.1038/s41598-022-10192-1

Koenker, R., & Hallock, K. (2001). *Quantile regression. 15*. Journal of economic perspectives.

Lee, K., & Lee, J.-S. (2016). Parental book reading and social-emotional outcomes for Head Start children in foster care. *Social Work in Public Health, 31*(5), 408–418. https://doi.org/10.1080/19371918.2015.1137523

Lee, K., & Ludington, B. (2016). Head Start's impact on socio-emotional outcomes for children who have experienced violence or neighborhood crime. *Journal of Family Violence, 31*(4), 499–513. https://doi.org/10.1007/s10896-015-9790-y

Lee, K., & Rispoli, K. (2016). Effects of individualized education programs on cognitive outcomes for children with disabilities in Head Start programs. *Journal of Social Service Research, 42*(4), 533–547. https://doi.org/10.1080/01488376.2016.1185075

Lipscomb, S. T., Pratt, M. E., Schmitt, S. A., Pears, K. C., & Kim, H. K. (2013). School readiness in children living in non-parental care: Impacts of Head Start. *Journal of Applied Developmental Psychology, 34*, 28–37. https://doi.org/10.1016/j.appdev.2012.09.001

Morris, P. A., Connors, M., Friedman-Krauss, A., McCoy, D. C., Weiland, C., Feller, A., Page, L., Bloom, H., & Yoshikawa, H. (2018). New findings on impact variation from the Head Start Impact Study: Informing the scale-up of early childhood programs. *AERA Open, 4*(2). https://doi.org/10.1177/2332858418769287

Petscher, Y., & Logan, J. A. R. (2014). Quantile regression in the study of developmental sciences. *Child Development, 85*(3), 861–881. https://doi.org/10.1111/cdev.12190

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., … Spier, E. (2010). *Head Start impact study final report*. U.S. Department of Health and Human Services, Administration for Children and Families.

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., … Spier, E. (2011). *Head Start impact study technical report*. U.S. Department of Health and Human Services, Administration for Children and Families.

Puma, M., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., & Downer, J. (2012). *Third grade follow-up to the Head Start impact study final report*. U.S. Department of Health and Human Services, Administration for Children and Families.

R Core Team. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing (4.0.3) https://www.r-project.org/.