Timing of High-Quality Child Care and Cognitive, Language, and Preacademic Development

Weilin Li, **George Farkas**, **Greg J. Duncan**, **Margaret R. Burchinal**, and **Deborah Lowe Vandell**
School of Education, University of California, Irvine.

Abstract

The effects of high- versus low-quality child care during 2 developmental periods (infant–toddlerhood and preschool) were examined using data from the National Institute of Child Health and Human Development Study of Early Child Care. Propensity score matching was used to account for differences in families who used different combinations of child care quality during the 2 developmental periods. Findings indicated that cognitive, language, and preacademic skills prior to school entry were highest among children who experienced high-quality care in both the infant–toddler and preschool periods, somewhat lower among children who experienced high-quality child care during only 1 of these periods, and lowest among children who experienced low-quality care during both periods. Irrespective of the care received during infancy–toddlerhood, high-quality preschool care was related to better language and preacademic outcomes at the end of the preschool period; high-quality infant–toddler care, irrespective of preschool care, was related to better memory skills at the end of the preschool period.

Keywords

child-care quality; propensity score matching; infant–toddler; preschool

Increases in women's educational and occupational opportunities have been accompanied by an expansion of nonmaternal child care in U.S. families (National Center for Education Statistics, 2010; Table 54) as well as questions about the effects of nonmaternal child care on child developmental outcomes. Despite evidence that higher quality care during the first 5 years is linked to child cognitive and academic outcomes (Camilli, Vargas, Ryan, & Barnett, 2010; Vandell & Wolfe, 2000), little work has partitioned the unique contributions of experiences in the infant–toddler versus the preschool period. Also, largely unexamined are associations between child outcomes and combinations of quality care during these two periods.

We used data from the multisite National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (henceforward NICHD Study of Early Child Care) to study child care quality effects on child cognitive, language, and preacademic

Correspondence concerning this article should be addressed to Weilin Li, School of Education, University of California, Irvine, 2000 Education, Irvine, CA 92697. weilinl@uci.edu.

skills shortly before kindergarten entry at 54 months. We used propensity score matching techniques to reduce selection bias from differences between families who placed their children in high- or low-quality child care. We proceeded by dividing early childhood into two developmental periods—the infant–toddler and the preschool years—and used a recognized quality threshold to distinguish between low- and high-quality child care. This enabled us to compare outcomes of four groups of children who received different combinations of quality care during the two developmental periods. By accounting for the propensity of families to select into each of these possible combinations, we obtained valuable information on the joint contribution of quality care at the infant–toddler and preschool periods to child outcomes.

Effects of Quality Care During the Infant–Toddler Period

Both developmental theory and empirical research suggest that high-quality child care during the infant–toddler period could enhance children's cognitive development and academic achievement in early childhood. The first 3 years of life have been identified by neuroscientists and developmental psychologists as important for domains such as early language and joint attention (Huttenlocher, 1979; Nelson & Sheridan, 2011; Shonkoff & Phillips, 2000).

Skill acquisition is enhanced when parents and other caregivers provide warm, responsive, and stimulating care involving frequent positive interactions (National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network [ECCRN], 2000). Consistent with this focus on the first 3 years are findings from economically diverse samples that showed that high-quality child care in this period, characterized by positive caregiver–infant interactions, was associated with higher cognitive and preacademic scores (Clarke-Stewart, Vandell, Burchinal, O'Brien, & McCartney, 2002; Vandell & Wolfe, 2000). Experimental studies such as the Infant Health and Development Program (IHDP; with its sample of low–birth-weight children) and the Early Head Start Research and Evaluation Project (with its sample of low-income children) revealed that high-quality care in centers improved cognitive development during infancy (Brooks-Gunn et al., 1994; Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Love et al., 2003, 2005) and, for a subset of children in the IHDP, through age 18 (McCormick et al., 2006).

Effects of High-Quality Care During the Preschool Period

Another set of studies have focused on the effects of high-quality child care during the preschool period (ages 3 to 5), a time when children are expanding their language and reasoning skills and learning how to interact effectively with both adults and other children. Interventions during this period have shown positive effects of high-quality preschool care on children's cognitive and academic skills (Garces, Thomas, & Currie, 2002; Ludwig & Miller, 2007; McCartney, Scarr, Philips, & Grajek, 1985), as have nonexperimental studies of pre-K programs (Gormley, Phillips, & Gayer, 2008; Howes et al., 2008; Mashburn et al., 2008).

Combined Effects of High-Quality Care During Both the Infant–Toddler and Preschool Periods

Several studies have documented effects of child care quality across the first 5 years on children's math and reading achievement prior to school entry (Campbell et al., 2001; NICHD ECCRN, 2005). However, surprisingly, the research literature provides relatively little guidance regarding the timing of quality care effects during the first 5 years, that is, how the quality of care during infant–toddler and preschool periods uniquely and in combination are associated with child cognitive, language, and preacademic outcomes.

With regard to how quality of care combines with the timing of that care, economist James Heckman and his colleagues provided a theoretical framework for understanding the relationship between human skill formation and human capital investment at different developmental stages (Cunha, Heckman, Lochner, & Masterov, 2006). In Heckman's models, the concepts of *self-productivity* and *complementarity* were formalized to explain how skill begets skill (Cunha, Heckman, & Schennach, 2010), or how human skill at an earlier stage augments skill attainment at later stages (self-productivity), and how human capital investment at the earlier stage raises the productivity of investments at subsequent stages (complementarity). Declining effect sizes with age have been reported in evaluations of infant–toddler and preschool child care interventions such as the IHDP (McCormick et al., 2006) and may have been caused by the fact that the children in the intervention group received similar types of preschooling as children in the control group after the program ended but in fact needed continued participation in high-quality programming to sustain impacts (Brooks-Gunn et al., 1994). The current article implicitly tested a prediction from Heckman's framework by estimating how different combinations of child care quality during the infant–toddler and the preschool periods were associated with different levels of child outcomes. Specifically, our study aimed to test the following hypotheses:

Hypothesis 1: High-quality infant–toddler care would improve cognitive outcomes at the end of the infant–toddler period (in these data, at 24 months of age). However, without subsequent high-quality care during the preschool period, children with high-quality infant–toddler care would not have higher cognitive, language, and preacademic scores at 54 months of age than children with low-quality infant–toddler care followed by low-quality preschool care.

Hypothesis 2: High-quality care during the preschool period would improve cognitive, language, and preacademic outcomes at the end of the preschool period (in these data, at 54 months of age). Moreover, the beneficial effects of high-quality preschool care would be augmented by high- as opposed to low-quality infant–toddler care. In other words, the combination of high-quality infant–toddler care and high-quality preschool care would be associated with higher cognitive, language, and preacademic scores at 54 months of age than the combination of high-quality care during the preschool period and low-quality infant–toddler care. These two set of hypotheses led to

Hypothesis 3: High-quality child care during both the infant–toddler and preschool periods would be associated with higher cognitive, language, and preacademic

performance at the end of the preschool period than any other child care quality combination during the two periods.

Beyond testing these three hypotheses, our results provided estimates of the associations between child outcomes and mixed combinations of care (i.e., high-quality infant care coupled with low-quality preschool care and low-quality infant care coupled with high-quality preschool care). Theoretical and empirical work is largely silent regarding which of these "mixed" conditions is likely to produce the larger effects on child outcomes.

The NICHD Study of Early Child Care was an observational study; hence, children and families self-selected into child care of various qualities. Children who experienced higher quality care were more likely to have families with greater income, higher maternal education, and more child-centered beliefs about child rearing—factors that were also related to higher cognitive and academic outcomes (NICHD ECCRN, 2005). Hence, children experiencing high- and low-quality child care had different family backgrounds, and those differences could bias estimates of the impact of quality care on child developmental outcomes.

To reduce possible biases, we employed Rosenbaum and Rubin's causal effect framework (Rosenbaum & Rubin, 1983; Rubin, 2001) in which inference about the impact of a treatment involves conjecture regarding what the outcome for targeted individuals would be if they had not received the treatment. Randomized experiments provide valid tests of this inference because, according to the law of large numbers, randomization implies that the treatment and control groups are, on average, balanced across all possible confounding variables. This balance implies that treatment assignment group is, on average, independent of all observed and unobserved factors.

Conventional multiple regression approaches, which would include the treatment indicator as a dummy variable (or, in our case, two treatment dummies—one each for quality infant care and quality preschool care) and selection factors such as child and family characteristics as covariates, reduce selection bias by controlling for measured confounders. However, these regression estimates can be problematic when there is insufficient overlap between treatment and control groups, that is, if two groups are not comparable in terms of demographics or child care characteristics indexed by control variables. An extreme example would be where the treatment group was largely composed of low-income families while the control group was largely composed of high-income families. Propensity score approaches match overlapping cases from the two comparison groups and thus reduce selection biases as well as avoid extrapolation beyond the region of the data that could occur in typical regression estimates (Caliendo & Kopeinig, 2008; Rosenbaum & Rubin, 1983). On the minus side, reliance on measured covariates still leaves propensity score matching open to the possibility of lingering selection bias due to imbalance on unmeasured confounding variables.

Method

Participants

In 1991, a socioeconomically diverse sample of children and their families were recruited at the children's birth at designated hospitals at 10 data collection sites: Little Rock, Arkansas; Irvine, California; Lawrence, Kansas; Boston, Massachusetts; Philadelphia, Pennsylvania; Pittsburgh, Pennsylvania; Charlottesville, Virginia; Morgantown, North Carolina; Seattle, Washington; and Madison, Wisconsin. A total of 1,364 families with full-term healthy newborns were enrolled in accordance with a conditionally random sampling plan, which was designed to ensure that recruited families reflected the diversity of the sites in terms of socioeconomic status, race, and ethnicity. The major exclusionary criteria were teenage mothers, families planning to leave the catchment area in the coming 3 years, children with obvious disabilities at birth, and mothers insufficiently conversant in English.

Beginning at 1 month after children's birth, families were scheduled for extensive periodic data collection by means of observations, interviews, questionnaires, and child assessments. Detailed family information was collected, as well as assessments of the home and of the child-care environments. These detailed measures and assessments provided unusually rich information to account for many potentially confounding factors in the process by which children were selected into low- or high-quality infant–toddler and preschool care.

Procedures

Child care quality—The Observationl Record of the Caregiving Environment (ORCE) was used to measure the quality of caregiving received at 6, 15, 24, 36, and 54 months. The ORCE could be used in different settings such as home care and center-based care to assess different types of caregivers such as relative, nanny, and teacher. Each assessment consisted of four 44-min child-focused observations across 2 days at 6, 15, 24, and 36 months and on 1 day at 54 months. Quality of caregiving was rated on 4-point scales. The final quality rating score was the mean of the subscales. Cronbach's alphas for the composite score ranged from .83 to .89, and reliabilities were greater than .80 at all ages.

We used a score of 3.0 on the averaged ORCE scores at 6, 15, and 24 months as the cutoff score to distinguish between low- and high-quality infant–toddler care, and 3.0 on the averaged ORCE scores at 36 and 54 months as the cutoff score for low- and high-quality care for the preschool period. Scores higher than 3.0 indicated that caregivers were more sensitive to children's behaviors, provided greater cognitive stimulation, had warmer and more sensitive interactions with children, fostered greater exploration, and were less emotionally detached (Vandell et al., 2010; Figure 1). Only children in child care settings with observed ORCE scores were included in the analysis.

Child outcomes—Our outcome measure at 24 months was the Bayley Mental Developmental Index (Bayley, 1993). This standardized test evaluates children's cognitive developmental status and has shown moderate to strong correlations with subsequent IQ measures. Hence, this index was viewed as a proxy for IQ (Nelson & Sheridan, 2011). The four child outcomes at 54 months were obtained from the Woodcock-Johnson Cognitive and

Achievement Batteries (Revised; Woodcock & Johnson, 1990) and the Preschool Language Scale (PLS; Zimmerman, Steiner, & Pond, 1979). The first outcome was the Woodcock-Johnson Memory for Sentences, which measured short-term memory (α = .84). The second outcome was a composite language score created from the Woodcock-Johnson Picture Vocabulary and the PLS Expressive and Receptive tests. The Woodcock-Johnson Picture Vocabulary scale assessed verbal comprehension by measuring the child's ability to recognize or name pictured objects (α = .73). The PLS measured a range of language behaviors, including vocabulary, morphology, syntax, and integrative thinking. The PLS has shown excellent correlations with other measures of early language development (Qi & Marley, 2011). The third outcome was the Woodcock-Johnson Letter-Word Identification, which measured symbolic learning and identification skills (α = .86). The fourth outcome was the Woodcock-Johnson Applied Problems, which measured skill in analyzing and solving practical mathematical problems (α = .85). The Woodcock-Johnson Applied Problems has demonstrated strong associations with subsequent academic achievement measures (Duncan et al., 2007). These four measures were standardized based on the overall NICHD Study of Early Child Care sample to a mean of 0 with a standard deviation of 1, so that analytic results could be compared across different outcomes.

Covariates—In our study, time-invariant demographic controls were measured at 1 month after birth. These covariates were dummy variables for study sites, child race (white or nonwhite), child gender, child's birth order, child temperament (maternal rating), maternal attitudes about raising children, maternal age, maternal educational level (in years), and paternal educational level (in years). Time-varying covariates measured at both 1 month and 24 months of age were child's health, maternal separation anxiety, maternal depression, maternal employment status, whether mother's partner was present in the household, and the family income-to-needs ratio.

Child difficult temperament was measured using the mother-reported 55-item Infant Temperament Questionnaire (Medoff-Cooper, Carey, & McDevitt, 1993). Maternal attitudes and beliefs about child rearing were measured with a 30-item questionnaire probing mothers' ideas about raising children (Schaefer & Edgerton, 1985). High scores indicated authoritarian child-rearing attitudes and beliefs. Maternal separation anxiety was assessed using the Separation Anxiety Scale (Hock, Gnezda, & McBride, 1983). High scores indicated high levels of maternal worry, sadness, and guilt during separation from the child and also indicated adherence to beliefs about the value of exclusive maternal care. Maternal depressive symptoms were measured using the Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977).

Analytic Approach

Propensity score matching in the current study involved four steps. First, we divided the sample into three sets of mutually exclusive groups defined by child care quality. The first set was based on infant–toddler care quality and was used to test the first part of Hypothesis 1, which concerned children with high-quality infant–toddler care regardless of preschool quality (labeled "early high") and children with low-quality infant–toddler care regardless of preschool quality (labeled "early low"). The second set was based on preschool quality and

was used to test the first part of Hypothesis 2, concerning children with high-quality care during the preschool period no matter the quality of infant–toddler care (labeled "late high") and children with low-quality care during the preschool period no matter the quality of infant–toddler care (labeled "late low").

In order to test the various hypotheses involving different combinations of infant–toddler and preschool care quality, four groups were defined by combinations of child care quality at both developmental periods: children with low-quality infant–toddler care and low-quality care during the preschool period (labeled "low-low"), children with high-quality infant–toddler care and low-quality care during the preschool period (labeled "high-low"), children with low-quality infant–toddler care combined with high-quality care during the preschool period (labeled "low-high"), and children with high-quality infant–toddler care and high-quality care during the preschool period (labeled "high-high"). This is the third set of mutually exclusive group comparisons.

In the second step, we built up contrasts to test the three hypotheses, which are shown in Table 1. The first column shows the hypothesis being tested. The second and third columns show the targeted and comparison groups in each contrast. Observations in comparison groups were selected to match observations in the corresponding targeted groups as described below. The fourth column shows the age at which the outcomes were measured.

The third step of the analysis employed logistic regressions for each contrast to predict a propensity score for each individual, defined as the conditional probability of being selected into the targeted group given the individual's value on a full set of covariates. Since we had six contrasts (note that in Table 1 the second contrast for Hypothesis 2 is the same as the third contrast for Hypothesis 3), a series of six binomial logistic regressions were used to generate propensity scores. For propensity scores of being selected into the early low, early high, low-low, high-low, low-high, or high-high groups, variables measured at 1 month after birth were included as predictors in the logistic models. For propensity scores of being selected into the late high or late low group, time-insensitive variables measured at 1 month after birth, as well as updated demographic variables and cognitive scores measured at 24 months of age, were used in the logistic models.

In the final step, for each contrast, observations in the comparison group were selected to match observations in the corresponding targeted group. For example, in the first contrast for Hypothesis 2, late high was the targeted group and late low was the comparison group. Thus, each individual in late high was matched to an individual in late low with almost the same propensity score. The match was conducted within each of the 10 research sites in an effort to reduce or eliminate unobserved differences in the selection process (Cook, Shadish, & Wong, 2008). A caliper width of 0.1 was used to ensure a sufficiently close match in propensity scores between targeted and comparison groups (Caliendo & Kopeinig, 2008). Balance checking was conducted to ensure the matching procedure was able to balance the distribution of the relevant covariates in both targeted and comparison groups. The results of balance checking are shown in the online supplemental material.

After propensity score matching, we employed two approaches to test each hypothesis. The first involved the most common propensity score methodology. We calculated the standardized mean difference (henceforward SMD) for each pair of balanced targeted and comparison groups and used that SMD to obtain inferences for the hypotheses. The coefficients could be viewed as effect size estimates (i.e., Cohen's *d*) because all outcomes and predictors were standardized to have a mean of 0 and *SD* of 1.

The second approach was regression on matched pairs. This was used because even after matching, small differences in distributions of covariates between the comparison and control groups may remain. Linear regression was applied to propensity score matched subsamples for the same reason it is used in the case of randomized experimental treatment and control group—to reduce variability and to increase the power of the comparison (Rubin & Thomas, 2000). Accordingly, the coefficients from the quality contrasts in our models estimated the mean difference between groups in terms of standard deviations of the outcome variable (i.e., they are comparable to the Cohen's *d* statistic).

Assessing the Matching Quality

Three methods were conducted to assess the matching quality. The first method was the most intuitive one, calculating SMDs of the control variables between matched pairs after propensity score matching. After matching, all SMDs should be between −.1 and .1, indicating balance between the matched targeted and comparison groups conditional on observed covariates. The second and third assessing methods were from Rubin (2001). These involved calculating the SMD of propensity scores of the targeted and comparison groups and calculating the ratio of the variances of the propensity scores in the two groups. Rubin (2001) illustrated that for the propensity score method to be trustworthy, the absolute SMD of propensity scores should be less than .25 and variance ratios should be between .5 and 2. These three measures were calculated to make sure that the inference would be drawn upon properly balanced samples.

Results

Table 2 presents descriptive statistics for the outcome and control variables in the analysis. The columns showed means and standard deviations for various groups with different child care quality during the infant–toddler period and the preschool period. General mental development was measured at 24 months of age, whereas specific competencies—memory, language, and preacademic skills (letter–word identification and applied problems)—were measured at 54 months. The control variables, including family and child characteristics, were collected at 1 month or at 24 months of age.

Variables in the "family characteristics" and "child characteristics" panels were used in binomial models to generate propensity scores. As expected, there was imbalance between certain pairs of groups on certain variables. For example, mean family income-to-needs at 1 month after birth was 3.88 for the high-high group compared with 2.66 for the low-low group—a highly significant difference of roughly half a standard deviation (*p* < .001). Another example of imbalance was the mother's usage of paid leave at 1 month after giving

birth— 59% of the group receiving low-quality care in the preschool period versus 49% of the group who received high-quality care in the preschool period ($p < .05$).

Also noteworthy is that for means of the 54-month outcomes (the second to the fifth rows and the four right-most columns of Table 2), the lowest average readiness was found for the low-low group and the highest was found for the high-high group, with the high-low and low-high groups relatively close together approximately midway between the two extremes. Comparisons across the four groups, after using propensity score methods to adjust for imbalanced covariates, were the primary focus of this article.

Propensity score matching is designed to adjust these kinds of imbalance in covariates; balance checking before and after matching is shown in Tables 1(a)–1(f) in the supplemental materials. Differences between the target and comparison groups on the covariates after propensity score matching were reduced, as evidenced by the values in each row. After matching, the standardized differences between groups ranged from −.1 to .1, and were not statistically significant, indicating balance between the matched targeted and comparison groups conditional on observed covariates. The last column of these tables demonstrated a substantial percentage reduction in bias after propensity score matching. Bias reduction could be negative for several covariates because of a compromise for balance across all covariates, that is, because of the compromise for all covariates to achieve an SMD between −.1 and .1. The last two rows of the tables also demonstrate Rubin's (2001) SMD of propensity scores and their variance ratios. All the absolute values of propensity score SMD were less than .001. This satisfied Rubin's "less than .25" rule, and all the variance ratios were between .989 and 1.016, which satisfied Rubin's "between .5 and 2" rule. The results in these supplemental tables demonstrate a successful propensity score matching procedure, increasing our confidence in the resulting estimates of group differences in outcomes.

The first two rows of each panel in Table 3 present the SMD of cognitive outcomes between targeted and comparison groups before (Model 1) and after (Model 2) propensity score matching. The third row shows comparison results from a regression run on matched samples (Model 3). Standardized mean differences in Model 2 and Model 3 were generally smaller than in Model 1, suggesting that preexisting differences among families who did and did not place their children in high-quality child care account for a considerable share of the simple differences in child outcomes. Accordingly, we focus on the findings from Model 2 and Model 3—the models based on propensity score matching without and with the covariates included as control variables.

Tests of Hypothesis 1

Results in the first column of Table 3 support Hypothesis 1 in showing significant positive associations between child care quality during the infant–toddler period and child cognitive outcomes at the end of this period (24 months). After propensity score adjustment (Models 2 and 3), results indicate that high-rather than low-quality infant–toddler care was associated with a .28 *SD* higher cognitive score at 24 months of age. Given that the difference of average ORCE scores between high-quality and low-quality infant care was 1.9 *SD*s, a 1-*SD* increase in ORCE increased cognitive outcome at 24 months by an effect size of .15 *SD*.

Hypothesis 1 also asserted that without subsequent high-quality care during the preschool period, children with high-quality infant–toddler care would not have higher cognitive and achievement scores at 54 months of age than children with low-quality infant–toddler care followed by low-quality preschool care. That is, the positive effect of high-quality infant–toddler care would decline by the end of the preschool period for children who received low as opposed to high-quality preschool child care. Results of this test can be seen in the second column of Table 3. Restricting attention to Model 3 results, we found support for this hypothesis except where memory was the outcome. At 54 months, the memory performance of the high-low group was 0.20 *SD* above that of the low-low group, and this is significant at the .05 level. This suggests that the short-run memory gained due to higher quality infant care continued through the preschool period, even if the child received low-quality care during this period. However, this hypothesis was not supported for language and reading (we found no statistically significant group differences for these outcomes) and was at best marginally supported for math (a group difference of .17 that is significant at only the .10 level).

Tests of Hypothesis 2

The second set of columns in Table 3 provide tests of Hypothesis 2, which asserts that high-quality care during the preschool period is associated with immediate improvement in cognitive and academic outcomes at 54 months and that the quality of infant–toddler care will affect 54 month outcomes even when preschool care was of high quality (high-high care will produce better outcomes than low-high care). The third column of this table shows the results of the first of these, by comparing outcomes for late-high-versus late-low-quality child care. Short-term memory outcomes at 54 months provided no support for the first hypothesis, but this hypothesis is strongly supported by the language, reading, and math results. Group differences for these outcomes were .19, .28, and .22, respectively, and are significant at the .001 level. Clearly, high-quality preschool care translated immediately into improved language, reading, and math outcomes, irrespective of whether the child received high-quality infant–toddler care.

Results in the fourth column of Table 3 test the second part of Hypothesis 2, that the quality of infant–toddler care would affect 54-month outcomes even when preschool care is of high quality. This was not supported for memory and language, but it was supported at the .05 level for reading and at the .10 level for math.

Tests of Hypothesis 3

The final three columns of Table 3 test the third hypothesis, that high-high care would produce better 54-month outcomes than any of the other combinations. In general, the Model 3 results support this hypothesis. Of the 12 coefficients in these columns, nine were positive and significant at the .10 level or better (5 at the .05 level or better), and only three were not significant. The magnitudes of the significant coefficients are generally similar to those discussed above, with the most regular pattern of significant positive effects occurring for the high-high versus low-low comparison. No strong pattern distinguishes between the magnitudes or statistical significance of the outcomes produced by the high-low versus the low-high child care quality pattern. That is, the high-high pattern produced the best

outcomes, the low-low pattern produced the worst outcomes, and there was little observable difference between the outcomes produced by high-low- versus low-high-quality child care.

Robustness Check

In the analyses reported in the previous section, we conducted pairwise propensity score matching to reduce selection bias and to draw inferences about the association between quality of child care and child outcomes. Our matching procedure achieved balance; that is, the comparison groups were equivalent at baseline on all covariates, supporting the validity of the estimated association. However, the method of pairwise propensity score matching had two limitations. The first limitation was in the matching procedure; all of the inferences were drawn from pairwise matched samples, which were a subset of the total analyzed sample. This limits the generalizability of the inference. The second limitation was that in the process of generating pairwise propensity scores, a series of binomial models for six different contrasts were estimated, and this might bias between-contrast comparisons (Bryson, Dorsett, & Purdon, 2002). For example, subsamples in one matched contrast could be more socioeconomically disadvantaged than subsamples in another matched contrast. Because we were concerned about these issues, we conducted two additional analyses to check the robustness of our results.

The first robustness check involved the estimation of pairwise propensity score weighted regression. This procedure enabled us to use all the observations in the analysis and weight each of them for unequal probability of selection (Lohr, 1999). Kang and Schafer (2007) called this a *doubly robust* method. For children in targeted groups, we generated weights as inverse propensity scores ($w = 1/p$). For children in comparison groups, we generated weights according to the formula $w = 1/(1 − p)$. These weights ensured that the children in the comparison group who were most like the children in the targeted group received larger weights and that those less like the children in the targeted group received smaller weights. These weights were applied to two separate regressions including all the control variables— the first regression for the targeted group and the second one for the comparison group. The two separate weighted regressions generated two sets of consistent estimates of regression coefficients. The first set, which was generated from weighted regression on the targeted group, indicated what the outcomes for the represented population would be if they had been assigned to the targeted group. So this first set of estimates was used to generate predictions for all the potential outcomes out of the full sample (including both targeted and comparison groups). Similarly, the second set of consistent estimates of regression coefficients was used to generate predictions of all the potential outcomes out of the full sample (including both targeted and comparison groups). The simple mean difference of the two predictions from the two sets of estimates was then used to obtain inferences for the hypotheses (Kang & Schafer, 2007; Schafer & Kang, 2008). The standard error for the mean difference was obtained by bootstrapping based on 2,000 replications.

Our second robustness check involved multinomial propensity score weighted regression estimation. This method was similar to our first robustness check except that propensity scores were generated from a multinomial model fit to the entire sample of all the four groups (high-high, low-low, high-low, and low-high). These *generalized propensity scores*

(Imbens, 2000) represented the probability of being assigned to a certain group rather than to the other three groups. This method could test the second limitation of pairwise propensity score matching discussed above by enabling between-contrast comparisons. For example, in this multinomial context, inference of implicit contrasts such as "high-low" versus "low-high," could be drawn by comparing contrast "high-high" vs. "high-low" with contrast "high-high" vs. "low-high." This multinomial method is not without its own limitations. It could be statistically less robust since a misspecification in one of the series would compromise all others (Lechner, 2001). A second limitation followed from multinomial regression's assumption of *independence of irrelevant alternatives* (i.e., the choice between any of the two alternatives was independent with the existence of the other alternatives; McFadden, 1984). For example, in our context, multinomial regressions were based on the assumption that the choice between low-high and low-low were independent of the existence of the other alternatives (i.e., high-high and high-low), and this assumption could have been violated if low-high and high-high were very close alternatives.

The results of robustness checking are presented in Table 4. Here the Model 3 results were replicated from Table 3 for ease of comparison with the robustness check estimates. Estimates for Model 4 involved pairwise propensity score weighted regression estimation for each of the six contrasts. Model 5 corresponded to multinomial propensity score weighted regression estimation for contrasts among high-high, high-low, low-high, and low-low. Hence, in Model 5 no estimates were obtained for the contrast of early high versus early low and the contrast of late high versus late low.

The point estimates strongly supported the Model 3 results shown in Table 3, although in some cases the standard errors in Model 4 and Model 5 were a little higher. In particular, similar to results in Table 3, Model 4 showed that high-quality infant care itself was associated with .23 *SD* higher mental development scores at 24 months. And if followed by low-quality care during the preschool period, children with high-quality infant care scored .17 to .21 *SD* higher than those with low-quality infant care.

Looking at the third column of Table 4 one can see that regardless of infant–toddler care quality, high-quality care in the preschool period was associated with .18 *SD* higher language scores, .16 *SD* higher reading scores, and .15 *SD* higher math scores. Also, if followed by high-quality care during the preschool period, children receiving high-quality infant care scored higher than children with low-quality infant care by .16 *SD* in reading scores. Children with consistently high-quality child care from birth to 54 months scored significantly higher in language, reading, and math than children with consistently low-quality child care and than those with high-quality infant care but low-quality care during the preschool period. They also achieved higher (by .16 *SD*) reading scores compared with children with low-quality infant care and high-quality care during the preschool period.

Since Model 5 (the multinomial propensity score weighted regression estimation) enabled between-contrast comparison, we summarized results from Model 5 in Figure 1. Figure 1 shows the means in standard deviations of the four 54-month outcomes, indexed by four bar groups for memory, language, letter–word, and math. Each bar group shows the standardized means for the four child care quality combinations (high-high, low-low, high-

low, and low-high). The height for the first bar of the memory bar group represents the standardized mean of memory scores of the high-high group. The heights of the other three bars in this memory group were calculated by subtracting mean differences in Model 5 from the standardized mean of the high-high group. For example, for the low-low group we subtracted .16, which was the standardized mean difference between the high-high group and the low-low group calculated by multinomial propensity score weighted regression, from .29 (last column in Table 2), the standardized mean of the high-high group, which gave us .13 as the estimated standardized mean of memory scores for the low-low group. Overall, the combination of high-quality care during both the infant–toddler and preschool periods was associated with the best child outcomes at 54 months; none of the other child care combinations had outcomes close to these.

Discussion

In this article, we analyzed data from the NICHD Study of Early Child Care to compare the cognitive, language, and preacademic outcomes of children with different combinations of child care quality during the infant–toddler and preschool periods. Detailed information about the families was used to implement propensity score matching to reduce selection bias. This was followed by propensity score weighted regression estimation for robustness checking. We found differentiated effects of quality of care in the infant–toddler period, in the preschool period, and in the combination of quality in these two periods on child outcomes. Results for each of our three hypotheses are discussed in turn.

Effects of Quality of Care in the Infant–Toddler Period

Testing Hypothesis 1, we found that the quality of child care in the infant–toddler period was positively and significantly related to two outcomes. At 24 months, high-quality care in the infant–toddler period was associated with higher cognitive development scores, as measured by the Bayley Mental Developmental Index. In addition, higher quality infant–toddler care was associated with better memory scores at 54 months for children in low-quality child care in the preschool period. This result was consistent with Huttenlocher's theory of brain development (Huttenlocher, 1979), which indicated that brain functions developed most rapidly during the infant–toddler period.

Effects of Quality of Care in the Preschool Period

Testing Hypothesis 2, we found that the quality of care in the preschool period was also related to child outcomes. Children who received high-quality child care in the preschool period obtained higher language, reading, and math scores at 54 months of age. Among students who received high-quality care in the preschool period, those who also received high-quality infant–toddler care scored better than those receiving low-quality infant–toddler care on reading (significant at the .05 level) and math (significant at the .10 level).

Joint Contributions of Infant–Toddler and Preschool Quality of Care

Unique in our investigation was its focus on the joint contribution of quality care in the infant–toddler and preschool periods (Hypothesis 3). Here our results suggest that the most robust differences were found between children who received high-quality care in both the

infant–toddler and preschool periods versus those who received such care in neither period. There was no clear winner when comparing 54-month outcomes for the high-low- versus low-high-quality child care patterns. Short-term memory outcomes at 54 months appear to be most sensitive to child care quality inputs during infancy–toddlerhood, whereas language, reading, and math outcomes at 54 months appear to be most sensitive to experiences in the preschool period.

Implications

Different patterns in memory versus in language, reading, and math were consistent with evidence for Huttenlocher's theory of brain development (Huttenlocher, 1979; Huttenlocher & Dabholkar, 1997), which asserts that cognitive functions (as well as other functions such as seeing and hearing) driven by synaptogenesis develop most rapidly during the first 3 years of life (although growth can continue well into adolescence). Following this line of argument, neurobiologists have relied on animal studies that have yielded strong evidence for such timing effects (Bornstein, 1989). However, animal studies may have limited generalizability to human development. Thus, the present article provided much needed evidence supporting the importance of timing (periods of differential growth and responsiveness) for human development.

In general, our results suggest that investing in high-quality infant–toddler care without subsequent high-quality care during the preschool period could be a productive strategy for cognitive domains such as memory but less productive for preacademic skills such as reading (letter–word identification) or math (applied problems). High-quality care during the preschool period without high-quality infant–toddler care was associated with increased kindergarten readiness, but not by as much as was achieved by consistent high-quality infant–toddler and preschool care (see the final column of Table 3). These findings provide further evidence that early education does not inoculate children from the impacts of subsequent experiences. Instead, as with education during the elementary years, experiencing high-quality education appears to translate into contemporaneous gains that may be maintained only with long-term (continued) exposure to high-quality education (Pianta, Barnett, Burchinal, & Thornburg, 2009).

Our findings support a strategy of distributing child care investment across early childhood periods, as opposed to front-loading investment on infant–toddler care or back-loading on preschool care. Of course, this recommendation needs to be considered within a benefit–cost framework. While expensive child care is not necessarily high-quality care, it is clear that high-quality care is expensive (Helburn, 1995). High-quality care occurs when caregivers provide frequent warm, stimulating, and responsive interactions with children, as well as clear intentional instruction. This is best achieved via low child/adult ratios and substantial investments in skilled caregivers (Pianta et al., 2009). If sufficient funds are available to provide high-quality care in one but not both of the time periods, then high-quality care during the preschool period may have an advantage, since it is less costly (because of higher child/teacher ratios for preschoolers than for infants) and yet is accompanied by relatively similar 54-month language, reading, and mathematics outcomes (compare high-low and low-high outcomes in Figure 1).

Limitations

Limitations of these policy suggestions include the following. First, the NICHD Study of Early Child Care is not a representative sample, and even with reduction in selection bias, the sample characteristics may limit generalization. Moreover, its cumulative response rate at the point of the 6-month interview was around 50%. The sample that used infant–toddler child care tended to be relatively advantaged economically and included disproportionately more white families than the U.S. sample in general. Second, our results are limited by observed variables, with omitted variable bias (i.e., selection bias) remaining linked to unmeasured variables or variables we could not include because they were assessed during the treatment period. Third, the impact of variation in quality may be truncated because we categorized quality of care so that we could conduct the propensity score matching. Furthermore, we could not include other child care characteristics when creating the propensity score groups, so we could not account for either type or quantity of child care in these analyses.

At least three types of studies might be undertaken to test the reliability of our conclusions. First, studies could be conducted using a more causal framework such as random assignment or instrumental variables to correct for unobserved selection bias. Second, efforts should be made to replicate these findings with measures of the quantity of high-quality care, as well as with study samples available in other data sets. In particular, low-income families constituted a relatively small share of the NICHD Study of Early Child Care sample, and teenage mothers and low birth weight babies were excluded altogether. Databases such as the Early Childhood Longitudinal Study—Birth Cohort better represent these important population subgroups. Future studies are greatly needed particularly for at-risk populations of children such as those attending Head Start or Early Head Start, since these groups of children have been found to benefit most from high-quality educational and child care experiences (Magnuson, Ruhm, Waldfogel, 2007). Third, future studies should include additional outcomes such as attention skills and socioemotional behaviors, including internalizing and externalizing problems and social skills. Attention skills have been found to be one of the strongest predictors for later achievement (Duncan et al., 2007), and socioemotional behaviors could also be linked to later academic achievement (Entwisle, Alexander, & Olson, 2005; NICHD ECCRN, 2004). These outcomes are important and may be positively affected by high-quality child care in the infant–toddler and/or preschool years.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Bayley, N. Bayley II scales of infant development. New York, NY: Psychological Corporation; 1993.

Bornstein MH. Sensitive periods in development: Structural characteristics and causal interpretations. Psychological Bulletin. 1989; 105:179–197. [PubMed: 2648441]

Brooks-Gunn J, McCarton CM, Casey PH, McCormick MC, Bauer CR, Tonascia J. Early intervention in low birth weight premature infants: Results through age 5 years from the Infant Health and Development Program. JAMA: Journal of the American Medical Association, 272. 1994; 16:1257–1262.

Bryson, A.; Dorsett, R.; Purdon, S. The use of propensity score matching in the evaluation of labour market policies (Working Paper No. 4). London, England: UK Department for Work and Pensions; 2002.

Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys. 2008; 22:31–72.

Camilli G, Vargas S, Ryan S, Barnett WS. Meta-analysis of the effects of early education interventions on cognitive and social development. Teachers College Record. 2010; 112:579–620.

Campbell FA, Pungello EP, Miller-Johnson S, Burchinal MR, Ramey C. The development of cognitive and academic abilities: Growth curves from an early intervention educational experiment. Developmental Psychology. 2001; 37:231–242. [PubMed: 11269391]

Clarke-Stewart KA, Vandell DL, Burchinal M, O'Brien M, McCartney K. Do regulable features of childcare homes affect children's development? Early Childhood Research Quarterly. 2002; 17:52–86.

Cook TD, Shadish WJ, Wong VC. Three conditions under which observational studies produce the same results as experiments. Journal of Policy Analysis and Management. 2008; 274:724–750.

Cunha, F.; Heckman, JJ.; Lochner, LJ.; Masterov, DV. Interpreting the evidence on life cycle skill formation. In: Hanushek, EA.; Welch, F., editors. Handbook of the economics of education. Amsterdam, the Netherlands: North-Holland; 2006. p. 697-812.

Cunha F, Heckman JJ, Schennach SM. Estimating the technology of cognitive and noncognitive skill formation. Econometrica. 2010; 78:883–931. [PubMed: 20563300]

Duncan GJ, Dowsett CJ, Claessens A, Magnuson K, Huston AC, Klebanov P, Brooks-Gunn J. School readiness and later achievement. Developmental Psychology. 2007; 43:1428–1446. [PubMed: 18020822]

Entwisle D, Alexander K, Olson L. First grade and educational attainment by age 22: A new story. American Journal of Sociology. 2005; 110:1458–1502.

Garces E, Thomas D, Currie J. Longer-term effects of Head Start. American Economic Review. 2002; 92:999–1012.

Gormley WT, Phillips D, Gayer T. Preschool programs can boost child outcomes. Scienceh. 2008 Jun 27.320:1723–1724.

Helburn, SW., editor. Cost, quality, and child outcomes in child care centers (Technical Report). Denver, CO: University of Colorado–Denver, Department of Economics, Center for Research in Economic and Social Policy; 1995.

Hock, E.; Gnezda, MT.; McBride, S. The measurement of maternal separation anxiety; Paper presented at the biennial meeting of the Society for Research in Child Development; Detroit, MI. 1983 Apr.

Howes C, Burchinal M, Pianta R, Bryant D, Early D, Clifford R, Barbarin O. Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. Early Childhood Research Quarterly. 2008; 23:27–50.

Huttenlocher PR. Synaptic density in human frontal cortex—Developmental changes and effects of aging. Brain Research. 1979; 163:195–205. [PubMed: 427544]

Huttenlocher PR, Dabholkar AS. Regional differences in synaptogenesis in human cerebral cortex. Journal of Comparative Neurology. 1997; 387:167–178. [PubMed: 9336221]

Imbens G. The role of the propensity score in estimating dose-response functions. Biometrika. 2000; 87:706–710.

Kang JDY, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating population means from incomplete data. Statistical Science. 2007; 22:523–539.

Lechner, M. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In: Lechner, M.; Pfeiffer, F., editors. Econometric evaluation of labour market policies. Heidelberg, Germany: Physica; 2001. p. 43-58.

Lohr, S. Sampling: Design and analysis. Pacific Grove, CA: Duxbury Press; 1999.

Love JM, Harrison L, Sagi-Schwartz A, van IJzendoorn MJ, Ross C, Ungerer JA, Chazan-Cohen R. Child care quality matters: How conclusions may vary with context. Child Development. 2003; 74:1021–1033. [PubMed: 12938696]

Love JM, Kisker EE, Ross C, Raikes H, Constantine J, Boller K, Vogel C. The effectiveness of Early Head Start for 3-yearold children and their parents: Lessons for policy and programs. Developmental Psychology. 2005; 41:885–901. [PubMed: 16351335]

Ludwig J, Miller DL. Does Head Start improve children's life chances? Evidence from a regression discontinuity design. Quarterly Journal of Economics. 2007; 122:159–208.

Magnuson KA, Ruhm C, Waldfogel J. Does prekindergarten improve school preparations and performance? Economics of Education Review. 2007; 26:33–51.

Mashburn AJ, Pianta RC, Hamre BK, Downer JT, Barbarin OA, Bryant D, Howes C. Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. Child Development. 2008; 79:732–749. [PubMed: 18489424]

McCartney K, Scarr S, Phillips D, Grajek S. Day care as intervention: Comparison of varying quality programs. Journal of Applied Developmental Psychology. 1985; 6:247–260.

McCormick MC, Brooks-Gunn J, Buka SL, Goldman J, Yu J, Salganik M, Casey PH. Early intervention in low birth weight premature infants: Results at 18 years of age for the Infant Health and Development Program. Pediatrics. 2006; 117:771–780. [PubMed: 16510657]

McFadden, D. Econometric analysis of qualitative response models. In: Griliches, Z.; Intriligator, M., editors. Handbook of econometrics. Vol. Vol. 2. Amsterdam, the Netherlands: North-Holland; 1984. p. 1396-1457.

Medoff-Cooper B, Carey WB, McDevitt S. The Early Infancy Temperament Questionnaire. Journal of Developmental and Behavioral Pediatrics. 1993; 14:230–235. [PubMed: 8408665]

National Center for Education Statistics. Digest of Education Statistics. 2010; 2009 Retrieved from http://nces.ed.gov/programs/digest/2010menu_tables.asp.

National Institute of Child Health and Human Development Early Child Care Research Network. The relation of child care to cognitive and language development. Child Development. 2000; 71:960–980. [PubMed: 11016559]

Nelson, CA.; Sheridan, M. Lessons from neuroscience research for understanding causal links between family and neighborhood characteristics and educational outcomes. In: Duncan, G.; Murnane, R., editors. Whither opportunity? Rising inequality and the uncertain life chances of low-income children. New York, NY: Russell Sage Foundation Press; 2011. p. 27-46.

NICHD Early Child Care Research Network. Trajectories of physical aggression from toddlerhood to middle childhood: Predictors, correlates, and outcomes. Monographs of the Society of Research in Child Development. 2004; 69(Whole No. 4, Serial No. 278):1–129.

NICHD Early Child Care Research Network. Child Care and Development: Results from the NICHD Study of Early Child Care and Youth Development. New York, NY: Guilford Press; 2005.

Pianta RC, Barnett WS, Burchinal MR, Thornburg KR. The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base, and what we need to know. Psychological Science in the Public Interest. 2009; 10:49–88.

Qi CH, Marley SC. Validity study of the Preschool Language Scale-4 with English-Speaking Hispanic and European American children in Head Start programs. Topics in Early Childhood Special Education. 2011; 31:89–98.

Radloff L. The CES-D scale: A self-report depression scale for research in the general population. Applied Psychological Measurement. 1977; 1:385–401.

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70:41–55.

Rubin DB. Using propensity scores to help design observational studies: Application to the tobacco litigation. Health Services & Outcomes Research Methodology. 2001; 2:169–188.

Rubin DB, Thomas N. Combining propensity score matching with additional adjustments for prognostic covariates. Journal of the American Statistical Association. 2000; 95:573–585.

Schaefer, ES.; Edgerton, M. Parent and child correlates of parental modernity. In: Sigel, IE., editor. Parental belief systems: The psychological consequences for children. Hillsdale, NJ: Erlbaum; 1985. p. 287-315.

Schafer JL, Kang J. Average causal effects from nonrandomized studies: A practical guide and simulated example. Psychological Methods. 2008; 13:279–313. [PubMed: 19071996]

Shonkoff, J.; Phillips, D., editors. From neurons to neighborhoods: The science of early childhood development. Washington, DC: National Academy Press; 2000.

Vandell DL, Belsky J, Burchinal M, Steinberg L, Vandergrift N. the NICHD Early Child Care Research Network. Do effects of early child care extend to age 15 years? Child Development. 2010; 81:737–756. [PubMed: 20573102]

Vandell, DL.; Wolfe, B. Child care quality: Does it matter and does it need to be improved? (Special Report No. 78). 2000. Retrieved from the Institute for Research on Poverty website: http://www.irp.wisc.edu/publications/sr/pdfs/sr78.pdf

Woodcock, RW.; Johnson, MB. Tests of achievement, WJ–R: Examiner's manual. Allen, TX: DLM Teaching Resources; 1990.

Zimmerman, IL.; Steiner, VG.; Pond, RE. Preschool Language Scale: Examiner's manual. San Antonio, TX: Psychological Corporation; 1979.
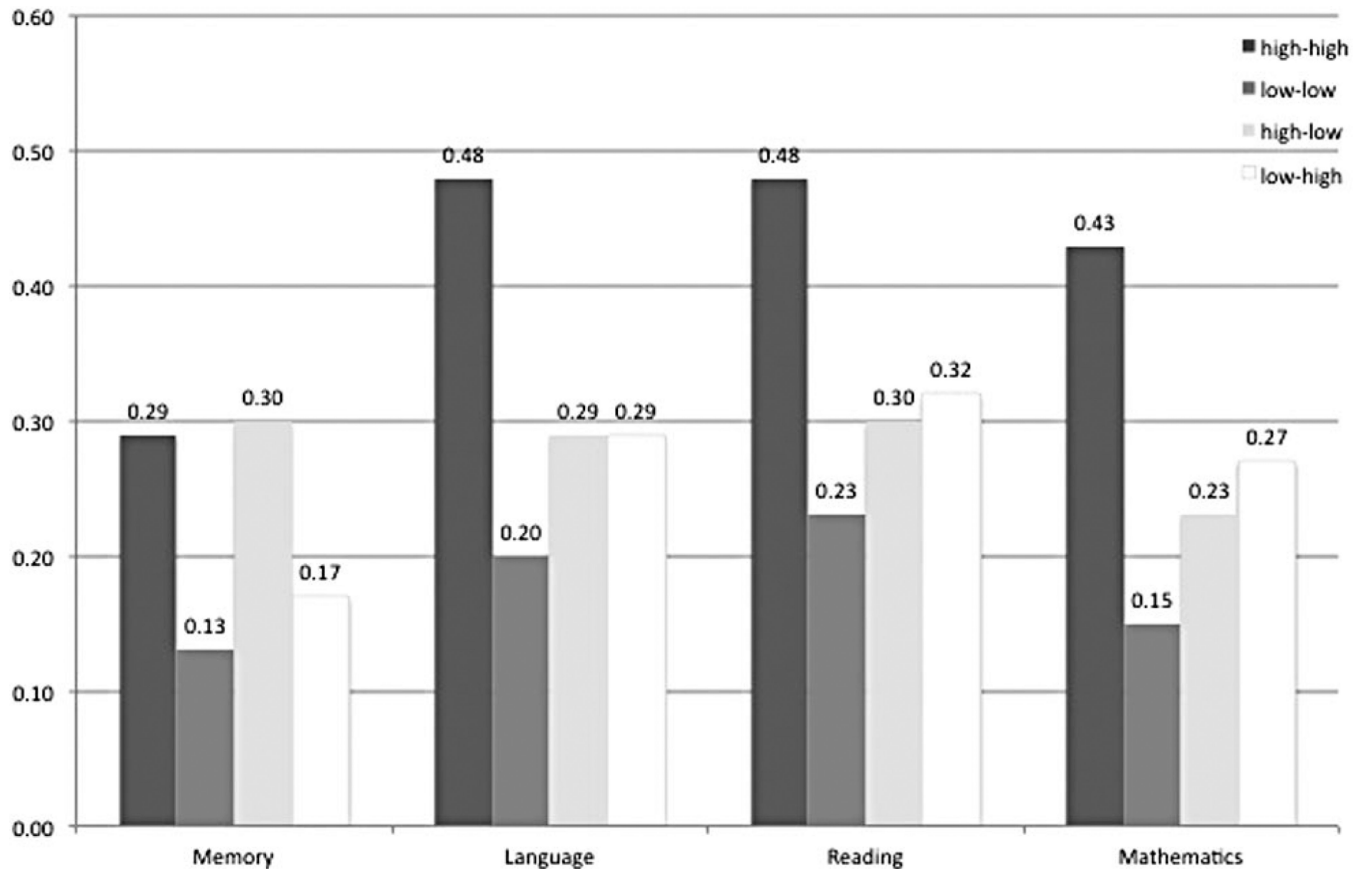
Figure 1.
Comparison among different combinations of child care quality during infant–toddler and preschool period. The height of the high-high group represents its standardized means. The heights of the other three groups were calculated by subtracting multinomial propensity score weighted regression estimation for differences (M5 in Table 4) from the standardized mean of the high-high group.

Table 1

Contrasts for Testing Hypotheses

Hypothesis	Targeted group	Comparison group	Time of outcome measurement
H1: High-quality infant–toddler care is associated with immediate improvement in cognitive scores at 24 months. However, without subsequent high-quality care for the preschool period, children with high-quality infant–toddler care will not have higher cognitive and achievement scores at 54 months of age than children with low-quality infant–toddler care followed by low-quality care for the preschool period.	Early high	Early low	24 months
	High-low	Low-low	54 months
H2: High-quality care during the preschool period is associated with immediate improvement in cognitive and academic scores at 54 months. In addition, the combination of high-quality infant–toddler and preschool care will produce higher cognitive and achievement scores at 54 months of age than the combination of high-quality care for the preschool period but low-quality infant–toddler care.	Late high	Late low	54 months
	High-high	Low-high	54 months
H3: Everything else the same, high-quality child care in both the infant–toddler and preschool stages is associated with the highest cognitive and academic scores at school entry compared with any other child care quality combination.	High-high	Low-low	54 months
	High-high	High-low	54 months
	High-high	Low-high	54 months

Note. H1–H3 = Hypotheses 1–3.

NIH-PA Author Manuscript

Table 2

Description of Sample and Analysis Variables

Variable	NICHD Study of Early Child Care	Focus on infant-toddler care		Focus on preschool care		Focus on both infant-toddler and preschool care			
		Early low	Early high	Late low	Late high	Low-low	High-low	Low-high	High-high
Child outcomes: M (SD)									
Cognitive development at 24 mo.	0.00 (1.00)	0.02 (0.94)	0.32 (0.92)	0.09 (0.98)	0.23 (0.92)	−0.04 (0.98)	0.34 (0.95)	0.15 (0.86)	0.39 (0.87)
Memory for sentences at 54 mo.	0.00 (1.00)			−0.05 (0.98)	0.27 (0.95)	−0.11 (0.96)	0.19 (0.94)	0.05 (0.96)	0.29 (0.97)
Language comprehension at 54 mo.	0.00 (1.00)			−0.14 (0.99)	0.36 (0.84)	−0.17 (0.97)	0.10 (0.92)	0.12 (0.95)	0.48 (0.82)
Letter–word identification at 54 mo.	0.00 (1.00)			−0.10 (0.97)	0.35 (0.95)	−0.17 (0.96)	0.10 (0.90)	0.14 (0.96)	0.48 (0.99)
Applied problems at 54 mo.	0.00 (1.00)			−0.11 (0.99)	0.30 (0.88)	−0.19 (0.94)	0.06 (0.97)	0.10 (0.96)	0.43 (0.89)
Family characteristics									
At 1 mo. after birth									
Maternal age (years): M (SD)	29.59 (4.77)	28.80 (5.25)	29.41 (4.91)	29.08 (5.15)	29.47 (5.29)	28.42 (5.04)	29.49 (4.69)	29.60 (5.37)	29.65 (4.99)
Maternal education (years): M (SD)	15.12 (2.28)	14.46 (2.16)	15.19 (2.53)	14.44 (2.45)	15.08 (2.28)	14.19 (2.09)	15.03 (2.61)	15.09 (2.15)	15.50 (2.34)
Maternal paid leave: %	67	60	65	59	49	68	65	57	68
Maternal child-rearing attitudes: M (SD)	70.62 (15.04)	73.83 (15.93)	70.73 (15.10)	74.42 (15.93)	70.05 (15.23)	74.91 (15.07)	73.00 (14.93)	70.01 (15.44)	68.18 (14.90)
Paternal education (years): M (SD)	15.06 (2.57)	14.37 (2.42)	15.07 (2.68)	14.40 (2.44)	15.32 (2.64)	14.05 (2.20)	14.73 (2.54)	15.15 (2.57)	15.49 (2.64)
Maternal separation anxiety: M (SD)	66.45 (12.39)	67.94 (13.01)	67.01 (12.33)			66.59 (12.66)	66.37 (12.77)	68.96 (12.76)	66.76 (11.78)
Mother has a job: %	81	74	81			82	85	71	82
Maternal depression: M (SD)	10.22 (8.19)	11.22 (8.59)	11.53 (8.78)			10.89 (8.61)	10.93 (9.69)	10.82 (7.89)	10.50 (8.36)
Partner in household: %	97	95	98			94	97	95	98
Family income-to-needs ratio: M (SD)	3.43 (2.80)	2.89 (2.58)	3.52 (2.69)			2.66 (2.24)	3.28 (2.33)	3.45 (3.14)	3.88 (3.02)
At 24 mo. of age									
Maternal separation anxiety: M (SD)	59.73 (12.42)			61.48 (13.00)	61.68 (12.92)				
Mother has a job: %	89					80	71		
Maternal depression: M (SD)	8.34 (7.38)			9.87 (9.14)	8.81 (7.81)				
Partner in household: %	93			88	93				
Family income-to-needs ratio: M (SD)	4.67 (3.38)			3.91 (3.15)	4.40 (3.18)				
Child characteristics									

At 1 mo. after birth

Variable	NICHD Study of Early Child Care	Focus on infant-toddler care		Focus on preschool care		Focus on both infant-toddler and preschool care			
		Early low	Early high	Late low	Late high	Low-low	High-low	Low-high	High-high
Gender (1 = male): %	48	50	49	48	48	51	43	44	51
Ethnicity: White/non-Hispanic: %	88	84	89	86	87	84	88	86	90
Difficult temperament: M (SD)	3.31 (0.63)	3.30 (0.62)	3.34 (0.67)	3.25 (0.65)	3.34 (0.64)	3.26 (0.60)	3.28 (0.69)	3.34 (0.66)	3.38 (0.62)
Child's birth order: M (SD)	1.66 (0.79)	1.78 (0.86)	1.63 (0.80)	1.82 (0.87)	1.77 (0.90)	1.78 (0.83)	1.70 (0.78)	1.71 (0.84)	1.54 (0.76)
Health of baby: M (SD)	3.72 (0.51)	3.68 (0.53)	3.73 (0.50)			3.65 (0.54)	3.70 (0.52)	3.75 (0.49)	3.75 (0.49)
At 24 mo. of age									
Health of baby: M (SD)	3.18 (0.71)			3.20 (0.68)	3.22 (0.72)				
Sample size	1,364	475	412	511	474	251	162	139	186
Proportion of sample	100%	35%	30%	37%	35%	18%	12%	10%	14%

Note. NICHD = National Institute of Child Health and Human Development.

Table 3

Standardized Mean Difference Before and After Matching

Variable	H1: Quality of infant care		H2: Quality of preschool care		H3: Quality of infant and preschool care		
	Early high vs. early low	High-low vs. low-low	Late high vs. late low	High-high vs. low-high	High-high vs. low-low	High-high vs. high-low	High-high vs. low-high
Age at outcome	24 months	54 months	54 months	54 months	54 months	54 months	54 months
Outcome and model							
Cognitive							
M1	0.36*** (0.07)						
M2	0.28*** (0.05)						
M3	0.28*** (0.05)						
Memory							
M1		0.34** (0.11)	0.15† (0.08)	0.17 (0.12)	0.38*** (0.10)	0.04 (0.11)	0.17 (0.12)
M2		0.25* (0.10)	0.06 (0.07)	0.15 (0.11)	0.14 (0.10)	−0.04 (0.11)	0.15 (0.11)
M3		0.20* (0.08)	0.05 (0.06)	0.10 (0.10)	0.16† (0.09)	−0.06 (0.09)	0.10 (0.10)
Language							
M1		0.34** (0.11)	0.34*** (0.08)	0.29** (0.10)	0.63*** (0.10)	0.29** (0.10)	0.29** (0.10)
M2		0.10 (0.10)	0.18** (0.06)	0.12 (0.09)	0.33*** (0.08)	0.19* (0.09)	0.12 (0.09)
M3		0.08 (0.07)	0.19*** (0.05)	0.10 (0.08)	0.35*** (0.08)	0.21** (0.08)	0.10 (0.07)
Reading							
M1		0.31** (0.11)	0.32*** (0.08)	0.29* (0.12)	0.61*** (0.11)	0.30** (0.11)	0.29* (0.12)
M2		0.11 (0.11)	0.26*** (0.06)	0.24* (0.11)	0.18† (0.09)	0.17 (0.11)	0.24* (0.11)
M3		0.09 (0.10)	0.28*** (0.06)	0.20* (0.09)	0.17* (0.08)	0.17† (0.10)	0.20* (0.09)
Math							
M1		0.29** (0.11)	0.31*** (0.08)	0.23* (0.11)	0.56*** (0.10)	0.27** (0.10)	0.23* (0.11)
M2		0.19† (0.11)	0.23*** (0.07)	0.19* (0.10)	0.27** (0.08)	0.15 (0.09)	0.19* (0.10)
M3		0.17† (0.09)	0.22*** (0.05)	0.16† (0.09)	0.28*** (0.08)	0.15† (0.08)	0.16† (0.09)

Note. Standard errors are given in parentheses. H1–H3 = Hypotheses 1–3; M1 = Model 1 (standardized mean difference before matching); M2 = Model 2 (standardized mean difference after pairwise matching); M3 = Model 3 (pairwise propensity score matched sample including covariates in linear regression).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$^{†}p < .1.$

$^{*}p < .05.$

$^{**}p < .01.$

$^{***}p < .001.$

Table 4

Robustness Check: Pairwise Propensity Score Weighted Regressions and Multinomial Propensity Score Weighted Regressions

Variable	H1: Quality of infant care		H2: Quality of preschool care		H3: Quality of infant and preschool care		
	Early high vs. early low	High-low vs. low-low	Late high vs. late low	High-high vs. low-high	High-high vs. low-low	High-high vs. high-low	High-high vs. low-high
Age at outcome	24 months	54 months	54 months	54 months	54 months	54 months	54 months
Outcome and model							
Cognitive							
M3	0.28*** (0.05)						
M4	0.23*** (0.07)						
M5							
Memory							
M3		0.20* (0.08)	0.04 (0.06)	0.10 (0.10)	0.16† (0.09)	−0.06 (0.09)	0.10 (0.10)
M4		0.21* (0.11)	0.04 (0.08)	0.13 (0.11)	0.15 (0.10)	−0.07 (0.10)	0.13 (0.11)
M5		0.17* (0.09)		0.12 (0.12)	0.16 (0.11)	−0.01 (0.11)	0.12 (0.12)
Language							
M3		0.08 (0.07)	0.18*** (0.05)	0.10 (0.08)	0.35*** (0.08)	0.21** (0.08)	0.10 (0.07)
M4		0.12 (0.10)	0.18** (0.06)	0.14 (0.09)	0.29*** (0.09)	0.15* (0.08)	0.14 (0.10)
M5		0.09 (0.10)		0.19 (0.11)	0.28** (0.09)	0.19* (0.10)	0.19 (0.11)
Reading							
M3		0.09 (0.10)	0.27*** (0.06)	0.20* (0.09)	0.17* (0.08)	0.17† (0.10)	0.20* (0.09)
M4		0.18 (0.10)	0.16* (0.08)	0.16* (0.08)	0.25* (0.11)	0.17 (0.11)	0.16* (0.08)
M5		0.07 (0.11)		0.16† (0.10)	0.25* (0.11)	0.18† (0.11)	0.16† (0.10)
Math							
M3		0.17† (0.09)	0.22*** (0.05)	0.16† (0.09)	0.28*** (0.08)	0.15† (0.08)	0.16† (0.09)
M4		0.12 (0.11)	0.15* (0.07)	0.14 (0.10)	0.26** (0.10)	0.15 (0.10)	0.14 (0.10)
M5		0.08 (0.11)		0.16 (0.11)	0.28** (0.10)	0.20† (0.11)	0.16 (0.11)

Note. Standard errors are given in parentheses. H1–H3 = Hypotheses 1–3; M3 = Model 3 (pairwise propensity score matched sample in linear regression); M4 = Model 4 (pairwise propensity score weighted regression estimation); M5 = Model 5 (multinomial propensity score weighted regression estimation).

$^\dagger p < .1.$

$^* p < .05.$

$^{**} p < .01.$

$^{***} p < .001.$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript