Child Development, September/October 2003, Volume 74, Number 5, Pages 1454–1475

Modeling the Impacts of Child Care Quality on Children's Preschool Cognitive Development

National Institute of Child Health and Human Development Early Child Care Research Network and Greg J. Duncan

The National Institute of Child Health and Human Development (NICHD) Study of Early Child Care compared 3 statistical methods that adjust for family selection bias to test whether child care type and quality relate to cognitive and academic skills. The methods included: multiple regression models of 54-month outcomes, change models of differences in 24- and 54-month outcomes, and residualized change models of 54-month outcomes adjusting for the 24-month outcome. The study was unable to establish empirically which model best adjusted for selection and omitted-variable bias. Nevertheless, results suggested that child care quality predicted cognitive outcomes at 54 months, with effect sizes of .04 to .08 for both infant and preschool ages. Center care during preschool years also predicted outcomes across all models.

Knowing whether and how much the quality of child care experienced by infants, toddlers, and preschool children affects their cognitive development is crucial for formulating policies regarding regulation of child care settings, family leave, public funding of child care through tax credits or subsidies for low-income families, and infant-care exemptions from welfare-to-work requirements. Research on this issue by developmental psychologists and economists has led to varying conclusions about the magnitude of quality effects, in part because investigators use different conceptual and analytic models. Our purpose is to compare various analytic approaches to the quality-outcome connections using data from the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care. The authors are an interdisciplinary team consisting of an economist and Study of Early Child Care developmental psychologists.

The study has followed a large and diverse sample of children and collected longitudinal data. The type and quality of child care settings experienced by the children between ages 6 and 54 months has been compared with their cognitive development at 54 months. We exploit three special features of these data. First, the study assesses quality using an intensive observational measure of caregiver–child interactions that is specific to each study child (rather than assessed at the arrangement or classroom level). Second, data are gathered on an unusually rich set of maternal and child characteristics that are likely to capture many plausible sources of selection bias. And third, the availability

This work is a collaborative work between the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Greg J. Duncan, Institute for Policy Research, Northwestern University. This study is directed by a Steering Committee and supported by NICHD through a cooperative agreement (U10), which calls for scientific collaboration between the grantees and the NICHD staff. Participating investigators, listed in alphabetical order, are: Virginia Allhusen, University of California, Irvine; Jay Belsky, Birkbeck University of London; Cathryn L. Booth, University of Washington; Robert Bradley, University of Arkansas, Little Rock; Celia A. Brownell, University of Pittsburgh; Margaret Burchinal, University of North Carolina, Chapel Hill; Bettye Caldwell, University of Arkansas; Susan B. Campbell, University of Pittsburgh; K. Alison Clarke-Stewart, University of California, Irvine; Martha Cox, University of North Carolina, Chapel Hill; Sarah L. Friedman, NICHD; Kathyrn Hirsh-Pasek, Temple University; Aletha Huston, University of Texas, Austin; Deborah J. Johnson, Michigan State University; Jean F. Kelly, University of Washington; Kathleen McCartney, Harvard University; Nancy Marshall, Wellesley College; Marion O'Brien, University of Kansas; Margaret Tresch Owen, University of Texas, Dallas; Chris Payne, University of North Carolina, Greensboro; Deborah Phillips, Georgetown University; Robert Pianta, University of Virginia; Suzanne M. Randolph, University of Maryland, College Park; Wendy Robeson, Wellesley College; Susan Spieker, University of Washington; Deborah Lowe Vandell, University of Wisconsin, Madison; Marsha Weinraub, Temple University. We express our appreciation to the study coordinators at each site who supervised the data collection, the research assistants who collected the data, and especially the families and child care providers who welcomed us into their homes and workplaces and cooperated willingly with our repeated requests for information.

An earlier version of this paper was presented at the biennial meetings of the Society for Research on Child Development, Minneapolis, MN, April 2001. We would like to thank Christina Gibson, Mark Appelbaum and members of the Brookings Roundtable on Children for comments on a prior draft, but in no way hold them responsible for remaining errors.

Correspondence concerning this article should be addressed to NICHD Early Child Care Research Network, OEP, Office of the Director, NICHD, 6100 Executive Blvd., 4A01, Rockville, Maryland 20852. Regression tables with coefficients for all covariates and child care variables are available on request.

 0009-3920/2003/7405-0014

of measures of children's cognitive ability, family environment, and child care quality between ages 6 and 54 months enables us to estimate change models of child care quality. We begin with a brief review of the literature and then outline our analytic approach for estimating child care impacts.

Background

Poor-quality child care is of concern because infants and preschoolers need responsive and stimulating interactions with adults—parents and other caregivers—to enhance social, cognitive, and language development in early childhood (Bronfenbrenner & Morris, 1998; Sameroff, 1983). Efforts to examine the contribution of child care quality to children's outcomes derive from a general interest in environmental influences on development. Drawing from the literatures on both home and child care environments, the proximal processes that influence cognitive outcomes involve interactions with adults characterized by ample talking, turn taking during play, contingent and focused attention on the child, and rich opportunities for exploration (Bradley et al., 1989; Hart & Risley, 1995; Hoff-Ginsburg, 1991; Katz & Snow, 2000; Tomasello & Farrar, 1986). The child care literature, in particular, has highlighted the importance of environments rich in spoken and written language experiences, where children are engaged in give-and-take conversations, afforded abundant opportunities to explore their environments, and provided with constructive models of adult language, reading, and learning toward which to aspire (Dickinson & Smith, 2001; McCartney, 1984; NICHD Early Child Care Research Network, 2000b). These are the primary mechanisms through which young children learn vocabulary and other emergent literacy skills, acquire knowledge about the world, sustain their eagerness to learn, and, most generally, become ready for school (Howes, 2000; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991).

Many psychologists and educators have concluded that the quality of children's nonparental care affects cognitive and language skills based on evidence from experimental and observational studies (cf. Clarke-Stewart, 1989; Lamb, 1998). The random-assignment experimental designs used in evaluations of center-based care (e.g., Barnett, 1995; Campbell & Ramey, 1994; Infant Health and Development Project, 1990; Schweinhart, Weikart, & Larner, 1986) provide definitive evidence that high-quality child care can enhance the cognitive and language development of children from impoverished families. For the most intensive programs, long-term effects were observed into adulthood in terms of higher IQ, educational level, and employment rates, and lower rates of contact with the law (Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Schweinhart et al., 1986; Yoshikawa, 1995).

Psychologists have also found that child care quality is associated with cognitive and language skills in nonexperimental studies (Vandell & Wolfe, 2000). Using standard measures of child care quality, researchers found that child care quality was related to language and cognitive development, even after they controlled for family selection factors such as socioeconomic status (SES), maternal education, parenting, or family structure in other large multisite studies—the Chicago Study (Clarke-Stewart, Gruber, & Fitzgerald, 1994); the Child Care and Family Study (Kontos, Howes, Shinn, & Galinsky, 1995); and the Cost, Quality, and Outcomes Study (Peisner-Feinberg & Burchinal, 1997)—and in smaller single-site studies (Burchinal, Roberts, Nabors, & Bryant, 1996; Burchinal, Roberts, et al., 2000; Dunn, 1993; McCartney, 1984; Phillips, McCartney, & Scarr, 1987; Schliecker, White & Jacobs, 1991). Published analyses of the current study—the NICHD Study of Early Child Care—demonstrate that observed child care quality measured from age 6 months onward was positively related to cognitive and language development at ages 2, 3, and 4½ years, even in the presence of substantial controls for selection factors (NICHD Early Child Care Research Network, 2000b, 2002a, 2002b).

In most of these observational studies, quality of child care was modestly related to children's development. Quality of care often accounted for less than 5% of the variance in children's developmental outcomes in analyses that adjusted for family selection factors. Child care quality was more strongly related to outcomes for children from low-income families than for middle-class children in many studies (Baydar & Brooks-Gunn, 1991; Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; Bryant, Burchinal, Lau, & Sparling, 1994; Caughy, DiPietro, & Strobino, 1994; Peisner-Feinberg & Burchinal, 1997; Vandell & Corasaniti, 1990), but not all (NICHD Early Child Care Research Network, 2000b). However, none of these studies implemented the most rigorous analytic methods designed to address biased estimation of quality effects resulting from parental selection of child care.

Finally, regulatable aspects of child care such as caregiver education and training, adult–child ratios, and group size have also been related to cognitive development in nonexperimental studies. Findings from the National Child Care Staffing Study (Howes, Phillips, & Whitebook, 1992), the Cost, Quality, and Outcomes Study (Howes, 1997), the Carolina Otitis Media Project (Burchinal, Roberts, et al., 2000) as well as the NICHD Study of Early Child Care (NICHD Early Child Care Research Network, 1999a, 2002b) show that children attending programs in which caregivers had more education and training, and in which child–staff ratios were smaller, performed better across a range of cognitive and social measures.

Nonetheless, because of potential selection effects in observational studies, not all psychologists and economists are convinced that child care quality is related to child outcomes (Blau, 1999; Scarr, 1998). Both developmentalists and economists worry that the quality impacts estimated in the nonexperimental studies can misstate true causal impacts (Blau, 1999; Burchinal, Ramey, Reid, & Jaccard, 1995; McCartney, 1984), but they have used different statistical methods to address the bias problem. These concerns focus on bias from failure to adjust for characteristics of families or children that are associated with both quality of care and children's cognitive development. Family selection is an important issue to consider in child care research because the type and quality of child care are related to demographic and family characteristics that predict child outcomes (cf. Lamb, 1998).

The selection variables included in this study are based on a growing literature linking family characteristics to use of child care that varies by type and quality. Specifically, children from more advantaged families, in which parents have more education and income, larger vocabularies, less authoritarian childrearing beliefs, more stimulating home environments, and are more responsive in interactions with their children are more likely to experience center-based child care as well as higher quality home-based care (NICHD Early Child Care Research Network, 1997, 1998; Phillips et al., 1987; Pungello & Kurtz-Costes, 1999). Family ethnicity and family structure have also been found to affect patterns of enrollment in child care (Capizzano, Adams, & Sonenstein, 2000; Ehrle, Tout, & Adams, 2000; Pungello & Kurtz-Costes, 1999). More psychological dimensions of the family environment, including maternal mental health (i.e., depression) and attitudes about work and parenting, are also associated with differing types and qualities of child care (NICHD Early Child Care Research Network, 1997). Finally, parents appear to make decisions about child care based on child characteristics such as temperament (Fox, Henderson, Rubin, Calkins, & Schmidt, 2001).

On the other hand, the associations between family and child care characteristics, especially quality, often are modest, and some studies question the degree of parental choice in the type and quality of care as well as parents' ability to judge child care quality (Helburn, 1995). Parents almost uniformly report their child is receiving high-quality care. However, in one of the few studies directly comparing parental and observational ratings, parental ratings were only modestly related to observed quality, despite very strong endorsement of the dimensions of care measured by the observational scale and close agreement between the parent and observational rating forms (Cryer & Burchinal, 1997). Two large child care studies reported only modest (i.e., $r = .10 – .25$) correlations between family characteristics and child care quality measures (NICHD Early Child Care Research Network, 1996; Peisner-Feinberg & Burchinal, 1997). Because it is logically impossible to measure all possible selection factors, our empirical efforts in this study are focused on contrasting methods for assessing bias from both measured and unmeasured characteristics of families and children.

Analytic Approach

Our goal is to estimate the causal impact of child care quality on children's cognitive development. By this we have in mind what Raudenbush (2001) and others have termed the "Rubin-Rosenbaum-Holland" theory of causation. When applied to our situation, the causal impact of an increment in child care quality for a given child is the difference between two potential outcomes: a child's cognitive development in the presence of an increment to existing child care quality and that same child's cognitive development in the absence of such an increment. The policy experiment we seek to address is the extent to which children's cognitive development would improve with increments to child care quality as it currently exists, but no other concurrent changes in the circumstances of the child or family.

Note that this definition of causation is child specific, can vary from child to child, is never observed directly, and amounts to a missing-data problem. An experiment in which children are allocated randomly to either quality increments or no such increments ensures that the unobserved counterfactual is missing at random. However, experiments in which children are randomly assigned to child care programs of varying typical quality (rather than planned intervention programs) have not been conducted. Statistical methods developed to address issues of selection bias have been applied to observation studies of child care quality and outcomes to account for the fact that family selection of child care quality is not a random process.

In the context of nonexperimental data, the task of causal inference requires a carefully specified model of the developmental process under investigation that includes all relevant predictors and correctly specifies all main effects and interactions among predictors. We consider three analytic approaches to estimating the relations between child care quality and a child's cognitive development. One is a level model relating age 54-month cognitive development to a child's past history of child care quality and characteristics of the family, child, and child care experiences. Change and residualized change analyses relate the change in a child's development between two distinct points in early childhood to the family and child care experiences between the 24- and 54-month assessments. The level and residualized change models have been widely used by psychologists to account for selection bias, whereas economists have relied on change models as well. Growth curve methods and other multivariate methods are widely used to describe change over time but have not been used as often to adjust for selection bias. We restricted our focus to the level, change, and residualized change models because the comparison of models becomes even more complex with multivariate approaches than with the selected cross-sectional approaches. Within each model, we examine varying levels of adjustment for observed and unobserved selection factors.

The general cognitive model assumed in all analyses in this paper represents cognitive development at the end of the preschool period as the product of child, parent, family, and child care characteristics. We assume that these characteristics can vary over time and divide the early childhood period into the infant and preschool periods. A simple version of this model, proposed by Blau (1999), assumes that child *i*'s cognitive development (Y_{it}) at the point of school entry (t = age 54 months in our data) is an additive function of the child's history of the blocks of variables representing the quality and quantity of home (HOME) and child care (CARE), plus time-invariant child (CHILD) and maternal and family (FAM) characteristics. In most of our empirical work we are able to distinguish between inputs before and after the child's second birthday, and we do so in our model with early (E) and late (L) notation:

$$\begin{aligned} Y_{it} = a_1 &+ \beta_{1\mathrm{E}}\mathrm{CARE}_{i\mathrm{E}} + \beta_{2\mathrm{E}}\mathrm{HOME}_{i\mathrm{E}} \\ &+ \beta_{1\mathrm{L}}\mathrm{CARE}_{i\mathrm{L}} + \beta_{2\mathrm{L}}\mathrm{HOME}_{i\mathrm{L}} + \beta_3\mathrm{CHILD}_i \\ &+ \beta_4\mathrm{FAM}_i + e_{it}. \end{aligned} \tag{1}$$

Our interest is in estimating the impact of the quality of both early ($\beta_{1\mathrm{E}}$) and later ($\beta_{1\mathrm{L}}$) child care on cognitive development. Threats to unbiased estimation of these impacts include error in measurement of predictors and unmeasured characteristics of the child, mother or family environment—elements of CHILD_i and FAM_i in Equation 1. The unmeasured characteristics will bias estimation to the extent that unmeasured variables are correlated with both choice of child care quality and children's cognitive development, and are independent of the other covariates included in the analysis model. Positive bias would occur if parents make sacrifices to obtain quality child care for their children and promote their children's development in other, unmeasured ways, resulting in overestimation of the quality effect. Downward bias may result if parents take actions on behalf of their children because of their perceptions of problems that are not well measured by the covariates. For example, a difficult-to-measure developmental delay or behavioral problem in early childhood might motivate a parent to seek unusually high-quality care to address the problem. Failure to adjust for child characteristics before entry into child care in this case will underestimate the estimated impact of quality on child outcomes.

One approach to the bias problem, used in earlier analyses of these data (e.g., NICHD Early Child Care Research Network, 2002a) is to estimate multiple regression models relating final level of cognitive ability to the child's history of child care and other inputs, controlling for a number of selection factors and taking care to include only selection factors that are unlikely to have been influenced by child care arrangements. We term these approaches to unbiased estimation of Equation 1 as level models because we are relating the final level of preschool cognitive ability to the history of child care quality and other inputs.

Our data provide us with an unusually extensive set of selection-factor variables, most of which are measured before the child's entry into child care settings, including: demographic characteristics, child temperament, parenting, maternal intelligence, maternal personality, maternal depressive symptoms, childrearing attitudes, and separation anxiety. We estimate models that include conventional measures of SES and progressively more of these typically unmeasured child care selection factors to gauge the extent to which quality impacts vary with the extent of both conventional and more extensive controls for selection factors (Altonji, Elder, & Taber, 2002). Altonji et al. (2002) argue that if one assumes that observable controls are a random subset of all possible controls, the changes in the key quality coefficients across models that add more observed covariates are informative of the possible bias that remains when all observables have been included. Small coefficient changes suggest little additional bias from unobservables; larger coefficient changes suggest more bias is possible.

We also attempt to control for unobservable sources of bias by estimating change models, where change in cognitive ability spans from ages 24 to 54 months. These models are based on the assumption that unobserved variables have similar impacts on both early and later outcomes, and error in assessment at both ages is random. To motivate our change model, suppose that an analogous relationship to Equation 1 describes a child's cognitive development at age 24 months (denoted by s) as an additive function of his or her history (from birth to time s) of the HOME and CARE inputs, plus CHILD and FAM influences. In this case there are only early (E) inputs:

$$Y_{is} = a_2 + \beta_{1E}\mathrm{CARE}_{iE} + \beta_{2E}\mathrm{HOME}_{iE} + \beta_3\mathrm{CHILD}_i + \beta_4\mathrm{FAM}_i + e_{is}. \quad (2)$$

Note that this equation includes many of the same parameters (β_{1E}, β_{2E}, β_3, and β_4) as the level model, Equation 1. This reflects the strong assumption that the impacts of increments to $\mathrm{CARE_E}$, $\mathrm{HOME_E}$, CHILD, and FAM on cognitive development are identical, conditional on covariates, at ages s (24 months in our data) and t (54 months). This assumption is dubious in light of the possible changing nature of the impacts of family and child characteristics and our differential ability to measure cognitive development at those two points. We ignore such problems for the time being but return to them later.

A simple difference model of Equations 1 and 2, using Δ to denote the $t-s$ difference, is:

$$\Delta Y_i = \Delta a + \beta_{1L}\mathrm{CARE}_{iL} + \beta_{2L}\mathrm{HOME}_{iL} + \Delta ei. \quad (3)$$

The obvious advantage of Equation 3 over Equations 2 or 1 is that the biases associated with unmeasured child and maternal or family characteristics have been eliminated by the differencing. A disadvantage is that Equation 3 differences out the quantity and quality of early child care and home environments and thus provides no way of estimating the developmental impacts of these factors.

It is crucial to note a common point of confusion over the change model, Equation 3: Even though the dependent variable is a change score, the independent variables are the average levels of child care quality and other inputs between the two measurement points. Furthermore, the key β_{1L} parameter in change Equation 3 corresponds precisely to the β_{1L} level parameter in Equation 1. Thus, the differencing in Equation 3 is merely a method to secure less biased estimates of the impacts of CARE_{iL} on the final preschool level of cognitive development.

If the β_{1E}, β_{2E}, β_3, or β_4 parameters do change over time and thus differ between Equations 1 and 2, then, continuing to use Δ to denote change and applying some simple algebra, the difference between Equations 1 and 2 becomes:

$$\begin{aligned}\Delta Yi = {} & \Delta a + \beta_{1L}\mathrm{CARE}_{iL} \\ & + \beta_{2L}\mathrm{HOME}_{iL} + \Delta\beta_{1E}\mathrm{CARE}_{iE} \\ & + \Delta\beta_{2E}\mathrm{HOME}_{iE} + \Delta\beta_3\mathrm{CHILD}_i \\ & + \Delta\beta_4\mathrm{FAM}_i + \Delta e_i.\end{aligned} \quad (4)$$

In this case, the role of early inputs and of the time-invariant child and family effects are not differenced out of the change equation because these factors are assumed to have a differential impact on cognitive development at ages 2 and 5. Equation 4 thus provides a rationale for including early inputs and parent and child characteristics in the change equation, but it also suggests that omitted elements of CARE_{iE}, HOME_{iE}, CHILD_i, and FAM_i may bias impacts of CARE_{iL} estimated in Equation 4.

Note, however, that the conditions for omitted-variable bias are different in the change model, Equation 4, compared with the level model, Equation 1. In Equation 1, bias arises when elements of CHILD_i or FAM_i have significant impacts on cognitive development and are correlated with child care quality. In Equation 4, bias arises when unmeasured elements of CHILD_i or FAM_i have significantly different impacts on development between the early and late periods and are correlated with child care quality. Note also that the coefficients on CARE_{iE}, HOME_{iE}, CHILD_i, and FAM_i in Equation 4 have a very different interpretation from their corresponding coefficients in either Equation 1 or 2: They represent the change in the impact of these measures on development at the two measurement points.

Equations 1 through 4 fail to account for the possibility that the impacts of quality may be nonlinear and that quality and quantity of child care may interact, or depend on the gender or early cognitive ability of the child, or on the education level of the mother. All of these possibilities are explored in our empirical work.

Psychologists and educators are reluctant to rely on simple change scores over level scores because of their greater measurement error (lower reliability). Typically, change scores are substantially less reliable than the original two scores when their true scores are moderately to highly correlated (Cronbach & Furby, 1970). One consequence of analyzing a less reliable outcome measure is increased standard errors for parameter estimates in a change equation such as Equation 3 as compared with level Equations 1 or 2.

Measurement error in dependent variables may or may not bias regression coefficients. If measurement error is uncorrelated with the predictors and the error term, the change model parameter estimates will be unbiased (Allison, 1990; Gujarati, 1995). However, bias may result if the measurement error in the change model is correlated with the true levels of the dependent and independent variables at Time 1 or 2 (Bound, Brown, Duncan, & Rodgers, 1994). The most likely causes of such bias are failure to include relevant interactions, differences in the impact of the omitted variable at the two time points, or use of assessment tools in which error is related to ability.

Any repeated-measures analysis is prone to bias when the ability to measure the outcome or the metric of the measure of that outcome varies over time (Bryk & Weisburg, 1977; Labouvie, 1980). This is of special concern in our case because omitted variables are likely to be more strongly associated with well-measured cognitive development at 54 months than with poorly measured cognitive development at 24 months. We are especially concerned because infant cognitive assessments provide more reliable assessment of children with lower than higher scores (McCall, 1977). Prediction of subsequent individual differences in cognitive skills is much better during the preschool years than during infancy (McCall, 1977). We examine our data for evidence of such bias for the level and change models.

The residualized change model (Cronbach & Furby, 1970; Diggle, Liang, & Zeger, 1994; Labouvie, 1980) is an alternative to a simple change model. In it, the later outcome assessment is used as dependent variable and the early assessment is used as a right-hand variable. As such, it is a variant of change Model 3, in which the implicit coefficient on the time s (for this study, 24 months) outcome is an estimated parameter rather than, as is the case for change Equation 3, being effectively fixed at 1.0. The residualized change version of Equation 3 is:

$$Y_{it} = a_3 + \delta_1 Y_{is} + \delta_2 \mathrm{CARE}_{i\mathrm{L}} + \delta_3 \mathrm{HOME}_{i\mathrm{L}} + \varepsilon_{it}. \quad (5)$$

The residualized change version of Equation 4 adds $\mathrm{CARE}_{i\mathrm{E}}$, $\mathrm{HOME}_{i\mathrm{E}}$, CHILD_i, and FAM_i to this equation.

This approach can provide considerably more power to detect associations when outcomes are highly correlated (Cronbach & Furby, 1970) and can be used when the earlier and later assessment of the outcome are not measured identically over time. On the other hand, including initial level as a predictor fails to take into account random variability in initial scores and therefore almost certainly builds in a biasing correlation between it and the error term of Equation 5. Furthermore, there is no longer a direct relationship between an explicit model of development and the parameters estimated in Equation 5, rendering the interpretation of δ_2 problematic. Finally, this model will result in biased estimation when the unobserved variables have differential impact on the early and late cognitive development.

In summary, there are a variety of analytic methods for addressing selection bias. The level model can adjust for observed but not unobserved selected factors. The simple change model adjusts for observed and unobserved selection factors but has less power and imposes additional strong assumptions about the data than the level model. The residualized change model also adjusts for observed and unobserved selection factors. It should provide more power and bias than the change model but less power and bias than the level model. Results emerging from the level, change, and residualized change models are contrasted using data from the NICHD Study of Early Child Care.

Method

Participants

Our data come from the NICHD Study of Early Child Care, which recruited mothers from hospitals near the following locations throughout 1991: Little Rock, Arkansas; Irvine, California; Lawrence, Kansas; Boston, Massachusetts; Philadelphia, Pennsylvania; Pittsburgh, Pennsylvania; Charlottesville, Virginia; Morganton, North Carolina; Seattle, Washington; and Madison, Wisconsin. The sample plan was not intended to provide a representative national sample but was designed to represent healthy births to nonteen parents at the selected hospitals. Potential participants were selected from among 8,986 mothers giving birth during selected 24–hour sampling periods.

The sample of 8,986 mothers was reduced to 5,416 mothers eligible for a phone call 2 weeks after the birth owing to both unplanned attrition (438 cases; mostly refusals) and planned sample exclusions (3,142 cases; mother $<$ 18 years old, multiple births, mother not fluent in English, family expects to move, medical complications, baby being put up for adoption, family lives too far away, family in another study, family lives in an unsafe neighborhood).

A conditional subsampling plan was next imposed to ensure that single-parent, low-maternal education, and minority distributional targets were met while continuing random selection of cases. Altogether, 3,015 families were targeted for recruitment. Unplanned (1,153 cases; refusals and lack of success with contacts at three different times of the day) and planned (151 cases; baby in hospital more than 7 days, planning to move within 3 years, 185 cases not contacted because enrollment quota was achieved before that family's name appeared on the contact sheets) reasons further reduced the sample from 3,015 screened mothers to the 1,364 recruited mothers who provided information at the 1-month interview. Further attrition between the 1-month and 6-month interviews reduced the 1,364 enrollments by 37 cases, for a cumulative response rate of 52.5%. Modest attrition reduced the sample of families providing information for the 15-, 24-, and 36-month and subsequent interviews. Thus, because of the attrition rate and the inclusion of the 10 sites selected nonrandomly, the Study of Early Child Care sample cannot be regarded as statistically representative of any a priori–defined population. Nevertheless, the sample is large and economically, geographically, and ethnically diverse, especially for an observational child care study.

Procedures

Child outcomes. Our outcome measure at 15 and 24 months is the Bayley Mental Developmental Index, which is normed to have a mean of 100 and standard deviation of 15. The Bayley test (Bayley, 1969) used at 15 months was based on 1969 norms, whereas the Bayley test used at 24 months was based on a 1993 revision of the test (Bayley, 1993). At 54 months we created cognitive and achievement composite scores. The cognitive score was computed as the mean of four scale scores (alpha = .83; $.46 < r < .70$): the Woodcock–Johnson Picture Vocabulary and Memory for Sentences tests and the Preschool Language Scale Expressive and Receptive tests. The achievement score was computed as the mean of three scale scores (alpha = .72; $.36 < r < .58$): Woodcock–Johnson Applied Problems (mathematical skill), Letter-Word Identification (reading skill), and Incomplete Words Scales (phonological knowledge). These scale scores are normed to have a mean of 100 and a standard deviation of 15. The 54-month cognitive composite score technically is not an intelligence score, but both language and memory are specific cognitive processes that are so highly correlated with IQ scores that they can be used to approximate cognitive scores (Neisser et al., 1996).

Child care type and quality. In all cases, our child care measures include: average quality, mean hours of care per week, proportion of occasions in center care at least 10 hr per week, whether exclusive mother care, and whether quality of care data are missing (yes = 1). Observational assessments of caregiver–child interaction were obtained for children who were in 10 or more hours per week of nonmaternal care. At least one such assessment was obtained between 6 and 24 months for 887 children and between 36 and 54 months for 985 children. Observations were conducted during two half-day visits scheduled within 2-week intervals at ages 6, 15, 24, and 36 months and during one half-day visit at 4½ years. At each visit, observers completed two 44-min cycles of the Observational Record of the Caregiving Environment (ORCE), during which they first coded the frequency of specific caregiver behaviors and then rated the quality of the caregiving. Positive caregiving composites were calculated for each age level observed by averaging the ratings (see NICHD Early Child Care Research Network, 1996, 2000a). At the end of each observation counts were made of the numbers of children and of child care providers in the setting from which measures of group size and adult–child ratio were derived. In addition, information on the educational levels of the child care providers were obtained as part of a caregiver interview.

At 6, 15, and 24 months, quality rating scores were the mean of five 4-point subscales (sensitivity to child's nondistress signals, stimulation of child's development, positive regard toward child, detachment [reflected], and flatness of affect [reflected]). Cronbach's alphas for the composite ranged from .87 to .89. At 36 months, these five scales plus two additional subscales—fosters child's exploration and intrusive (reflected)—were included in the composite (alpha = .83). At 4½ years, the positive caregiving composite was the mean of 4-point ratings of caregivers' sensitivity and responsivity, stimulation of cognitive development, intrusiveness (reflected), and detachment (reflected; alpha = .72). To ensure that observers at the 10 sites were making comparable ratings, all observers were certified before beginning data collection and tested for observer drift every 3 to 4 months. Reliability estimates were computed for both the master-coded videotapes and live observations using Pearson correlations and repeated-measures analysis of variance (ANOVA). Reliability exceeded .80 at all ages.

Child care type and quantity were collected during interviews with the mother beginning when the child was 1 month of age. Every 3 months when the child was between 3 and 36 months, mothers were asked about the type and hours per week spent in up to three care arrangements. From 42 to 54 months, mothers were interviewed every 4 months. To be consistent with prior reports (NICHD Early Child Care Research Network, 2002a), center care was coded if the mother reported the child spent at least 10 hr per week in center care, and total hours of care was tallied across all arrangements. The proportion of measurement occasions in which children were in center care was the index of type of care.

The demographic controls included study site, child gender, child ethnic group (non-Hispanic African American, non-Hispanic European American, Hispanic American, or other), maternal years of education at child's birth, average family income-to-poverty threshold ratio from 6 to 54 months, and the percentage of measurement occasions when a partner lived in the household (1–54 months). Each of these has been related to child care experiences (Pungello & Kurtz-Costes, 1999).

A great advantage of the NICHD study data is its wealth of measures of maternal, family, and child conditions measured before the child's first entry into child care. Child difficult temperament was measured by a 55-item Infant Temperament Questionnaire (Medoff-Cooper, Carey, & McDevitt, 1993) completed by mother. A composite measure, difficult temperament, was formed from the subscales, approach, activity, intensity, mood, and adaptability. Maternal sensitivity (positive, nonintrusive, responsive, and supportive maternal care) was coded from videotaped 15-min mother–child observations during semistructured play at 6 months. The maternal sensitivity score was a composite of 4-point ratings of sensitivity to nondistress, intrusiveness (reverse scored), and positive regard. Videotapes from all sites were coded at one location (see NICHD Early Child Care Research Network, 1999b, for details).

Quality of home environment was measured with the Infant/Toddler version of the Home Observation for Measurement of the Environment (HOME; Caldwell & Bradley, 1984), which is an assessment of the overall quality of the physical and social resources available to the child in the family context. Maternal depressive symptoms were measured using the Center for Epidemiological Studies Depression Scale (CES–D, Radloff, 1977) administered at 6 months. Maternal personality was measured with a composite of the neuroticism, extroversion, and agreeableness scales from the NEO Five-Factor Inventory, a short form of the NEO Personality Inventory (Costa & McCrae, 1985). Maternal separation anxiety was assessed using Subscale I of the Separation Anxiety Scale (Hock, Gnezda, & McBride, 1983). High scores indicate high levels of maternal worry, sadness, and guilt during separation from her child, and adherence to beliefs about the value of exclusive maternal care. Maternal attitudes and beliefs about childrearing were measured with a 30-item questionnaire probing mothers' ideas about raising children (Schaefer & Edgerton, 1985). High scores indicate authoritarian childrearing attitudes and beliefs. Maternal beliefs about costs and benefits of maternal employment for children were measured using the Attitudes Toward Maternal Employment questionnaire (Greenberger, Goldberg, Crawford, & Granger, 1988). High scores indicate positive beliefs. Maternal vocabulary was assessed by the Peabody Picture Vocabulary Test–Revised (PPVT–R; Dunn & Dunn, 1981), which was administered to mothers when their children were 36 months old. Each of these family characteristics was included because they had been linked theoretically or empirically to both child outcomes and family selection of child care arrangements (Pungello & Kurtz-Costes, 1999).

Results

The analysis plan involved fitting a series of level, change, and residualized change models to 54-month cognitive and achievement composite scores. For these analyses, we transformed the ORCE quality scores to have a mean of 0 and standard deviation of 1 so coefficients can be interpreted as estimated changes in outcomes associated with a standard deviation change in quality. All models included early or later ORCE quality scores, average hours of care per week, proportion of occasions in center care, and dummy variables that indicated whether the child received exclusive maternal care or had missing quality data. The variables representing early quality of care were computed as the mean of observed quality at 6, 15, and 24 months, and early hours of care and amount of center care were averaged from maternal reports collected every 3 months from 3 to 36 months. The later quality variable was computed from observed quality at 36 and 54 months, and later hours and amount of care variables from maternal report from interviews every 4 months from 42 to 54 months. Missing quality data occurred for children whose care setting should have been observed but was not. Our missing data procedures include the use of a dummy variable indicating who has missing data and assigning the mean value on the quality variable for those individuals. We also included a dummy variable for exclusive maternal care. The resulting model produces coefficients for quality that describes the linear association between quality and cognitive development based only on data from individuals observed in care. The coefficient for maternal care describes the mean difference in cognitive outcomes for children in care and not in care, and the coefficient for missing care describes the mean difference in cognitive outcomes for children observed and not observed in care. Missing value dummy variables were created for all covariates, and individuals with missing values were assigned the mean score on that measure. These missing value variables were added to the analyses to ensure that no child was deleted from the analysis because of missing child care variables or covariates.

Within each set of analyses (level, change, residualized change), four models were fit that correspond to models used in previous studies relating child care quality to cognitive development. The first model included only site as a covariate. The second model added ethnicity, maternal education, and gender. The third model added early and later family income and poverty threshold, maternal sensitivity and HOME enrichment scores at 6 months, maternal depression, maternal vocabulary, proportion of time with a partner in the household, and missing value dummy variables for each covariate. The final model added the 1-month childrearing attitudes score, 1-month maternal separation anxiety, 1-month rating of benefits of work, 6-month temperament rating, and 6-month maternal psychological adjustment ratings. Covariates were selected if we viewed them as reflecting potential selection bias and as exogenous influences. Any covariates that could be causally influenced by child care experiences (e.g., subsequent maternal sensitivity) were excluded as potentially endogenous influences.

Table 1 presents descriptive statistics for the longitudinal data used in these analyses. The columns in which variables are listed correspond to the study child's age when data were gathered. With average maternal education between a high school and college degree and income-to-poverty threshold ratios slightly above 3.0, the sample is slightly more affluent than the nation as a whole. Mean ORCE ratings are close to 3.0 on its 4-point scale. A score of less than 2 on the ORCE indicates low-quality care. When averaged between 6 and 24 months of age or 36 and 54 months, very few children were observed in low-quality care (*n*s = 43 and 31, respectively). The child care providers in the sample had, on average, more than 13 years of education (standard deviations ranged from 2.09 to 2.25 years). As expected, group sizes and child–adult ratios increased with age and were characterized by substantial variation given the range of types of care observed in this study. The increasing use of center-based care with age follows a familiar pattern.

Cognitive Development

Level model analysis. Table 2 presents regression results from level models in which the cognitive score at 24 months or 54 months is regressed on a linear measure of ORCE quality ratings and other controls. In Model 1, the child's 24-month cognitive score is related to child care quality between 6 and 24 months and a full set of birth-to-24-month parental and child controls. It produces an estimated impact of 1.7 for a 1 *SD* increase in ORCE quality. This can be viewed as an effect size of 1.66/14.6 = .11, when the sample standard deviation on the Mental Developmental Index (MDI; 14.6) was used, or as an effect size of .13 when the pooled standard deviation from the model was used.

Models 2 through 5 provide estimates of level models that relate cognitive development at 54 months to child care quality averaged over both the 6- to 24-month and 36- to 54-month intervals, with progressively more comprehensive control variables included in the model. In all cases the child care measures listed in the first column are included in the models. It is interesting that the coefficients on the early ORCE quality measure are generally similar to the coefficients on the later ORCE quality measure.

Looking across the cognitive outcome results for Models 2 through 5, it can be seen that controls for maternal education and the child's sex and gender in Model 3 cut the two ORCE child care quality coefficients in Model 2 roughly in half. The addition

Table 1
Description of Sample and Analysis Variables

	1 month			6 months			15 months			24 months			36 months			54 months		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Child outcomes																		
Cognitive development[a]							1,180	108.6	14.1	1,162	92.1	14.6				1,078	98.2	15.1
Academic achievement[b]																1,056	99.7	11.7
Family and mother characteristics																		
Maternal education (years)	1,363	14.2	2.5															
Partner in household	1,364	85%		1,279	86%		1,243	85%		1,207	85%		1,216	83%		1,084	83%	
Income/poverty threshold				1,270	3.7	3.1	1,234	3.7	3.2	1,187	3.7	3.0	1,208	3.6	3.1			
Ethnicity: White/non-Hispanic[c]	1,364	79%																
Maternal depression	1,363	11.4	9.0	1,278	9.0	8.3	1,241	9.0	8.2	1,119	9.4	8.6	1,202	9.2	8.3	1,077	9.8	8.7
Maternal agreeableness				1,272	46.3	5.3												
Maternal extroversion				1,272	42.5	5.8												
Maternal neuroticism				1,272	29.8	7.2												
Maternal separation anxiety	1,351	70.3	13.3															
Maternal benefits of work	1,363	19.2	3.2															
M – IQ (PPVT)													1,167	99.0	18.4			
HOME Total				1,279	36.5	4.6												
Maternal sensitivity				1,272	9.2	1.8												
Child-rearing attitudes	1,358	32.7	3.5															
Child characteristics																		
Gender (male = 1)	1,353	50%																
Difficult temperament				1,279	3.2	.4												
Child care																		
Quality: ORCE rating				593	2.98	.57	656	2.93	.57	669	2.81	.56	707	2.80	.46	854	2.98	.56
Hours/week				1,320	22.4	20.6	1,276	24.6	20.5	1,251	25.2	20.5	1,230	26.3	20.1	1,057	35.2	17.5
Any center care (1 = yes)				1,364	9%		1,269	12%		1,239	20%		1,229	36%		1,136	74%	
Center care, 10+ hr/week				1,364	8%		1,269	11%		1,239	17%		1,229	27%		1,136	54%	
Exclusive maternal care				1,316	43%		1,269	37%		1,239	34%		1,229	31%		1,136	20%	
Caregiver education				583	13.35	2.25	607	13.34	2.21	584	13.5	2.09	636	13.8	2.14	737	15.1	2.4
Observed child – adult ratio				591	2.77	2.02	608	2.99	1.90	668	3.70	2.23	707	4.86	2.97	854	6.90	3.59
Observed group size				591	3.29	3.21	668	3.77	3.28	669	5.06	4.51	707	7.34	5.99	854	12.71	7.37

Note. PPVT = Peabody Picture Vocabulary Test; HOME = Home Observation for Measurement of the Environment; ORCE = Observational Record of the Caregiving Environment.
[a]The 54-month cognitive composite includes Preschool Language Scale (PLS) Auditory Comprehension ($M = 98.3$, $SD = 19.9$), PLS Expressive Language ($M = 100.6$, $SD = 20.0$), Woodcock – Johnson (WJ) Picture Vocabulary ($M = 101.0$, $SD = 14.7$), and WJ Memory for Sentences ($M = 93.0$, $SD = 18.6$).
[b]The 54-month academic achievement composite includes WJ Applied Problems ($M = 103.3$, $SD = 15.9$), WJ Letter-Word Identification ($M = 99.3$, $SD = 13.9$), and WJ Incomplete Words ($M = 96.7$, $SD = 13.6$).
[c]Ethnicity includes 11% African American, 6% Hispanic, and 4% other.

Table 2
Level Model Analyses: Child Outcomes Predicted From Child Care in Analyses That Adjust for Increasing Numbers of Child and Family Characteristics as Potential Selection Factors

Independent variables	24-month score ($N = 1{,}162$)	54-month cognitive score ($N = 1{,}078$)				54-month achievement score ($N = 1{,}056$)			
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 2	Model 3	Model 4	Model 5
R^2	0.34***	0.18***	0.36***	0.42***	0.43***	0.13***	0.29***	0.34***	0.35***
Coefficients	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)
ORCE quality 6–24 months	1.66***	3.35***	1.64**	1.34**	1.36**	2.26***	0.96*	0.85*	.90*
(mean *z* score)	(0.48)	(0.55)	(0.50)	(0.48)	(0.49)	(0.45)	(0.41)	(0.41)	(0.41)
ORCE quality 36–54 months		2.72***	1.57***	1.19**	1.17**	2.15***	1.33***	1.17**	1.12**
(mean *z* score)		(0.50)	(0.44)	(0.43)	(0.43)	(0.40)	(0.36)	(0.36)	(0.36)
Mother care 6–24 months	− 0.27	− 1.50	− 1.35	− 1.84	− 1.51	− 0.87	− 0.81	− 0.98	− 0.64
(1 = yes; 0 = no)	(1.15)	(1.38)	(1.23)	(1.18)	(1.19)	(1.11)	(1.01)	(0.98)	(0.99)
Mother care 36–54 months		− 4.41*	− 1.84	− 0.46	− 0.33	− 3.16*	− 1.25	− 0.45	− 0.42
(1 = yes; 0 = no)		(1.74)	(1.55)	(1.49)	(1.49)	(1.41)	(1.29)	(1.26)	(1.26)
Quality 6–24 months missing	− 0.64	− 6.88***	− 4.11**	− 3.93**	− 4.19**	− 3.67*	− 1.61	− 1.89	− 2.15
(1 = yes; 0 = no)	(1.40)	(1.61)	(1.44)	(1.40)	(1.40)	(1.31)	(1.19)	(1.18)	(1.18)
Quality 36–54 months missing		− 1.57	− 1.19	0.06	− 0.02	− 1.79	− 1.48	− 0.71	− 0.69
(1 = yes; 0 = no)		(1.69)	(1.50)	(1.44)	(1.45)	(1.37)	(1.24)	(1.21)	(1.21)
Mean hours of care/week 3–24	0.04	0.14**	0.07	0.06	0.06	0.06	0.01	− 0.01	− 0.01
months	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Mean hours of care/week 27–54		− 0.15***	− 0.07	− 0.07	− 0.06	− 0.06	0.01	0.02	0.02
months		(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
Proportion center care 3–24 months	5.41***	1.20	− 1.29	− 1.39	− 1.09	− 0.47	− 2.30	− 2.07	− 2.07
	(1.58)	(2.27)	(2.02)	(1.93)	(1.94)	(1.83)	(1.67)	(1.62)	(1.63)
Proportion center care 27–54 months		7.19***	5.28***	4.34**	4.10**	5.15***	3.60***	3.21**	3.13**
		(1.65)	(1.48)	(1.43)	(1.44)	(1.33)	(1.22)	(1.20)	(1.21)

Note. Model 1 includes all covariates; Model 2 includes site-only covariates; Model 3 adds gender, ethnicity, and maternal education; Model 4 adds income/poverty thresholds, partner in household, 6-month parenting, maternal depressive symptoms, and maternal vocabulary; and Model 5 adds child temperament, maternal personality, maternal child-rearing attitudes, maternal separation anxiety, and 1-month maternal attitudes about benefits of work. ORCE = Observational Record of the Caregiving Environment; *B*(*se*) = regression coefficient and standard error.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

of many more selection factors in Models 4 and 5 produces proportionately smaller reductions in the estimated impacts of ORCE-based quality. This suggests that a readily available maternal measure—completed schooling—accounts for most but not all of the adjustment to the quality impact estimate obtained in a regression with a much more extensive set of observable maternal and home characteristics.

The most inclusive level model (Model 5) suggests that a 1 *SD* increase in either early or later child care quality raises cognitive scores by 1.4 and 1.2 points, which in turn suggests that higher quality across the entire period is associated with 2.6-point increase. Both of these coefficients are statistically significant at conventional levels. Each coefficient amounts to an effect size of about .08 to .09 (1.3/15.1 = .086), whereas both combined produce an effect size of about .17 for a 1 *SD* change in ORCE-based quality maintained across the entire 4-year period of measurement. Using the standard deviation from the analysis model, the effect sizes range from .09 to .18.

Also noteworthy in the 54-month cognitive outcome regression models in Table 2 is that the proportion of time spent in center care between ages 27 and 54 months (but not earlier) appears to have a positive impact on cognitive development, with the difference between 100% of occasions versus no occasions in such care associated with a 4.1-point increase in cognitive test scores in Model 5 (effect size = .27). In addition, the cases with missing quality assessments (primarily because the child care provider refused to be observed) in the early but not later period have lower scores. Prior analyses indicated that child care settings with missing data in infancy probably provided lower quality care because caregivers tended to have less education and training (NICHD Early Child Care Research Network, 1996).

Change model analysis. Even the inclusion of the extensive set of characteristics in Model 5 does not rule out the possibility that yet more controls might reduce the estimated impact further. If one accepts their assumptions, change models remove the biasing impacts of all persistent characteristics, both observed and not observed.

Table 3 presents results from various formulations of simple and residualized change models for cognitive outcome (top half of the table). The first such model (Model 6) includes only the earlier and later child care variables listed in the first column of Table 3. Recalling that the interpretation of a child care quality coefficient in our change model is identical to its interpretation in our level model, it produces an estimated impact of child care quality between 36 and 54 months on the cognitive outcome (1.17 points) that is similar to estimated impacts of 36- to 54-month quality in the most comprehensive level models and is statistically significantly different from zero ($p < .05$).

The addition of site, gender, and ethnicity reduces the ORCE's coefficient further, and the addition of extensive family characteristics in Model 8 reduces it to a statistically and substantively insignificant. The coefficient in the final model is .59, which translates

Table 3
Simple and Residualized Change Models: Child Outcomes Predicted From Child Care, Adjusting for Child and Family Characteristics

	Simple change models					Residualized change models				
	Model 6	Model 7	Model 8	Model 9	Model 10	Model 6′	Model 7′	Model 8′	Model 9′	Model 10′
	Cognitive score ($N = 1{,}032$)									
R^2	0.02*	0.08***	0.10***	0.13***	0.13***	0.46***	0.48***	0.53***	0.56***	0.57***
Coefficients	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)
ORCE quality 36–54	1.17**	0.87*	0.72	0.58	0.59	1.73***	1.55***	1.11**	0.88*	0.86*
months (mean *z* score)	(0.44)	(0.44)	(0.44)	(0.45)	(0.45)	(0.40)	(0.40)	(0.39)	(0.38)	(.38)
Mother care 36–54 months	– 2.18	– 2.11	– 1.83	– 1.50	– 1.52	– 3.20*	– 3.20*	– 2.13	– 1.45	– 1.38
(1 = yes, 0 = no)	(1.61)	(1.58)	(1.57)	(1.57)	(1.58)	(1.45)	(1.43)	(1.37)	(1.34)	(1.34)
Quality 36–54 months	– 0.93	– 0.93	– 1.12	– 0.19	– 0.31	– 1.76	– 1.40	– 1.39	– 0.27	– 0.38
missing (1 = yes)	(1.53)	(1.50)	(1.50)	(1.51)	(1.51)	(1.38)	(1.36)	(1.30)	(1.28)	(1.29)
Mean hours of care/week	– 0.03	– 0.02	– 0.02	– 0.00	– 0.01	– 0.10**	– 0.08*	– 0.05	– 0.04	– 0.04
27–54 months	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Proportion center care	3.41*	4.04**	3.82**	3.26**	3.22**	4.56***	5.37***	4.70***	3.95**	3.84**
27–54 months	(1.47)	(1.46)	(1.47)	(1.48)	(1.49)	(1.33)	(1.32)	(1.28)	(1.26)	(1.26)
	Academic achievement ($N = 1{,}014$)									
R^2	0.02*	0.08***	0.09***	0.10***	0.11***	0.34***	0.36***	0.41***	0.43***	0.44***
Coefficients	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)
ORCE quality 36–54	0.38	0.25	0.39	0.45	0.43	1.23***	1.28***	0.95**	0.88**	0.84*
months (mean *z* score)	(0.45)	(0.45)	(0.45)	(0.46)	(0.46)	(0.35)	(0.35)	(0.34)	(0.34)	(0.34)
Mother care 36–54 months	– 0.38	– 0.51	– 0.84	– 0.95	– 1.04	– 1.83	– 2.12	– 1.26	– 0.87	– 0.89
(1 = yes, 0 = no)	(1.66)	(1.63)	(1.63)	(1.64)	(1.64)	(1.27)	(1.26)	(1.22)	(1.21)	(1.21)
Quality 36–54 months	– 0.13	– 0.62	– 0.90	– 0.54	– 0.57	– 1.26	– 1.47	– 1.41	– 0.79	– 0.78
missing (1 = yes)	(1.56)	(1.54)	(1.54)	(1.56)	(1.57)	(1.19)	(1.19)	(1.15)	(1.15)	(1.15)
Mean hours of care/week	0.08	0.07	0.06	0.08	0.07	– 0.01	– 0.00	0.02	0.04	0.04
27–54 months	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Proportion center care	1.54	1.73	1.73	1.66	1.82	3.15**	3.57**	2.87*	2.54*	2.57*
27–54 months	(1.51)	(1.50)	(1.50)	(1.55)	(1.54)	(1.16)	(1.16)	(1.13)	(1.13)	(1.13)

Note. Model 6 outcome is change score, no covariates, and Model 6′ outcome is 54-month scores and 24-month cognitive scores. Models 7 and 7′ add site. Models 8 and 8′ add gender, ethnicity, and maternal education. Models 9 and 9′ add income/poverty thresholds, partner in household, 6-month parenting, maternal depressive symptoms, and maternal vocabulary. Models 10 and 10′ add child temperament, maternal ratings of personality, child-rearing attitudes, separation anxiety, and benefits of work. ORCE = Observational Record of the Caregiving Environment; *B*(*se*) = regression coefficient and standard error.
*$p < .05$. **$p < .01$. ***$p < .001$.

into an effect size of .04 based on sample standard deviation (.59/15.1) or .05 based on the standard deviation estimated with the model. Taken together, alternative formulations of the simple change model in the left half of Table 3 suggest that ORCE-based quality between 36 and 54 months has a modest to null impact on cognitive development**.**

Estimates from various residualized change models (in which 54-month cognitive development is regressed on 24-month cognitive development plus the quality measures) are presented in the right half of Table 3. The quality impacts retain statistical significance in all of the models and imply that a 1 *SD* increase in ORCE-based quality is associated with between .9 and 1.7 points change in cognitive development, with an effect size for the most comprehensive model equal to .06.

Academic Achievement

Level models. Table 2 also presents regression results from level models in which academic achievement at 54 months is regressed on a linear measure of ORCE quality ratings and other controls. In the case of site-only-covariate Model 2, the coefficients on quality imply that a 1 *SD* change in ORCE quality in either early or late childhood is associated with a 1.0- to 1.3-point increase in children's achievement scores. Controls for additional family and child characteristics reduced the ORCE child care quality coefficients in Model 2 only slightly. As with the level analyses of cognitive scores, the addition of many more selection factors in Models 4 and 5 produced smaller reductions in the estimated impacts of ORCE-based quality.

The most inclusive level model (Model 5) suggests that a 1 *SD* increase in either early or later child care quality raises achievement scores by .9 to 1.1 points (suggesting that higher quality across the entire 4-year period is associated with a 2.0-point increase). As before, the proportion of time spent in center care between ages 27 and 54 months (but not earlier) is positively related to achievement, with the difference between 100% of time versus no time in such care associated with a 3.1-point increase in achievement test scores in Model 5 (effect size of .27).

Change models. The bottom half of Table 3 presents results from analyzing the 54-month achievement outcome with the simple and residualized change models. We included the change model analyses as a point of comparison, although we recognize that the 24-month cognitive assessment and 54-month achievement measures cannot be regarded as repeated measures. The first such model (Model 6) includes only child care variables and suggests that a 1 *SD* change in quality is significantly associated with a .38-point change in achievement scores at 54 months. This estimate is about one third that from the most restrictive level model (Model 5) and is substantively and statistically nonsignificant.

The residualized change models yielded quality coefficients that were larger than those from the corresponding simple change models but smaller than those from the level models. Quality impacts retain statistical significance in all of the models. The most comprehensive model suggests that a 1 *SD* increase in ORCE-based quality is associated with a .84-point change in cognitive development, which implies an effect size of .07.

Extensions

Nonlinear relations of quality to outcomes. Our worry that a linear ORCE might miss important thresholds in the impacts of child care quality on cognitive development led us to estimate level, change, and residualized change models in which we categorized child care quality and compared means across groups. Although all analyses indicated that children in the highest quality care scored higher than children in the lowest quality care, there was no consistent pattern of evidence regarding thresholds at either low- or high-quality levels. In part, this resulted from a lack of sample observations of children in very low quality settings (e.g., only 32 children were in settings with ORCE scores between 1 and 2.)

Moderators of relations of quality to outcomes. We next tested for various interactions between: early and later child care quality, child care quality and child gender, child care quality and mother's schooling level, and child care quality and type of care (center vs. not). Again, we estimated level, change, and residualized change models. Only one statistically significant interaction emerged—a Gender $\times$ Quality interaction in the change model analysis of achievement—and it was not interpreted because it was not replicated across outcomes or models.

Child's initial cognitive level as moderator. We also tested whether the impacts of child care quality depended on the initial cognitive ability of the child. There are two reasons to believe this might be the case. On a substantive level, it could well be that children with early cognitive deficits might profit the most from sensitive and stimulating interactions with caregivers or be most adversely affected by the absence of such interactions. On a methodological level, there was evidence that the Bayley scores used in our change model are likely more reliable and thus better predictors of later performance for infants who score in the lower range than for those who score in the normal to high range.

To investigate these possibilities, we created a dummy variable indicating whether the child had scored in a problematic cognitive range at 15 months, defined as scores less than or equal to .5 *SD* below the mean. We chose the 15-month measure because it preceded the 24-month measure used in our level and change models. This dummy variable and its interactions with all child care and family variables were added to the final level (Model 5 in Table 2), change (Model 10 in Table 3), and residualized change (second Model 10 in Table 3) models.

The resulting coefficients are shown in Table 4. The first two rows in Table 4 repeat coefficients shown in Tables 2 and 3. The next rows show the corresponding coefficients for children who did and did not score in the low range on the Bayley MDI at 15 months. The coefficients for later child care quality were significantly different for these two groups of children in the change model analyses of cognitive outcomes, $F(1, 940) = 4.16$, $p = .04$, and achievement outcomes, $F(1, 924) = 6.09$, $p = .01$, but not in the level model analyses of cognitive outcomes, $F(1, 972) = 0.00$, $p = .99$, and achievement outcomes, $F(1, 955) = 1.15$, $p = .28$, or the residualized change model analyses of cognitive outcomes, $F(1, 939) = .99$, $p = .32$, or achievement outcomes, $F(1, 925) = 2.01$, $p = .16$. Later quality is related to cognitive development in a similar manner for both groups of children in the level model, but is positively and significantly related only for children who initially scored low in the change model analysis. Similar results were observed when achievement scores were analyzed.

The change models both make the assumption that error in assessments of outcomes at both time points is random, whereas the level model makes that assumption only about the later assessment. Accordingly, these analyses suggest that the lower reliability of measuring cognitive ability for children scoring higher on the test at the earlier assessment and not the later assessment may be biasing estimates from our change models, especially because this interaction was not observed in the level analysis that did not include the 24-month cognitive score in the model. Alternatively, these results may

Table 4
Child Care Quality and Child Outcomes, Testing for Different Associations for Children With Low and Normal Early Cognitive Scores

	54-month cognitive score			54-month achievement score		
Independent variable	Level model	Change model	Residual change	Level model	Change model	Residual change
	B(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)	*B*(*se*)
	Models with all covariates[a]					
Quality 6–24 months	1.36**			0.90*		
	(0.49)			(0.41)		
Quality 36–54 months	1.17**	0.59	0.86*	1.12**	0.43	0.84*
	(0.43)	(0.45)	(0.38)	(0.36)	(0.46)	(0.34)
	Models with interactions between low MDI and child care					
15-month MDI low ($n = 265$)						
Quality 6–24 months	2.33*			2.39**		
	(1.02)			(0.83)		
Quality 36–54 months	1.12	2.02*	1.50*	1.60*	2.08*	1.54*
	(0.77)	(0.85)	(0.75)	(0.65)	(0.86)	(0.64)
15-month MDI normal ($n = 757$)						
Quality 6–24 months	1.26*			0.63		
	(0.54)			(0.46)		
Quality 36–54 months	1.11*	– 0.00	0.66	0.78	– 0.40	0.48
	(0.50)	(0.53)	(0.45)	(0.42)	(0.54)	(0.40)

Note. *B*(*se*) = regression coefficient and standard error; MDI = Mental Developmental Index.
[a]These coefficients are also reported in Tables 2 and 3. All covariates are included in the model.
*$p < .05$. **$p < .01$. ***$p < .001$.

reflect the fact that children with early cognitive deficits profit the most from high-quality care. Further work on these issues is an important research priority.

Other quality measures. Finally, we estimated quality impacts using alternative measures of quality: staff-to-child ratio and group size, both as reported by the mother and as observed in the child care setting; and child-care-setting-reported teacher education and training (Table 5). We employed the range of level and change models listed at the bottom of the table. The first model included type of observed care (center: yes or no) and hours of child care as covariates. The second model added maternal education, gender, and ethnicity, and the final model included all of the family and child covariates included in the models described earlier.

Teacher education showed relatively consistent associations with children's 54-month achievement outcomes (i.e., math and reading skill, and phonological knowledge) in the change but not level models. Group size also showed consistent, although weaker, negative associations with 54-month cognitive outcomes in all change model analyses but not in any level model analyses. Ratio showed no consistent pattern of association. We find these results puzzling. In the ORCE-based results, simple change models generally produced the smallest and least consistently significant quality coefficients, which led us to speculate that faulty model assumptions may be leading to a systematic underestimation of quality effects. However, Table 5 shows that the simple change models generally produce the largest and most consistently significant coefficients for these alternative measures of quality. Further research on the range of quality measures available in the NICHD Study of Early Child Care is another priority.

Discussion

We employed a variety of approaches to address the problem of unobservable sources of bias in estimates of the impacts of child care quality on children's cognitive development. We found considerable evidence of bias from models that do little to control

Table 5
Level, Change, and Residualized Change Analyses of Caregiver Education, Group Size, and Adult – Child Ratios and Child Outcomes

	Level model			Change model			Residualized change		
	1	2	3	1	2	3	1	2	3
Caregiver education									
24-month cognitive development	1.56***	0.72**	0.61*						
	(0.28)	(0.27)	(0.30)						
54-month cognitive development	0.64*	0.10	0.00	0.46*	0.46*	0.42+	0.48*	0.24	0.17
	(0.26)	(0.23)	(0.22)	(0.22)	(0.22)	(0.23)	(0.20)	(0.20)	(0.20)
54-month academic achievement	0.61**	0.22	0.19	0.50*	0.64**	0.65**	0.52**	0.34*	0.32+
	(0.20)	(0.18)	(0.18)	(0.23)	(0.23)	(0.23)	(0.17)	(0.17)	(0.17)
Observed group size									
24-month cognitive development	0.05	0.11	0.07						
	(0.18)	(0.16)	(0.16)						
54-month cognitive development	– 0.02	– 0.04	– 0.01	– 0.27***	– 0.27***	– 0.27**	– 0.19**	– 0.17**	– 0.15*
	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)
54-month academic achievement	0.03	0.01	0.03	– 0.21**	– 0.20*	– 0.22**	– 0.07	– 0.06	– 0.05
	(0.07)	(0.06)	(0.07)	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)	(0.06)
Observed child – adult ratio									
24-month cognitive development	– 0.11	0.03	0.03						
	(0.30)	(0.27)	(0.27)						
54-month cognitive development	0.22	0.07	0.08	– 0.27+	– 0.26+	– 0.24+	– 0.10	– 0.11	– 0.08
	(0.17)	(0.15)	(0.14)	(0.14)	(0.14)	(0.14)	(0.13)	(0.13)	(0.13)
54-month academic achievement	0.25*	0.16	0.17	– 0.22	– 0.18	– 0.16	0.04	0.05	0.05
	(0.13)	(0.12)	(0.11)	(0.14)	(0.14)	(0.15)	(0.11)	(0.11)	(0.11)

Note. Model 1 covariates are type of care (center = yes or no) and hours of care per week; Model 2 adds gender, ethnicity, and maternal education; and Model 3 add income/poverty thresholds, partner in household, child temperament, parenting, maternal vocabulary, maternal ratings of psychological adjustment, child-rearing attitudes, separation anxiety, and benefits of work.
+.1 < p < .05. *p < .05. **p < .01. ***p < .001.

for child care selection factors. It is not surprising that the various approaches to adjusting for selection bias yielded different conclusions, and we were not able to determine definitively which model provided the least biased estimates of the impact of child care quality on child outcomes. We were, however, able to establish a fairly narrow range of effects sizes (.04 to .08) on cognitive ability and achievement for a 1 *SD* increment in observed child care quality sustained between 24 and 54 months of age. Child care quality increments sustained during the first 2 years of life may have similar positive impacts, although we were unable to establish the robustness of these estimates using our full range of methods.

Until recently, most of the existing literature relied on far fewer covariates than some psychologists and economists feel are necessary to provide the needed bias adjustments (Duncan & Gibson, 2000; Scarr, 1998). Compared with level models that controlled only for site, level models that also controlled for a few demographic and child characteristics (notably, mother's schooling and the child's gender and ethnicity) reduced the estimates of the impact of 36- to 54-month quality on 54-month cognitive development by close to half. Adding covariates included in various prior analyses (e.g., family income or poverty level, maternal sensitivity, HOME scores) reduced the coefficients by an additional 14%. Adding all baseline maternal and child characteristics available in the NICHD study data had virtually no additional effect on the quality impact estimates. Coefficient reductions were smaller for the achievement outcomes than for the cognitive outcomes, and they were smaller in the residualized change model analysis than in the level model analysis.

Change models further reduced the coefficient on child care quality, which is consistent with the possibility that unobserved selection factors that are correlated with both choice of child care quality and children's cognitive development may still be biasing coefficients even in the most complete level models. For cognitive skills, our least comprehensive change model yielded similar coefficients to those reported for the level models; the most comprehensive change model reduced the quality coefficients to statistical nonsignificance.

It is important to appreciate that the three types of models we employed vary in the number of assumptions on which the model depends and in the degree to which the model adjusts for selection bias due to omitted variables. All three models assumed that included covariates are exogenous (i.e., are not changed by quality of care in a way that affects how they relate to the outcome measure). Efforts were made to include only family and child measures that would not be influenced by child care experiences, although it is possible that some covariates might be at least partially endogenous. If so, estimates of the quality impact from the models with extensive covariates may be biased.

Most of our models assumed that the relation between child care quality and child outcomes is linear. This assumption was supported in follow-up analyses for the cognitive outcomes, but the absence of large numbers of children in very low quality care leaves open the possibility that the impacts of quality increments are different for children at the low end of the child care quality distribution. Follow-up analyses also suggested no important interactions between child care quality and family or child characteristics.

The change and residualized change models, as distinct from the level models, made additional assumptions as they attempted to adjust for omitted-variable bias. For example, estimates of the quality impact from the residualized change model will be biased if important selection variables are not included and they have a similar impact across time. The simple change model provides the strongest adjustment for omitted-variable bias by subtracting the earlier assessment from the outcome of interest. However, the assumptions of this model will be violated if omitted variables correlated with child care quality have differential impacts on child outcomes at different ages. Our attempts to test whether model assumptions were met in these change analyses suggested that there might be reason for concern.

The Role of Selection Bias

What do these analyses imply about bias due to omitted variables in analyses relating child care quality and child outcomes? A starting point for answering this question is to note that the range of coefficients estimating the impact of a 1 *SD* increment in child care quality from the least sophisticated level model to the most complete change models is large for both cognitive outcomes (raw score coefficients range from 2.72 to .59) and achievement outcomes (1.33 to .43). A small set of easily measured covariates (most important, maternal schooling) reduces the upper bound of this range considerably (from 2.72 to 1.57 for cognitive outcomes and from 1.33 to 1.17 for achievement outcomes). Adding an extensive set of covariates in the level and residualized change models produces small additional coefficient changes, which suggests that there may be relatively little additional bias from unobserved variables (Altonji et al., 2002).

Adding the earlier and time-invariant covariates in the change models markedly reduced the child care quality coefficients. These results indicate that concern regarding the possible impact of omitted variables is legitimate, but they raise additional methodological concerns. Adding these variables can change the quality coefficients only if they have a stronger shared association between child care quality and cognitive outcomes at one age than at the other age. This raises questions about the underlying assumption that each omitted variable also has a comparable effect at both times. Testing this assumption provided some evidence suggesting that estimates of ORCE-based child care quality coming from our change model were biased downward. On the other hand, our change models of impacts of alternative measures of child care quality generally produced larger estimates than did corresponding level and residualized change models.

All in all, our analyses do not clearly answer the question about the degree to which omitted-variables bias has led to faulty conclusions regarding whether and how much child care quality affects children's cognitive outcomes. The general convergence in estimated effect sizes across the level, residualized change, and change models, at least for the children for whom we have reliable infant scores, suggests that child care quality is a modest, but reliable, predictor of cognitive development and academic achievement during early childhood. Thus, these results indicate that the frequently observed link between child care quality and cognitive development cannot be completely explained by selection bias, although the estimated effect sizes vary with the methods used to control for selection factors.

Child Care and Cognitive Development

What about the child care quality effect sizes implied by the coefficients in our models? Although there is no way to know definitively which model provides the best estimates, almost all of them employing a reasonable number of covariates suggest modest quality effect sizes, that is, a .04 to .08 *SD* increment to cognitive ability and achievement for a 1 *SD* increment in child care quality sustained between 24 and 54 months of age. Effect sizes for standard deviation child care quality increments before age 24 months were around .08, although in this case we were unable to use change models to investigate the robustness of this estimate. Previous analyses from the NICHD Study of Early Child Care (NICHD Early Child Care Research Network, 1999b), which relied on level models and included many, but not all, of the same controls for selection bias reported a considerably higher effect size of .19 for a 1 *SD* change in quality sustained over a 1- to 3-year period on the school readiness scores of 3-year-olds.

Models tested separately on children with low early MDI scores produced larger effect sizes, which ranged from .07 to .18. This higher range either challenges the change model assumptions (it was not found in the level models) or indicates that child care quality has a bigger impact for these developmentally at-risk children, as has been found in the early intervention literature (cf. Ramey, Bryant, & Suarez, 1985), and the change models provided the best estimates of those effects.

Although our primary focus has been on ORCE-based quality impacts, it is important to point out that time in center-based child care in the third and fourth years of life had the most consistently significant associations with both cognitive and achievement outcomes across all of our various models. Our most complete models produced effect size estimates for spending all versus none of this developmental period in center-based care ranged from .09 to .27 for cognitive outcomes and .22 to .33 for achievement outcomes. In contrast, center-based care earlier in childhood did not have consistently significant associations with cognitive and achievement outcomes. Comparable findings regarding center care for cognitive outcomes have previously been reported with the NICHD data (NICHD Early Child Care Research Network, 2000b, 2002a).

Putting the Effect Sizes in Context

McCartney and Rosenthal (2000) urged investigators to place effect sizes from any given analysis in the context of other relevant effect sizes as a means of assessing their practical significance. In nonexperimental studies, for example, a .10 effect size is judged small (Cohen, 1988) but still could have both statistical and policy significance. The most complete models reported in this study suggest that the effect sizes of a 1 *SD* increase in child care quality between 24 and 54 months of age ranges from .04 to .08 for the full sample of children and from .07 to .18 for initially low MDI children. How do these compare with the range of effect sizes from (a) other naturalistic studies of variation in child care quality;

(b) analyses that combine child care and early intervention programs; and (c) experimental studies of high-quality early intervention programs, welfare reform interventions, and elementary class size interventions?

Taking into account the duration of the quality increments, the effect sizes obtained in the present analysis for a 2.5-year quality enhancement between 24 and 54 months are considerably less than those obtained in other nonexperimental studies of variation in child care quality. For example, the Cost, Quality, and Outcome Study, which encompassed a broader range of quality than did the NICHD study, reported an effect size for a 9-month increment in child care quality of approximately .20 on a measure of language and math skills at entry to kindergarten in a sample of more than 700 predominantly middle-class children (Peisner-Feinberg et al., 2001). A recent, secondary analysis combining results from the Cost, Quality, and Outcome Study, a study of Head Start programs in North Carolina, and study of public preschool programs in North Carolina (Burchinal, Peisner-Feinberg, et al., 2000) reports effect sizes for 1 *SD* change in quality sustained over a 6- to 9-month period of .21 to .24 for prereading and math skills.

Our effect sizes are also substantially smaller than those reported in experimental studies of early preschool intervention programs offering levels of quality that routinely exceed those of more typical community-based child care programs and focusing on children at risk because of both economic and developmental factors (Layzer, Goodson, & Moss, 1993; Phillips, Voran, Kisker, Howes, & Whitebook, 1994). For example, treatment effect sizes on IQ were 2.0 at age 5 in the 2-year Milwaukee Project; 1.0 at 3 years and .75 at age 5 for the 3- and 5-year treatment of the Abecedarian Project, .60 for the 1- to 2-year Perry Preschool Project; and .50 at age 5 for 5-year Project CARE (Ramey et al., 1985). These projects provided high-quality care to disadvantaged children randomly assigned to treatment for a minimum of 2 years. The much greater intensity and duration of these treatments, as well as their target population of at-risk children, are the most likely explanations for their larger effect sizes. In addition, the effect sizes in these randomized experimental studies may be more precise because extensive covariates are not needed to adjust for potential selection bias.

The range of child care quality effect sizes obtained in the current study is also in the lower half of the range of effect sizes (.01 to .31) for parent and teacher reports on early elementary school achievement reported in recent studies of the effects of welfare reform experiments on child outcomes for children who were preschool age at random assignment (Morris, Huston, Duncan, Crosby, & Bos, 2001). Effect size estimates ranging from .15 to .35 were obtained in experimental studies of the impacts of class-size reduction on elementary children's math and reading scores (Krueger, 1999).

One crucial policy question is whether our estimated effects are economically significant, which depends on a comparison of the benefits associated with a .04 to .08 *SD* increase in test scores and the cost of producing a 1 *SD* (half-point) change in the ORCE for 3 years. This is an important policy issue because quality child care seems to provide a more effective means of improving children's cognitive outcomes, especially for at-risk children, than other types of widely used programs such as home visit programs (St. Pierre, Layzer, & Barnes, 1995). Although it is commonly assumed that the cost differences between high- and low-quality care are substantial because quality is associated with caregiver education and wages (Phillips et al., 1994), parents in this sample reported relatively modest differences in the fees they paid for low- and high-quality care, and similar increments in costs were estimated in another large study of child care centers (Helburn, 1995). If quality could somehow be increased by 1 *SD* for the estimated increment in wages ($15 per week, or about $2,000 for the interval between 24 and 54 months), our results suggest that the quality impacts might well be cost effective, particularly for children with low initial cognitive ability. Another promising line of intervention research is to focus on increasing the use of center-based care for 3- and 4-year-olds given our findings of consistently larger effect sizes associated with that kind of change. A final judgment on the economic significance of the effect sizes we estimate requires more information than we have at hand.

Limitations of the Study

There are important qualifications to these conclusions. First, even our most comprehensive level models omit potentially important variables, and we do not know how including them would affect our estimates. The change model provides less biased estimates with larger standard errors when model assumptions are met, but possibly more biased estimates when the impact of the omitted variables vary over time (Wang & Burchinal, 2001). In our case, it is likely that at least some omitted variables operate differently over time. For example, entitlement programs such as Head Start or prekindergarten programs may be primarily responsible for placing low-income children in higher quality care at older ages, but differential parental selection of higher quality care among more educated parents may be primarily responsible for placing children in higher quality care at younger ages. If this is the case, the change model may underestimate the impact of quality.

Second, few children were observed in low-quality settings, and quality impacts may be larger for children in the worst child care settings. Our attempts to isolate the impacts of improving on the lowest quality observed were not successful, perhaps because the very worst settings are underrepresented in the data, rendering it difficult to assess the benefits of improving them. Another consequence of the truncated distribution of child care quality is that we may be underestimating effect sizes because of restriction in variability in our primary predictor of interest, child care quality. Efforts to simulate nationally representative data with our sample suggest that quality of care observed in this study is higher and less variable than overall quality of care in the nation (NICHD Early Child Care Research Network, 2000a).

Third, center care was coded only if the child spent at least 10 hr in that setting. Follow-up analyses indicated similar quality and center effects when any center care was coded. Finally, our quality measure does not explicitly reflect educational dimensions of the child care settings, which may be particularly important for the outcomes assessed in this study. Most of our conclusions relate to child care quality measured by the ORCE when the child is between 24 and 54 month old. The ORCE focuses on the sensitivity or responsivity and affective quality of caregiver–child dyadic interactions. As such, it may be missing dimensions of quality more directly related to the educational content of the child care setting and to cognitive achievement outcomes for children, such as specific learning-focused exchanges, the curriculum, or available learning materials. The fact that being in center-based care, independent of quality, has a consistent positive relation to cognitive outcomes suggests that there may be features in the structure and organization of child care centers, and the typically stronger educational qualifications of center-based providers, that are important influences not captured in the ORCE.

Fourth, quality effect sizes are underestimated with all three models to the extent that quality and related covariates were measured with error. The NICHD Study of Early Child Care has maintained high standards for interrelater reliability within and across sites, but it is inevitable that our ratings from 1 to 2 days of observation do not completely describe the quality of care. Indeed, because the quality measurement was based on sampling slices of time, the resulting sample error can be thought of as a source of measurement error. In addition, our quality measure does not include measures of curriculum or activities that were included in other quality measures that showed stronger associations with child care quality in other studies (Peisner-Feinberg & Burchinal, 1997). Measurement error usually decreases estimated effect sizes, and downward bias due to measurement error may pose a bigger threat to unbiased estimation than the omitted-variables bias. Further work is needed to address this question before we can conclude that we have unbiased estimates of quality effect sizes.

Conclusions

There has been substantial debate about the adequacy of commonly used statistical methods and nonexperimental data for securing unbiased estimates of the effects of child care quality on child outcomes (e.g., Duncan & Gibson, 2000). The current study used a variety of methods for controlling for selection bias, derived from both the economics and the developmental literatures, to gain a better understanding of this important scientific and policy issue. Our findings do not identify the best method, although they do provide a fairly narrow range of likely effect sizes. They suggest that although the prevailing developmental literature on child care may have overestimated the developmental consequences of child care quality, studies that control for demographic, family, and maternal variables produce estimates that are close to the high end of our range of preferred estimates. There remains a compelling need for experimental studies of child care that encompass the typical range of quality in community-based arrangements.

At the same time, policy decisions continue to be made regarding the desirability of investing in efforts to improve the quality of child care for the nation's children. Our findings suggest that such efforts are likely to improve children's cognitive and achievement outcomes, although we did not assess whether the benefits of such efforts exceed the cost. We find more consistent support for the likely benefits of providing more opportunities for preschoolers to attend center-based child care. Finally, our results also suggest that efforts to improve quality may be especially important for young children with low initial cognitive skills. All of our conclusions, although often consistent with the existing research, require replication using even stronger research designs and samples with strong representation of children in low-SES families.

References

Allison, P. (1990). Change scores as dependent variables in regression analysis. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 93 – 114). Oxford, England: Blackwell.

Altonji, J., Elder, T., & Taber, C. (2002). *Selection on observable and unobservable variables: Assessing the effectiveness of Catholic schools*. Mimeographed, Northwestern University.

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, *5*(3), 25 – 50.

Baydar, N., & Brooks-Gunn, J. (1991). Effects of maternal employment and child care arrangements on preschoolers' cognitive and behavioral outcomes: Evidence from the children of the national longitudinal survey of youth. *Developmental Psychology, 27*, 932 – 945.

Bayley, N. (1969). *Bayley Scales of Infant Development*. New York: Psychological Corporation.

Bayley, N. (1993). *Bayley II Scales of Infant Development*. New York: Psychological Corporation.

Blau, D. (1999). The effects of child care characteristics on child development. *Journal of Human Resources*, *34*, 786 – 822.

Bound, J., Brown, C., Duncan, G., & Rodgers, W. (1994). Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics*, *12*, 345 – 368.

Bradley, R. H., Caldwell, B. M., Rock, S. L., Ramey, C. T., Barnard, K. E., Gray, C., et al. (1989). Home environment and cognitive development in the first three years: A collaborative study involving six sites and three ethnic groups in North America. *Developmental Psychology, 25*, 217 – 235.

Bronfrenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 1 Theoretical models of human development* (5th ed., pp. 993 – 1028). New York: Wiley.

Bryant, D. M., Burchinal, M. R., Lau, L. B., & Sparling, J. J. (1994). Family classroom correlates of Head Start children's developmental outcomes. *Early Childhood Research Quarterly, 9*, 289 – 309.

Bryk, A., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Bryk, A., & Weisburg, H. (1977). Use of nonequivalent control group design when subjects are growing. *Psychological Bulletin*, *84*, 950 – 962.

Burchinal, M. R., Peisner-Feinberg, E., Bryant, D. M., & Clifford, R. (2000). Children's social and cognitive development and child care quality: Testing for differential associations related to poverty, gender, or ethnicity. *Journal of Applied Developmental Sciences*, *4*, 149 – 165.

Burchinal, M. R., Ramey, S. L., Reid, M. K., & Jaccard, J. (1995). Early child care experiences and their association with family and child characteristics during middle school. *Early Childhood Research Quarterly, 10*, 33 – 61.

Burchinal, M. R., Roberts, J. E., Nabors, L., & Bryant, D. M. (1996). Quality of center infant care and infant cognitive and language development. *Child Development*, *67*, 606 – 620.

Burchinal, M. R., Roberts, J. E., Riggins, R., Zeisel, S., Neebe, E., & Bryant, M. (2000). Relating quality of center child care to early cognitive and language development longitudinally. *Child Development*, *71*, 339 – 357.

Caldwell, B. M., & Bradley, R. H. (1984). *Home observation for measurement of the environment*. Little Rock: University of Arkansas Press.

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M. R., & Ramey, C. (2001). The development of cognitive and academic abilities: Growth curves from an early intervention educational experiment. *Developmental Psychology, 37*, 231 – 242.

Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development*, *65*, 684 – 698.

Campbell, F. A., Ramey, C. T., Pungello, E. P., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science*, *6*, 42 – 57.

Capizzano, J., Adams, G., & Sonenstein, F. (2000). Child care arrangements for children under five: Variation across states. *New Federalism: National Survey of America's Families, Series B (No. B-7)*. Washington, DC: Urban Institute.

Caughy, M. O., DiPietro, J. A., & Strobino, D. M. (1994). Day-care participation as a protective factor in the cognitive development of low-income children. *Child Development*, *65*, 457 – 471.

Clarke-Stewart, K. (1989). Infant day care: Maligned or malignant? *American Psychologist*, *44*, 266 – 273.

Clarke-Stewart, K. A., Gruber, C. P., & Fitzgerald, L. M. (1994). *Children at home and in day care*. Hillsdale, NJ: Erlbuam.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Costa, P. T., & McCrae, R. R. (1985). *The NEO personality inventory manual*. Odessa, FL: Psychological Assessment Resources.

Cronbach, L. J., & Furby, L. (1970). How should we measure ''change''—Or should we? *Psychological Bulletin, 74*, 68 – 80.

Cryer, D., & Burchinal, M. (1997). Parents as child care consumers. *Early Childhood Research Quarterly, 12*, 35 – 58.

Dickinson, D. K., & Smith, M. W. (2001). Supporting language and literacy development in the preschool classroom. In D. K. Dickinson & P. O. Tabors (Eds.),

Beginning literacy with language: Young children learning at home and school. Baltimore: Brookes.
Diggle, P., K. Liang, & S. Zeger (Eds.). (1994). *Analysis of longitudinal data*. Oxford, England: Clarendon Press.
Duncan, G. J., & Gibson, C. (2000). *Selection and attrition in the NICHD Childcare Study's analyses of the impacts of childcare quality on child outcomes*. Evanston, IL: Institute for Policy Research, Northwestern University.
Dunn, L. (1993). Proximal and distal features of day care quality and children's development. *Early Childhood Research Quarterly, 8*, 167 – 192.
Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test – Revised*. Circle Pines, MN: American Guidance Service.
Ehrle, J., Tout, K., & Adams, G. (2000). *Who's caring for our youngest children? Child care patterns for infants and toddlers*. New Federalism: National Survey of America's Families Series. Washington, DC: The Urban Institute.
Fox, N. A., Henderson, H. A., Rubin, K. H., Calkins, S. D., & Schmidt, L. A. (2001). Continuity and discontinuity of behavioral inhibition and exuberance: Psychophysiological and behavioral influences across the first four years of life. *Child Development*, *72*, 1 – 21.
Greenberger, E., Goldberg, W. A., Crawford, T. J., & Granger, J. (1998). Beliefs about the consequences of maternal employment for children. *Psychology of Women Quarterly, 12*, 35 – 59.
Gujarati, D. M. (1995). *Basic econometrics*. New York: McGraw-Hill.
Hart, B., & Risley, T. R. (1995). Meaningful experiences in the everyday experiences of young American children. Baltimore: Brookes.
Helburn, S. W. (Ed.). (1995). *Quality and child outcomes in child care centers*. (Technical report) Denver: University of Colorado, Department of Economics.
Hock, E., Gnezda, M. T., & MacBride, S. (1983, April). *The measurement of maternal separation anxiety*. Paper presented at the biennial meeting of the Society for Research in Child Development, Detroit, MI.
Hoff-Ginsburg, E. (1991). Mother-child conversation in different social classes and communicative settings. *Child Development*, *62*, 782 – 796.
Howes, C. (1997). Children's experiences in center-based child care as a function of teacher background and adult: child ratio. *Merrill Palmer Quarterly, 43*, 404 – 425.
Howes, C. (2000). Social development, family, and attachment relationships of infants and toddlers: Research to practice. In D. Cryer & T. Harms (Eds.), *Infants and toddlers in out-of-home care* (pp. 87 – 116). Baltimore: Brookes.
Howes, C., Phillips, D. A., & Whitebook, M. (1992). Thresholds of quality: Implications for the social development of children in center-based child care. *Child Development*, *63*, 449 – 460.
Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*, 236 – 248.
Infant Health and Development Project. (1990). Enhancing the outcomes of low-birth-weight, premature infants-—A multisite, randomized trial. *Journal American Medical Association*, *263*, 3035 – 3042.
Katz, J. R., & Snow, C. E. (2000). Language development in early childhood: The role of social interaction. In D. Cryer & T. Harms (Eds.), *Infants and toddlers in out-of-home care* (pp. 49 – 86). Baltimore: Brookes.
Kontos, S., Howes, C., Shinn, M., & Galinsky, E. (1995). *Quality in family child care and relative care*. New York: Teachers College Press.
Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, *114*, 497 – 532.
Labouvie, E. W. (1980). Measurement of individual differences in intraindividual changes. *Pychological Bulletin*, *88*, 54 – 59.
Lamb, M. (1998). Nonparental child care: Context, quality, correlates, and consequences. In I. Sigel & A. Renninger (Eds.), W. Damon (Series Ed.), *Handbook of child psychology: Vol 4. Child psychology in practice* (5th ed., pp. 73 – 133). New York: Wiley.
Layzer, J. I., Goodson, B. D., & Moss, M. (1993). *Life in preschool: Final report, Vol. 1. Observational Study of Early Childhood Programs*. Washington DC: U.S. Dept of Education.
McCall, R. B. (1977). Childhood IQs as predictors of adult educational and occupational status. *Science*, 482 – 483.
McCartney, K. (1984). The effect of quality of day care environment upon children's language development. *Developmental Psychology, 20*, 244 – 260.
McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, *71*, 173 – 180.
Medoff-Cooper, B., Carey, W. B., & McDevitt, S. (1993). The Early Infancy Temperament Questionnaire. *Journal of Developmental and Behavioral Pediatrics*, *14*, 230 – 235.
Morris, P. A., Huston, A. C., Duncan, G. J., Crosby, D. A., & Bos, J. M. (2001). *How welfare and work policies affect children: A synthesis of research*. New York: Manpower Demonstration Research Corporation.
National Institute of Child Health and Human Development Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly, 1*, 269 – 306.
National Institute of Child Health and Human Development Early Child Care Research Network. (1997). Familial factors associated with characteristics of nonmaternal care for infants. *Journal of Marriage and the Family, 59*, 389 – 408.
National Institute of Child Health and Human Development Early Child Care Research Network. (1998). *Parental selection of child care in the second and third year.* Washington, DC: Author.
National Institute of Child Health and Human Development Early Child Care Research Network. (1999a). Child outcomes when child-care center classes meet

recommended standards for quality. *American Journal of Public Health*, *89*, 1072 – 1077.

National Institute of Child Health and Human Development Early Child Care Research Network. (1999b, April). *Effect sizes from the NICHD study of early child care.* Paper presented at the biennial meeting of the Society for Research in Child Development, Albuquerque, NM.

National Institute of Child Health and Human Development Early Child Care Research Network. (2000a). Characteristics and quality of child care for toddlers and preschoolers. *Journal of Applied Developmental Sciences*, *4*, 116 – 135.

National Institute of Child Health and Human Development Early Child Care Research Network. (2000b). The relation of child care to cognitive and language development. *Child Development*, *71*, 960 – 980.

National Institute of Child Health and Human Development Early Child Care Research Network. (2002a). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Education Research Journal*, *39*, 133 – 164.

National Institute of Child Health and Human Development Early Child Care Research Network. (2002b). Structure > process > outcome: Direct and indirect effects of caregiving quality on young children's development. *Psychological Science*, *13*, 199 – 206.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77 – 101.

Peisner-Feinberg, E., & Burchinal, M. (1997). Concurrent relations between child care quality and child outcomes: The study of cost, quality, and outcomes in child care centers. *Merrill-Palmer Quarterly, 43*, 451 – 477.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., et al. (2001). The relation of preschool child care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, *72*, 1534 – 1553.

Phillips, D. A., McCartney, K., & Scarr, S. (1987). Child care quality and children's social development. *Developmental Psychology, 23*, 537 – 543.

Phillips, D., Voran, M., Kisker, E., Howes, C., & Whitebook, M. (1994). Child care for children in poverty: Opportunity or inequity. *Child Development*, *65*, 472 – 492.

Pungello, E. P., & Kurtz-Costes, B. (1999). Why and how working women choose child care: A review with a focus on infancy. *Developmental Review, 19*, 31 – 96.

Radloff, L. (1977). The CES – D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385 – 410.

Ramey, C. T., Bryant, D. M., & Suarez, T. M. (1985). Preschool compensatory education and the modifiability of intelligence: A critical review. In D. K. Detterman (Ed.), *Current topics in human intelligence. Vol. 1. Research methodology* (pp. 247 – 296). Norwood, NJ: Ablex.

Raudenbush, S. W. (2001). Toward a coherent framework for comparing trajectories of individual change. In L. Collins & A. Sayers (Eds.), *New methods for the analysis of change* (pp. 35 – 64). Washington, DC: American Psychological Association.

Sameroff, A. J. (1983). Developmental systems: Contexts and evolution. In W. Kessen (Ed.), *Handbook of child development: Vol. 1 History, theories, and methods* (4th ed., pp. 237 – 294). New York: Wiley.

Scarr, S. (1998). American child care today. *American Psychologist*, *53*, 95 – 108.

Schaefer, E. S., & Edgerton, M. (1985). Parent and child correlates of parental modernity. In I. E. Sigel (Ed.), *Parental belief systems: The psychological consequences for children*. Hillsdale, NJ: Erlbaum.

Schliecker, E., White, D. R., & Jacobs, E. (1991). The role of day care quality in the prediction of children's vocabulary. *Canadian Journal of Behavioral Science*, *23*, 12 – 24.

Schweinhart, L. J., Weikart, D. P., & Larner, M. B. (1986). Consequences of three preschool curriculum models through age 15. *Early Childhood Research Quarterly, 1*, 15 – 45.

Singer, J. D., Fuller, B., Keiley, M. K., & Wolf, A. (1998). Early child-care selection: Variation by geographic location, maternal characteristics, and family structure. *Developmental Psychology, 34*, 1129 – 1144.

St.Pierre, R. G., Layzer, J., & Barnes, H. V. (1995). Two-generation programs: Design, cost, and short-term effectiveness. *The Future of Children*, *5(3)*, 76 – 93.

Tomasello, M., & Farrar, J. (1986). Joint attention and early language. *Child Development*, *57*, 1454 – 1463.

Vandell, D. L., & Corasaniti, M. A. (1990). Child care and the family: Complex contributors to child development. In K. McCartney (Ed.), *New directions in child development* (pp. 23 – 38). San Francisco: Jossey-Bass.

Vandell, D. L., & Wolfe, B. (2000). *Child care quality: Does it matter and does it need to be improved?* Madison: Institute for Research on Poverty, University of Wisconsin-Madison.

Wang, C., & Burchinal, M. (2001). *Modeling behavioral science data with unobserved covariate*s. Unpublished manuscript, University of North Carolina.

Woodcock, R. W., & Johnson, M. B. (1990). *Tests of achievement, WJ – R. Examiner's manual*. Allen, TX: DLM Teaching Resources.

Yoshikawa, H. (1995). Long-term effects of early childhood programs on social outcomes and delinquency. *The Future of Children*, *5*(3), 51 – 75.