A Tale of Two Methods: Comparing Regression and Instrumental Variables Estimates of the Effects of Preschool Child Care Type on the Subsequent Externalizing Behavior of Children in Low-Income Families

By: Danielle A. Crosby, Chantelle J. Dowsett, Lisa A. Gennetian, and Aletha C. Huston

Crosby, D.A., Dowsett, C.J., Gennetian, L.A., & Huston, A.C. (2010). A tale of two methods: Comparing regression and instrumental variables estimates of the effects of preschool child care type on the subsequent externalizing behavior of children in low-income families. *Developmental Psychology, 46*, 1030-1048. DOI:10.1037/a0020384.

*****Note: Made available courtesy of the American Psychological Association. This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.**
Link to Journal: http://www.apa.org/pubs/journals/dev/index.aspx

Abstract:
We apply instrumental variables (IV) techniques to a pooled data set of employment-focused experiments to examine the relation between type of preschool childcare and subsequent externalizing problem behavior for a large sample of low-income children. To assess the potential usefulness of this approach for addressing biases that can confound causal inferences in child care research, we compare instrumental variables results with those obtained using ordinary least squares (OLS) regression. We find that our OLS estimates concur with prior studies showing small positive associations between center-based care and later externalizing behavior. By contrast, our IV estimates indicate that preschool-aged children with center care experience are rated by mothers and teachers as having fewer externalizing problems on entering elementary school than their peers who were not in child care as preschoolers. Findings are discussed in relation to the literature on associations between different types of community-based child care and children's social behavior, particularly within low-income populations. Moreover, we use this study to highlight the relative strengths and weaknesses of each analytic method for addressing causal questions in developmental research.

Article:
Changes in U.S. policy toward low-income families over the past two decades have contributed to unprecedented employment rates among single mothers and equally dramatic increases in the use of nonparental care for young children (Schoeni & Blank, 2000). An estimated one million low-income children under the age of 6 years entered regular care arrangements in the wake of welfare reform (Fuller & Kagan, 2000). Interest in identifying the effects of various early care and education settings on child well-being has intensified amidst mounting evidence that the first five years of life form an important foundation for later development (Chase-Lansdale & Votruba-Drzal, 2004; Shonkoff & Phillips, 2000).

Numerous studies suggest that center-based preschool programs can promote school readiness, especially among low-income children (Gormley, Gayer, Phillips, & Dawson, 2005; Love et al., 2003; Magnuson, Ruhm, & Waldfogel, 2007; NICHD Early Child-care Research Network,

2005), prompting many states to expand their efforts to fund such programs (see Magnuson & Waldfogel, 2005). Long-standing concerns exist, however, about the consequences of nonmaternal care—particularly center care—for children's socioemotional development (see Belsky, 2001). Several large-scale studies have demonstrated a link between center care experience and behavior problems during middle childhood (Halle et al., 2006; Magnuson, Meyers, Ruhm, & Waldfogel, 2004; NICHD Early Childcare Research Network, 2004), and early childhood professionals have reported increased rates of behavior difficulties in structured settings (Gilliam, 2005). Yet other studies have failed to find an association between center care and problem behavior, particularly within low-income samples (Loeb, Fuller, Kagan, & Carrol, 2004; Votruba-Drzal, Coley, & Chase-Lansdale, 2004).

Isolating whether developmental effects are attributable to child care or other factors is a major challenge in this work. Many of the empirical techniques used in developmental science make it difficult to separate correlation from causation (Blau, 1999; Duncan, Magnuson, & Ludwig, 2004). In this study, we evaluated the strengths and weaknesses of two methods for estimating the effects of different preschool child care settings on children's subsequent behavior. Specifically, we compare the typical approach of ordinary least squares (OLS) regression, controlling for a set of co variates, with instrumental variables (IV) estimation. By exploiting exogenous variation in families' use of different types of care—this case, arrangements induced by experimental treatments—IV controls for biases that arise when child or family characteristics affect both care selection and child functioning. In addition to its methodological contribution, this study informs the literature on the relation of community-based child care to young children's social behavior within low-income populations.

THE PROBLEM OF ENDOGENEITY: TWO ANALYTICAL APPROACHES

Ordinary Least Squares Regression

OLS regression is a common analytic tool in social science research and, under classical assumptions, produces the best un biased linear estimates (Berry, 1993). If these assumptions are violated, however, predictors can be correlated with the error term (or *endogenous*), resulting in biased parameter estimates. For example, omitting variables related to both the predictor and the outcome from the analytic model induces a spurious correlation between them. This is a significant concern in the child care literature given that the determinants of care and child well-being are not randomly distributed across families (Burchinal & Nelson, 2000; Cleveland, Wiebe, van den Oord, & Rowe, 2000; Lamb, 1998; National Institute of Child Health and Human Development, Early Childcare Research Network & Duncan, 2003). Such factors as parental education, income, ethnicity, and beliefs are associated with child care quality, quantity, and type, as well as with children's development (Early & Burchinal, 2001; Fuller, Kagan, Caspary, & Gauthier, 2002; Huston, Chang, & Gennetian, 2002; NICHD Early Childcare Research Network, 2004; Pungello Kurtz-Costes, 1999). In addition, *simultaneity bias* is possible the association between child care and behavior is attributed entirely to the effect of care on behavior when the relationship actually bidirectional in nature (Duncan et al., 2004). For example, parents' selection of child care likely reflects their assessment children's behavior or maturity, and parents of children with extensive behavior problems often face considerable constraints their care options (Gilliam, 2005).

The most common method for addressing endogeneity concerns in child care research (and developmental science in general) multivariate regression with controls for a set of observed characteristics. This technique has several merits and is often driven by strong conceptual models; however, it is impossible to guarantee empirically that all potential sources of bias are accounted for, particularly those that are difficult to measure or observe. Moreover, the extensive use of covariates can approximate situations that do not exist in the real world and may be problematic if covariates are determined by, rather than determinants of, an outcome (Duncan et al., 2004; Newcombe, 2003).

Instrumental Variables Analysis

Instrumental variable (IV) techniques, commonly used in the field of economics, have the potential to remove endogeneity bias from regression estimates (Angrist, Imbens, & Rubin, 1996; Moffitt, 2005). The central strategy in IV estimation is to find a variable, or "instrument," that produces exogenous variation in the predictor of interest, variation that can then be used to cleanly estimate the relationship between the predictor and outcome. IV has been used only occasionally in developmental research (see Arnold, McWilliams, & Arnold, 1998; Foster & McLanahan, 1996; Gennetian, Magnuson, & Morris, 2008; Loeb et al., 2004; Magnuson et al., 2004), primarily as a result of challenges in finding viable instruments. In this study, we required an instrument that predicted child care type but was otherwise uncorrelated with child functioning.

A relatively novel application of IV with random-assignment data from social policy experiments has tremendous potential for addressing casual questions about development and family process and resolves some of the challenges faced in prior IV work (Gennetian, Morris, Bos, & Bloom, 2005; Riccio & Bloom, 2002). Random assignment status is an ideal candidate for use as an instrument, because it is, by definition, exogenous to child, family, and community characteristics (Angrist et al., 1996). Several studies have used this method to explore mechanisms by which intervention effects occur (Gennetian, Crosby, Dowsett, Huston, & Principe, 2006; Gibson-Davis, Magnuson, Gennetian & Duncan, 2005; Liebman, Katz, & Kling, 2004; Ludwig, Duncan, & Hirschfield, 2001; Magnuson & McGroder, 2002; Morris, Duncan, & Rodrigues, 2006; Morris & Gennetian, 2003). Here, we applied this technique to data from several welfare and employment experiments to assess the effects of child care arrangements induced by the various policies tested in these studies. We used this analysis to illustrate the conditions, assumptions, and hazards of IV estimation and its potential utility for addressing developmental questions.

Two-stage least squares estimation (2SLS) is the most common form of IV analysis. In the first stage, the endogenous predictor (in this case, type of care) is regressed on the instrument (random assignment status) and a set of covariates to obtain coefficients that reflect the amount of variation in child care attributable to program group membership. The first-stage coefficients are used to generate predicted values (probabilities) for each type of care for each child. Free from selection bias, these values are then used in the second stage equation to obtain a "clean" estimate of the relation of child care type to behavior; that is, child behavior is regressed on the predicted values and covariates from the first-stage. The substitution of predicted scores for actual scores in the second stage requires an adjustment to the standard errors; this correction is included in most statistical packages offering 2SLS estimation.[1]

There is an important distinction between the effects captured by OLS regression and IV estimation. Whereas OLS estimates rely on all of the natural variation that exists across the entire sample, IV estimates are derived only from the variation attributable to the (exogenous) instrument—in this case, parents who were induced by the experiment to use care arrangements they would not have otherwise used. For example, random assignment to a treatment that includes additional resources for child care may induce parents to use arrangements they would not be able to afford in the absence of the experiment. In economics, this effect is referred to as the local average treatment effect and differs from both intent-to-treat estimates, which measure overall program-control group differences regardless of take-up or participation rates, and treatment-on-the-treated estimates, which capture effects for those who actually receive the treatment (Angrist et al., 1996; Moffitt, 2005).

One of the benefits of instrumental variable techniques is the tendency to generate parameter estimates that are consistent (i.e., convergent to population parameters as sample size grows), an improvement on conventional OLS estimates, which are never consistent in the presence of endogeneity regardless of sample size. At the same time, IV estimates are not impervious to bias, particularly in the context of very small samples (Angrist & Krueger, 2001; Gennetian et al., 2005). The amount of bias depends largely on how well several key assumptions of IV are met (Angrist et al., 1996; Gennetian, Magnuson, & Morris, 2008). First, instruments must represent a truly exogenous source of variation and must have a meaningful effect on both the endogenous variable and the outcome of interest. Instrumental variable analyses (with experimental or nonexperimental data) often suffer from low power to detect effects because of a weak relationship between the instrument and endogenous predictor. Weak instruments tend to produce estimates that are unreliable and biased toward the OLS estimate (Bound, Jaeger, & Baker, 1995).

A second assumption, the stable unit treatment value assumption (SUTVA) requires that there be no overall community or displacement effects and that individuals within a particular program group have equal exposure to the treatment. In this study, we assume that individual values of child care and child behavior were unaffected by others' values on these variables, a reasonable assumption given that the number of participants in any one community was relatively small. It is unlikely that crowding out occurred in child care availability or that a single care arrangement or classroom contained multiple sample members. (For other scenarios in which SUTVA would not hold, see discussion in Sobel, 2006).

Well-designed and well-implemented social policy experiments satisfy the aforementioned assumptions (Gennetian et al., 2005). Two additional assumptions are more difficult to meet and deserve further discussion. Instrumental variables are assumed to move the endogenous predictor in one direction only (i.e., effects are *monotonic*). In an experimental context, the expectation is that participants do not behave opposite to their group assignment; for example, we assume that program group parents were not induced by the treatments to use less child care than they would have if they belonged to the control group. Although this assumption should be taken seriously, simulations in which participants act counter to their assignment suggest that violations do not substantively alter our findings.

Finally, and perhaps most important to our analysis, the *exclusion restriction* requires that an instrumental variable be unrelated to the outcome variable except through its effect on the predictor being instrumented— that is, that it be totally mediated. In our case, instrumenting for type of care alone implies that all of the program effects on child behavior occur through their impacts on type of care. This assumption is untenable given that the experiments were designed primarily to influence parents' economic outcomes and may also have affected maternal well-being and parenting. Our analysis therefore instruments for these variables as well. IV estimation requires at least one instrument for each endogenous regressor; moreover, to produce meaningful results, the instruments should have differential effects on the variables in question (see Gennetian et al., 2008). To disentangle the developmental effects of child care and income, for example, one ideally would have access to an instrument predicting child care (but not income) and another instrument predicting income (but not child care).

Studies with multiple-group research designs (i.e., multiple treatments) and those with site differences in impacts can yield multiple instruments (Gennetian et al., 2005). Pooling microdata across random assignment studies testing different policies is another viable option for accessing multiple instruments with the added benefit of increasing sample size. In the current study, we pool across 21 experiments to estimate the effects of center- and home-based child care on children's behavior net of the effects of maternal employment, earnings, income, and depressive symptoms. The mechanics of this technique are delineated below.

RELATIONS OF TYPE OF CARE TO SOCIAL BEHAVIOR

Rich longitudinal data on children's experiences in care (e.g., the NICHD Study of Early Child Care) have greatly advanced the field by simultaneously examining amount, timing, quality, and stability of care. These dimensions vary greatly within any type of care, but several structural differences between center- and home-based settings may affect social development independently of other aspects of care. As compared with home-based care, centers are more likely to have a predictable daily schedule, a structured curriculum, an environment designed for children, trained caregivers, larger groups of similarly aged children, and higher ratios of children to adults (e.g., Fuller, Kagan, Loeb, & Chang, 2004; Kisker, Hofferth, Phillips, & Farquhar, 1991; Kontos, Hsu, & Dunn, 1994). Larger group sizes can provide greater opportunity for interaction with same-age peers, a potentially important experience for the development of social competence. At the same time, large groups can also increase the likelihood of interpersonal conflict and reduce the opportunity for interaction with adults (Dowsett, Huston, Imes, & Gennetian, 2008).

Overall differences in care quality may also contribute to different effects for center- and home-based arrangements. Although both types of settings exhibit a wide range of quality (e.g., Fuller et al., 2004), observational studies of settings serving low-income samples indicate higher average quality for centers than for home-based care, particularly with respect to cognitive stimulation (Coley, Chase-Lansdale & Li-Grining, 2001; Dowsett et al., 2008; Fuller et al., 2004; Loeb et al., 2004). Indeed, numerous studies report better school readiness skills among children of all income levels who attend centers during the preschool years (see Magnuson & Waldfogel, 2005).

Findings regarding children's social behavior, however, have been mixed. On the one hand, high-quality, center-based preschool programs for low-income children can promote social skills (Barnett, 1995; Schweinhart & Weikart, 1997; Yoshikawa, 1995). In the national Head Start Impact Study, three-year-olds assigned to Head Start were rated by their mothers as having fewer behavior problems than children in the control group, though no differences were found for four-year-olds (U.S. Department of Health and Human Services, Administration for Children and Families, 2005). Two nonexperimental comparisons of different types of community-based child care for low-income children show no relation between center attendance and maternal ratings of problem behavior or social competence, controlling for care quality and family selection factors (Loeb et al., 2004; Votruba-Drzal et al., 2004).

On the other hand, analyses of community-based child care for two large samples of children linked center experience and negative social behavior. Children in the NICHD Study of Early Child Care who spent more time in centers during their early years were rated higher on externalizing problems by teachers at age 4 and 1/2 years and throughout elementary school (Belsky et al., 2007; National Institute of Child Health and Human Development, Early Childcare Research Network, 2004). Subsequent analyses indicated that the number of other children in the setting rather than center care per se predicted externalizing problems; moreover, several approaches to testing the causal effects of child care within this sample have produced mixed results (McCartney et al., 2010). Nationally representative data from the Early Childhood Longitudinal Study—Kindergarten Cohort (ECLS–K) also indicated higher levels of teacher-rated externalizing behavior among children who attended center-based programs prior to school entry (Halle et al., 2006; Loeb, Bridges, Bassok, Fuller, & Rumberger, 2007; Magnuson et al., 2007).

THE PRESENT STUDY

The present study has a dual purpose. First, we compare two analytic methods—OLS regression and IV estimation—for estimating the relations between child care type and externalizing problems in order to illustrate the strengths and weaknesses of each. Second, we seek to use these analyses to inform the literature on the developmental effects of different care settings for children in low-income, single-mother families experiencing welfare-to-work policies. Both purposes uniquely contribute to child care research.

Debate about child care effects on behavior has focused primarily on externalizing problems, in part because demonstrated relations of child care to externalizing behavior are pronounced and also because it is a potential precursor to deviant and aggressive behavior. Yet, evidence that group care settings may contribute to stress and anxiety in young children (Ahnert, Gunnar, Lamb, & Barthel, 2004; Watamura, Dowzella, Alwin, & Gunnar, 2003) suggests that internalizing problems should not be overlooked. In the current study, we conducted all analyses for both externalizing and internalizing behavior but concentrated on the former given space constraints and concerns about precision in the estimates of the latter (full results available on request). Measures of child behavior typically vary across reporters and contexts (e.g., home and school). Although maternal ratings provide valuable information, they are confounded with maternal well-being. Mothers with high levels of depressive symptoms tend to rate children higher on problem behaviors than do other mothers (Loeb et al., 2004; NICHD Early Childcare Research Network, 1999; Yeung, Linver & Brooks-Gunn, 2002). Moreover, mothers' and

teachers' reports of child behavior tend to be only modestly correlated (usually in the .20–.30 range), and teacher ratings generally are not associated with maternal depression (Chang, 2003; Eisenberg et al., 2001; Runions & Keating, 2007; West, Denton, & Germino-Hausken, 1999). In this study, we examined both mother and teacher reports and controlled for maternal depressive symptoms.

METHOD

Data Sources

We use data from six random-assignment studies of welfare and employment programs for low-income parents (see Table 1). Because some studies included multiple sites and others had multiple treatments, these studies provide information on 21 programs: Connecticut Jobs First, New Hope, two tests of the Minnesota Family Investment Program (MFIP) in urban counties, one test of the MFIP in rural counties, three tests of labor force attachment models in National Evaluation of Welfare-to-Work Strategies (NEWWS), one test of Canada's Self-Sufficiency Program (SSP) in British Columbia, two tests of SSP in New Brunswick, and 10 sites in the New Chance study. Although a common set of policies was tested across the 16 sites in New Chance, prior work indicates that program impacts on child care use varied by site (Yoshikawa, Rosman, & Hsueh, 2001). To capitalize on this variation in effects, we analyzed the New Chance data at site level, excluding six sites that had fewer than 10 observations in our targeted child age range. In sensitivity analyses, we include New Chance as a single instrument (collapsing across sites) to test the robustness of our results.

Table 1: Study and Program Descriptions

Study	Programs/Sites	Key program features	Primary Sources
Connecticut Jobs First Evaluation	New Haven and Manchester, CT	Welfare recipients required to enroll in employment and training services, with a generous financial incentive offered in the form of income disregard; a 21-month time limit imposed on receipt of benefits.	Bloom et al., 2002
New Hope Project	Milwaukee, WI	Earnings supplements, child care subsidies, and health care subsidies offered to low-income adults working 30+ hr per week.	Bos et al., 1999
Minnesota Family Investment Program (MFIP)	MFIP–urban (three counties) MFIP–rural (three counties) MFIP incentives only–urban	Long-term welfare recipients to participate in employment services; financial incentive provided through an income disregard. Financial incentive offered for employment with no requirement for employment and training services.	Gennetian & Miller, 2000
National Evaluation of Welfare-to-Work Strategies (NEWWS), Labor Force Attachment (LFA)	NEWWS–Atlanta LFA NEWWS–Grand Rapids LFA NEWWS–Riverside LFA	Participants required to look for work immediately, usually through a job club that lasted from 1–3 weeks. Those unable to find a job were often enrolled in adult basic education, vocational training, or work experience.	Hamilton et al., 2001
Canada's Self-Sufficiency Project (SSP)	SSP–New Brunswick SSP–British Columbia SSP Plus–New Brunswick	Earnings supplement offered to long-term welfare recipients who left welfare for full-time work within a year of program entry. SSP Plus offered employment services in addition to earnings supplements.	Michalopoulos et al., 2002
New Chance Evaluation	Ten sites in eight states[a]	Mix of educational, personal development, employment-related, and support services provided to 16- to 22-year-old mothers receiving welfare.	Quint et al., 1997

[a]Sites were located in Allentown, Pennsylvania; Bronx, New York; Chula Vista, California; Denver, Colorado; Harlem, New York; Lexington, Kentucky; Minneapolis, Minnesota; Philadelphia, Pennsylvania; Pittsburgh, Pennsylvania; and San Jose, California.

Key policies tested in these experiments included the following: (a) earnings supplements, which provided cash bonuses for a set level of employment or allowed continued welfare receipt as earnings increased; (b) mandatory employment services that required participants to engage in job search activities or move directly into employment; (c) time limits on welfare receipt; and (d)

enhanced child care assistance, which offered families additional resources to facilitate the use of licensed care. Although none of the programs constituted direct interventions for children, developmental impacts could occur through changes in family economic circumstances, family functioning, and experiences in various care environments.

Program impacts on parents' economic outcomes and child care use for the larger samples in these studies are well documented (D. Bloom & Michalopoulos, 2001; Gennetian, Crosby, Huston, & Lowe, 2004; Schoeni & Blank, 2000). Earnings supplements increased income and employment, whereas employment mandates in isolation typically increased only the latter because additional earnings were offset by lost welfare benefits. Programs offering enhanced child care assistance increased the use of center-based child care; programs without enhanced child care assistance increased the use of home-based care (Crosby, Gennetian, & Huston, 2005).

Our pooled data set has several advantages, including random assignment data, comparable measures across studies, longitudinal follow-ups of children into elementary school, and a large sample (N = 3,290). Ideally, the data for this study would be derived from one large random assignment evaluation across multiple sites and testing numerous treatments. Pooling microdata from several experiments has been described as the next best option (Gennetian et al., 2005; Riccio & Bloom, 2002). We determined this to be a reasonable approach with the current set of studies given their comparability in design, sampling, and measurement. Each study included a survey approximately two to three years after random assignment as well as administrative records on employment, earnings, and welfare receipt throughout the follow-up period Across the experiments, similar measures of baseline socioeconomic and demographic characteristics and subsequent employment, income, earnings, parent psychological well-being, child care, and children's behavior problems were used. Our analyses controlled for a variety of baseline characteristics and study/site indicators to account for any unmeasured time-invariant characteristics of particular programs.

The current data set has already generated useful information about the well-being of low-income children across various sites in the United States and Canada (Clark-Kauffman, Duncan, & Morris, 2003; Gennetian, 2004).

Participants

In each study, one or two "focal" children per family were selected to be evaluated in the follow-up interview. Our analysis sample was limited to focal children who were 3 years 0 months to 4 years 11 months old at the beginning of the child care data collection period. We focus on this age group because children of this age are likely to spend more time in care compared with school-age children. We exclude children younger than three years of age because the studies lack adequate data for this age group. Furthermore, issues surrounding care (e.g., availability, quality, parental preferences) and its effects on development are likely to differ for infants and toddlers in comparison with preschoolers.

Descriptive information for the analysis sample appears in Table 2. Baseline characteristics of parents and children by study are presented in Appendix A. As expected with random assignment, any differences between program and control groups (8 of the 156 contrasts shown were statistically significant at the $p < .05$ level) can be attributed to chance. Survey response

rates were quite high (from 75% to 85%) given the disadvantaged nature of this sample; attrition analyses appear in the individual reports. Approximately 15% of our preschool-aged sample lacked child care and/or social behavior data, more than a third of which is attributable to a priori design decisions in two of the studies. New Hope participants completed behavior ratings for only one of their (randomly selected) children even though up to two children per household were included in the sample. In SSP, participants were asked about child care for their youngest child only. Rather than impute values for variables with incomplete item-level data, our main analyses were restricted to cases with valid child care and behavior data. All cases in this subsample had complete baseline information (standard forms were administered at study entry, usually in the context of applying for or receiving public assistance); more than 99% had administrative economic data, and 97% had valid data on maternal depressive symptoms.

Table 2: Descriptive Characteristics of Sample

Variable	Full sample (N = 3,290) M (SD)	Program group (N = 1,601) M (SD)	Control group (N = 1,689) M (SD)	Significance of –program–control difference[a]
Measured at or before study entry				
Parent characteristics				
Age (years)	27.1 (5.8)	26.7 (5.7)	27.5 (5.8)	$p < .01$
Race (%)				
Black	37.5	36.6	38.4	
White	45.9	47.3	44.6	
Latino	11.3	10.6	12.0	
Other	5.2	5.5	4.9	
Never married (%)	65.3	66.7	64.0	
Parent education, employment, and income				
High school graduate (%)	59.5	58.8	60.0	
Employed in year prior to random assignment (%)	41.5	40.4	42.5	
Earnings in year prior to random assignment ($)	2,090	2,003	2,174	
On Aid to Families With Dependent Children for 2 or more years prior to random assignment (%)	0.7	0.7	0.7	
Family composition				
No. of children in family	1.9 (1.1)	1.9 (1.1)	2.0 (1.1)	
Age of child (years)	3.4	3.4	3.5	
Child is male (%)	50.3	50.0	50.6	
Measured during 2-year follow-up				
Child care outcomes				
Any care (%)	79.2	82.6	76.0	$p < .01$
Only center-based care (%)	21.0	20.8	21.3	
Only home-based care (%)	24.1	25.9	22.3	$p < .05$
Mix of center- and home-based care (%)	33.9	35.9	31.9	$p < .05$
Economic outcomes				
Average quarterly employment (%)	44.3	49.0	39.9	$p < .01$
Average quarterly income ($)	2,903.1	3,042.1	2,771.7	$p < .01$
Average quarterly earnings ($)	1,120.5	1,206.7	1,038.9	$p < .01$
Maternal depressive symptoms, CES–D	14.2 (11.8)	14.8 (11.9)	13.7 (11.7)	$p < .01$

Note. CES–D = Center for Epidemiological Studies—Depression scale; range = 0–60, with higher numbers indicating greater depressive symptomatology.
[a]Two-tailed t tests were applied to differences between the experimental and control group covariates.

To further consider the potential implications of missing data for our estimates, we examined patterns of missing data; conducted sensitivity analyses excluding studies with higher amounts of missing data (described in more detail below); and, re-ran our OLS analyses using multiple imputation methods. In this final check, we used the Imputation by Chained Equations program in STATA 10.0) to impute values for missing child care, child behavior, maternal depression, and economic outcome data based on an extensive list of baseline covariates including study site.

All of the strategies described above indicated no qualitative or significant differences in the estimates.

Measures

Family and child baseline covariates. The following demographic and socioeconomic indicators assessed at baseline (i.e., prior to random assignment) were included as covariates in the analyses: child's age and gender, mother's age and ethnicity, number of children in the household, mother's marital history, mother's completion of high school or a graduate equivalency degree, mother's employment and earnings in the prior year, and family welfare receipt of two or more years. We also included a series of dummy variables representing each study site.

Child care experiences. As part of the follow-up surveys, mothers provided retrospective reports of child care use for the prior 18 months to 2 years. Those who used at least one regular care arrangement (10+ hr per week) at any point during this period were asked additional questions about each arrangement. The child care measure covered all or most of the period between random assignment and the follow-up assessment, with one exception: In New Chance, detailed child care information was collected for the first 18 months of the study (when children were approximately 3 to 5 years old), whereas behavior outcomes were measured during the 42-month interview (when children were approximately 5 to 7 years old).

We categorized care arrangements into center-based care (including child care centers, preschool programs, and Head Start), and home-based care (including unregulated care by relatives nonrelatives and family child care homes that may be licensed certified). Our main analyses compared outcomes for children who experienced four mutually exclusive patterns of care: (a) only center-based care, (b) only home-based care, (c) a mix of center and home-based arrangements, or (d) no nonmaternal care over the follow-up period. Exclusive center- or home-based care may proxy for more exposure to that type of care or may indicate more stable patterns of care than mixed arrangements. The frequency of each type of care is shown in Table 2. Almost 80% of preschoolers this sample were in some form of child care during the assessment period; one third of these children experienced mixed care, 21% were in center-based care exclusively, and 24% were in home based arrangements exclusively.

Four of the studies (Connecticut Jobs First, MFIP, New Chance, and New Hope) collected information about duration of care arrangements; we used these data to examine the association between amount and type of care. Children who experienced center care exclusively and home care exclusively were in care for approximately 60% of the follow-up period. Children in mixed care were in center and home care (either concurrently or consecutively) for 45% and 52% of the follow-up, respectively. Assessments of care quality and information about licensing were unavailable in these studies. Zero-order relationships between the baseline covariates and types of care are shown in Table 3. Correlations were very similar for program and control groups (not shown).

Table 3: Correlation of Predictor Variables and Baseline Covariates With Behavioral Outcomes

Variable	Only center-based care	Only home-based care	Mix of center- and home-based care
Measured at or before study entry			
Parent characteristics			
Age	0.05**	−0.06**	−0.1**
Race			
Black	0.16**	−0.12**	0.09**
White	−0.14**	0.10**	−0.06**
Latino	−0.01	0.00	0.02
Other	−0.03	0.03	−0.08**
Marital status			
Never married	0.02	0.00	−0.01
Parent education, employment, and income			
High school graduate	0.01	0.01	0.09**
Employment during prior year	−0.04*	0.09**	0.13**
Earnings during prior year	−0.02	0.04*	0.10**
Receiving Aid to Families With Dependent Children for 2 or more years prior to random assignment	−0.02	0.00	−0.08**
Family composition			
No. of children in family	0.04*	0.00	−0.06**
Age of child	0.11**	−0.10**	0.03
Child is male	0.00	0.02	−0.01
Measured during 2-year follow-up			
Economic outcomes			
Average quarterly employment	−0.07**	0.16**	0.19**
Average quarterly income	−0.02	0.02	0.14**
Average quarterly earnings	−0.04*	0.08**	0.16**
Maternal CES–D score	−0.05**	0.04*	0.07**

Note. CES–D = Center for Epidemiological Studies—Depression scale; range = 0–60, with higher numbers indicating greater depressive symptomatology.
*$p < .05$. **$p < .01$.

Parent economic and psychological well-being during the follow-up

Employment, income, and earnings. Unemployment insurance and public assistance records provided information on mothers' employment and earnings from the time of random assignment through the follow-up period. From these records, we constructed average quarterly employment, average quarterly income, and average quarterly earnings across the quarters for which we had child care data. The income measure was the sum of welfare income (from public assistance records), earnings, and earnings supplements when appropriate. Income from other household members or noncustodial parents (e.g., child support) was not included.

Maternal depressive symptoms. As part of the follow-up interview, parents completed a version of the Center for Epidemiological Studies—Depression scale (CES–D; Radloff, 1977). The CES–D assesses the frequency of such symptoms as crying or feeling lonely in the prior week. Two of the studies used a subset of the original 20 items, so standardized *z* scores (created within study) were used in the analysis.

Children's externalizing behavior. Maternal ratings of children's externalizing behavior were collected at follow-up in each study with three comparable scales—items from the Behavior Problem Index (Achenbach & Edelbrock, 1981; Peterson & Zill, 1986) in the Connecticut Jobs First, MFIP, New Chance, and NEWWS studies; the Problem Behavior Scale of the Social Skills Rating System (Gresham & Elliot, 1990) in New Hope; and a four-item Externalizing subscale (Morris & Michalopoulos, 2000) in SSP. For the pooled data set, we converted behavior scales to standardized *z* scores using study-specific control group means and standard deviations. Three studies (Connecticut Jobs First, New Chance, and New Hope) collected teacher ratings of

externalizing behavior for children who had entered school by the time of the follow-up assessment (N = 379) with measures that were similar or identical to those completed by parents. Maternal and teacher reports of externalizing behavior were moderately correlated, $r(366)$ =.34, p <. 001. Zero-order correlations of the covariates and predictors with mother- and teacher-reported outcomes are shown in Table 4.

Table 4: Correlation of Predictor Variables and Baseline Covariates With Externalizing Behavior

	Externalizing behavior	
Measured at or before study entry	**Parent report**	**Teacher report**
Parent characteristics		
Age	−.01	−.17**
Race		
Black	−.01	.10
White	.03	−.09
Latino	−.04*	−.02
Other	.00	−.01
Marital status		
Never married	.01	.05
Separated/divorced		
Parent education, employment, and income		
High school graduate	−.05**	−.03
Employment during prior year	−.02	.12*
Earnings during prior year	−.03	−.02
Receiving Aid to Families With Dependent Children for 2 or more years prior to random assignment	.08***	−.02
Family composition		
No. of children in family	.06**	−.11
Age of child	−.01	.10
Child is male	.11**	.20**
Measured during 2-year follow-up		
Maternal CES-D score	.22***	.02
Child care outcomes		
Only center-based care	.02	−.02
Only home-based care	−.05**	−.03
Mix of center- and home-based care	.03	.07
Economic outcomes		
Average quarterly employment	−.05**	−.03
Average quarterly income	−.04*	−.02
Average quarterly earnings	−.04*	−.04

*p < .05. **p < .01.

Analytic Plan

We performed parallel OLS and IV analyses estimating the effects of different types of preschool child care on mother and teacher reports of children's externalizing problems during the early school years. In all of the analyses, behavior ratings were compared for children who experienced only center care, only home-based care, mixed care, or no regular child care during the study period. The no-care group served as the omitted category because it is considered the norm for the population relevant to these studies (i.e., single mothers receiving public assistance prior to federal welfare reform).

Ordinary least squares regression models. In the first OLS model, we predicted maternal ratings of externalizing behavior as a function of the types of care children experienced during the follow-up period, controlling for baseline (pretreatment) covariates. A second model included maternal depressive symptoms, employment, income, and earnings as additional predictors. A similar pair of OLS models was evaluated for teacher ratings of child behavior; however, because teacher data are available in only three studies and for a relatively small

number of children, these models contain a reduced list of covariates and include maternal earnings as the only additional predictor. Three covariates (mother's age, mother never married, and employment in year prior to study entry) as well as maternal depressive symptoms were omitted because they contribute little to the prediction of teacher-rated behavior.

Indicators of experimental treatment status were not included in the OLS models. To detect systematic differences by treatment group status in the relations between care and behavior, we conducted OLS analyses for the program and control groups separately as well as jointly and found no notable differences (results not shown).

Instrumental variables analysis. We used instrumental variables (with 2SLS) to evaluate a set of models similar to those described above for OLS. Instead of relying on naturally occurring variation in child care, however, the IV models exploited experimentally induced variation to estimate the relationship between type of care setting and children's behavior.

In the first-stage equation, each child care category (i.e., exclusive center care, exclusive home-based care, and mixed care) was regressed on a set of baseline covariates and 21 dummy variables representing random assignment status (experimental = 1; control = 0) for each of the programs/sites. Thus, program–control group differences in the occurrence of each type of care within a given study/site were used to generate predicted child care values for children in that study; these values reflect the likelihood of children experiencing each type of care given their family's assignment to the program or control group in a particular study. Differential impacts on child care use across the studies created variability in the predicted levels for individual children. The second-stage IV equation was identical to the OLS regression with the exception that predicted child care values from the first stage were used instead of the observed values. To estimate these models, we use the "ivreg" procedure in STATA, which automatically provides the correct second-stage standard errors (assuming that linearity in the dependent variable does not affect our results).

As in the OLS analysis, we first tested the effects of child care type and then evaluated a second model that considers the potential role of maternal depressive symptoms, employment, income, and earnings. To satisfy the exclusion restriction in IV, we instrumented for these additional predictors in the same way we instrumented for child care type (see Gennetian et al., 2008). We took advantage of 21 unique instruments available in the pooled data set. The IV models with teacher data included a reduced set covariates and instrument for type of care and maternal earnings only. Detailed empirical specifications for the IV models are provided in Appendix B.

RESULTS

Individual Study Impacts

We began by examining individual study impacts on child care, employment, income, earnings, and depressive symptoms (shown as program-control group differences in Table 5) and on mother and teacher ratings of children's externalizing behavior (shown Table 6). These "reduced form" effects underlie our ability estimate the IV models. The important information in these tables is not whether there were impacts in individual studies but whether there was variation in impacts on both child care type and behavior across studies—variation that allowed us to

estimate the effects program-induced variability to variation in children's behavior Indeed, we found that our instruments statistically related to the predictors of interest.

Table 5: Reduced Form Effects of Programs on Child Care, Employment, Income, Earnings, and Scores on the Center for Epidemiological Studies—Depression Scale (CES–D)

Variable	Only center-based care (%) β (*SE*)	Only home-based care (%) β (*SE*)	Mix of center- and home-based care (%) β (*SE*)	Average quarterly employment (%) β (*SE*)	Average quarterly income (%) β (*SE*)	Average quarterly earnings (%) β (*SE*)	CES–D β (*SE*)
Connecticut Jobs First	1.37 (3.18)	1.59 (5.04)	6.66 (4.78)	0.01 (0.04)	−0.18 (0.18)	−0.23 (0.20)	0.13 (0.11)
New Hope	13.25 (8.52)	−14.65 (7.35)	9.51 (8.77)	0.13 (0.05)	0.58 (0.25)	0.49 (0.27)	−0.15 (0.18)
MFIP–urban	−3.71 (3.89)	−0.39 (5.36)	4.99 (5.50)	0.10 (0.04)	0.30 (0.15)	0.00 (0.19)	0.01 (0.11)
MFIP incentives only–urban	−3.66 (4.07)	−3.91 (5.55)	6.93 (5.88)	0.03 (0.04)	0.29 (0.15)	−0.19 (0.19)	0.06 (0.12)
MFIP–rural	12.66 (6.10)	−12.36 (8.54)	1.79 (9.07)	0.04 (0.06)	0.26 (0.21)	−0.14 (0.27)	0.34 (0.18)
NEWWS–Atlanta LFA	7.03 (4.17)	5.15 (2.51)	−6.00 (3.68)	0.03 (0.03)	−0.01 (0.08)	0.05 (0.09)	0.01 (0.08)
NEWWS–Grand Rapids LFA	−1.25 (4.07)	−0.25 (4.46)	1.85 (5.72)	0.14 (0.04)	0.05 (0.11)	0.40 (0.12)	0.26 (0.13)
NEWWS–Riverside LFA	−5.59 (4.10)	7.23 (4.10)	14.20 (4.82)	0.17 (0.03)	0.05 (0.15)	0.37 (0.14)	0.02 (0.11)
SSP–New Brunswick	−3.78 (2.94)	2.37 (5.07)	3.07 (3.47)	0.12 (0.04)	0.57 (0.11)	0.27 (0.11)	−0.03 (0.10)
SSP–British Columbia	6.70 (4.29)	4.36 (5.24)	−2.66 (3.40)	0.08 (0.05)	0.56 (0.19)	0.36 (0.19)	−0.06 (0.11)
SSP–Plus	9.69 (7.01)	9.32 (8.88)	−2.01 (6.02)	0.27 (0.07)	0.84 (0.21)	0.70 (0.20)	−0.20 (0.16)
New Chance–Allentown	−38.54 (27.79)	−0.78 (1.90)	41.37 (29.71)	0.00 (0.09)	−1.24 (1.03)	0.07 (0.12)	−0.50 (0.63)
New Chance–Bronx	−16.34 (28.39)	24.07 (10.47)	−14.70 (28.59)	−0.11 (0.15)	−0.82 (1.58)	−0.17 (0.16)	0.44 (0.38)
New Chance–Chula Vista	11.16 (10.86)	3.24 (23.13)	−13.85 (22.71)	−0.05 (0.10)	0.69 (1.24)	−0.07 (0.27)	−0.16 (0.22)
New Chance–Denver	36.37 (17.31)	−13.67 (13.04)	−5.78 (23.70)	0.18 (0.12)	2.54 (1.11)	0.20 (0.27)	0.33 (0.51)
New Chance–Harlem	22.23 (13.85)	−5.91 (19.79)	2.11 (24.75)	−0.25 (0.16)	−2.58 (1.40)	−0.36 (0.29)	0.64 (0.55)
New Chance–Lexington	30.02 (20.29)	−10.45 (13.30)	−13.91 (23.83)	−0.02 (0.14)	0.84 (0.68)	−0.08 (0.20)	0.58 (0.54)
New Chance–Minneapolis	12.81 (21.99)	5.30 (6.94)	−5.51 (23.14)	−0.17 (0.13)	0.44 (0.82)	−0.02 (0.18)	−0.03 (0.27)
New Chance–Philadelphia	−0.56 (27.10)	39.33 (17.22)	−35.22 (29.66)	−0.30 (0.25)	0.45 (0.83)	0.02 (0.28)	0.23 (0.63)
New Chance–Pittsburgh	8.49 (18.96)	−2.30 (12.64)	5.77 (20.56)	0.12 (0.10)	0.84 (0.62)	−0.17 (0.21)	0.09 (0.29)
New Chance–San Jose	−4.75 (16.24)	21.25 (13.12)	42.94 (21.86)	0.01 (0.06)	−0.38 (1.15)	0.31 (0.22)	0.05 (0.33)

Note. N = 3,290.

Comparison of Ordinary Least Squares Regression and Instrumental Variables Estimation Models

In parallel OLS and IV models estimating the effects of care type on externalizing behavior, we considered children who experienced exclusive center-based care, exclusive home-based care, mixed arrangements, in comparison with those not in regular nonmaternal care. For mother-reported behavior, Model 1 tested the effect of the child care variables, controlling for baseline covariates; Model 2 added mothers' CES–D scores, employment, income, and earnings. For

teacher-reported behavior, Model included the baseline covariates and child care variables, and Model 2 added maternal earnings. In the OLS models, maternal depression and economic variables were included as additional predictors; in the IV models, these variables were instrumented by deriving predicted values from the experimental impacts on each.

Table 6: Reduced Form Effects of Programs on Externalizing Behavior

	Externalizing behavior	
Program	**Parent report β (*SE*)**	**Teacher report β (*SE*)**
CT Jobs First	−0.03 (0.09)	−0.35 (0.15)
New Hope	−0.06 (0.18)	−0.36 (0.22)
MFIP–urban	−0.11 (0.11)	—
MFIP incentives only–urban	−0.04 (0.12)	—
MFIP–rural	0.05 (0.17)	—
NEWWS–Atlanta LFA	−0.13 (0.08)	—
NEWWS–Grand Rapids LFA	0.16 (0.11)	—
NEWWS–Riverside LFA	0.02 (0.10)	—
SSP–New Brunswick	0.12 (0.11)	—
SSP–British Columbia	0.24 (0.12)	—
SSP–plus	−0.18 (0.17)	—
New Chance–Allentown	0.90 (0.34)	0.91 (0.51)
New Chance–Bronx	0.27 (0.37)	0.35 (0.73)
New Chance–Chula Vista	0.12 (0.36)	0.87 (0.43)
New Chance–Denver	−0.22 (0.34)	−0.82 (0.60)
New Chance–Harlem	−0.71 (0.37)	0.74 (0.36)
New Chance–Lexington	0.15 (0.58)	−1.04 (0.42)
New Chance–Minneapolis	−0.11 (0.31)	−0.28 (0.50)
New Chance–Philadelphia	0.59 (0.41)	0.23 (0.54)
New Chance–Pittsburgh	0.52 (0.29)	−0.27 (0.44)
New Chance–San Jose	−0.74 (0.50)	−1.13 (0.48)

Note. MFIP = Minnesota Family Investment Program; NEWWS = National Evaluation of Welfare-to-Work Strategies; LFA = Labor Force Attachment; SSP = Canada's Self-Sufficiency Program. Dashes indicate that teacher report was not collected.

Mother-rated behavior. Results for mother-reported problems appear in the first four columns of Table 7; in each pair coefficients, OLS estimates appear first, followed by the corresponding IV estimates.

The OLS estimates indicate a positive association between center care (either exclusively or combined with home-based arrangements) and externalizing behavior and no effects for exclusive home-based care. Coefficients for exclusive center care and mixed care were similar in magnitude, and both changed little when maternal depressive symptoms and economic variables were added to the model. The magnitude of the effects was small: 10% increase in the probability of experiencing either type of care was associated with .01–.02 of a standard deviation more externalizing behavior compared with children who were not in child care.

A different pattern of relations between care arrangement and externalizing behavior emerged from the IV analyses; these models indicate lower externalizing scores for children who experienced only center care arrangements as preschoolers. This effect was statistically significant in the model that simultaneously instrumented for maternal depressive symptoms and economic outcomes. The larger coefficient in Model 2 suggests that the center care effect cannot be explained by variation in mothers' depressive symptoms, employment, income, and earnings. The coefficient of .015 indicates that a 10% increase in the probability of exclusive center care was associated with a .15 standard deviation decrease in externalizing problems (compared to no care). The IV estimates for exclusive home-based care and mixed care are not significantly different from the reference category of no care; post hoc comparisons of exclusive center care

with the other types of care indicate differences that do not reach statistical significance (center only vs. mixed, $\chi^2(1) = 2.28$, $p = .13$; center only vs. home only, $\chi^2(1) = 1.43$, $p = .23$).

Table 7: Effects of Types of Care on Parent- and Teacher-Reported Externalizing Behavior: Ordinary Least Squares (OLS) Regression and Instrumental Variable (IV) Estimation Results

	Parent-reported externalizing problems				Teacher-reported externalizing problems			
	Model 1		Model 2		Model 1		Model 2	
Type of care	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**
Only center-based care (%)	0.001† (.001)	−0.008 (.005) [16.42]	0.001* (.001)	−0.015* (.008) [16.60]	0.002 (.002)	−0.025† (.013) [3.61]	0.002 (.002)	−0.025† (.013) [3.61]
Only home-based care (%)	0.000 (.001)	−0.003 (.005) [18.59]	0.000 (.001)	−0.004 (.008) [18.64]	0.002 (.002)	−0.015 (.019) [4.27]	0.002 (.002)	−0.013 (.021) [4.27]
Mix of center- and home-based care (%)	0.002** (.001)	−0.003 (.005) [14.91]	0.002** (.001)	−0.003 (.008) [14.85]	0.003† (.002)	−0.016† (.009) [2.69]	0.003 (.002)	−0.016† (.009) [2.69]
Maternal depressive symptoms (CES–D)[a]			0.257** (.019)	0.442 (.322) [2.54]				
Average quarterly employment			−0.135* (.064)	−1.031 (.994) [43.89]				
Average quarterly income			−0.012 (.018)	0.235 (.150) [15.92]				
Average quarterly earnings			0.021 (.020)	0.374 (.296) [32.35]			0.008 (.034)	0.084 (.442) [12.69]
R^2	.04		.11		.11		.11	
Sample size	3,290				379			

Note. Standard errors are presented in parentheses; first-stage *F* statistics are presented in brackets. Each empirical model included the following covariates measured at or prior to study entry: age and gender of child; whether mother has high school degree, mother's prior employment and earnings; whether mother received Aid to Families With Dependent Children for 2 years or more; whether mother never married; whether mother is Black, Latino, White (omitted), or of other ethnicity; and indicators for each random assignment study. The instruments identifying the first stage of the independent variable model for parent-reported behavior included random assignment status in the following: CT Jobs First, MFIP–rural, MFIP–urban, MFIP incentive only, 10 New Chance sites, New Hope, NEWWS–LFA (separately for Atlanta, Grand Rapids, and Riverside), SSP–British Columbia, SSP–New Brunswick, and SSP plus. The instruments identifying the first stage in the teacher model include random assignment status in CT Jobs First, 10 New Chance sites, and New Hope.
[a]Maternal CES–D scores are standardized ($M = 0$, $SD = 1$). $^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

We performed several tests to evaluate the validity of the IV estimates. First stage *F* tests on the instruments (shown in brackets in Table 7) indicated sufficient strength for each endogenous predictor, with the exception of maternal depressive symptoms (not surprising given that this outcome was not targeted by the programs and that few impacts occurred). We further evaluated our IV results by employing two commonly used tests. First, when instruments outnumber regressors, an overidentification test (e.g., Sargan statistic) can be used to measure the association between instruments and the error term in the second-stage equation; valid instruments should not have any such association as they are assumed to reflect random variation in the predictor. For all models in the current analysis, Sargan's statistic confirmed the exogenous nature of our instrument set. Second, we also computed the Durbin–Wu–Hausman (DWH) chi-square statistic to evaluate whether OLS and IV estimators were equivalent; rejection of the null hypothesis is generally taken as an indication of non-equivalence and endogeneity (though other conditions may lead to rejecting the null as well, including model misspecification; Davidson & MacKinnon, 1993). Statistically significant DWH tests for exclusive center care,

$\chi^2(1) = 5.29$, $p < .05$, income, $\chi^2(1) = 5.77$, $p < .05$, and earnings, $\chi^2(1) = 4.49$, $p < .05$, suggested bias in the OLS estimates for these variables.

Teacher-rated behavior. According to the OLS estimates shown in Columns 5 and 7 of Table 7, children who experienced mixed care as preschoolers had more teacher-rated externalizing problems in elementary school than those without preschool child care experience. A 10% increase in the probability of using mixed care was associated with a .02 standard deviation increase in externalizing behavior.

By contrast, the IV analyses indicated effects (of marginal statistical significance, $p < .10$) in the opposite direction, specifically, children who experienced either exclusive center care or mixed care received lower externalizing scores from teachers. The IV coefficient for exclusive center care indicates that a 10% increase in the probability of this care was associated with a .25 standard deviation reduction in teacher-reported externalizing problems. This effect was larger than that for parent-reported problems but was also less precise (indicated by the larger standard error), partly because few studies collected teacher data. As in the parent models, post hoc tests revealed no statistically significant differences between the three types of care. At the same time, the DWH test of consistency between OLS and IV estimates indicates that some endogeneity existed, $\chi^2(1) = 9.80$, $p < .05$; the only individual variable for which this test approached significance was exclusive center care, $\chi^2(1) = 2.21$, $p = .14$.

The finding that both exclusive center care and mixed care reduce behavior problems at school tentatively suggests a benefit from having at least some exposure to center care during the preschool years. A follow-up IV analysis indicates that experiencing *any* center care was associated with significantly lower teacher ratings of externalizing behavior ($\beta = -.015$, $SE = .008$, $p < .05$). A similar analysis of home-based care revealed no significant effects on teacher-rated behavior.

Sensitivity Analyses

We conducted follow-up analyses with alternative model specifications to ensure that results were not disproportionately influenced by one or two studies. In one analysis, we excluded the New Chance study because its program and sample differed somewhat from those of other studies. In a second test, we collapsed across the New Chance sites, treating them as a single experiment (and using treatment status as one instrument), given concerns about small samples in the individual sites. In a third analysis, we excluded the Canadian SSP sites, because differences in local and national contexts could affect our estimates. The OLS and IV coefficients for child care type remained qualitatively similar across all of these analyses (results available upon request).

DISCUSSION

This study joins several recent articles that highlight the usefulness of comparing results across methods within a single data set (e.g., Magnuson et al., 2007; McCartney et al., 2010; National Institute of Child Health and Human Development, Early Child-care Research Network & Duncan, 2003). Our approach differs from earlier work by examining experimentally induced differences in the types of childcare arrangements used by low-income mothers affected by policies designed to increase their employment. We contrast OLS and IV estimates of the

relations between type of child care and children's externalizing behavior with pooled data from 21 random-assignment policy experiments that increased maternal employment, changed family economic resources, and increased children's exposure to a variety of nonmaternal care settings.

Cross-Method Comparison of Findings

Consistent with prior research linking center care experience to behavior problems (e.g., Halle et al., 2006; Magnuson et al., 2007; NICHD Early Childcare Research Network, 2004), our OLS results indicate higher maternal ratings of externalizing problems for school-age children who had attended centers as preschoolers (either exclusively or combined with home-based care) than for children who had not had regular care arrangements. Teachers also reported more problems for children with prior center experience (at a marginal level of significance), particularly for those who had been in mixed care. The OLS results are robust to controls for baseline covariates, maternal depression, employment, income, and earnings.

By contrast, IV techniques indicate that children who had been in only center arrangements during their preschool years had fewer subsequent externalizing problems, according to both mothers and teachers. Teachers also reported fewer problems for children who had experienced mixed care arrangements. That is, according to teachers, children who had experienced any center care as pre-schoolers exhibited fewer problems in their early school years than did their peers with no such experience. Having at least some exposure to organized group care prior to school entry may help children acquire appropriate classroom behaviors. These results should be considered suggestive rather than conclusive. Although the IV coefficients for center care were consistently larger than those for home-based and mixed care, estimates for the three types of care were not statistically different from one another.

Reconciling Discrepant Results Across Methods

Interestingly, our OLS results mirror those obtained in correlational studies, whereas our IV results are similar to findings from random-assignment studies of early childhood interventions. This pattern suggests that an instrumental variables approach may address biases not easily controlled for in correlational designs. A number of attempts to reproduce experimental impact estimates with nonexperimental regression methods demonstrate the difficulty of doing so, suggesting that the threat posed by selection bias is nontrivial (see H. S. Bloom, Michalopoulos, Hill, & Lei, 2002; Ludwig et al., 2001; Wilde & Hollister, 2002;). Two methods that yield estimates similar to those from experiments are regression discontinuity and careful matching of groups on pretest measures (Cook, Shadish, & Wong, 2008), but neither of these methods is common in child care research (see Gormley et al., 2005, for an exception).

Child care history prior to study entry is one unmeasured factor that might help to explain the discrepant findings. The present data captured care experiences across the preschool period (i.e., ages to 5 years) and social behavior ratings when children were 5 to years old. Random assignment presumably equates program and control groups on amount and type of child care experienced prior to three years of age, but different histories could affect the OLS comparisons, which capture the cumulative effects of child care Extensive hours of care (especially in centers) for infants and toddlers have been shown to predict relatively high levels externalizing behavior in preschool and elementary school (Belsky et al., 2007; Waldfogel, 2006). During the early

1990s when these data were collected, low-income mothers of infants and toddlers were much less likely to be in the labor force (and to be using child care) than presently.

We used multiple treatment effects as instruments to separate the effects of child care type from those of employment, income, earnings, and maternal depressive symptoms. The child care coefficients change little when additional variables are added to the model, but is possible that treatments that increased center care also changed some unmeasured feature of children's lives (e.g., parents' employment schedule or stability) in ways that influenced behavior.

In reconciling the IV and OLS results, we questioned whether center care induced by the experiments was somehow different than center care received by children in the control group (and by extension, center care accessible to the general population). Several programs that increased the use of centers offered more generous, seamless, or comprehensive child care assistance than the assistance available in the control condition (Crosby et al. 2005), which may have allowed families to access higher quality care or maintain more stable care than they would have otherwise We doubt, however, that OLS–IV discrepancies reflect systematic differences in quality or stability within type because (a) OLS estimates of the relations between type of care and behavior do not differ by treatment group status (i.e., center care is associated with more problems in both groups), and (b) there are no experimental impacts on crude measures of structural characteristics of care (e.g., child-to-adult ratios) in the subset of studies that collected this information (Crosby et al., 2005).

Notably, our OLS and IV results are not discrepant for all predictors (i.e., maternal depression and economic variables). Moreover, parallel analyses examining the effects of child-care type on school achievement, we find consistent results across the two methods—both the OLS and IV estimates indicate academic benefits for children attending centers in the year or two prior to school entry (Gennetian et al., 2006). These findings provide some assurance that the OLS–IV differences reported here are not an artifact of method. Finally, the robustness of our IV coefficients to changes in the sample and instrument set boosts our confidence in the results.

Limitations and Strengths

This study relies on a global measure of children's care experiences (i.e., types of arrangements during a 2-year period); thus, we cannot determine the extent to which measured effects reflect care quality, stability, or intensity, nor do we know whether children in different settings experienced different amounts of care in their lifetimes. Second, small samples of preschoolers in some of the programs and sites prevented us from examining whether the observed relationships varied for particular subgroups of children (e.g., by child gender or ethnicity). Third, this set of studies does not contain large enough samples of children younger than the age of 3 years to estimate the effects of care type on infants and toddlers, a group for which there are particular concerns about the effects of organized care (Belsky et al., 2007). Finally, our method of complete-case analysis could lead us to underestimate or over. estimate the true effects of care on behavior if study attrition led to a restricted range of behavior problem reports. However, our confidence in the estimates is strengthened given that we also conducted missing-at-random–based analyses of multiply imputed data; in that case, any selective nonresponse would have to be over and above patterns of missingness related to the covariates. Further research is necessary

both to replicate the current set of results and to identify particular circumstances in which center care can promote or impede socioemotional adjustment (e.g., see McCartney et al., 2010).

Although this broad look at child care type stands in contrast to the rich information about child care in other studies, our data set has several strengths not available in most other studies, including longitudinal, random-assignment designs with multiple sources of information on large samples of low-income children who experienced substantial shifts in child care as a result of mothers' interactions with welfare and employment programs. Moreover, the effects captured by the IV analysis—the impact of child care for families whose arrangements would change as a result of the interventions tested in these studies—are particularly relevant for policy. Given these advantages, the information produced here is, intended to complement that from more intensive naturalistic studies in the pursuit of understanding how child care experiences relate to children's development.

Benefits and Challenges of Using Instrumental Variables in Developmental Research

Both OLS and IV represent useful approaches for understanding developmental issues; under the right conditions, however, the latter may offer some advantage in estimating causal relations. By relying on exogenous variation (vs. naturally occurring variation), IV techniques can purge some of the problematic components from the predictor variable, producing less-biased estimates and strengthening causal claims. The fact that IV relies on only a portion of the total variance, however, often leads to less precise (as indicated by large standard errors) and more conservative estimates than those obtained using OLS.

Identifying appropriate instruments (i.e., those that satisfy the assumptions of IV and have sufficient predictive power) can be challenging. Social policy experiments are ideal for this purpose in many respects but generally only useful if multiple treatments within or across studies are available to permit estimation of different predictors (Gennetian et al., 2008). In the current study, we have access to several instruments based on between-site differences in program impacts on employment, income, maternal depression, and child care that resulted both from different programs and from differences in program implementation across sites. For example, the type of child care support provided to New Chance participants varied by site, leading, in essence, to different treatments; some sites provided on-site child care, whereas others helped families access providers in the community (for details, Yoshikawa et al., 2001).

Many topics of interest to developmental scientists may not amenable to random-assignment designs, particularly given large number of cases needed to perform IV. In the absence experimental data, state or local variations in child care regulation and policy may provide viable instruments; however, states often differ in multiple ways that are important for children's development, making the exclusion restriction more difficult to satisfy Large national data sets (e.g., ECLS-B) lend themselves to approach, allowing investigators to use sources of exogenous variation in child care experiences to gain a better understanding the effects on children (for an example of this approach, Magnuson et al., 2007).

CONCLUSIONS

Associations between center care and behavior problems have appeared in research findings for more than 25 years (see Belsky, 2001), but they occur less consistently for low-income samples

(Loeb et al., 2004; Votruba-Drzal et al., 2004). The overhaul of U.S. welfare system in the 1990s raised many questions about how children would fare in the face of policies designed to reduce welfare reliance and increase employment among low-income mothers. Our findings address some of these questions using techniques to isolate the effects of policy-induced changes preschool child care on children's behavior during the early school years. We find tentative evidence that community-based center care used under these circumstances may reduce, rather than increase, externalizing problems. These modest effects occur the range of programs (of unknown quality) that were available and selected by parents; larger benefits might be expected from policies designed specifically to provide children with high-quality care (see Rigby, Ryan, & Brooks-Gunn, 2007).

Given the growth of various forms of preschool experience the U.S.—and their apparent advantages for academic skills, particularly for low-income children—it is critical to identify ways which these settings can also foster social skills and self-control Developmental scientists have an expanding set of theoretical methodological tools to examine such topics. We hope to see added to this toolbox; particularly when combined with random assignment designs, this technique has the potential to greatly improve researchers' ability to infer causality.

ACKNOWLEDGEMENTS

We thank Joshua Angrist, David Card, Janet Currie, Rebekah Levine Coley, Howard Bloom, Greg Duncan, Virginia Knox, and Cybele Raver for comments on preliminary findings of this study and drafts of this article. This article is part of the Next Generation project, a study led by MDRC that examines the effects of welfare, antipoverty, and employment policies on children and families. Funding for this project was provided by the David and Lucile Packard Foundation, the William T. Grant Foundation, the John D. and Catherine T. MacArthur Foundation, and National Institute of Child Health and Human Development 5RO1HD45691-2. Many thanks go to the funders of these studies and our partners in their evaluation for access to the data, in particular, Connecticut's Department of Social Services, Florida's Department of Children and Families, Human Resources Development Canada, Minnesota's Department of Human Services, and the U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation and Administration for Children and Families.

Appendix A: Baseline Characteristics by Experimental Group Status, by Study and Site

Variable	Connecticut job first		MFIP rural full		MFIP urban full		MFIP urban incentives only		New chance		New hope		NEWWS Atlanta	
	C	E	C	E	C	E	C	E	C	E	C	E	C	E
Parent characteristics														
Age	26.8	26.3	26.9	25.5	25.9	25.9	25.9	26.2	19.6	19.9	27.7	26.6	29.3	29.2
Race (%)														
Black	42.5	41.6	2.4	1.5	31.2	36.9	31.2	33.6	49.2	45.1	48.3	60.3	96.2	95.5
White	37.2	35.4	90.7	92.3	57.0	47.5	57.0	56.3	16.9	15.9	15.0	12.1	3.2	3.2
Latino	20.2	21.5	0.2	1.5	3.4	2.5	3.4	0.9	28.8	36.3	35.0	25.9	0.3	0.8
Other	0.0	1.3	7.4	4.6	8.6	13.1	8.6	9.4	5.1	2.7	1.7	1.7	0.0	0.4
Never married (%)	78.8	75.3	40.0	33.8	64.4	68.9	64.4	68.1	79.7	85.8	68.3	77.6	71.4	75.7
Parent education, employment, and income														
High school graduate (%)	67.8	54.5**	80.0	89.2	76.5	67.7	76.5	73.6	8.5	8.8	65.0	62.1	67.4	59.9
Employed in year prior to random assignment (%)	56.7	46.6*	47.3	52.3	52.3	49.1	52.3	46.5	40.7	29.2	63.3	70.7	38.6	33.2
Earnings in year prior to random assignment ($)	3,214	2,682	2,806	3,311	3,706	3,433	3,706	2,039*	1,762	1,155	3,783	3,993	1,565	1,017*
Receiving Aid to Families With Dependent Children for 2 or more years prior to random assignment (%)	69.5	61.2	63.6	56.9	56.4	62.7	56.4	67.2	0.0	0.0	70.0	70.7	74.1	77.4
Family composition														
No. of children in family	1.9	1.9	2.0	1.8	1.8	2.0	1.8	2.1*	1.5	1.7	2.5	2.3	2.2	2.3
Age of child (years)	3.0	2.9	3.0	3.0	3.0	3.0	3.0	2.9	3.8	3.7	3.8	3.8	4.0	4.0
Child is male (%)	49.5	46.6	56.4	49.2	55.6	48.4	55.6	52.7	45.8	50.4	50.0	50.0	51.0	54.3

NEWWS Grand Rapids		NEWWS Riverside		SSP British Columbia		SSP New Brunswick		SSP plus New Brunswick		SSP plus New Brunswick	
C	E	C	E	C	E	C	E	C	E	C	E
26.4	27.6*	28.9	29.7	28.4	27.5	26.5	25.9	26.5	25.5	26.5	25.5
37.1	39.3	16.9	18.2	1.3	0.7	1.9	1.1	1.9	0.0	1.9	0.0
55.0	50.3	49.7	48.3	75.3	77.6	86.5	90.4	86.5	94.4	86.5	94.4
5.3	7.6	31.3	33.6	2.7	4.8	0.1	0.6	0.1	0.0	0.1	0.0
2.6	2.8	2.1	0.0**	20.7	17.0	11.7	7.9	11.7	5.6	11.7	5.6
59.6	56.2	46.7	35.7*	58.7	57.1	73.1	77.5	73.1	80.6	73.1	80.6
65.6	65.8	49.4	60.8*	49.3	56.5	55.6	53.9	55.6	75.0	55.6	75.0*
60.3	51.4	32.8	29.4	24.0	29.9	35.1	32.6	35.1	75.0*	35.1	44.4
2,557	2,248	1,922	2,322	1,319	1,182	1,056	1,213	1,056	963	1,056	963
73.5	69.2	66.0	65.7	74.0	76.9	84.2	85.4	84.2	88.9	84.2	88.9
2.1	2.1	2.0	2.2	1.6	1.6	1.6	1.5	1.6	1.5	1.6	1.5
4.0	3.9	3.8	3.8	3.0	2.9	2.9	3.0	2.9	3.2	2.9	3.2
50.3	47.9	49.4	49.7	48.7	47.6	50.9	50.6	50.9	50.0	50.9	50.0

Note. MFIP = Minnesota Family Investment Program; NEWWS = National Evaluation of Welfare-to-Work Strategies; SSP = Canada's Self-Sufficiency Program; C = control group; E = experimental group. $^{*}p < .05$. $^{**}p$ -< .01.

Appendix B: Instrumental Variables (Two-Stage Least Squares) Model Specifications

Our primary interest in using these data was to estimate the effects of three types of care on children's externalizing behavior (net of the effects of maternal employment, income, and earnings), as represented by the following equation:

$$Y_{it}=X_{it}^{'}\beta_1+\alpha_1\sum_{j=1}^{3}CareType_{jit}+\alpha_2 CESD+\alpha_3 Employment_{it}+\alpha_4 Income_{it}+\alpha_5 Earnings_{it}+\delta t+\zeta_{it}$$

where Y_i is behavior for child *i*; CareType$_i$ is child *i*'s participation in only center-based care, only home-based care, or mixed care in study/site *t; CESD*$_i$ is child *i*'s mother's score on the *CESD* measure of depressive symptoms; *Employment*$_i$ is child *i*'s mother's average quarterly employment rate; *Income*$_i$ is child *i*'s mother's average quarterly income (as the sum of earnings, public assistance, and earnings supplements); *Earnings*$_i$ is child *i*'s mother's average quarterly earnings; and *X* is a standard set of baseline characteristics, including focal child's age and gender; mother's age, race/ethnicity (i.e., White, Black, Latino, or other), and level of education (high school degree or not); whether the mother has ever been married; total number of children; maternal employment and earnings in the year prior to study entry; and, family's welfare history (i.e., more or less than 2 years of receipt). In addition, because of the design of our data and unobserved static study-site-level characteristics, we include the main effects for each welfare or employment study through a series of indicators depicted as δ*t*.

Given the high potential for endogeneity in the relations of child care arrangements, maternal depression, and mothers' economic behavior to children's behavior, we used an instrumental-variables approach to generate predicted values for these variables, which were purged of omitted-variables bias. Specifically, we used program-control group contrasts for the 21 program/site combinations in our data to predict each of the seven potentially endogenous regressors. Denoting experimental group dummies as $Z_1, Z_2, \ldots . Z_{21}$, our first-stage equations were as follows:

$$Centercare_{it1}=X_{it}^{'}\gamma_1+\sum_{k=1}^{21}\pi_{k1}Z_{itk}+\delta t+\eta_{it1} \quad (1)$$

$$Mixedcare_{it2}=X_{it}^{'}\gamma_2+\sum_{k=1}^{21}\pi_{k2}Z_{itk}+\delta t+\eta_{it2} \quad (2)$$

$$Homecare_{it3}=X_{it}^{'}\gamma_3+\sum_{k=1}^{21}\pi_{k3}Z_{itk}+\delta t+\eta_{it3} \quad (3)$$

$$CESD_{it4} = X'_{it}\gamma_4 + \sum_{k=1}^{21} \pi_{k4} Z_{itk} + \delta t + \eta_{it4} \quad (4)$$

$$Employment_{it5} = X'_{it}\gamma_5 + \sum_{k=1}^{21} \pi k_5 Z_{itk} + \delta t + \eta_{it5} \quad (5)$$

$$Income_{it6} = X'_{it}\gamma_6 + \sum_{k=1}^{21} \pi_{k6} Z_{itk} + \delta t + \eta_{it6} \quad (6)$$

$$Earnings_{it7} = X'_{it}\gamma_7 + \sum_{k=1}^{21} \pi_{k7} Z_{itk} + \delta t + \eta_{it7} \quad (7)$$

In the second stage of the instrumental-variables analysis, actual values for the regressors in Equation 1 were replaced with the predicted values generated in the first-stage (Equations 2–7), and the standard errors were adjusted accordingly.

FOOTNOTES

[1]We note that the linear two-stage procedure described here can be used regardless of whether the endogenous variable (i.e., the dependent variable in the first stage) is continuous or binary; however, when the outcome of interest (i.e., the dependent variable in the second stage) is nonlinear, a slightly different approach (e.g., generalized method of moments) is needed to obtain instrumental variable estimates (Amemiya, 1990; Foster, 1997; Gennetian et al., 2008).

REFERENCES

Achenbach TM, Edelbrock CS. Behavioral problems competencies reported by parents of normal and disrupted children aged four through sixteen. Monographs of the Society for Research in Child Development. 1981;46:1–82.

Ahnert L, Gunnar MR, Lamb ME, Barthel M. Transition to child care: Associations with infant–mother attachment, infant negative emotion, and cortisol elevations. Child Development. 2004;75:639–650.

Amemiya Y. Two-stage instrumental variable estimators for nonlinear errors-in-variables model. Journal of Econometrics. 1990;44:311–332.

Angrist JD, Imbens GW, Rubin DB. Identification causal effects using instrumental variables. Journal of the American Statistical Association. 1996;91:444–455.

Angrist JD, Krueger AB. Working Paper No. 8456. Cambridge, MA: National Bureau of Economic Research; 2001. Instrumental variables and the search for identification: From supply and demand to natural experiments.

Arnold DH, McWilliams L, Arnold EH. Teacher discipline and child misbehavior in day care: Untangling causality with correlational data. Developmental Psychology. 1998;34:276–287.

Barnett WS. Long-term effects of early childhood programs on cognitive and school outcomes. The Future of Children. 1995;5:25–50.

Belsky J. Developmental risks (still) associated with early child care. Journal of Child Psychology and Psychiatry. 2001;42:845–859.

Belsky J, Vandell DL, Burchinal M, Clarke-Stewart KA, McCartney K, Owen MT. the NICHD Early Child Care Research Network. Are there long-term effects of early child care? Child Development. 2007;78:681–701.

Berry WD. Understanding regression assumptions. Newbury Park, CA: Sage; 1993.

Blau DM. The effect of child care characteristics on child development. Journal of Human Resources. 1999;34:786–822.

Bloom D, Michalopoulos C. How welfare and work policies affect employment and income: A synthesis of research. New York: MDRC; 2001.

Bloom D, Scrivener S, Michalopoulos C, Morris P, Hendra R, Adams-Ciardullo D, Walter J. Jobs first: Final report on Connecticut's welfare reform initiative. New York: MDRC; 2002.

Bloom HS, Michalopoulos C, Hill CJ, Lei Y. Can nonexperimental comparison group methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs? New York: MDRC; 2002.

Bos JM, Huston AC, Granger RC, Duncan GJ, Brock TW, McLoyd VC. New Hope for people with low incomes: Two-year results of a program to reduce poverty and reform welfare. New York: MDRC; 1999.

Bound J, Jaeger D, Baker R. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association. 1995;90:443–450.

Burchinal M, Nelson L. Family selection and child care experiences: Implications for studies of child outcomes. Early Childhood Research Quarterly. 2000;15:385–411.

Chang YE. Unpublished dissertation. University of Texas; Austin: 2003. Mothers' attitudes toward maternal employment, maternal well-being, maternal sensitivity and children's socioemotional outcomes when mothers engage in different amounts of employment.

Chase-Lansdale PL, Votruba-Drzal E. Human development and the potential for change from the perspective of multiple disciplines. In: Chase-Lansdale PL, Kiernan K, Friedman RJ, editors. Human development across lives and generations: The potential for change. New York: Cambridge University Press; 2004. pp. 343–366.

Clark-Kauffman E, Duncan GJ, Morris P. How welfare policies affect child and adolescent achievement. The American Economic Review. 2003;93:299–303.

Cleveland HH, Wiebe RP, van den Oord EJCG, Rowe DC. Behavior problems among children from different family structures: The influence of genetic self-selection. Child Development. 2000;71:733–751.

Coley RL, Chase-Lansdale PL, Li-Grining CP. Welfare, Children and Families: A three-city study. Baltimore, MD: Johns Hopkins University; 2001. Child care in the era of welfare reform: Quality, choices, and preferences (Policy Brief 01–4)

Cook TD, Shadish WR, Wong VC. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. Journal of Policy Analysis and Management. 2008;27:724–750.

Crosby DA, Gennetian LA, Huston AC. Child care assistance policies can affect the use of center-based care. Applied Developmental Science. 2005;9:86–106.

Davidson R, MacKinnon JG. Estimation and inference in econometrics. New York: Oxford University Press; 1993.

Dowsett CJ, Huston AC, Imes AE, Gennetian LA. Structural and process features in three types of child care for children from high and low income families. Early Childhood Research Quarterly. 2008;23:69–93.

Duncan GJ, Magnuson K, Ludwig J. The endogeneity problem in developmental studies. Research in Human Development. 2004;1:59–80.

Early DM, Burchinal MR. Early childhood care: Relations with family characteristics and preferred care characteristics. Early Childhood Research Quarterly. 2001;16:475–497.

Eisenberg N, Cumberland A, Spinrad TL, Fabes RA, Shepard SA, Reiser M, Guthrie IK. The relations of regulation and emotionality to children's externalizing and internalizing problem behavior. Child Development. 2001;72:1112–1134.

Foster EM. Instrumental variables for logistic regression: An application of the generalized method of moments. Social Science Research. 1997;26:487–504.

Foster EM, McLanahan S. An illustration of the use of instrumental variables: Do neighborhood conditions affect a young person's chance of finishing high school? Psychological Methods. 1996;1:249–260.

Fuller B, Kagan SL. Remember the children: Mothers balance work under welfare reform: Growing up in poverty project, Wave 1 findings–California, Connecticut, Florida. Berkeley, CA: University of California; 2000.

Fuller B, Kagan SL, Caspary GL, Gauthier CA. Welfare reform and child care options for low-income families. The Future of Children. 2002;12:97–119.

Fuller B, Kagan SL, Loeb S, Chang Y. Child care quality: Centers and home settings that serve poor families. Early Childhood Research Quarterly. 2004;19:505–527.

Gennetian LA. How sibling composition affects adolescent schooling outcomes when welfare reform policies increase maternal employment. Eastern Economic Review. 2004;30:81–100.

Gennetian LA, Crosby DA, Dowsett C, Huston AC, Principe D. Maternal employment, early care settings and the achievement of low-income children [Mimeo] New York: MDRC; 2006.

Gennetian LA, Crosby DA, Huston AC, Lowe T. How child care assistance in welfare and employment programs can support the employment of low-income families. Journal of Policy Analysis and Management. 2004;23:723–743.

Gennetian LA, Magnuson K, Morris PA. From statistical associations to causation: What developmentalists can learn from instrumental variables techniques coupled with experimental data. Developmental Psychology. 2008;44:381–393.

Gennetian LA, Miller C. Reforming welfare and rewarding work: Final report on the Minnesota Family Investment Program. New York: MDRC; 2000.

Gennetian LA, Morris PA, Bos JM, Bloom HS. Constructing instrumental variables from experimental data to explore how treatments produce effects. In: Bloom HS, editor. Learning more from social experiments: Evolving analytic approaches. New York: Russell Sage; 2005. pp. 75–114.

Gibson-Davis C, Magnuson K, Gennetian L, Duncan G. Employment and risk of domestic abuse among low-income single mothers. Journal of Marriage and the Family. 2005;67:1149–1168.

Gilliam WS. Prekindergartners left behind: Expulsion rates in state prekindergarten programs. 2005 Retrieved from http://www.fcd-us.org/PDFs/NationalPreKExpulsionPaper03.02_new.pdf.

Gormley WT, Gayer T, Phillips D, Dawson B. The effects of universal pre-k on cognitive development. Developmental Psychology. 2005;41:872–884.

Gresham FM, Elliot SN. Social skills rating system manual. Circle Pines, MN: American Guidance Service; 1990.

Halle TG, Hair EC, Lavelle B, Martin L, Scott L, McNamara M, Zaslow M. Effect of type and extent of child care on low-income children's school readiness and academic growth through fifth grade. Paper presented at the biennial Head Start Research Conference; Washington, DC. 2006. June.

Hamilton G, Freedman S, Gennetian LA, Michalopoulos C, Walter J, Adams-Ciardullo D, Brooks J. Five-year adult and child impacts for eleven programs. Washington, DC: U.S. Department of Health and Human Services; 2001. National evaluation of welfare-to-work strategies: How effective are different welfare-to-work approaches?

Huston AC, Chang YE, Gennetian LA. Family and individual predictors of child care use by low-income families in different policy contexts. Early Childhood Research Quarterly. 2002;17:441–469.

Kisker EE, Hofferth SL, Phillips DS, Farquhar E. A profile of child care settings: Early education and care in 1990. Vol. 1. Princeton, NJ: Mathematica Policy Research; 1991.

Kontos S, Hsu H, Dunn L. Children's cognitive and social competence in child-care centers and family day-care homes. Journal of Applied Developmental Psychology. 1994;15:387–411.

Lamb ME. Nonparental child care: Context, quality, correlates, and consequences. In: Sigel I, Renninger KA, editors. Child psychology in practice. 5. New York: Wiley; 1998. pp. 73–134.

Liebman JB, Katz LF, Kling JR. IRS Working Paper 493. Princeton, NJ: Princeton University, Industrial Relations Section; 2004. Beyond treatment effects: Estimating the relationships between neighborhood poverty and individual outcomes in the MTO experiment.

Loeb S, Bridges M, Bassok D, Fuller B, Rumberger RW. How much is too much? The influence of preschool centers on children's social and cognitive development. Economics of Education Review. 2007;26:52–66.

Loeb S, Fuller B, Kagan SL, Carrol B. Child care in poor communities: Early learning effects of type, quality, and stability. Child Development. 2004;75:47–65.

Love JM, Harrison L, Sagi-Schwartz A, van IJzendoorn MH, Ross C, Ungerer JA, Chazan-Cohen R. Child care quality matters: How conclusions may vary with context. Child Development. 2003;74:1021–1034.

Ludwig J, Duncan GJ, Hirschfield P. Urban poverty and juvenile crime: Evidence from a randomized housing-mobility experiment. Quarterly Journal of Economics. 2001;116:655–680.

Magnuson KA, McGroder SM. Working Paper No. 280. Joint Center for Poverty Research; 2002. The effect of increasing welfare mothers' education on their young children's academic problems and school readiness.

Magnuson KA, Meyers MK, Ruhm CJ, Waldfogel J. Inequality in preschool education and school readiness. American Educational Research Journal. 2004;41:115–157.

Magnuson KA, Ruhm CJ, Waldfogel J. Does prekinder-garten improve school preparation and performance? Economics of Education Review. 2007;26:33–51.

Magnuson KA, Waldfogel J. Early childhood care and education: Effects on ethnic and racial gaps in school readiness. Future of Children. 2005;15:169–196.

McCartney K, Burchinal M, Clarke-Stewart A, Bub KL, Owen MT, Belsky J. Testing a series of causal propositions relating time in child care to children's externalizing behavior. Developmental Psychology. 46:1–17.

Michalopoulos C, Tattrie D, Miller C, Robins PK, Morris P, Gyarmati D, Ford R. Making work pay: Final report on the self-sufficiency project for long-term welfare recipients. New York: MDRC; 2002.

Moffitt R. Remarks on the analysis of causal relationships in population research. Demography. 2005;42:91–108.

Morris P, Duncan G, Rodrigues C. Unpublished manuscript. 2006. Does money really matter? Estimating the impacts of family income on children's achievement with data from social policy experiments.

Morris PA, Gennetian LA. Identifying the effects of income on children's development using experimental data. Journal of Marriage and the Family. 2003;65:716–729.

Morris P, Michalopoulos C. The Self-Sufficiency Project: Effects on children of a program that increased parental employment and income. New York: MDRC; 2000.

National Institute of Child Health and Human Development, Early Child Care Research Network. Type of child care and children's development at 54 months. Early Childhood Research Quarterly. 2004;19:203–230.

National Institute of Child Health and Human Development, Early Child Care Research Network. Duncan GJ. Modeling the impacts of child care quality on children's preschool cognitive development. Child Development. 2003;74:1454–1475.

Newcombe NS. Some controls control too much. Child Development. 2003;74:1050–1052.

Peterson JL, Zill N. Marital disruption, parent–child relationships, and behavior problems in children. Journal of Marriage and the Family. 1986;48:295–307.

Pungello EP, Kurtz-Costes B. Why and how working women choose child care: A review with a focus on infancy. Developmental Review. 1999;19:31–96.

Quint JC, Bos JM, Polit DF. New chance: Final report on a comprehensive program for young mothers in poverty and their children. New York: MDRC; 1997.

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

Riccio J, Bloom H. Extending the reach of randomized social experiments: New directions in evaluations of American welfare-to-work and employment initiatives. Journal of the Royal Statistical Society. 2002;165:13–30.

Rigby E, Ryan RM, Brooks-Gunn J. Child care quality in different state policy contexts. Journal of Policy Analysis and Management. 2007;26:887–907.

Runions KC, Keating DP. Young children's social information processing: Family antecedents and behavioral correlates. Developmental Psychology. 2007;43:838–849.

Schoeni R, Blank R. NBER Working Paper No. 7627. Cambridge, MA: National Bureau of Economic Research; 2000. What has welfare reform accomplished? Impacts on welfare participation, employment, income, poverty and family structure.

Schweinhart LJ, Weikart DP. The High/Scope preschool curriculum comparison study through age 23. Early Childhood Research Quarterly. 1997;12:117–143.

Shonkoff J, Phillips D, editors. From neurons to neighborhoods. Washington, DC: National Academy of Sciences; 2000.

Sobel ME. What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. Journal of the American Statistical Association. 2006;101:1398–1407.

U.S. Department of Health and Human Services, Administration for Children and Families. Head Start Impact Study: First-year findings. Washington, DC: Author; 2005.

Votruba-Drzal E, Coley RL, Chase-Lansdale PL. Child care and low-income children's development: Direct and moderated effects. Child Development. 2004;75:296–312.

Waldfogel J. What do children need? Public Policy Research. 2006;13:26–34.

Watamura SE, Dowzella B, Alwin J, Gunnar MR. Morning-to-afternoon increases in cortisol concentrations for infants and toddlers at child care: Age differences and behavioral correlates. Child Development. 2003;74:1006–1020.

West J, Denton K, Germino-Hausken E. NCES Pub. No. 2000–070. Washington, DC: U.S. Department of Education, National Center for Educational Statistics; 1999. America's kinder-gartners. Retrieved from http://nces.ed.gov/pubs2000/2000070.pdf.

Wilde ET, Hollister R. Discussion Paper No. 1242–02. Madison, WI: Institute for Research on Poverty; 2002. How close is close enough? Testing nonexperimental estimates of impact against experimental estimates of impact with education test scores as outcomes.

Yeung WJ, Linver MR, Brooks-Gunn J. How money matters for young children's development: Parental investment and family processes. Child Development. 2002;73:1861–1879.

Yoshikawa H. Long-term effects of early childhood programs on social outcomes and delinquency. The Future of Children. 1995;5:51–75.

Yoshikawa H, Rosman EA, Hsueh J. Family, school, and community variation in teenage mothers' experiences of child care and other components of welfare reform: Selection processes and developmental consequences. Child Development. 2001;72:229–317.