Privileging the Participant:
The Importance of Take-Up Rates In Social Welfare Evaluations

Christina Gibson

Princeton University

November 30, 2001

Author Contact Information

Christina Gibson
Postdoctoral Fellow
Center for Research on Child Wellbeing
Princeton University
286 Wallace Hall
Princeton, NJ 08540
(609) 258 2772
cgibson@princeton.edu

I thank Greg Duncan, Thomas Cook, Lindsay Chase-Lansdale, Jens Ludwig, Hiro Yoshikawa, Chris Jepsen, Katherine Magnuson, Jennifer Romich, Isaac McFarlin Jr., and Sara McLanahan for their assistance and suggestions. I also gratefully acknowledge financial support from Northwestern University/University of Chicago's Joint Center for Poverty Research and the MacArthur Network on Successful Pathways Through Middle Childhood. Additional funding was received from the National Institute of Child Health and Human Development, which funded grant 5 R01 HD36916 for the Bendheim-Thoman Center for Research on Child Wellbeing at Princeton University, grant 5 P30 HD32030 for the Office of Population Research at Princeton University, and grant 1 R01 HD41141-01 for the Science of Ecology and Early Development at Princeton University.

Privileging the Participant:
The Importance of Take-Up Rates In Social Welfare Evaluation

Abstract

This paper analyzes how variation in participant take-up rates affected the impacts of the New Hope project, a random-assignment, anti-poverty program. New Hope offered experimental members four benefits - child care subsidies, wage subsidies, health insurance, and, if needed, a temporary community service job - that were available to families working full time. Take-up of the benefits was far from universal and experimental participants who used one of the benefits rarely used all of them. Clustering and propensity score methods are used to analyze take-up sub-groups and to estimate program impacts within each. All take-up patterns adopted by experimental members were associated with at least one positive program impact. However, the primary beneficiaries were those parents who used the community service jobs. They increased their employment effort, felt less stressed, and had children with higher teacher-related academic accomplishment scores.

Key Words: Program evaluation, anti-poverty programs, family well-being

Introduction

During the 1990s, a number of evaluations were conducted on several randomly-designed anti-poverty experiments in both the United States and Canada, including programs in Florida, Connecticut, Minnesota, Georgia, and New Brunswick (Bloom and Michalopoulos, 2001; Knox, Miller, and Gennetian, 2000; Morris and Michalopoulos, 2000). Although program components varied by project, evaluators hypothesized that family and child well-being would be enhanced by increases in human capital, income, and use of high quality child care (Sherman, 2001; Hamilton, Freedman, and McGroder, 2000; Huston et al., 2001). A recent synthesis of the evaluation efforts indicated that many programs had modest impacts; the most successful programs were those that offered a combination of financial and work-based supports (Morris et al., 2001; Duncan and Chase-Lansdale, in press).

One evaluated program was the New Hope project, an experimental, anti-poverty program, which was designed to make work pay for its low-income participants. The program, conducted in Milwaukee, Wisconsin, enrolled approximately 1300 participants beginning in August of 1994. If employed for thirty hours or more a week in a given month, experimental group members were eligible for four benefits. First, they were eligible to receive a wage supplement that ensured that the net income of families increased as they earned more on the job. The supplement did not begin to phase out until families were above the poverty line. Second, the program offered subsidized health insurance through a Health Maintenance Organization (HMO). Third, participants received job search assistance regardless of employment status. If a suitable job could not be found they were offered a community service job (CSJ) for at least six months, although the pay was minimum wage. The final benefit was a childcare subsidy. New Hope participants who had at least one dependent child under age thirteen and worked 30 hours per week were able to choose any state licensed or county-certified child care provider, including providers of both preschool programs for young children and extended day programs for school-age children. The two-year evaluation indicated that New Hope had generally modest impacts on earnings and employment, and significantly boosted children's academic behavior, especially that of boys (Bos et al., 1999; Huston et al., 2001).

The evaluation of New Hope, like evaluations of the other social welfare programs, exploit the randomized design of the project and concentrate on average differences between experimental and control members. Yet New Hope ethnographic evidence and administrative and survey-based evidence have indicated wide disparities in experimental group member's use, or take-up, of the four benefits. Even though experimental participants were free to choose and use benefits as they saw fit, take-up of services was sporadic and far from universal (Gibson and Weisner, in press; Bos et al., 1999). No evaluation of New Hope, though, has systematically analyzed the disparity in take-up rats to understand how these variations in benefit use may influence program impacts.

In this paper, I augment previous evaluations of New Hope by classifying experimental participants according to benefit use, and estimate how program effects varied by these participant take-up patterns. I find that enrollees in the program varied widely on demographic characteristics prior to the program, and that this variation influenced subsequent program responses. I also find that most of the program's impacts are concentrated within one take-up sub-group – those who used a community service job - although my results also indicate that all experimental members who were formally involved with New Hope show a positive program impact on at least one measure of employment, parenting, or child well-being.

Analyses of take-up patterns within a multi-faceted program like New Hope are important because they offer program insight not possible with more traditional evaluations.

First, my analyses link take-up strategies with program impacts; it is thus possible to see *which* strategy is responsible for driving *which* program effects. Relatedly, it is a way of evaluating the relative value of the four benefits. It is possible, for example, that one component of the program contributed little to participant well-being. This knowledge is useful for those who may be planning further interventions similar to that of New Hope's. Second, my analyses associate socio-demographic attributes with take-up sub-groups, illuminating the interaction between personal characteristics and program participation. This is particularly useful in the case of experimental group members who did not take up the New Hope offer, as it provides evidence for why they did not participate.

To analyze service utilization, I use a three-stage methodological approach. The first stage is to examine patterns of benefit take-up. To do this, I use a clustering methodology that classifies experimental group members according to their take-up patterns of New Hope benefits. Once these have been identified, the second stage is to simulate a random assignment experiment within each of these sub-groups. This is done using propensity scores, which associate baseline socio-demographic characteristics with the probability of sub-group membership for experimental and control group members. Differences in outcomes between experimental and control group members should then be due to the program, and not self-selection bias. The final stage is estimating New Hope's program impacts within each of these service use sub-groups.

My results indicate that New Hope benefit use was heterogeneous and impacts were conditional on participant involvement. I estimate the presence of five distinct service use sub-groups: multi-supplement users, child care and wage supplement users, health insurance and wage supplement users, community service job users, and non-participants. Those who were doing relatively well economically at baseline only needed New Hope to satisfy their child care or health insurance needs. With those needs met, parents indicated that they felt less harried and were more hopeful that they could accomplish their day-to-day tasks. Those who had poor work histories at baseline used New Hope to access the labor market and as a result, they significantly increased their employment and felt less parenting stress. Their children also scored consistently higher on measures of school-based achievement and behavior. However, another sizeable group who also had poor work histories never participated in the program at all; socio-economic characteristics indicated that they might have simply had too many barriers to sustain the necessary employment effort. Given this heterogeneity, I conclude that variation in program response is an important component in an evaluation of multi-faceted anti-poverty programs.

Literature Review

In order to situate my evaluation of New Hope in the larger social welfare evaluation literature, I will first briefly review the impacts of three potential pathways by which New Hope could affect family and child well-being: increases in income, maternal employment, and child care use. I will then discuss participation in voluntary anti-poverty programs, and conclude with a review of how personal characteristics may affect take-up of program components.

Pathways of Influence of Anti-Poverty Programs

Programs like New Hope that mandate a work requirement could positively affect families and children through increases in income, maternal employment, and use of formal child care (Guo and Harris, 2000; Bos et al., 1999; Brock et al., 1997; Morris et al., 2001; Huston et al., 2001). Research has shown, for example, that poorer mothers are more stressed, adopt harsher parenting styles, provide a less cognitively stimulating home environment, and suffer from more mental health problems than do those of a higher socio-economic status (Guo and

Harris, 2000; McLoyd, 1990; McLoyd et al., 1994; Elder et al., 1995; Seccombe, 2000; Magnuson and Duncan, in press). Likewise, living in poverty is associated with a host of detrimental outcomes for children, including increased risk of health problems, lower cognitive functioning, fewer years of completed schooling, less labor market success, and increases in teen-age fertility (Parker, Greer and Zuckerman, 1988; Haveman and Wolfe, 1995; Duncan et al., 1998; Seccombe, 2000).

Some researchers argue that the impacts of maternal employment on family well-being are inconclusive (Magnuson and Duncan, in press). While recent reviews of six work promotion programs found no negative effects on parenting, other research has shown that jobs that are low in complexity and autonomy may increase feelings of stress and decrease feelings of control and independence (Perry-Jenkins, Repetti and Crouter, 2000; Chase-Lansdale and Pittman, in press; Parcel and Menaghan, 1994, 1995). Other reviews have concluded that additional research on maternal employment is needed, particularly for the impacts on children. Another review found that studies of maternal employment on child well-being, impacts differed by the method and data used and results were inconclusive (Harvey, 1999). Important questions about maternal employment still need to be addressed, including more research into the effects of non-traditional work hours and other "non-normal" job characteristics (Magnuson and Duncan, in press; Perry-Jenkins et al., 2000).

Previous studies indicate that non-maternal care may be beneficial for low-income children, because it provides a more stimulating environment than children would otherwise receive (Gamoran, Mare and Bethke, 1999; Posner and Vandell, 1999; Currie and Thomas, 1995). Intensive, high-quality intervention programs have had positive impacts on the cognitive and behavioral outcomes of young children, with some impacts lasting well into adolescence (Currie, 2000; Barnett, 1995). However, low-income children are more likely to be in lower quality after-school and day care programs, and thus not reap the benefits of high quality care (Currie, 2000; Hofferth et al., 1994; Vandell and Wolfe, 2000).

Participation in Voluntary Social Welfare Programs

As is typical with most evaluations of randomized social programs, New Hope was evaluated in terms of its "Intent to Treat" (ITT) effects (Shadish, Cook and Campbell, 2001; Heckman and Robb, 1985; Katz, Kling and Liebman, 2001). ITT estimates are the average experimental versus control differences across the full participant population, without regard to how the treatment was received or used (Shadish et al., 2001). However, as non-participation among experimental group members is often quite high, an alternative is the "Treatment on the Treated" (TOT) estimate (Katz et al., 2001; Heckman and Robb, 1985; Bloom, 1984; Quint, Bos and Polit, 1997). Instead of estimating an average effect for all experimental participants, TOT parameters measure the average effect for all experimental group members who actually took up the treatment.

However, even TOT estimates may obscure experimental impacts for multi-component programs like New Hope. Yoshikawa and his colleagues (2001) demonstrated that analyses of take-up patterns in programs with multiple benefits could reveal impacts not found in conventional ITT or TOT analyses. They used clustering and Heckman selection models in their analyses of the New Chance program, a welfare-to-work program for teenage welfare mothers. The researchers found that of the seven strategies adopted, experimental mothers who stressed the development of their human capital through education and job training had children who scored higher on school readiness measures than did comparable control members. This was encouraging to the designers of New Chance, as ITT estimates of the program's impact had shown that children of mothers in the experimental group actually fared worse than those in the control group (Quint et al., 1997). Besides the work of Yoshikawa and his colleagues, though, very little research has been done to link patterns of service utilization to program impacts, even though randomized programs with multiple elements are common and even encouraged by policy makers (Bloom and Michalopoulos, 2001; Morris et al., 2001).

Qualitative work indicates that an analysis strategy, such as the one adopted by Yoshikawa and colleagues, may be a fruitful one for an evaluation of New Hope. Data from the New Hope Ethnographic Study (NHES), a randomly selected set of 44 families from the larger New Hope sample, indicates that there was great heterogeneity in the take-up of New Hope benefits (Gibson and Weisner, in press). Both rational choice and ecocultural models were used to explain participants' choices within the New Hope project. The rational actor model explains families who make benefit-use choices based on a cost-benefit calculus, which is primarily financial in nature. The ecocultural model explains benefit use for families who try to construct a meaningful, sustainable daily routine. Neither theoretical framework alone was sufficient to fully explain take-up, but instead participants were motivated by a combination of each. The heterogeneity of responses indicates the importance of variation for evaluating program impacts.

Personal Characteristics and Take-Up Patterns of Benefits

Previous research has indicated that harder-to-employ populations are more likely to have weaker labor force attachment, lower previous earnings, less educational attainment, and higher occurrences of mental and physical health problems (Danziger et al., in press; Meyer and Cancian, 1999; Harris, 1996). This could characterize the group of New Hope participants who use the community service job benefit (CSJ) (because they cannot find employment on their own) or those who do not participate (because they cannot sustain the work requirement). Those likely to use the health insurance subsidy are more likely to come from low-wage jobs or earn insufficient amounts to obtain insurance through other means (Royalty, 2001; Currie and Yelowitz, 1999). Research has also shown that those most likely to use non-maternal care (and hence need the child care subsidy) are more likely to be single parents, have higher maternal incomes, and smaller families with older children (Singer et al., 1998; NICHD, 1997; Brayfield and Hofferth, 1995; Hofferth et al., 1994).

For other types of participants, however, connections between socio-demographic characteristics and service use are less clear. Those who do not participate may either have significant initial obstacles that prohibit them from sustaining the necessary work effort. Conversely, their employment and family situations may be such that they do not need New Hope's assistance (Danziger et al., in press; Magnuson, 1999; Weisner et al., 1999). Furthermore, almost nothing is known about baseline demographic conditions and the use of any two benefits in combination with each other.

Methodology

Data and Measures

Data were gathered from six quantitative sources. First, just prior to random assignment, 1,357 volunteer families completed a baseline survey form that provided information on an array of socio-demographic characteristics. Second, a survey administered two years after random assignment measured employment experiences and work-related outcomes of the program and control groups; about 1,086 individuals completed the two-year survey (a response rate of 80 percent). Third, a subset of 812 experimental and control parents also answered additional questions about their family practices and children's well-being for the Child and Family Survey

(CFS). Every family who had at least one child between the ages of one and ten at baseline qualified; up to two children were chosen from each family. Of the 812 families, 578 were used in the final analysis (demographic characteristics of the full and the CFS sample are presented in Table 1).[1] Fourth, teachers of New Hope children ages five to twelve were asked to rate the children on a variety of academic behaviors and skills. Of the 557 possible responses, 418 were returned for a response rate of 75 percent. Fifth, a database maintained by New Hope as part of its management information system provided data on the use of benefits by all program participants. Sixth, evaluators obtained administrative data on receipt of benefits from the state TANF welfare and food stamp programs, earnings data reported by employers to the Social Security system and, in aggregated form, state Earned Income Tax Credit payments.

To measure the impact of the program, I examine fourteen outcomes in four areas: parental employment, parental emotional well-being, child care use, and child academic achievement and behavior. All scales have adequate internal consistency and reliability (Huston et al., 2001). A detailed description of each appears below, and means and standard deviations appear in Table 2.

<u>Employment and earnings</u> Parental income is the summation of two years of post-baseline quarterly earnings reported to the Social Security system. Number of quarters employed is the number of quarters over two years that the parent has earnings that were reported to the Social Security system.

<u>Parental well-being</u> Time pressure was a two-item measure, which asked respondents how often they felt rushed in general and how often they had extra time, using a five-point scale (1= never; 5 = all of the time). The State Hope Scale (Snyder et al., 1996) is comprised of two sub-scales. The agency subscale measures "belief in one's capacity to initiate and sustain actions" (e.g., "I am meeting the goals I set for myself") and the pathways subscale measures "belief in one's capacity to generate routes" (e.g., "I can think of many ways to reach my current goals"). The response scale was five points (1=strongly disagree, 5=strongly agree). Test-retest reliability over one month ranges from .48 to .93, suggesting that the scale is responsive to changes in individuals' lives (Huston et al., 2001). Stress is a one-item question that asked how often the respondent felt stressed in the last month, using a five point scale (1=none of the time, 5=almost all of the time). The calculated score was a dichotomy - much or almost all of the time as opposed to the other categories. The Pearlin Mastery Scale (Pearlin et al., 1981) is intended to measure internal locus of control. It contains seven items (e.g., "There is really no way I can solve some of the problems I have") and uses a four point response scale (1=strongly disagree, 4=strongly agree). Self-esteem was assessed using the Rosenberg Self-esteem Scale (Rosenberg, 1979), which consists of ten self-evaluative items (e.g., "On the whole, I am satisfied with myself") rated on a four-point scale (1=strongly disagree, 4=strongly agree).

<u>Parenting Skills</u> Degree of parenting control is a five-item scale from the Self-Sufficiency Project (Statistics Canada, 1995). It includes items about consistency and effectiveness of discipline (e.g., how often the child ignores the parent's punishment), which used a six-point response scale (1=never, 6=all of the time). Parenting stress is an eight-item scale (Quint et al., 1997) that contains two subscales: three items about the overall stress associated with parenting

[1] Some 67 Hmong participants in the program were not included in the family study because of cultural and language differences. Disregarding this portion of the sample, the final response rate was 78 percent.

(e.g., whether they felt “trapped by responsibilities as parents”) and five items about the stresses associated with parenting the target child (e.g., “My child seems to be much harder to care for than most”). Both sets of items used a five-point response scale (1=not at all true, 5=very true). Monitoring was measured with four items that were taken from the five-year follow-up to the JOBS parent and child assessment. Items ask parents about their familiarity with the child’s friends and their knowledge of the child’s whereabouts and companions when away from home using a five-point response scale (1=never/almost never; 5 = always).

Use of Child Care Formal care includes the number of months that the child spent in Head Start, a pre-school, nursery school or child care center, school-based extended day care, or any other program that was not in someone’s home. Informal care is the number of months that the child was cared for by a family member or someone outside their household.[2]

Children’s Academic Outcomes Academic achievement is the academic subscale of the Social Skills Rating System (SSRS) (Gresham and Elliott, 1990). On this ten-item measure, the teacher rates the child’s performance on a five-point scale (1 = bottom 10 percent, 5 = top 10 percent) in comparison to others in the same classroom on reading skill, math skill, intellectual functioning, motivation, oral communication, classroom behavior, and parental encouragement.
The positive behavior scale is 25 items that are divided into three subscales: compliance/self control (e.g., thinks before he/she acts, usually does what I tell him/her); social competence and sensitivity (e.g., gets along well with other children, shows concern for other people’s feelings); and autonomy (e.g., tries to do things for him/herself, is self-reliant). The five-point response scale ranges from 1 = never to 5 = all of the time. Problem behavior consists of three subscales of the SSRS: externalizing, internalizing, and hyperactivity (Gresham and Elliott, 1990). Externalizing problems include aggression and lack of behavior control (e.g., “is aggressive toward people or objects,” “has temper tantrums”). Internalizing problems include social withdrawal and excessive fearfulness (e.g., “appears lonely,” “acts sad or depressed”). Hyperactivity measures the child’s ability to pay attention and cooperate (e.g., “is easily distracted” and “disturbs ongoing activities”), and how often teachers had to discipline the child for misbehavior (1=never to 5=several times a week).

Overview of Analysis

I used a three-part methodological approach to determine patterns of benefit use and to link those patterns of program use. First, I used clustering methodology to distinguish among different program take-up strategies for the experimental group members (Magnusson, 1998; Yoshikawa, Rosman and Hsueh, 2001; Bergman, 1998). Second, using propensity score methodology, I match similar controls to experimentals within each take-up strategy, thereby creating a proxy for a randomized experiment within each service use sub-group (Rosenbaum and Rubin, 1984). Finally, I compare regression-adjusted differences on parent and child outcomes between the experimentals and controls within each sub-group. The difference should represent the New Hope impact associated with each program strategy. Each of these three steps is now explained in further detail.

[2] In cases where the number of months of formal or informal care exceeded 24 (reflecting simultaneous uses of different types of care), the number of months was imputed back to 24.

Cluster Analysis

To analyze patterns of New Hope take-up, I created clusters representative of people's possible choices within the program (Milligan, 1996). I used the number of wage supplements, child care supplements, and health insurance supplements experimental participants received during the first two years of New Hope. I also included the number of months that people were involved in community service jobs (CSJ). All estimates come from New Hope's Management Information System (MIS).

Following the recommendation of cluster methodologists, clusters were created using a two-step process (Bergman and Magnusson, 1997; Milligan, 1996). First, Ward's algorithm was used to form hierarchical clusters on standardized variables. Under hierarchical clustering, each observation begins as its own cluster. As the algorithm progresses, clusters are compared using squared Euclidean distance to group observations that result in the smallest increase in the overall sums of squares. This results in a minimization of within-cluster variation. The second step uses the centroids obtained from the hierarchical cluster method and, using a k-means iterative process, reclassifies observations, assigning them to the closest centroid. The clusters are recalculated using the new centroid, and the process continues until either no other changes are possible or the maximum number of iterations has been reached. Because of the sensitivity of this method to outliers, a preliminary step taken before clustering was the removal of outliers from the analysis (based on an average squared Euclidean distance greater than .5). After the removal of three outliers, a total of 276 experimental families were included in this sample. All clustering was done using the SLEIPNER software program (Bergman and El-Khouri, 1998).

Propensity Scores

To account for self-selection associated with different take-up strategies, I used propensity score methods to create comparable control groups for each service use sub-group (Rosenbaum and Rubin, 1984; Dehejia and Wahba, 1998). Propensity scores use baseline characteristics to predict the likelihood of experimental group members belonging to a particular sub-group. The probability estimates are then calculated for each control group member, using the same set of baseline characteristics. Assuming that all covariates related to program status are observed (known as the "ignorable treatment assumption"), any given sub-group will have experimental and control group members who are equivalent in regards to baseline characteristics (Dehejia and Wahba, 1999). A comparison between the two should provide an estimate of the program impact, because a randomized experiment has been simulated within each service use sub-group.[3]

Given the number of possible combinations, exact matching based on multiple baseline covariates seems unlikely (Rosenbaum and Rubin, 1984; Rubin, 1979). However, propensity scores can account for the inherent dimensionality problem in comparing two individuals. They reduce what is actually compared by constructing a cumulative index for each individual that associates their observable characteristics with the likelihood of sub-group membership. This index represents the association between all characteristics and program response; matching is done on the index score and not on individual characteristics. Matching on one number (instead of on *n* dimensions) greatly increases the likelihood of a matched pair (Dehejia and Wahba, 1998).

[3] This, of course, only approximates a randomly-designed experiment where differences on outcomes can be attributed to program status and not systematic differences between the experimental and control groups.

The first step, then, is to construct the propensity scores for each of the service use sub-groups. Membership in a sub-group *j* for program member *p* (SG_{jp}) is a function of the probability score of baseline characteristics (bc), as modeled in equation 1:

$$\text{Prob}\ (SG_{jp} = 1) = \Phi(bc_p\beta) \qquad (1)$$

Where the probability of *p* being in *j* is a function of the cumulative normal distribution of *p'*s vector of baseline characteristics ($bc_p\beta$). The baseline characteristics used in parent outcome models are: earnings and number of quarters employed the three months prior to New Hope enrollment, if the participant was working full time at baseline or had ever worked full time (full time defined as thirty hours a week or more), age, race/ethnicity, gender, family structure, education, owning a car, resident location (north or south side), currently receiving any aid (includes AFDC, food stamps, Medicaid, social security, unemployment, and general assistance), and if the parent had lived in a household that had received aid when the parent was a child. For models using child outcomes, I also included the sex and age of the child.[4]

Estimates of β obtained in (1) are then used to predict the likelihood of any program group member belonging to a service sub-group as presented in equation 2:

$$\text{Prob}(SG_{ji}=1) = bc_i\,\hat{\beta} \qquad (2)$$

The predicted value of belonging to a sub-group for any New Hope enrollee *i* [as opposed to just program participant *p*, as was the case in (1)] is the product of their baseline characteristics (bc_i) and $\hat{\beta}$, the estimates of β from (1). This is tantamount to estimating how the control group would have responded to the offer, even though this response is not observed. The end result is that every participant, regardless of treatment assignment status, has a propensity score that predicts the likelihood of their belonging to each sub-group ($SG_{j,\, j=1\ldots n}$) based on their baseline characteristics.

In order to construct matches for each of the sub-groups $SG_{1\ldots}SG_n$, I followed procedures recommended by Dehejia and Wahba (1998). For sub-group SG_k, I ranked in descending order the propensity scores of each participant. Propensity scores were then divided into deciles, so that each decile had the same number of observations (but not necessarily the same number of experimentals and controls). Each decile was tested for significant differences on individual covariates between experimental and control group members using Hotelling's T-squared generalized means test. If the means were different, then interaction terms were added until there were no longer significant differences within the deciles (Dehejia and Wahba, 1998; Rosenbaum and Rubin, 1984). Non-significant differences within the decile indicate that, based on observed characteristics, the experimental and control group members are similar, approximating a random assignment (Rosenbaum and Rubin, 1985).

---

[4] Although the same set of covariates was used for all models, the type of variable (continuous versus dummy) did differ. The decision to use a continuous versus a dummy variable was based on the goodness-of-fit test for the probability model. Based on analyses not shown, the use of either type of variable does not significantly affect the results. Furthermore, for the community service job and non-participant models, the dummy variable "having ever worked full time" was substituted for "working full time at baseline" and an additional control was added for having had no prior employment in the past three months. Both of these variables explained significantly more variance in these models than if they had been omitted.

Using a nearest match algorithm, I then matched experimental and controls on their probability scores. I matched with replacement, so that controls could be matched to more than one experimental. Otherwise, if there were only a few comparable controls, then the remaining experimentals would have to be matched to very different controls. Any non-matched controls were discarded from my sample (Dehejia and Wahba, 1998).

Unlike previous uses of propensity, I do not include all matched pairs in my regression analyses (Dehejia and Wahba, 1998; Lalonde, 1986; Agodini and Dynarski, 2001). Program impacts should be calculated for a sample size proportionate to the size of the take-up sub-group. Therefore, I stratified my sample before it was matched by including only participants whose propensity score met a certain threshold. For example, if non-participants account for 40 percent of the experimental sample, I would include in non-participant sub-group models only New Hope members whose propensity score was in the top 40 percent of the non-participant probability distribution (I am calling this "meeting the mean probability threshold", labeled as "PT" in the tables).[5] Those whose propensity scores were in the lowest 60 percent were discarded. As a robustness check, I also stratified my sample on plus or minus five points from the mean probability threshold (e.g., PT + 5 percent, PT – 5 percent).[6]

After my sample has been restricted to the appropriate probability threshold, I can then model the New Hope impact, using the propensity scores obtained in (2):

Where $\text{Prob}(SG_{ji}=1) > Z$:

$$NHO_i = \alpha + \delta BASECHAR_i + \beta_1 EGS_i + e_i \qquad (3)$$

In (3), given that the propensity score of person *i* to be in SG_j is greater than *Z* (where *Z* is proportionate to group sample size), a New Hope measured outcome (NHO) is regressed on a vector of baseline characteristics (BASECHAR) and the dummy variable for experimental group status (EGS). Baseline covariates used to create the propensity scores also comprise BASECHAR. The parameter on EGS represents the regression-adjusted difference between the control and experimental groups within sub-group SG_j.[7] All models were weighted by the inverse of the number of control observations that were matched with more than one experimental observation.

Results

Cluster Solution

Figure 2 presents the results for a five-cluster solution. Milligan (1996) argues that a cluster solution should explain at least 65 percent of the variance, and should be significantly different than that same cluster solution for randomly formed sub-sets of the data. The five-cluster solution explained 69.3 percent of the variance and had significantly higher sum of squares than other five-cluster solutions generated in twenty random data sets (13.95, p-value < .01). Although a six cluster solution explained slightly more variance (73.1 percent), the clusters

---

[5] It is important to note that this does not mean that they had a propensity score of 40 percent or higher to be a non-participant. Rather, their scores were in the top 40 percent of the non-participant probability distribution.

[6] Due to the extremely small sample sizes that result from a probability threshold of mean level plus five (PT+5 percent), these results are not presented in the main tables. However, they, with the other two threshold cut-offs, can be found in Appendix Tables B and C.

[7] Although bootstrapping of the propensity score estimates may be indicated because of the non-randomness of the samples, recent empirical work has demonstrated that bootstrapping does not change the standard error estimates of propensity score parameters (Agodini and Dynarski, 2001).

it created were too small for analysis. A four cluster solution explained less variance (63.1 percent), falling below the recommended 65 percent.[8]

Each bar on this figure represents the number of months a benefit was received, arranged by benefit take-up groups.[9] The first set of bars, with the label "Health insurance/wage supplement" under it, for example, indicates that this group heavily utilized the health insurance and wage supplement components of the program, but had little to no involvement with child care subsidy or the community service job benefit.

Figure 2 shows variation in the response to the New Hope program offer. The largest of the supplement-based groups is the child care/wage supplement (CCWS) group (n = 55). Equal in size is the multi-supplement (MS) group (n = 30) and the health insurance/wage supplement (HIWS) group (n = 33). The HIWS group used more health insurance than did any other group, but did not take-up the child care subsidy. The MS group used all three supplements at a higher rate than did the other two supplement-based groups. The largest single group of experimentals was the non-participant (NP) group, which had no formal involvement with New Hope (n = 113). The CSJ group used New Hope primarily for access to employment, although they did receive a few wage and health insurance supplements as well. They were similar in size (n = 45) to the CCWS group.

Demographic Characteristics of the Five Sub-Groups

Table 3 presents the baseline demographics for the full CFS sample and each of the service use sub-groups groups, using matched experimentals and controls.[10] The first column of the first row, for example, indicates that the prior earnings in the last three quarters for the full CFS sample (both experimentals and controls) was $2780. The second column in the second row indicates that past earnings of experimental and controls whose propensity score met the mean probability threshold for the health insurance/wage supplement group was $8164.

This analysis indicates that New Hope participants differed in baseline characteristics both economically and by family structure. Economically, the health insurance/wage supplement (HIWS) and multi-supplement (MS) groups are relatively advantaged at baseline, having earned $8164 and $5930, respectively, in the three quarters prior to New Hope enrollment. More than four-fifths of both of these groups were working full-time at baseline. The least advantaged groups are the non-participants (NP) and the CSJ users. They have very low earnings (about $1300 for each) and less than one-fifth of the NP and one-twentieth of CSJ users were working full-time at baseline. The CCWS group had earnings that were less than the most advantaged (the HIWS and MS groups) but more than the least advantaged (the NP and CSJ groups). As for family structure, almost one in three members of both the HIWS and NP groups were married, rates that were 20 percentile points higher than any of the other groups. And almost three-fourths of the MS families had a child that was under the age of two – a rate that was three times as high of that of the HIWS families.

Program Impacts by Service Use Sub-Group

Results of the New Hope impacts using the five service use sub-groups are presented in Tables 4 through 6. For all groups except the non-participatory group, parent outcomes are

---

[8] It is an artifact of cluster creation that the larger the cluster solution, the higher the amount of variance explained. That is why this criterion must be used in combination with significance tests and interpretability criteria.

[9] Although the analyses were done on standardized variables, the figure presents the results in months to ease in interpretation.

[10] Appendix Table A presents the three probability thresholds and sample sizes for each of the sub-groups.

presented in Table 4 and child outcomes in Table 5.[11] Impacts for non-participants on both parents and children are presented in Table 6. In order to ease comparison across measures with different scales, coefficients on parental emotional well-being, parental skills, and child academic outcomes represent effect sizes.[12]

In all tables, the first column represents the regression-adjusted differences between New Hope experimental and control members for the full CFS sample.[13] For each of the sub-groups, there are two columns, representing the regression-adjusted differences between likely experimental and control members for that group. The first column for each sub-group (labeled "PT") represents enrollees whose propensity score met the mean probability threshold, as explained above. The second ("PT +5") is for enrollees whose propensity score met the mean probability threshold plus 5 percent. In Table 4, for example, the first row of the first column indicates that in the CFS sample, the average experimental member earned $1224 more than did the average control member (a statistically insignificant difference). In the first row, second column, experimental participants whose propensity scores met the mean probability threshold for the health insurance/wage supplement (HIWS) sub-group distribution earned a non-significant difference of $136 more than did controls in the same circumstances. And those whose propensity score for the HIWS groups was at mean level plus five (in this case, PT+5 percent = 0.16) had a non-significant program impact of $811.

The group that displayed the biggest earnings impact was the CSJ users (first panel). They worked an average of two quarters more over controls, an increase of 25 percent. No other group showed a significant change in quarters employed, and the child care and wage supplement (CCWS) and the health insurance and wage supplement (HIWS) group showed non-significant decreases. Unfortunately for the CSJ users, however, this increase in work hours did not translate into an increase in earned income. In fact, no sub-group showed a significant increase in earnings, although MS and HIWS group showed non-significant increases.

There were a number of noteworthy impacts on parental emotional well-being (second panel). On the positive side, New Hope increased selected parents' feelings of being able to accomplish their goals, as indicated by the significant increase in the agency sub-scale score for two of the sub-groups. The finding was slightly larger for HIWS members (effect size .70, p-value <.10) than for CCWS members (.51, <.05). The HIWS group also felt less stressed (-.48, p-value <.10). CSJ parents reported higher scores on the measures of self-esteem and the path sub-scale, although these results were only significant at the PT+5 percent threshold level. However, the coefficients increase from the PT to PT + 5 percent models, and the standard errors decrease, indicating that the sample size may simply have been too small in the PT model. There was one negative finding: the MS group had a showed a significant decrease in their feelings of mastery (.50, p-value <.05).

Turning to Table 5, top panel, there are few impacts of the program on parenting skills. New Hope did decrease the parenting stress of likely CSJ users (-.29; p-value <.05). It also decreased feelings of control and increased feelings of monitoring for this sub-group, but neither

---

[11] As the parenting skill measures have the child as the unit of analysis, they are presented on Table 5 instead of on Table 4. Parental emotional well-being outcomes have the parent as the unit of analysis.

[12] Effect sizes are the difference between experimental and control outcomes expressed as a proportion of the standard deviation of the outcome for both groups combined. As these outcomes have been standardized to have a mean of zero and a standard deviation of one, the effect sizes represent the change relative to one standard deviation. Effect sizes of 0.20 are considered small, 0.50 are medium, and 0.80 are large (Cohen, 1988).

[13] The New Hope program impacts that I report are marginally different than those found in the two-year evaluation (Bos et al,, 1999), as I use a slightly different set of covariates.

of those results was significant across thresholds. No other sub-group showed any significant impacts.

Looking at the child care impacts (middle panel), the most striking result is the increased use of formal child care for experimentals in every participatory group, except for the HIWS group. Impacts ranged from an increase of three months (the CCWS group; p-value <.10) to seven months (the MS group; p-value <.01). Bigger impacts for the MS group relative to the CCWS group mirror the usage of the subsidy (as presented in Figure 1), as the parents in the MS group received more child care subsidies than did those in the CCWS group. Experimental children in the CCWS group, however, were the only experimental children who spent significantly less time in informal care than did their control counterparts.

As for teacher reports of academic behavior (the bottom panel), New Hope had the biggest impacts on children in CSJ families. They showed vary large impacts on all three measures, ranging from an increase in positive behavior of .82 (p-value <.01) to an increase in academic achievement of .51 (p-value <.05). The models for all teacher reported measures, however, have extremely small samples, so the large effect sizes should be viewed cautiously. No other consistent impacts were found for the other three sub-groups.

There were few significant impacts on non-participants, except for a marginal increase in employment and one measure of parental emotional well-being (Table 6). Likely experimental non-participants, relative to likely non-experimental controls, report working one half of one quarter more over the two year period, but this was only marginally significant at the PT + 5 percent level (p-value < .10). They also report decreases in their feelings of stress (-.24, p-value <.10). There were no program impacts for non-participants on earnings, parenting skills, use of child care, or child's academic achievement.

Discussion

The demographic characteristics of the five take-up sub-groups indicate important heterogeneity in the New Hope population. The sub-groups show important differences in earnings, education rates, prior labor force attachment, and household structure at baseline. Those likely to use the wage-based benefits (the MS, HIWS, and CCWS sub-groups) had higher employment rates and prior earnings, lower rates of welfare receipt, and higher degrees of educational attainment. Conversely, those who were likely not to participate or to use New Hope only for job access (the NP and CSJ sub-groups) had poor labor ties, lower educational attainment, and higher rates of receiving aid. This finding of New Hope enrollees' heterogeneity is consistent with previous research that recognizes the "diversity in the poverty experience" (Seccombe, 2000). Low-income populations differ by household structure, race, education, labor force attachment, geographical region, and physical and mental health (Bumpass and Lu, 2000; Danziger et al., in press; Harris, 1996; Blank, 1997). In this regard, the population enrolling in New Hope was typical of the low-income and working poor populations, who have divergent life circumstances and may therefore require different types of assistance.

Given this heterogeneity, New Hope was wise to offer a "menu" of services that allowed participant flexibility. Two groups who were already working full-time at baseline used New Hope to fill a specific need: one group took advantage of the child care subsidy and another access to health insurance. With these needs provided for, parents in these groups, relative to similarly situated controls, showed a decrease in stress and an increase in their feelings of agency. The strategy adopted by these two sub-groups is consistent with prior qualitative and quantitative research, which indicated that some families in New Hope were only one "boost" away from economic self-sufficiency (Gibson and Weisner, in press; Magnuson, 1999). These parents did not need much assistance from New Hope, but effectively utilized the services that were offered.

Overall, however, there were very few impacts on parental well-being and parenting skills. This finding is consistent with previous research, which indicates that parenting mental health is rarely impacted by welfare-to-work evaluation programs (Chase-Lansdale and Pittman, in press). Duncan and Chase-Lansdale (in press), in their examination of five randomized-design welfare to work programs, find no significant changes on maternal mental health, parental control, parenting styles, or cognitive stimulation in the home. They hypothesize that even the best programs cannot provide enough support for parents to cope with the challenges of working full-time, finding and maintaining child care, and being an effective parent.

Yet clearly the use of CSJ made a difference in the lives of people who used them. Of the four sub-groups who were involved in the program, those in the CSJ group were associated with the most program impacts. Not only did they increase their employment relative to similarly situated controls, they felt decreases in parental stress and increases in their ability to monitor their children. Their children spent 25 percent more time in formal care than did control group children. All three of these factors – decreases in parental stress, increases in gate keeping, and time in formal care – have been linked to improved child well-being for low-income children (Magnuson and Duncan, in press; Duncan and Chase-Lansdale, 2001; Seccombe, 2000; McLoyd, 1990). It is not surprising, then, that children of CSJ users showed positive program impacts on academic skills and behavior. However, it is also true that these families did not have a significant increase in earnings – an artifact, perhaps, due to the minimum wages that they were earning. It is also curious that they did not avail themselves of the child care subsidy, which would also represent a significant drain on their resources.

Yet the positive program impacts for this sub-group are noteworthy because the characteristics of the CSJ users mirror those in the welfare-reliant population who are unable to make a successful transition to employment (Danziger et al., in press; Pavetti and Acs, 2001; Kalil et al., 1998). If this program element could be replicated, then it may be possible to increase the labor market ties of a population that otherwise is struggling to do so.

However, the CSJ element of the New Hope program was more supportive and nurturing than other job access programs (see Poglinco et al., 1998, for details of New Hope CSJs). First, an attempt was made to "match" CSJ users to the job, so that the skills of the participants could be developed. Second, case representatives monitored the placements and provided both pragmatic and emotional support. Third, almost all of the placements were in other social service agencies, which had similar goals to that of New Hope in promoting the economic well-being of disadvantaged populations. All of these factors combined meant that CSJ users felt that they were engaged in "real" jobs learning valuable skills, and nearly 60 percent of them were able to transition their placements into full-time employment (Brock et al., 1997; Bos et al., 1999).[14] This highly personalized and supportive approach is not only time-intensive, it may not be feasible on a larger scale.

Yet experimental non-participants, who at baseline looked very similar to community service job users, did not utilize New Hope for employment access. This raises the issue of what differentiates these two groups. One possible explanation is the differences in family structure: while one in three likely non-participants was married, only one in twenty likely CSJ users had a

[14] The distribution of the CSJ placements was as follows: 35 percent were in office support, 29 percent building construction or property maintenance, 9 percent child care, 8 percent food service, and the rest were in other types of occupations (Poglinco, Brash, and Granger, 1998).

spouse. These low marriage rates for CSJ users are similar to those found in an analysis of New Hope community service jobs by Poglinco, Brash, and Granger (1998), who speculate that fewer family ties may have made it easier for single parents to use a CSJ. An alternative explanation is that CSJ users were limited in their outside resources, as they could not rely on spouses for additional income.[15]

Unfortunately, participant choice of benefits resulted in a sizeable portion of the experimental group choosing not to formally participate. Non-participation is a chronic problem in social service evaluation (Bloom, 1984); one randomized experiment that offered a generous wage subsidy had experimental participation rates as low as 30 percent (Quint et al., 1997). Perhaps non-participation was especially likely in a program like New Hope, where access to benefits was tied to a demonstrated work effort. Studies of low-income and welfare populations have documented families with multiple barriers to employment, including mental health problems, substance abuse, and domestic violence (Olson and Pavetti, 1996; Pavetti and Acs, 2001; Danziger et al., in press; Meyer and Cancian, 1999). Baseline measures indicate that non-participants may have been more disadvantaged; only one in five likely non-participants were working more than thirty hours a week and nearly 90 percent were receiving some kind of aid. It is likely that some families simply could not consistently maintain the necessary employment.

As mentioned above, the validity of the propensity score method relies on the ignorable treatment assumption, which assumes that assignment to treatment groups is based only on observed characteristics (Dehejia and Wahba, 1999; Rosenbaum and Rubin, 1984). If there were program impacts on the experimental non-participants, then this may be an indication that there are important characteristics that were not observed at baseline that were related to program assignment status. Although there were no impacts on employment, parenting, use of child care, or school-based behavior, New Hope did significantly improve the emotional well-being of experimental non-participants. Does this indicate that the ignorable treatment assumption is violated?

Not necessarily. Both of the program impacts – decreases in stress and feelings of hope – could have been mediated through the presence of the New Hope caseworkers. In analyses not shown, experimental non-participants reported receiving significantly more emotional support from case workers than did control members in their sub-group. Although the measure did not refer to New Hope staff in particular, ethnographic evidence indicates that New Hope case workers were extremely supportive and caring (Weisner et al., 1999). Anecdotal evidence indicates that most case workers tried to contact participants on a regular basis, even if the participants had no formal ties to the program. Research has shown that that social support is an important moderator of the detrimental consequences of poverty (Seccombe, 2000; Magnuson and Duncan, in press; McLoyd, 1997), indicating that the dedicated nature of the New Hope staff may have even benefited participants who had no formal involvement.

Nevertheless, all of the participatory take-up sub-groups showed a positive, significant experimental versus control difference on at least one of the four dimensions examined: employment, parental emotional well-being, child care use, and child academic behavior. The

[15] Another possible explanation is that not everyone who was out of work qualified for a CSJ. Participants were eligible for a community service job if they a) had looked unsuccessfully for a job for eight weeks; b) had lost a regular job and had sought employment unsuccessfully for three weeks; or c) needed additional employment in order to meet the full-time work requirement. Although difficult to verify empirically, it could be that non-participants were less likely to qualify for a CSJ. I do not have information on other problems that may have been prevalent in participants' lives, such as domestic abuse, substance abuse, or legal problems, so it is almost impossible to know if these were more common in the lives of non-participants as opposed to those of CSJ users.

degree of these impacts varied widely across service use sub-groups, reflecting the diverse population that comprised New Hope's experimental group. Thus, policy makers should be aware of demographic heterogeneity in low-income populations, and the importance this heterogeneity has in relationship to take-up of benefits. Differentiation in program use led to differentiation in program outcomes, indicating the important pathways by which New Hope affected the lives of participants.

Bibliography

Agodini, R. and M. Dynarski (2001). Are experiments the only option? A look at dropout prevention programs. Twenty-Third Annual APPAM Research Conference, Washington, DC.

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

Bergman, L.R. (1998). "A pattern-oriented approach to studying individual development" in Methods and models for studying the individual. (Eds) R.B. Cairns, L.R. Bergman and J. Kagan. (pp. 83-121). Thousand Oaks, CA: Sage Publications.

Bergman, L.R. and B.M. El-Khouri (1998). Sleipner: A statistical package for pattern oriented analyses (version 2). Stockholm: Department of Psychology, Stockholm University.

Bergman, L.R. and D. Magnusson (1997). "A person-oriented approach in research on developmental psychology." Development and Psychopathology 9, 291-320.

Blank, R.M. (1997). It takes a nation: A new agenda for fighting poverty. New York: Russell Sage.

Bloom, H. (1984). "Accounting for no-shows in experimental evaluation designs." Evaluation Review 8, 225-246.

Bloom, H. and C. Michalopoulos (2001). "How welfare and work policies affect employment and income: A synthesis of research." Next Generation working paper. New York: Manpower Demonstration Research Corporation .

Bos, J., A.C. Huston, R.C. Granger, G.J. Duncan, T.W. Brock and V.C. McLoyd (1999). New Hope for people with low incomes. New York: Manpower Demonstration Research Corporation.

Brayfield, A. and S.L. Hofferth (1995). "Balancing the family budget: Differences in child care expenditures by race/ethnicity, economic status, and family structure." Social Science Quarterly 76, 158-177.

Brock, T., F. Doolittle, V. Fellerath and M. Wiseman (1997). Implementation of a program to reduce poverty and reform welfare. New York: Manpower Demonstration Research Corporation.

Bumpass, L. and H.H. Lu (2000). "Trends in cohabitation and implications for children's family contexts in the United States." Population Studies 54, 29-41.

Chase-Lansdale, P.L. and L. Pittman (in press). "Parenting and the home environment." Future of Children 11.

Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.

Currie, J. (2000). "Early childhood intervention programs: What do we know?" National Bureau of Economic Research working paper.

Currie, J. and D. Thomas (1995). "Does Head Start make a difference?" The American Economic Review 85, 341-364.

Currie, J. and A.S. Yelowitz (1999) "Health insurance and less skilled workers." Joint Center for Poverty Research Working Paper No. 63, Evanston, IL: Northwestern University.

Danziger, S., M. Corcoran, S. Danziger, C. Heflin, A. Kalil, J. Levine, D. Rosen, K.S. Seefeldt, K. Siefert and R. Tolman (in press). "Barriers to the employment of welfare recipients" in Prosperity for all? The economic boom and African Americans. (Eds) R. Cherry and W.M. Rodgers. New York: Russell Sage Foundation.

Dehejia, R.H. and S. Wahba (1998). "Propensity score matching methods for non-experimental causal studies." National Bureau of Economic Research Working Paper No. 6829.

Dehejia, R.H. and S. Wahba (1999). "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs." Journal of the American Statistical Association 94, 1053-1062.

Duncan, G.J., J. Brooks-Gunn, W.J. Yeung and J. Smith (1998). "How much does childhood poverty affect the life chances of children?" American Sociological Review 63, 406-423.

Duncan, G.J. and P.L. Chase-Lansdale (2001) "Welfare reform and child well-being." ." Joint Center for Poverty Research Working Paper No. 217, Evanston, IL: Northwestern University.

Elder, G.H.J., J. Eccles, M. Ardelt and S. Lord (1995). "Inner-city parents under economic pressure: Perspectives on the strategies of parenting." Journal of Marriage and the Family 57, 771-784.

Gamoran, A., R. Mare and L. Bethke (1999). "Effects of nonmaternal child care on inequality in cognitive skills." Institute for Research on Poverty Discussion Paper No. 1186-99. Madison, WI: University of Wisconsin-Madison.

Gibson, C. and T. Weisner (in press). "'Rational'" and ecocultural circumstances of program take-up among low-income working parents." Human Organization

Gresham, F.M. and S.N. Elliott (1990). Social skills rating system manual. Circle Pines, MN: American Guidance Service, Inc.

Guo, G. and K.M. Harris (2000). "The mechanisms mediating the effects of poverty on children's intellectual development." Demography 37, 431-447.

Hamilton, G, Freedman, S. and S. McGroder (2000). National evaluation of welfare-to-work strategies: Do mandatory welfare-to-work programs affect the well-being of children? A synthesis of child research conducted as part of the national evaluation of welfare-to-work strategies. Washington, D.C.: U.S. Department of Health and Human Services

Harris, K.M. (1996). "Life after welfare: Women, work, and repeat dependency." American Sociological Review 61, 407-426.

Harvey, E. (1999). "Short-term and long-term effects of early parental employment on children of the National Longitudinal Survey of Youth." Developmental Psychology 35, 445-459.

Haveman, R. and B. Wolfe (1995). "The determinants of children's attainment: A review of methods and findings." Journal of Economic Literature 33, 1829-1878.

Heckman, J. and R. Robb (1985). "Alternative methods for evaluating the impact of interventions" in Longitudinal analysis of labor market data. (Eds) J. Heckman and B. Singer (pp. 156-246). Cambridge, MA: Harvard University Press.

Hofferth, S.L., J. West, R.W. Henke and P. Kaufman (1994). Access to early childhood programs for children at risk Washington, D.C.: National Center for Education Statistics.

Huston, A.C., J. Bos, R. Granger, G.J. Duncan, T. Brock, V.C. McLoyd, D. Crosby, C. Gibson, V. Fellarath, K. Magnuson, R. Mistry, S. Poglinco, J. Romich and A. Ventura (2001). "Work-based anti-poverty programs for parents can enhance the school performance and social behavior of children." Child Development 72, 318-336.

Kalil, A., M. Corcoran, S. Danziger, R. Tolman, K.S. Seefeldt, D. Rosen and Y. Nam (1998) Getting jobs, keeping jobs, and earning a living wage: Can welfare reform work? Institute for Research on Poverty Discussion Paper, No. 1170-98. Madison, WI, University of Wisconsin-Madison.

Knox, V., Miller, C., and L. Gennetian. 2000. Reforming welfare and rewarding work: A summary of the final report on the Minnesota family investment program. New York: Manpower Demonstration Research Corporation.

Katz, L., J. Kling and J. Liebman (2001). "Moving to opportunity in Boston: Early results of a randomized mobility experiment." Quarterly Journal of Economics 116, 607-654.

Lalonde, R. (1986). "Evaluating the econometric evaluations of training programs." American Economic Review 76, 604-620.

Magnuson, K. (1999). "Appendix k - the barrier indicator index" in New Hope for people with low incomes: Two-year results of a program to reduce poverty and reform welfare. J. Bos, A.C. Huston, R. Granger, G.J. Duncan, T. Brock and V.C. McLoyd (pp. 346-355). New York: Manpower Demonstration Research Corporation.

Magnuson, K. and G.J. Duncan. (in press). "Parents in poverty" in Handbook of parenting, 2nd edition. (Ed) M. Bornstein. Mahwah, N.J.: Lawrence Erlbaum Associates.

Magnusson, D. (1998). "The logic and implications of a person oriented approach" in Methods and models for studying the individual. (Eds) R.B. Cairns, L.R. Bergman and J. Kagan (pp. 83-121). Thousand Oaks, CA: Sage Publications.

McLoyd, V.C. (1990). "The impact of economic hardship on black families and children: Psychological distress, parenting, and socioemotional development." Child Development 61, 311-346.

McLoyd, V.C. (1997). "The impact of poverty and low socioeconomic status on the socioemotional functioning of African-American children and adolescents: Mediating effects" in Social and emotional adjustment and family relations in ethnic minority families. (Eds) R.D. Taylor and M.C. Wang (pp. 7-34). Mahwah, N.J.: Lawrence Erlbaum Associates.

McLoyd, V.C., T. Jayaratne, R. Ceballo and J. Borquez (1994). "Unemployment and work interruption among African American single mothers: Effects on parenting and adolescent socioemotional functioning." Child Development 65, 562-589.

Meyer, D. and M. Cancian (1999). "Economic well-being following an exit from aid to families with dependent children." Journal of Marriage and the Family 60, 479-492.

Michalopoulos, C. and P. Morris (2000). The Self-Sufficiency Project at 36 months: Effects on children of a program that increased parental employment and income. Ottawa, Canada: Social Research and Demonstration Corporation.

Milligan, G.W. (1996). "Clustering validation: Results and implications for applied analysis" in Clustering and classification. (Eds) P. Arabie, L.J. Habert and G.D. Soete (pp. 341-375). River Edge, NJ: World Scientific.

Morris, P., A.C. Huston, G.J. Duncan, D. Crosby and J. Bos (2001) “How welfare and work policies affect children: A synthesis of research.” Next Generation Working Paper. New York: Manpower Demonstration Research Corporation .

NICHD [National Institute of Child Health and Human Development] Early Child Care Research Network. (1997). "The effects of infant child care on infant-mother attachment security; results of the NICHD study of early child care." Child Development 68, 860-879.

Olson, K.K. and L. Pavetti (1996). “Personal and family challenges to the successful transition from welfare to work: Final report.” Working Paper. Washington, D.C: Urban Institute.

Parcel, T. and E. Menaghan (1994). "Early parental work, family social capital, and early childhood outcomes." American Journal of Sociology 99, 972-1009.

Parcel, T. and E. Menaghan (1995). Parent's jobs and children's lives. New York: Aldine De Gruyter.

Parker, S., S. Greer and B. Zuckerman (1988). "Double jeopardy: The impact of poverty on early child development." Children at Risk 35, 1227-1240.

Pavetti, L. and G. Acs (2001). "Moving up, moving out, or going nowhere? A study of the employment patterns of young women and the implications for welfare mothers." Journal of Policy Analysis and Management 20, 721-736.

Pearlin, L.I., M.A. Lieberman, E. Menaghan and J.T. Mullan (1981). "The stress process." Health & Social Behavior 22, 337-256.

Perry-Jenkins, M., R.L. Repetti and A.C. Crouter (2000). "Work and family in the 1990s." Journal of Marriage and the Family 62, 981-998.

Posner, J.K. and D.L. Vandell (1999). "After-school activities and the development of low-income urban children: A longitudinal study." Developmental Psychology 35, 1-12.

Quint, J., J. Bos and D. Polit (1997). New Chance: Final report on a comprehensive program for disadvantaged young mothers and their children New York: Manpower Demonstration Research Corporation.

Rosenbaum, P.R. and D.B. Rubin (1984). "Reducing bias in observational studies using subsclassification on the propensity score." Journal of the American Statistical Association 70, 516-524.

Rosenbaum, P.R. and D.B. Rubin (1985). "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score." Journal of the American Statistician 39, 33-38.

Rosenberg, M. (1979). Conceiving the self New York: Basic Books.

Royalty, A.B. (2001). "Do minimum wage increases lower the probability that low-skilled workers will receive fringe benefits?" Joint Center for Poverty Research Working Paper No. 222. Evanston, IL: Northwestern University.

Rubin, D. (1979). "Using multivariate matched sampling and regression adjustment to control bias in observational studies." Journal of the American Statistical Association 74, 318-328.

Seccombe, K. (2000). "Families in poverty in the 1990s: Trends, causes, consequences, and lessons learned." Journal of Marriage and the Family 62, 1094-1113.

Shadish, W., T. Cook and D.T. Campbell (2001). Experimental and quasi-experimental designs for generalized causal inference New York: Houghton-Mifflin.

Sherman, A. 2001. "How children fare in welfare experiments appears to hinge on income". Child Defense Fund Working Paper Washington, D.C.: Child Defense Fund.

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

Snyder, C.R., S.C. Sympson, F.C. Ybasco and T.F. Borders (1996). "Development and validation of the State Hope Scale." Journal of Personality and Social Psychology 72, 321-335.

Statistics Canada (1995). Self-sufficiency project: Self-complete questionnaire, parents Montreal: Statistics Canada.

Vandell, D.L. and B. Wolfe (2000). Child care quality: Does it matter and does it need to be improved? Washington, D.C.: US Department of Health and Human Services.

Weisner, T., L. Bernheimer, C. Gibson, E. Howard, K. Magnuson, J. Romich and E. Lieber (1999). "From the living rooms and daily routines of the economically poor: An ethnographic study of the new hope effects on families and children" Biannual meeting of the Society for Research in Child Development, Albuquerque, NM.

Yoshikawa, H., E.A. Rosman and J. Hsueh (2001). "Variation in teenage mothers' experiences of child care and other components of welfare reform: Selection factors and developmental consequences." Child Development 72, 299-317.