Institute for Research on Poverty
Discussion Paper no. 1388-10

Have Welfare-to-Work Programs Improved Over Time In Putting Welfare Recipients To Work?

David H. Greenberg
University of Maryland-Baltimore County
E-mail: dhgreenberg@umbc.edu

Philip K. Robins
University of Miami
E-mail: probins@miami.edu

November 2010

The authors are grateful to Andreas Cebulla for contributing to the early stages of the research through discussions and helping to construct the database used in the study and to Gayle Hamilton for providing information about the Employment Retention and Advancement experiments. Helpful comments were also received by participants in a seminar at the University of Miami. Original funding to construct the database was provided by the British ESRC, the British Academy, and the Rockefeller Foundation. Later funding to refine the database was provided by the Administration for Children and Families at the U.S. Department of Health and Human Services.

Abstract

Data from 76 experimental welfare-to-work programs conducted in the United States between 1983 and 1998 are used to investigate whether the impacts of such programs on employment had been improving over time and whether specific program features influencing such changes can be identified. Over the period, an increasing percentage of control group members received services similar to those offered to program group members. As a result, differential participation in program service activities between program and control group members decreased steadily over time. This reduction in the net receipt of program services tended to reduce the impact of these programs on employment. However, the negative influence of the reduced incremental services was offset by other factors that resulted in program impacts remaining essentially constant from 1983 to 1998. Suggestions are made for possibly improving program impacts in future experiments.

JEL Classification: J21, I38

Keywords: welfare-to-work, program evaluation, AFDC, TANF

Have Welfare-to-Work Programs Improved Over Time In Putting Welfare Recipients To Work?

I. INTRODUCTION

Beginning in the late 1960s, welfare agencies in the United States started to introduce "welfare-to-work" programs with, as the name suggests, the objective of getting as many welfare recipients as possible into employment. Although the designs of these programs have varied across the states and over time, they typically incorporate such features as assessment of basic skills, structured job search, and training and education. They also sometimes provide subsidized jobs in the private or public sector and, more recently, financial incentives to work (earnings disregards and supplementary payments for achieving certain employment goals). Most have also been mandatory—that is, welfare recipients who do not cooperate could have their grants reduced or, in some cases, terminated.

Welfare-to-work programs have played an especially important role in the Aid to Families with Dependent Children (AFDC) program (now called Temporary Assistance to Needy Families or TANF), which is the major cash public assistance program for families with children in the United States. Over time, increasing funding has been channeled to welfare-to-work programs for AFDC recipients and increasing pressure has been put on states to have AFDC recipients partake of the services provided by the programs.[1] Consequently, the mandatory feature of these programs has been increasingly enforced. Today, mandatory welfare-to-work programs for TANF recipients are found throughout the United States.

---

[1]Mandatory programs for AFDC recipients were first established in 1967 under the Work Incentive (WIN) Program, but WIN never received sufficient funding to establish an effective mandate for more than a small minority of AFDC recipients. The 1981 Omnibus Budget Reconciliation Act (OBRA) provided states with considerable flexibility in designing welfare-to-work programs, resulting in considerable increases in enrolment in these programs. The 1988 Family Support Act replaced the WIN program with the Job Opportunities and Basic Skills Training (JOBS) Program, established minimum participation rate targets for state welfare-to-work programs, required participation by mothers with children as young as three (and, at state option, as young as one), increased the sanction for nonparticipation, and, for the first time, committed federal funds to education in welfare-to-work programs. Finally, among other things, the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) of 1996, which replaced AFDC with TANF, required states to meet a specified schedule of minimum work participation rates. In addition to unsubsidized jobs, this requirement could be met by job search, job training and vocational education, and subsidized jobs. Thus, state use of welfare-to-work programs was encouraged. PRWORA also established lifetime time limits on how long AFDC payments could be received.

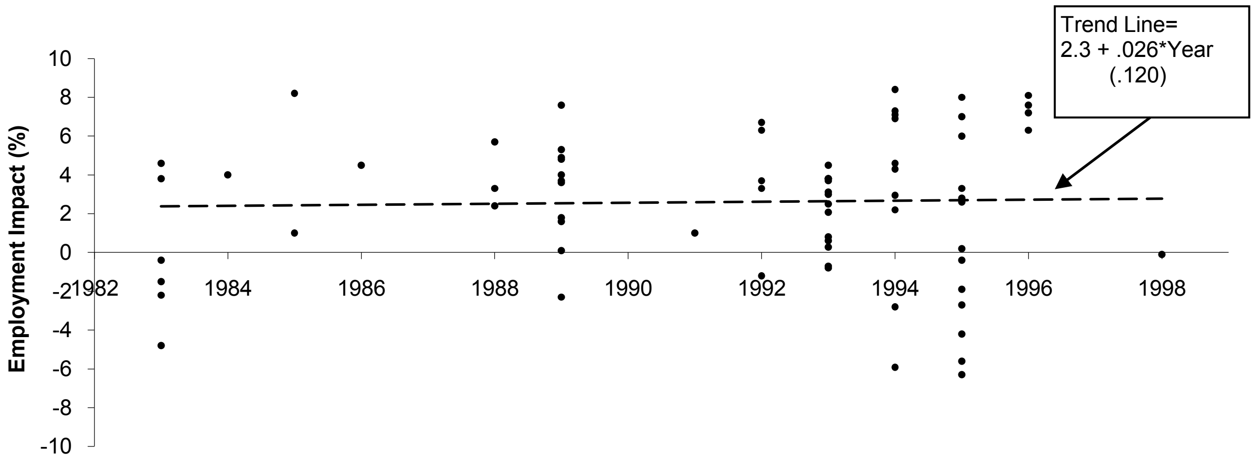
Given the increasing emphasis on mandatory welfare-to-work programs in the AFDC program, it seems reasonable to anticipate that they are effective in meeting their goal of putting AFDC recipients to work and that they have become progressively more effective over time. Some light on whether this is actually the case is shed by a number of random assignment experimental tests of welfare-to-work programs for AFDC recipients that were initiated in various states between 1983 and 1998.[2] Averaging 76 estimates from these experiments, the rate of employment in the seventh quarter after random assignment is found to have increased by 2.6 percentage points among those assigned to the program group, from 37 percent to nearly 40 percent. While this increase is not large, it is not trivial either. However, as shown by Figure 1, program impacts on employment essentially remained constant over time, changing by a statistically insignificant .026 percentage points each year, on average.[3]

There are several potential explanations for this lack of discernable improvement in program impacts on employment over time. The first and most obvious is that not much was learned over time about how to run welfare-to-work programs more effectively. In other words, there is a steep learning curve. The second is that welfare-to-work programs did improve over time, but earlier programs tended to be implemented at sites or among population groups where success came relatively easily, while later programs were run in sites or among population groups where success was more difficult. For example, labor markets may have been tighter at the early sites or the welfare populations at these sites may have been more job-ready. A third possible explanation reflects the growth in the use of welfare-to-work programs over time. It seems likely that earlier programs were introduced in environments where

[2]These experiments are described in some detail in the following section. For purposes of this study, it would have been helpful if the welfare-to-work experiments covered a longer time span. However, random assignment welfare-to-work experiments of mandatory programs did not begin until after OBRA passed in 1981. At this point, the federal government usually made random assignment experiments a condition for providing the states with the waivers they needed to modify their welfare-to-work programs. It is these changes that were tested experimentally. Once PRWORA passed in 1996, states no longer required these waivers to change their welfare-to-work programs, and welfare-to-work experiments became rare.

[3]In terms of the program evaluation literature, these impacts represent impacts of the "intent to treat." Impacts of the "treatment on the treated," which are often studied in the evaluation literature, are not identified from these experiments because the treatments were mandatory and, hence, impacts on the non-treated (nonparticipants) cannot be assumed to be zero. Impacts of the "treatment on the treated" can be identified when studying voluntary programs for which it can be safely assumed that impacts on nonparticipants are zero. For a discussion of the differences between voluntary and mandatory programs, see Friedlander, Greenberg, and Robins (1997).

Figure 1
Employment Impacts Over Time,
Estimated Seven Quarters After Random Assignment



employment-orientated services for AFDC services were not readily available, while later programs tended to replace already existing programs that also offered employment-orientated services, even though later programs may have provided more services or a different service configuration. If later programs produced less of an increment in the receipt of employment-orientated services than earlier programs, it would not be surprising if program effects on employment failed to grow. In fact, given these circumstances, it would be somewhat surprising if their effects did not shrink.

Some evidence supporting the third possibility is given in Figure 2. This figure shows how net overall program participation changed over time, where "overall program participation" refers to partaking in at least one program activity (e.g., job search assistance, basic education, vocational education, or work experience).[4] Net participation is defined as the participation rate of program group members minus the participation rate of control group members.[5]

As Figure 2 shows, overall net participation declined markedly over time (the 2.8 percentage point decline per year is statistically significant at the one percent level), implying that later welfare-to-work programs resulted in smaller increments in the treatment given to the program group than earlier programs. Thus, the fact that the impacts are roughly constant over time (from Figure 1) seems to imply that each percentage point increase in net overall program participation produced a greater increase in the impact in employment among later programs than among earlier programs. It is important, therefore, to account for changing net participation rates when examining trends in program impacts.

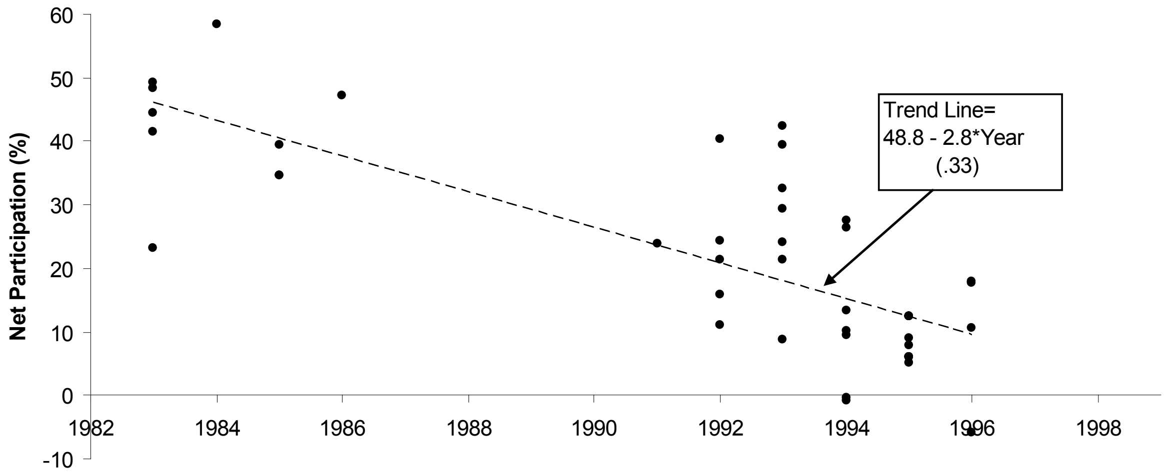
This paper uses the previously mentioned data on random assignment experimental tests of welfare-to-work programs for AFDC recipients to investigate the three possibilities suggested above. The following section describes these data. The third section discusses the methods we use in our investigation. Findings concerning why program impacts on employment have not grown are presented in the fourth section. The final section gives our conclusions.

---

[4]The activities may be same or different for program and control group members. Later in the paper, we define net participation rates separately by type of activity.

[5]There are fewer data points in Figure 2 than in Figure 1 because net overall participation rates are not available for all the programs for which employment impacts were estimated.

Figure 2
Net Participation Over Time,
Estimated Seven Quarters After Random Assignment



2. DATA

This study relies on data from 21 random assignment evaluations of mandatory U.S. welfare-to-work programs for AFDC recipients conducted in various localities (often counties) between 1983 and 1998. These evaluations are listed in Appendix Table A. Although welfare-to-work programs similar to those evaluated continue to be widespread, random assignment evaluations of them became rare after 1996.[6] Thus, these evaluations provide, perhaps, the best available information from which to learn about welfare-to-work programs for the AFDC population. While the potential for generalizing the results of any given individual study is limited, the variation in program content and population characteristics arising from combining studies into an integrative review enormously increases the scope for generalization (Hall et al., 1994).

To ensure comparability across the evaluations, inclusion criteria were established relating both to the kind of program being evaluated and the evaluation strategy. First, all the evaluated welfare-to-work programs had to include an active intervention (e.g., job search, work experience, remedial education, or training) that was intended to assist welfare recipients in increasing their employment. Second, all the programs were mandatory in the sense that recipients who did not participate in job search, vocational training, remedial education, or work experience as required were potentially liable for sanctions through the reduction or removal of their welfare benefit. Third, all the programs were directed at persons receiving AFDC or TANF benefits. Thus, welfare-to-work programs aimed at food stamp, disability, and unemployment compensation beneficiaries or at transfer recipients outside the United States were excluded.

[6] As previously discussed, between 1982 and 1996, the U.S. Department of Health and Human Services (DHHS) usually required random assignment evaluations of state changes in their AFDC program as a condition for receiving waivers permitting the changes. We have checked with DHHS and verified that we located all the random assignment evaluations of mandatory welfare-to-work programs that began in the U.S. between 1982 and 1996. However, six random assignment evaluations could not be used because they did not estimate the program effect on employment in the seventh quarter after random assignment (the program effect estimate used in our empirical analysis). In addition, two studies were excluded because of severe problems with their random assignment designs.

Finally, the evaluations were restricted to those that assigned AFDC or TANF recipients to program and control groups on a random basis. Those recipients assigned to the program group were required to participate in the welfare-to-work program being evaluated, while those assigned to the control group were eligible to receive any services that existed prior to the introduction of the program. By comparing outcomes such as employment for the two groups, program effects (often called "impacts") can be measured. Not only is randomized assignment considered by many to be the model or "gold standard" of evaluation research by providing unbiased estimates of program effects, this restriction effectively standardized methodological procedures. Moreover, as indicated by Appendix Table A, all but one of the 21 evaluations were conducted by just three research organizations. Each of these three organizations has over three decades of experience in implementing and monitoring random assignment procedures and each has a strong reputation for performing random assignment evaluations efficiently and effectively.

All the evaluations that met the criterion listed above were included in the study sample. The 21 evaluations provide information about 76 welfare-to-work programs that operated in 45 sites (i.e., separate counties or metropolitan areas). The multi-site evaluations assessed programs that varied across the sites to a greater or lesser degree. One reason the number of programs exceeds the number of sites is because two experimental programs were run simultaneously in some sites so that outcomes for participants in each program could be compared to one another, as well as to a control group. In addition, some of the evaluations conducted separate analyses of one- and two-parent families. Because programs in which these two family-types were enrolled often differed from those for one-parent families in some of their features and were evaluated separately, we treat them as distinct programs.

For each of the 76 programs included, our database contains estimated program impacts on employment in each available quarter after random assignment, as well as the levels of statistical significance for each of these impact estimates. In this study, we only use the impacts estimated for the seventh quarter after random assignment.

The database also contains a number of explanatory variables. These include the year during which the mid-point of random assignment occurred and measures of rates of participation in various program activities (job search, basic education, vocational training, and work experience), rates of sanctioning, the characteristics of the program population (gender, age distribution, family structure, employment prior to random assignment, and so forth), and socioeconomic information for each of the program sites and for each of the evaluation years (e.g., the site unemployment and poverty rates, the percentage of the workforce engaged in manufacturing employment, the annual rate of change in manufacturing employment, and so forth). Although most of the study data were extracted directly from the reports on each of the 21 evaluations, the site social-economic information was obtained from various government sources, such as U.S. Census Bureau and the U.S. Bureau of Labor Statistics Web sites. Because members of the control group often had access to services similar to those received by the program group and were also subject to sanctions, separate estimates of rates of participation in the sorts of services provided by the program and of sanction receipt for both the program and control groups are available in our database, as are estimates of *net* program effects on these rates, which are measured as differences between the program and control group rates.

3. A MODEL EXPLAINING VARIATION IN WELFARE-TO-WORK PROGRAM EFFECTS

The data in Figures 1 and 2, though suggestive, do not represent a formal analysis of how program effects vary over time. The program effects in Figures 1 and 2 are based on estimates from evaluations with varying sample sizes and this variation in sample size needs to be taken into account in examining variation in program effects over time. Furthermore, other features of the evaluations vary across studies, including types of services offered, characteristics of the tested samples, and the economic environment in which the evaluations took place.

A procedure well suited to examining the problem at hand is meta-analysis. In the meta-analysis literature, two types of statistical models have been commonly used. These models are termed “fixed effects" and "random effects" models, although as will be discussed below, the latter is really a generalization of the former and is more appropriately termed a "mixed effects" model.

Both the fixed effects and random effects models take into account the fact that the individual underlying estimated employment effects are based on different sample sizes, and hence have different levels of statistical precision. It would not make sense to weigh two studies equally that produce estimates having very different levels of statistical precision. For example, suppose one study produced an estimated employment effect of a particular training program of 10 percent, but this estimate was very imprecise and not statistically significant because of a small sample size of, say, only 500 persons. Suppose another study produced an estimated effect of 2 percent for the same program, but was very precisely estimated because of a much larger sample of, say, 4,500 persons. If we did not take into account the sample sizes, we might conclude that the effect of the program was 6 percent (the unweighted mean of the estimates produced by the two studies). However, the true effect is probably closer to 2 percent because of the total sample used in the two studies (5,000), 90 percent was from the latter study.

In order to account for sampling variation across studies with varying sample sizes, the following statistical model is specified:[7]

(1) $T_i = T_i^* + e_i$, where

T_i is the estimated welfare-to-work program effect, T_i^* is the "true" program effect (obtained if the entire target population was evaluated), and e_i is the error due to estimation on a sample smaller than the population. It is assumed that e_i has a mean of zero and a variance of v_i.

In order to provide an estimate of the mean effect that takes into account the fact that v_i varies across studies (that is, v_i is smaller for studies with larger samples), a weighted mean can be calculated, with the weight being the inverse of the v_i, $1/v_i$. If sampling variation were the only source of variation in the training program effects, weighting in this manner produces the most precise estimate of the mean program effect.

[7]Much of the remainder of this section is drawn from Raudenbush (1994).

Using the estimated variances from each study produces a weighted mean employment impact estimate of 2.8 percentage points (compared to an unweighted mean impact estimate of 2.6 percentage points). Thus, the weighted mean is close to the unweighted mean. Both the unweighted and weighted mean impact estimates are statistically different from zero at the 1-percent level.

Sampling variation is not the only source of variation in estimates across studies, however. There are two other sources of variation that are taken into account in meta-analysis. One source has to do with the fact that the estimates are produced for different programs, over different time periods, for different population groups, in different locations, and so forth. The other source arises because there are unmeasured factors that cause variation in program effects. These could be related to staff attitudes toward welfare recipients and other features of the welfare-to-work program or environment that were not measured in conducting the program evaluations.

Each of these sources of variation may be identified by extending the model described by equation (1) in the following way:

$$(2) \qquad T_i^* = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots\ldots \beta_p X_{pi} + u_i,$$

where β_0 is the model intercept, the X_is are observed characteristics of the studies that cause variation in the true program effects T_i^*, the βs are coefficients representing the marginal effects of the characteristics on the true program effect, and u_i is a random error term with variance σ^2, representing unmeasured factors causing variation in program effects. Equation (2) is sometimes termed a "structural" model in the meta-analysis literature.

Together, equations (1) and (2) constitute a statistical model of the variation in program effects. Substituting equation (2) into equation (1) yields the mixed effect model we estimate:

$$(3) \qquad T_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots\ldots \beta_p X_{pi} + e_i + u_i.$$

In equation (3), there are three potential sources of variation in T_i – sampling error (the e_i), observed characteristics of the studies (the X_is), and random error (the u_i). If the βs are not zero, but u_i is identically zero for all studies, then the model is referred to as a "fixed effects" model. In the fixed effects model, there are two sources of variation in the estimated program effects—sampling error and variation in observed characteristics. The weight used in estimating the fixed effects model is the inverse of the sampling variance ($1/v_i$), because the only source of variation in the estimates, other than the X_is, is the sampling variance. If the βs are not zero and u_i, as well as e_i, varies across studies, then the model is referred to as a "mixed effects" model. In the mixed effects model, there are three sources of variation in the estimated program effects—sampling error, variation in observed characteristics, and random error caused by variation in unobserved characteristics. The weight used in estimating the mixed effects model is the inverse of the sum of the sampling error plus the random effects error ($1/[v_i + \sigma^2]$). Clearly, the fixed effects model is a special case of the mixed effects model. It is possible to test statistically for the significance of the fixed and random effects.

To estimate the mixed effects model, an estimate of σ^2 is obviously needed. Raudenbush (1994) describes a variety of procedures for estimating the model, including method of moments estimators and maximum likelihood estimators. One procedure, based on a method of moments estimator, involves the following steps. First, equation (2) is estimated using ordinary least squares (OLS). Then, the mean square residual variance from the regression is used to calculate an estimate of σ^2, based on the following formula:

(4) $s^2 = \text{MSR} - k/(n-p-1)$,

where MSR is the mean square residual from the OLS regression[8] and k is a constant given by the following formula (see Raudenbush, 1994, p. 319):

(5) $k = \Sigma v_i - \text{trace}[\mathbf{X r V X (X r X)}^{-1}]$,

where the boldface refers to matrix notation for the vector of p explanatory variables (the X_i) and the n sampling variances (the v_i), and trace is the sum of the diagonal elements of the resulting matrix. Essentially, the estimate of σ^2 is based on the total residual variance from the OLS regression less an adjustment term based on a weighted average of the sampling errors (v_i) for each observation. After

[8]The MSR is calculated by dividing the residual sum of squares by the number of degrees of freedom in the regression, which is the number of observations (n) minus the number of βs estimated in the model (p+1).

obtaining the estimate of σ^2, the model is re-estimated by weighted least squares, using $1/[s^2 + v_i]$ as weights.

Using the estimated variances (v_i) from each study and the method of moments estimator of σ^2 described by equations (4) and (5) as weights produces a mixed employment impact estimate of 2.84, which is very close to the fixed effect impact estimate. Like the unweighted and fixed effect impact estimates, the mixed effect mean impact estimate is statistically significant at the 1-percent level.

In addition to the fixed and mixed effects models, there is a third model, called the “unweighted model,” in which it is assumed that there is no variation in the v_i across studies. If all studies have the same sample sizes, then the unweighted model is appropriate and can be estimated by a simple ordinary least squares regression of the program effects on the observed characteristics. Of course, if standard errors of the program effects are not available for the studies, the unweighted model must be used. Sometimes, however, the unweighted model is appropriate if there is uncertainty about the accuracy of the estimated standard errors from the underlying studies.

For completeness, we estimated all three models (the unweighted, fixed effects, and mixed models). Using the test suggested by Raudenbush (1992, p. 314), the unweighted and fixed effect models were emphatically rejected in favor of the mixed model for almost every model specification.[9] Accordingly, in the remainder of this paper, we only present results from the mixed model (estimates of the unweighted and fixed effects models are available from the authors on request).

4. FINDINGS

It was suggested in Section 1 that the apparent lack of a positive time trend in the impacts on employment of the welfare-to-work programs in our sample could be due to the earlier programs being located in places or among population groups where success came relatively easier than it did for later programs. This possibility is investigated in the first of the two mixed effect model regressions reported in

[9]The test for the mixed effect model is a test of the hypothesis that sigma-square is zero. The test statistic is given by $Q=\Sigma w_i(T_i - \beta_0 - \beta_1X_1 - \beta_2X_2 -..... \beta_pX_p)^2$, where $w_i = 1/v_i$. This statistic is approximately distributed as chi-square, with n-p-1 degrees of freedom.

Table 1. This regression uses the impact on employment in the seventh quarter after random assignment as the dependent variable and a time trend variable that equals one for programs initiated in 1983 (the year in which random assignment occurred in the earliest of our 76 programs), two for programs started in 1984, and so forth. The remaining variables attempt to control for several site characteristics measured during the seventh quarter after random assignment (the poverty rate, the unemployment rate, and the annual rate of change in manufacturing employment) and several characteristics of the samples used in the program evaluations (the percentage of each sample that worked the year prior to random assignment, the average age of the sample members at the time of random assignment, and whether the evaluated program enrolled one- or two-parent families).

As discussed in Section 1, without these controls, the time trend is positive, although very small and statistically insignificant. If the time trend was understated because the earlier programs in our sample were implemented in sites or among population groups where impacts on employment were likely to be relatively large, the coefficient on the time trend variable should become more positive once site and sample characteristics are controlled for. However, it actually moves in the opposite direction, becoming negative, although remaining very small and statistically insignificant. As indicated in Table 1, none of the site and population characteristics are statistically significant either. Thus, the findings do not appear to suggest that the time trend is understated.

The second regression in Table 1, adds a number of explanatory variables that measure the characteristics of the experimental welfare-to-work programs. Four of these variables are estimates of net (i.e., program group minus control group) participation rates in various program service components (job search, basic education, vocational training, and work experience). A fifth variable measures the net program sanction rate for each program. [10] There were some missing values for the participation and sanction rate variables, generally fewer than 20 percent, which were predicted by regression equations that are described later. The remaining two variables are dummy indicators of whether each program

[10]There were some missing values for the participation and sanction rate variables. These missing values were predicted by regression equations that are described later. The number of missing values varies by variable but is generally under 20 percent.

TABLE 1
Mixed Effect Regression Estimates of Program and Contextual Characteristics on Program Impacts on Employment in the 7th Quarter after Random Assignment

Characteristic	Coefficient	Standard Error	Coefficient		Standard Error
Number of years since 1982	-0.004	0.114	0.054		0.189
Two-parent family target group = 1	-1.004	1.266	-2.150		1.388
Average age of target group	0.078	0.130	0.171		0.182
Percentage of target group with recent employment	-0.045	0.037	-0.013		0.039
Annual percentage change in local manufacturing employment	0.052	0.132	0.035		0.129
Poverty rate (%)	-0.070	0.104	-0.007		0.115
Unemployment rate (%)	-0.195	0.212	-0.182		0.212
Percentage Sanctioned (net)			0.148	*	0.085
Percentage ever participated in job search (net)			0.077	**	0.039
Percentage ever participated in basic education (net)			0.039		0.054
Percentage ever participated in vocational education (net)			-0.187		0.145
Percentage ever participated in work experience (net)			0.035		0.107
Financial incentive tested =1			0.002		1.105
Time Limit tested =1			1.887		1.506
Constant	4.872	5.160	-3.537		6.208

**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

tested financial incentives and time limits.[11] That is, they indicate whether members of the program group were eligible for a financial incentive under certain conditions or whether they were subject to time limits and members of the control group were not.

As seen in Table 1, with the exception of the coefficient on vocational training, all the coefficients on the seven program characteristic variables are positive, implying that greater incremental use of most program components tends to increase the impact of welfare-to-work programs on employment. However, only the coefficients on the net rate of participation in job search and on the net sanction rate are statistically significant at conventional levels.[12] Once program characteristics are held constant, the coefficient on the time trend variable becomes positive and considerably larger in magnitude than in either Figure 1 or in the first regression. One possible interpretation of a positive time trend once program characteristics are held constant is that the administration of welfare-to-work programs improved over time. However, the coefficient on the time trend variable is still not close to being statistically significant.[13]

The positive coefficients in Table 1 on all of the net participation rates except the one for vocational education suggest why the impact of welfare-to-work programs on employment may not have increased over time. If these participation rates declined over time, then they would retard any growth in program impact on employment. Some evidence on this issue is presented in Table 2.

[11]The word "tested" is used to indicate that members of the program group were eligible for financial incentive or subject to time limits and members of the control group were not.

[12]The coefficient on the testing of financial incentives is extremely small, implying that financial incentives have had no effect on the employment impacts of welfare-to-work programs. This may appear surprising because financial incentives are specifically designed to encourage employment. However, the evidence from cost-benefit studies of these welfare-to-work-programs is that financial incentives increase the incomes of welfare recipients by roughly the amount they cost the government (Greenberg, Deitch, and Hamilton, 2009; Greenberg and Cebulla, 2008). The regular AFDC and TANF programs, in contrast, raise the incomes of welfare recipients by considerably less than they cost the government because they create incentives for recipients to work less. Thus, financial incentives appear to be a relatively efficient means of transferring income to the working poor. Moreover, some types of financial incentive programs, such as the Earned Income Tax Credit (EITC) and Canada's Self-Sufficiency Project (SSP), have been found to significantly increase employment of single-parent mothers on welfare (Michalopoulos et al., 2002; Meyer and Rosenbaum, 2001). However, the financial incentives in the EITC and SSP programs are quite different than the financial incentives tested in the U.S. welfare-to-work experiments.

[13]Estimates of the fixed effects model (available from the authors on request) indicate a positive time trend that is somewhat larger than the time trend from the mixed effects model (.115 compared to .054), but it is still not statistically significant.

TABLE 2
Mixed Effect Regression Estimates of the Annual Change in Program Participation and Sanction Rates
(Standard Errors Appear in Parentheses)

	Program Group Rates		Control Group Rates		Net Rates		Number of Programs
Overall Participation	0.023		2.724	***	-2.861	***	49
	(0.445)		(0.455)		(0.367)		
Participation in Job Search	-0.710	*	1.508	***	-2.041	***	65
	(0.437)		(0.287)		(0.431)		
Participation in Basic Education	-1.421	***	0.202		-0.644		65
	(0.492)		(0.220)		(0.399)		
Participation in Vocational Education	-0.341		0.043		-0.255	*	65
	(0.376)		(0.374)		(0.144)		
Participation in Work Experience	-0.386	*	0.362	***	-0.712	***	73
	(0.202)		(0.070)		(0.172)		
Receipt of Sanctions	0.186		0.080		0.093		60
	(0.326)		(0.157)		(0.235)		

***Statistically significant at the 1 percent level.
*Statistically significant at the 10 percent level.

Each of the estimates in Table 2 is a coefficient on the time trend variable from a different mixed effect model regression in which the dependent variable is a participation rate or a sanction rate and the explanatory variables are identical to those used in the first regression in Table 1.[14] Each value can be interpreted as an estimate of the change in a participation rate or the sanction rate over one year.

The first column in Table 2 provides estimates of time trends for the program groups in the evaluated welfare-to-work programs. Three of the four coefficients on participation in particular services are statistically significant at the 10-percent level or less. Perhaps surprisingly, these estimates imply that these persons were *less* likely to receive program services if they were enrolled in a later program than if they were enrolled in an earlier program. Although not statistically significant, the positive coefficient for the sanction rate implies that members of the program group were more likely to be sanctioned in later programs. Estimates of time trends for the randomly assigned control groups appear in the second column. Unsurprisingly, the time trends for members of the control group are positive. Moreover, the time trends for overall participation, job search, and work experience are highly significant. Estimates of time trends for the net participation and net sanction rates are shown in the third column. Given the declining participation in program services by the program group and rising participation by the control group, it is not surprising that the net rates of participation in program services are higher for earlier programs than for later programs, and except for basic education, significantly so. Although the estimated coefficient on the net sanction rates is not statistically significant, the positive point estimate suggests that there may have been an upward time trend in net sanction rates. The fourth column in Table 2 shows the number of net participation and sanction rate estimates that are available in our database. As indicated, these numbers are somewhat smaller than the 76 available estimates of program impacts on employment.

Table 3 presents estimates of the change in the participation and sanction rates during the 16 years between the time random assignment of the earliest welfare-to-work program occurred (early 1983) and the time random assignment of the most recent program occurred (late 1998). The values in the first

[14] The full regression results are available from the authors upon request.

TABLE 3
Estimated Changes in Program Participation and Sanction Rates between 1983 and 1998
(Standard Errors Appear in Parentheses)

	Program Group (A)		Control Group (B)		Net (C)		Alternative Net Col (A) - Col (B) (D)		Mixed Effect Mean Net Rates (E)	
Overall Participation	0.36		43.59	***	-45.78	***	-43.23	***	21.29	***
	(7.12)		(7.28)		(5.87)		(10.18)		(1.32)	
Participation in Job Search	-11.36	*	24.14	***	-32.65	***	-35.49	***	19.91	***
	(6.99)		(4.60)		(6.90)		(8.37)		(1.52)	
Participation in Basic Education	-22.74	***	3.24		-10.30		-25.98	***	7.27	***
	(7.87)		(3.52)		(6.38)		(8.62)		(1.42)	
Participation in Vocational Education	-5.46		0.68		-4.09	*	-6.15		2.52	***
	(6.01)		(5.99)		(2.30)		(8.49)		(0.51)	
Participation in Work Experience	-6.17	*	5.79	***	-11.39	***	-11.96	***	3.07	***
	(3.24)		(1.12)		(2.75)		(3.48)		(0.60)	
Receipt of Sanctions	2.98		1.29		1.48		1.69		6.06	***
	(5.22)		(2.51)		(3.75)		(5.79)		(0.86)	

***Statistically significant at the 1 percent level.
*Statistically significant at the 10 percent level.

three columns were computed by simply multiplying the estimates in Table 2 by 16. Column (D) was computed by subtracting Column (B) from Column (A). The resulting values provide an alternative to the estimates appearing in Column (C) of the change between 1983 and 1998 in the net participation and sanction rates. Except for basic education, the estimates of net participation rates are very similar in Columns (C) and (D), suggesting that the estimates in the first two columns are reasonably accurate. Column (E) provides the mean values of the participation and sanction rate variables from the mixed effect model. As a comparison of either Column (C) or Column (D) with Column (E) suggests, the negative changes in net participation rates between 1983 and 1998 were quite large relative to the mean net participation rates. The net sanction rate, in contrast, was relatively stable over time.

How did the rather large negative changes in net participation in welfare-to-work program services affect the success of the evaluated programs in increasing employment? This can be seen by multiplying each of these changes (as shown in Columns (C) and (D) of Table 3) by its corresponding coefficient estimate in Table 1, recalling that each of these coefficients provide an estimate of how a one percentage point change in a net participation rate affects the impact of welfare-to-work programs on employment. The results of these calculations are presented in the first two columns of Table 4. They suggest that the reduction in net participation in program services tended to retard growth in the impact of welfare-to-work programs on employment. Most of this effect results because net participation in job search, which has a substantial and statistically significant positive influence on employment impacts, shrank considerably between 1983 and 1998. Viewed somewhat differently, the results in the first two columns of Table 4 imply that had net participation in program activities not diminished between 1983 and 1998, the impact of welfare-to-work programs on employment would have grown by about two-and-half percentage points—that is, the mean mixed employment impact estimate of 2.8 percentage points, which was mentioned earlier, would have been approximately twice as large.

As previously discussed, the shrinkage in net participation in the services provided by welfare-to-work programs occurred because participation among those not assigned to the programs (the control groups) tended to "catch up" with the program groups. This resulted both because control group members tended over time to have received more of the sorts of services offered by the experimental programs and

TABLE 4
Estimates of the Effect of Changes in Participation Rates on Program Impacts on Employment

	Based on Net Change		Based on Alternative Net Change		Assumes Control Participation Unchanged
Participation in Job Search	-2.50	*	-2.72	*	-0.87
	(1.41)		(1.57)		(0.75)
Participation in Basic Education	-0.40		-1.02		-0.89
	(0.70)		(1.51)		(1.33)
Participation in Vocational	0.76		1.15		1.02
Education	(0.81)		(2.20)		(1.63)
Participation in Work Experience	-0.40		-0.42		-0.22
	(1.25)		(1.33)		(0.75)
	-2.54		-3.01		-0.96
TOTAL	(2.17)		(3.37)		(2.36)

*Statistically significant at the 10 percent level.

because program group members tended over time to have received fewer of the services. What would have happened if receipt of program services by members of the control groups had remained at their 1983 level? This question is addressed by multiplying each of the changes in participation rates for program group members (as shown in Column (A) of Table 3) by its corresponding coefficient estimate in Table 1. The findings from this calculation appear in the third column of Table 4. They imply that, by itself, the reduction in the receipt of program services by members of the program groups had a relatively modest effect of slightly less than one percentage point on program impacts on employment. Thus, the increase in the receipt of these sorts of services by control group members was probably relatively more important, especially their highly statistically significant increase in receiving job search services.

The findings in the first two columns of Table 4 suggest that the reductions in net participation in welfare-to-work program services tended to cause the impact of the programs on employment to fall over time. However, in the introduction to this paper, we observed that there was virtually no time trend in these impacts. Thus, the influence of the negative changes in net participation in welfare-to-work program services must have been offset by other factors.

What were these other factors? The evidence here is far weaker than that for the effects of the net reductions in participation in welfare-to-work program services. However, some hints appear in the second regression reported in Table 1. For example, the positive and marginally statistically significant coefficient on the net sanction rate implies that the rise in this rate between 1983 and 1998, which is shown in Table 3, should have increased the impact of welfare-to-work programs on employment over time. However, if these two values are multiplied by one another, the positive effect on the employment impact is less than one quarter of one percentage point (.219=.148 x 1.48). Although the net sanction rate probably grew over time, it did not grow by very much. A second possibility is suggested by the positive coefficient on the time trend variable in the second regression in Table 1. Earlier, we suggested that this positive sign may indicate that the administration of welfare-to-work programs improved over time. Although very imprecisely estimated, the time trend point estimate implies that the impact on employment could have grown by nearly a percentage point (0.864=16x.054) between 1983 and 1996 for this reason. A third possible factor is suggested by the positive, although statistically insignificant, coefficient in Table 1 on the testing of time limits. Time limits were not tested experimentally until 1994, but over 40 percent of the experimentally evaluated programs that were initiated between 1994 and 1998 (13 of 30) did test this provision.[15] The regression coefficient of 1.887 on time limits in Table 1implies that growth in the use of time limits from zero to around 40 percent would have caused the impact of welfare-to-work programs on employment to increase by about three quarters of a percentage point. Although the analysis of two of the three factors just considered is based on statistically insignificant regression coefficients, together, the three appear to offset much of the negative influence of the reduction over time in net participation in welfare-to-work program services.

5. CONCLUSIONS

In this paper we have used data from 76 experimental welfare-to-work programs conducted in the United States between 1983 and 1998 to investigate whether impacts of such programs on employment have been improving over time and whether specific program features influencing such changes can be identified. Over the period covered by our data, an increasing percentage of control group members were receiving services similar to those offered to program group members. As a result, net participation in program service activities, and hence the "intensity" of the treatment, decreased steadily over time. This reduction in the net receipt of program services tended to reduce the impact of these programs on employment. However, program impacts on employment were essentially constant from 1983 to 1998, although there was a very small upward trend. This may have occurred because the negative influence of the reduced incremental services was offset by an increase in the use of sanctions, the introduction of time limits, and perhaps improved administration of the programs, all of which tended to increase program impacts.

Findings from this study imply that there could have been a considerably larger upward trend in program impacts on employment had more effort been made to engage program group members in the

[15]Although these 13 programs tested provisions that limited the amount of time individuals remained eligible for AFDC benefits, they differed in various respects from one another and from the lifetime time limit included in the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) of 1996.

welfare-to-work experiments in job search. However, although providing this service is relatively inexpensive,[16] participation by program group members appears to have actually fallen over time.

Since 1998, several additional welfare-to-work experiments have been conducted. One of the largest of these is the so-called ERA (Employment Retention and Advancement) experiments that took place in 14 sites throughout the United States (Hendra et al., forthcoming). The results from ERA have only recently become available and they indicate very small impacts on employment in almost every site tested (out of 14 sites, only 3 had statistically significant impacts on employment). One of the reasons for the small impacts may be fairly large receipt of services by control group members and hence, small impacts on net participation rates among program group members. In only 4 out of the 14 sites, for example, was there a statistically significant difference between program group and control group members in the receipt of job search services, and in only 3 out of the 14 sites was there a statistically significant difference between program group and control group members in the receipt of education and training services.[17] Given this relatively small "intensity" of treatment in the ERA experiments, it is perhaps not surprising that the impacts on employment were also small. From this perspective, the findings from ERA are consistent with the results presented in this paper.

Our results have important implications for the design and conduct of future experimental welfare-to-work programs. First, as has been pointed out by others but bears repeating, it is important to closely monitor the behavior of the control group because their experiences in the welfare system can have important implications for the interpretation of estimated program impacts. Second, designers of future experiments should ensure that the features of the programs they are testing represent significant departures from the features being received by control group members. Otherwise, the "intensity" of the treatment may not be large and the experimental program may have relatively little impact on

[16]For evidence, see Table 7 of Greenberg and Cebulla (2008).

[17]The generally small differences between the program and control groups in the receipt of these services were by intention. These were pre-employment services, but as its names implies, the objective of the ERA experiments was to test services that were designed to improve the post-employment experience of the program group such as career assessment, planning, assistance in finding a better job while working, and advice about problems on the job. Although few control group members received these post-employment services, receipt was also low among ERA program group members and differences in receipt between program and control group members were statistically significant in less than half the sites for most of the post-employment service measures.

employment and other outcomes of policy interest. Finally, most of the recent welfare-to-work programs being tested in the United States have been based on models designed mainly by welfare agencies, with limited input from the research organization conducting the evaluation. In cases where the evaluators have had a stronger hand in the design of the experimental treatment, impacts have tended to be larger.[18] In future experiments, evaluators need to play a more significant role in the design of the treatment, to help ensure that, at least in principle, the treatment has the potential to be cost-effective.

[18]For example, a recent welfare-to-work program conducted in Canada, the Self-Sufficiency Project (SSP), was designed jointly by the evaluators and the government agency sponsoring the evaluation, and the "treatment" represented a significant departure from what the control group was receiving. As a result, the experiment yielded sizable employment impacts (see Michalopoulos et al., 2002).

APPENDIX TABLE A
U.S. Welfare-to-Work Evaluations Included in the Database

Program Title	Short Program Name	Evaluator	Mid-Point of Random Assignment
Greater Avenues for Independence Program	GAIN (California)	MDRC	1989
Job Search and Work Experience in Cook County	Cook County	MDRC	1985
Community Work Experience Demonstrations	West Virginia	MDRC	1983
WORK Program	Arkansas	MDRC	1983
Employment Initiatives	Baltimore	MDRC	1983
Saturation Work Initiative Model	SWIM (San Diego)	MDRC	1985
Employment Services Program	Virginia	MDRC	1984
Project Independence (Florida's JOBS Program)	Florida	MDRC	1991
Jobs First	Connecticut	MDRC	1996
The Family Transition Program	FTP (Florida)	MDRC	1994
The Los Angeles Jobs-First GAIN Evaluation	Los Angeles	MDRC	1996
The San Diego Job Search and Work Experience Demonstration	San Diego	MDRC	1983
National Evaluation of Welfare-to-Work Strategies	NEWWS	MDRC	1993
Minnesota Family Investment Program	MFIP	MDRC	1994
Vermont's Welfare Restructuring Project	Vermont	MDRC	1995
Teenage Parent Demonstration	Teenage Parents	Mathematica Policy Research (MPR)	1988
Indiana's Initial Welfare Reform Program	Indiana I	Abt Associates	1995
Indiana's Modified Welfare Reform Program	Indiana II	Abt Associates	1998
To Strengthen Michigan Families	TSMF(Michigan)	Abt Associates	1993
Family Investment Program	FIP (Iowa)	MPR	1994
California Work Pays Demonstration Program	CWPDP	UCLA School of Public Policy and Social Research	1993

References

Friedlander, Daniel, David H. Greenberg, and Philip K. Robins. 1997. "Evaluating Government Training Programs for the Economically Disadvantaged." *Journal of Economic Literature* 35(4, December): 1809–1855.

Greenberg, David, Victoria Deitch, and Gale Hamilton. 2009. *Welfare-to-Work Program Benefits and Costs*. New York: MDRC, February.

Greenberg, David, and Andreas Cebulla. 2008. "The Cost-Effectiveness of Welfare-to-Work Programs: A Meta-Analysis." *Public Budgeting & Finance* Summer 2008: 112–145.

Hall, Judith A., Linda Tickle-Degnen, Robert Rosenthal, and Frederick Mosteller. 1994. "Hypotheses and Problems in Research Synthesis." In *The Handbook of Research Synthesis*, eds. Harris Cooper and Larry V. Hedges. New York: Russell Sage Foundation.

Hendra, Richard, et al. Forthcoming. *The Employment Retention and Advancement Project: Implementation Findings and Final Impact Results for 12 Models in the Employment Retention and Advancement (ERA) Project*, New York: MDRC.

Meyer, Bruce D., and Dan T. Rosenbaum. 2001. "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." *Quarterly Journal of Economics* 116(3, August): 1063–1114.

Michalopoulos, Charles, Doug Tatrie, Cynthia Miller, Philip K. Robins, Pamela Morris, David Gyarmati, Cindy Redcross, Kelly Foley, and Reuben Ford. 2002. *Making Work Pay: Final Report on the Self-Sufficiency Project for Long-Term Welfare Recipients*. Ottawa: Social Research and Demonstration Corporation, July.

Raudenbush, Stephen W. 1994. "Random Effects Models." In *The Handbook of Research Synthesis*, eds. Harris Cooper and Larry V. Hedges. New York: Russell Sage Foundation.