

WHEN WELFARE-TO-WORK PROGRAMS SEEM TO WORK WELL: EXPLAINING WHY RIVERSIDE AND PORTLAND SHINE SO BRIGHTLY

DAVID GREENBERG, KARL ASHWORTH,
ANDREAS CEBULLA, and ROBERT WALKER*

Of welfare-to-work programs evaluated by random assignment, two stand out as having exceptionally large estimated effects: one in Riverside, California, and the other in Portland, Oregon. The authors use data from 24 evaluations and the tools of meta-analysis to examine why. The findings indicate that the apparently superior performance of these two programs in increasing the earnings of participants is only partly attributable to program design (for example, the type of services provided, the use of sanctions, and the quality of program administration). Differences in caseload characteristics and site characteristics are probably more important. However, Riverside and Portland's relatively large effects in reducing the percentage of participants on the AFDC rolls appear mainly attributable to the design of the programs run in these sites.

Over the past several decades, there has been increasing emphasis on getting welfare recipients into work. One way this has been done is through mandatory "welfare-to-work programs" in which welfare recipients are required to participate in job

search, vocational training, and remedial education, or are assigned to "work experience" positions at government or non-profit agencies. Recipients who do not fulfill the mandate to participate in welfare-to-work programs can be sanctioned through reductions in, or elimination of, their welfare grants.

Since the mid-1980s, a number of formal evaluations of welfare-to-work programs have used an experimental design in which welfare recipients are randomly assigned to either a program group or a control group. These two groups are then compared in terms of such post-program outcomes as

*David Greenberg is Professor Emeritus of Economics at the University of Maryland, Baltimore County; Karl Ashworth is a statistician at the U.K. Office for National Statistics; Andreas Cebulla is a Research Director at the National Centre for Social Research, U.K.; and Robert Walker is a Professor in the Department of Sociology and Social Policy at the University of Nottingham and a Research Fellow at the Institute for Fiscal Studies, U.K. The authors are indebted to Larry Mead for helpful comments on earlier drafts of this article. Much of the work on this paper took place at the Centre for Research in Social Policy at Loughborough University, U.K. Grants from the British ESRC (Award L216 25 2042), the British Academy, and the Rockefeller Foundation supported work on this project.

The data used to generate the results presented in this article are available through the first author at the Department of Economics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250; dhgreenberg@umbc.edu.

earnings and the percentage in each group that remain on welfare.

Most random assignment evaluations have indicated that welfare-to-work programs produce the desired results—the earnings of members of the program group are larger than those of controls, while the percentage remaining on welfare is lower—but that these effects are at best “modest” (see Friedlander, Greenberg, and Robins 1997; Ashworth et al. 2004). Two welfare-to-work programs, one run in Riverside, California, and the other in Portland, Oregon, have stood out, however, as producing evaluation results that are markedly larger than the others. Because of these evaluations, the Riverside and Portland programs have been considered models worthy of imitation by welfare agencies throughout the country, as well as by foreign governments. This paper considers the question of whether they should, in fact, be emulated.

Using a database that contains information on 24 random assignment evaluations of welfare-to-work programs and analytic tools from meta-analysis, we examine two competing explanations for the larger effects on earnings and on the receipt of welfare found for Riverside and Portland. The first is that they are due to the design of the programs—for example, the combination of services provided and the administration of sanctions. In other words, the favorable effects result from factors under the control of those running welfare-to-work programs. The second explanation is that the superior performances of the Riverside and Portland welfare-to-work program are attributable to factors that are not under the control of program administrators—for instance, especially favorable socio-economic conditions or the characteristics of the client populations. In addition, the positive findings for Riverside and Portland could, in part, result from sampling error. The goal of the research reported in this article is to sort out the relative contributions of the program designs used in Riverside and Portland and of factors that are not under the control of program administrators.

Data

This study uses a unique database, assembled specifically for synthesizing findings from evaluations of welfare-to-work programs. We use data from all 24 random assignment evaluations of mandatory welfare-to-work programs for persons in receipt of Aid for Families with Dependent Children (AFDC; now called Transitional Aid for Needy Families, TANF), which were implemented between 1982 and 1996 and for which findings were available by the year 2000. The 24 evaluations are listed in Appendix Table A. During this period, AFDC was the major cash public assistance program for families in the United States. The sample population in each of the 24 studies was composed entirely or almost entirely of single parents, over 90% of whom were female. These evaluations provide information about 64 welfare-to-work programs in over 50 sites. The number of programs exceeds the number of sites because two experimental programs were run simultaneously in some sites. All the financial information we use has been inflated to year 2000 dollars, using the U.S. Consumer Price Index (CPI-U).

The 24 evaluations were conducted similarly. AFDC applicants and recipients were first randomly assigned either to a program group, with mandatory participation in the welfare-to-work program being evaluated, or to a control group that could not enter the program being evaluated but could receive any services that existed prior to the introduction of the welfare-to-work program. Relying mainly on administrative data, the evaluators collected various measures of outcomes, including earnings and the percentage receiving AFDC, on the members of the program and control groups over time. The program effect was estimated as the difference in mean outcome between the program group and the control group, a measure that is often referred to as the “program impact.” To maintain the random assignment experimental design and not introduce selection bias, the evaluators included *all* persons who were assigned to the program and control groups

in computing these mean values, regardless of whether they were employed and regardless of whether they were still on the welfare rolls. The fact that all but two of the evaluations were conducted by three organizations and 15 of the 24 by one (MDRC, a New York City-based research firm) helps ensure that they were conducted similarly.

For each of the 24 evaluations, the database contains measures of program impacts on average earnings and the percentage in receipt of AFDC for each available quarter and year after random assignment, as well as the levels of statistical significance for each of these impact measures. The database also contains the values of a number of explanatory variables. These include the characteristics of the program population (gender and ethnic mix, age distribution, family structure, education levels, and so forth), measures of program effects on participation in various activities (job search, basic education, vocational training, and work experience), program effects on sanctioning, and socio-economic data for each of the program sites and for each of the evaluation years (site unemployment and poverty rates, the percentage of the work force in manufacturing employment, median household income, and the maximum AFDC payment for which a family of three was eligible). Although most of the study data were extracted directly from the evaluation reports, the site socio-economic data were obtained from various government sources, such as the U.S. Census Bureau and the U.S. Bureau of Labor Statistics websites. 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, *net* program effects on rates of service and sanction receipt, which were measured as differences between the program and control groups, are used as explanatory variables in the study.

Although the data in the database used in this study are quite "rich," they are not ideal. In particular, program impact measures for most of the evaluated programs are available for two years or less. Conse-

quently, we focus on the third and seventh quarters after random assignment.

Bright Stars: Riverside and Portland

Much of the recent U.S. welfare debate can be characterized as a dispute between advocates of human capital investment and supporters of a "work-first" approach. Under the human capital approach, education (particularly adult basic education) and training are used to upgrade the skills of welfare recipients *before* they seek work, which, it is hoped, will result in stable, well-paying jobs. The work-first approach, in contrast, places emphasis on job search so that participants can find a job as quickly as possible. It is often accompanied by a strong message that employment, even at poor-paying jobs, should be sought expeditiously and, hence, training and education should only be provided as a last resort. Both the Riverside and Portland programs are usually characterized as emphasizing a "work-first approach," as opposed to the "human capital approach" followed by many of the other programs included in our database. However, as discussed below, both the Riverside and Portland programs differed from the strict work-first model in key respects.

Riverside County was one of six counties included as part of the evaluation of GAIN (Greater Avenues for Independence), a job search and job training program for AFDC recipients in California, which was initiated state-wide in 1986. Because California gave its counties considerable latitude in how they implemented GAIN, the programs that developed in the six evaluation counties varied considerably. In particular, Riverside placed much greater emphasis on a work-first approach than did the other five counties, which, to varying degrees, adopted a human capital approach. Thus, GAIN is widely perceived as a test of these two competing approaches, and the strong positive findings for Riverside, which are discussed below, have often been attributed to its approach. However, unlike most work-first programs, Riverside expended considerable resources on training those who did not

Table 1. Program Effects on Earnings and AFDC
Receipt for a “Typical Site,” for Riverside, and for Portland.

Description	Number	“Typical Site”		Effects		Differences from “Typical Site”	
		Effects	Standard Deviation	Riverside	Portland	Riverside	Portland
Earnings (\$)							
3rd Quarter after Random Assignment	53	79.32	80.04	300.05	189.04	220.73	109.72
7th Quarter after Random Assignment	52	101.15	103.33	340.32	337.90	239.17	236.75
AFDC Receipt (%)							
3rd Quarter after Random Assignment	41	1.66	3.57	3.90	6.80	2.24	5.14
7th Quarter after Random Assignment	49	1.95	4.73	5.30	11.60	3.35	9.65

find jobs immediately, as well as on case management and job search for those who did (Riccio, Friedlander, and Freedman 1994).

The Portland program was one of 11 welfare-to-work programs included in the National Evaluation of Welfare-to-Work Strategies (NEWWS), which was conducted in six different states. The Portland program produced by far the largest effects of the 11 NEWWS programs, which varied considerably in their designs. Portland’s welfare-to-work program resembled other work-first programs because it emphasized to participants that the goal for them was to obtain a job; but rather than being told to take a job immediately, participants were encouraged to wait until they could find a “good” job. In addition, as in the human capital approach, clients who were in need of more skills were encouraged to enroll in education or training (Scrivener et al. 1998).

Table 1 demonstrates why the Riverside and Portland programs are considered so exceptional by comparing the effects of these programs with those of a “typical site” during the 3rd and the 7th quarters after random assignment. By “typical site,” we simply mean the average program impact estimates across our entire sample, includ-

ing Riverside and Portland.¹ Because a reduction in the receipt of AFDC is a goal of most welfare-to-work programs, such reductions are treated as a positive program impact.

As shown, program effects on earnings and on the receipt of AFDC are considerably greater in Riverside and Portland than in a “typical site.” With the exception of Riverside’s impact on the receipt of AFDC, the differences between Riverside’s and Portland’s effects and those for a “typical site” substantially exceed the standard deviations of the sample means. It is these relatively large effects that caught the eyes of numerous policy-makers and analysts.

¹A few evaluations reported program impact estimates for either quarter 2 or 4, but not quarter 3, and for either quarter 6 or 8, but not quarter 7. These values were included in the 3rd and 7th quarter averages, respectively. In addition, a few evaluations reported program effects on annual earnings, but not program effects on quarterly earnings. In these instances, the annual earnings impact estimates for the first and the second year after random assignment were divided by four and assigned to quarter 3 and quarter 7, respectively.

Approach

Our approach to decomposing the factors that account for Riverside's and Portland's high relative success involves four steps. (1) Regression equations are computed for the 3rd and 7th calendar quarters after random assignment. The dependent variables in these regressions are the estimates of each program's effects on earnings and the receipt of AFDC. The explanatory variables include measures, for each program, of the effects on the utilization of sanctions and the receipt of services, the characteristics of the participants, and the characteristics of the socio-economic environment. Because separate regressions are computed for two different program impact estimates in two different calendar quarters, four regressions are estimated in all. (2) The value of each explanatory variable in the meta-regression models is obtained for a "typical site" by averaging across all the sites for which 3rd quarter and 7th quarter program impact estimates are available. (3) Each of the mean values obtained in step 2 is subtracted from the corresponding value for Riverside and for Portland to determine how the former differed from the latter in terms of program, participant, and site characteristics. (4) The value of each of the differences obtained from step 3 is multiplied by its corresponding regression coefficient, which was obtained from step 1. Thus, the regressions are used to predict how the effects in Riverside and in Portland differ from the effects in a "typical site." The values obtained from step 4 indicate how each explanatory variable contributed to Riverside's and Portland's exceptionally large effects.

As discussed by Hedges (1994), Shadish and Haddock (1994), and Greenberg, Michalopoulos, and Robins (2003), in running regressions in which estimates of program effects are the dependent variable and thus subject to sampling error, it is standard practice to compute weighted least squares regressions in which the weights are the inverse of the variance of each

estimate.² We follow this convention. The weights used in the regressions on program effects on welfare receipt also include a random error term to account for the effects of variables omitted from the regression. The procedure we followed in deriving the purely random error term is given in Raudenbush (1984).³ We do not

²Typically, the evaluations used in this study did not report the exact value of the variance of the program impact estimates, but instead reported that estimates of program effects were not statistically significant or were significant at the 1%, 5%, or 10% level. Thus, most of the standard errors had to be imputed. For AFDC effects, the imputation of the standard errors was done as follows:

$$\sigma^2 = \sqrt{(P_t(1-P_t)/N_t) + (P_c(1-P_c)/N_c)}$$

where σ^2 is the standard error of the program impact, P_t the proportion receiving AFDC in the treatment group, N_t the number of people in the treatment group, P_c the proportion receiving AFDC in the control group, and N_c the number of people in the control group. For earnings effects, imputation of the standard error was considerably more complex. First, for effects that were significant at the 5% or 10% levels, it was assumed that the p-value was distributed at the midpoint of the possible range, that is, if $0.1 > p > 0.05$, p was assumed to equal 0.075, and if $0.05 > p > 0.01$, p was assumed to equal 0.03. Second, cases for which effects were significant at the 1% level have an unbounded t-value, and cases for which effects were non-significant can have extremely small t-values and large standard errors. Therefore, for these cases we used the following procedure: (1) we multiplied each of the standard errors imputed as described above for effects that were significant at the 5% or 10% levels by the square root of the sample on which the impact estimate was based; (2) we computed the average of the values derived in (1); and (3) for cases in which effects were significant at the 1% level or were non-significant, we imputed the standard error by dividing the constant derived in (2) by the square root of the sample size on which the impact estimate was based.

³Raudenbush (1994) described several methods for obtaining an estimate of the random effects error. One he suggested that provides a good approximation of random effects error involves the following steps. First, an unweighted ordinary least squares (OLS) regression is estimated. Then the mean square residual variance from the regression is used to calculate an estimate of the random effects error, $\hat{\sigma}^2$, based on the formula

$$\hat{\sigma}^2 = \text{MSR} - \bar{v},$$

where MSR is the mean square residual from the OLS regression and \bar{v} is the mean variance of the estimated

include this term in weighting the regressions on program effects on earnings because, based on a test in Raudenbush (1994:314), we were unable to reject the hypothesis that the purely random error term equals zero.⁴ Thus, the earnings regressions only adjust for sampling error. In addition to the weighted regressions, we ran (unreported) unweighted regressions. As it turns out, the regression estimates are not very sensitive to whether the regressions are weighted.

Why Riverside and Portland Shine So Brightly

Regression Results

The regressions for the estimates of program effects on earnings and the receipt of welfare are reported in Tables 2 and 3, respectively. Because our primary interest in this study is to assess the extent to which program design accounts for the exceptionally large program effects found in Riverside and Portland, the explanatory variables include program effects on sanctioning and participation in job search and training activities. Decisions concerning which additional variables to include were largely driven by the fact that only 41 to 53 observations were available to estimate the

regressions, and, as a consequence, multicollinearity was a concern. Thus, we attempt to be parsimonious, selecting explanatory variables when (a) there are strong conceptual reasons to anticipate that they would influence program effects and (b) they actually appear to do so.⁵

In Tables 2 and 3, the mean of the observed value for a variable is assigned to all observations for which the value is missing. However, none of the values for site characteristic variables are missing, and fewer than 5% of the values for any of the program participation and sanction rate variables and no more than 15% of any of the values for the client characteristic variables are missing. Sensitivity tests were carried out using other procedures, but the findings, in general, were little affected by alternative treatments of missing values.

There is a possibility that the measures of program participation rates that are used as explanatory variables in the regressions are endogenously determined. This could occur, for example, if programs that have a client population of individuals who are mostly job-ready (for example, high school graduates with considerable previous work experience) tend to stress job search, while programs with large fractions of clients who are not job-ready tend to emphasize basic education. Similarly, programs that are located at sites with low unemployment rates might tend to emphasize job search, and those with high unemployment rates might make more use of vocational training. Under these circumstances, program characteristic measures would, in part, reflect client and site characteristics, causing

impact in the sample of program impact estimates—that is, $\sum v_i/n$, where v_i is the variance of the sampling error of each of the estimated program effects and n is the number of program impact estimates. The weighted least squares regression is estimated by using $1/[\hat{\sigma}^2 + v_i]$ as weights. Obviously, if σ^2 does not differ significantly from zero, as in the case of the earnings regressions, only the variances of the sampling errors are used as weights.

⁴The test for the random effects model is a test of the hypothesis that sigma square is zero. The test statistic, Q , is given by $Q = \sum (1/v_i)(T_i - \beta_0 - \beta_1 X_1 - \beta_2 X_2 - \dots - \beta_p X_p)^2$, where T_i are estimates of a program impacts, which are taken from evaluation reports of the programs, the X 's are covariates of the program impact estimates, and the β 's are coefficients estimated in a regression in which only the variances of the sampling errors are used as weights. Q is approximately distributed as chi-square, with $n-p-1$ degrees of freedom, where n is the number of program impact estimates and p is the number of covariates in the regression.

⁵We also estimated regressions that include most of the explanatory variables available in our database to see if less parsimonious regressions could better “explain” why program effects in Riverside and Portland differ from those in a typical or average site. Although these more inclusive regressions do a slightly better job of explaining the differences, they do not change our key conclusions. Thus, because the estimates from the more parsimonious regressions are less subject to multicollinearity and, hence, more readily interpreted, we focus on them.

Table 2. Weighted Regression Estimates for Program Effects on Earnings During the 3rd and 7th Quarters after Random Assignment.

Variable	Quarter 3	Quarter 7
Constant	227.18* (122.62)	333.01* (153.61)
Program Effect on % Receiving		
Sanctions	4.46*** (1.03)	5.75*** (1.34)
Job Search	3.12*** (0.95)	1.99* (0.97)
Basic Education	-1.63 (1.11)	-0.48 (1.11)
Vocational Training	-7.26*** (2.50)	-4.42 (2.54)
Work Experience	-2.08 (2.11)	1.24 (2.39)
Program Includes		
Financial Incentives	-45.53 (28.54)	-66.79* (33.36)
Participant Characteristics		
% White	1.63*** (0.56)	1.02 (0.58)
Average Age < 30 = 1	-47.89** (22.31)	-76.27*** (23.42)
Site Characteristics		
Unemployment Rate	3.20 (4.40)	-10.36* (4.91)
Poverty Rate	-37.20** (15.21)	-28.59 (17.82)
Poverty Rate Squared	1.19*** (0.45)	0.82 (0.52)
R ²	.337	.413
Number of Observations	53	52

*Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

estimates of the relation between these measures and program effects to be biased. It should be borne in mind, however, that the regressions control directly for client and site characteristics. Moreover, the program characteristic variables that we actually use in the regressions are measured in terms of the degree to which each program *changes* the pre-program regime—that is, the difference between service participa-

tion rates and sanction rates for the program group and the control group. Although program designs may reflect the characteristics of the available client population, it is not apparent that *changes* in how programs are run would be affected by client and site characteristics, assuming that these characteristics remain fairly stable.

Although the regressions were mainly estimated to determine the sources of Riverside's and Portland's exceptional performance, a useful by-product is the information they provide about the design features of welfare-to-work programs that are most effective and the circumstances under which these programs work best. Thus, we briefly discuss the regression findings. The key finding is that while factors that contribute to the design of welfare-to-work programs appear to influence their effects, so do the characteristics of program participants and the socio-economic characteristics of the program sites.

Earnings

Table 2 indicates that three program activities have a statistically significant relationship with program effects on earnings during the 3rd quarter: sanctions, the extent of job search, and the amount of vocational training. The first two of these relationships are positive. The coefficients on participation in vocational training activities, however, are negative, albeit statistically significant only during the 3rd quarter after random assignment. The negative relationship during this quarter may result from persons who participate in vocational training being unlikely to work at the same time. The coefficients on basic education and work experience are statistically insignificant and small for both the 3rd and 7th quarters, implying that increases in the provision of these services relative to controls have little or no effect on program effects on earnings.

Some recent welfare-to-work programs not only provide services, but also incorporate financial incentives that reward program participants who take jobs. The coefficient on the dummy variable represent-

ing programs with financial incentives is negative, implying that programs with such incentives have earnings effects that are smaller than earnings effects of programs without such incentives. Granted that the 3rd quarter coefficient is statistically insignificant and the 7th quarter coefficient is only marginally significant, such a negative effect is nonetheless theoretically possible. Financial incentives should, in theory, unambiguously encourage work among welfare recipients who otherwise would not work; but their impact on the hours and, hence, earnings of welfare recipients who are already working can be either positive or negative (see Blank, Card, and Robins 2000).

Two participant characteristics independently affect program effects on earnings.⁶ Programs in which the average age of participants was less than 30 have a smaller

impact during both the 3rd quarter and the 7th quarter than programs with older participants. In addition, the earnings effects of programs increase with the proportion of participants who were white (that is, neither black nor Hispanic), but this relation is statistically significant in only the 3rd quarter.

How local conditions may mediate program effects on earnings is suggested by the influence of unemployment and poverty rates. The findings suggest, for example, that program effects on earnings shrank during the 7th quarter by about \$10 for each percentage point increase in the unemployment rate. However, the 3rd quarter coefficient on the unemployment rate is positive, small, and statistically insignificant. The poverty rate appears to have a U-shaped relation with program effects on earnings, first causing them to fall as the poverty rate increases and then causing them to rise. Although the coefficient estimates that imply this U-shaped relationship are statistically significant only for the 3rd quarter, the coefficients for the 7th quarter are similar. This U-shaped relationship, the bottom of which occurs just below the sample mean of 15%, suggests that welfare-to-work programs are especially effective in areas with exceptionally plentiful or exceptionally limited economic opportunities. In the former case, they may help better prepare individuals for a relatively large number of available jobs; and in the latter case, they may help disadvantaged persons compete for the scarce opportunities that exist.

Receipt of AFDC

Table 1 indicates that the impact of welfare-to-work programs on the receipt of AFDC is typically small, less than two percentage points. However, the standard deviations of these mean effects are over twice as large as the means themselves, suggesting that there is considerable variation across programs. Table 3 indicates that during the 3rd quarter, little of this variation is attributable to program differences in job search, training, or sanction

⁶Like us, Michalopoulos and Schwartz (2001) and Bloom, Hill, and Riccio (2003) assembled data from multiple random assignment evaluations of welfare-to-work programs. Unlike us, however, they did not find evidence that program effects on earnings differed by the characteristics of those who participated in the evaluated programs. Although the reasons our findings differ from the findings of these two other studies are not entirely clear and merit further examination, there are at least two possible explanations. First, both Michalopoulos and Schwartz (2001) and Bloom, Hill, and Riccio (2003) are based on a subset of fewer than half the programs included in our study. It is possible that their findings and ours differ because different samples of programs were examined. Second, both studies are based on the characteristics of individuals (for example, whether a given individual is white or black), while our study is based on impact estimates drawn from evaluation reports and aggregate measures of participant characteristics (for example, the percentage of each study sample that was white or black). In principle, the former data should be superior to the latter. However, it is possible that our aggregate measures reflect certain program- or site-level influences that are missed by variables measured on individuals. For example, a high proportion of blacks or Hispanics among program participants would suggest that the AFDC population is mainly located in inner cities where relatively few jobs are located. Such an effect should be picked up by aggregate measures of the sort we use, but may not be picked up by data on individuals.

*Table 3. Weighted Regression
Estimates for Program Effects on
AFDC Receipt During the 3rd and 7th
Quarters after Random Assignment.*

<i>Variable</i>	<i>Quarter 3</i>	<i>Quarter 7</i>
Constant	16.716 (11.131)	21.660** (10.220)
Program Effect on % Receiving		
Sanctions	0.023 (0.052)	0.172*** (0.062)
Job Search	0.064 (0.043)	0.083* (0.049)
Basic Education	0.010 (0.047)	-0.046 (0.059)
Vocational Training	0.071 (0.135)	0.354*** (0.162)
Work Experience	-0.058 (0.100)	-0.140 (0.120)
Program Includes		
Financial Incentives	-7.735*** (1.478)	-7.787*** (1.517)
Time Limit	5.739*** (1.861)	2.581 (2.447)
Participant Characteristics		
% White	0.065** (0.028)	0.078*** (0.029)
% Female	-0.172 (0.114)	-0.209* (0.106)
% Employed in Year Prior to R.A.	-0.105** (0.045)	-0.133*** (0.050)
% under Age 25	0.082** (0.031)	0.012 (0.034)
<i>R</i> ²	.360	.392
Number of Observations	41	49

*Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

rates; the regression coefficient estimates on these variables are small and statistically insignificant. Given that the regressions imply that sanction and participation rates influence earnings in the 7th quarter after random assignment, the lack of influence in the 3rd quarter may reflect a lag in redetermining AFDC eligibility after a job is obtained. Indeed, coefficients that are associated with program effects on sanctioning and participation in job search and

vocational training are all sizable, positive, and statistically significant by the 7th quarter. Although the sign on vocational training is negative in the earnings regressions and positive in the receipt of AFDC regressions, it is statistically significant only in the earnings regressions for the 3rd quarter and in the AFDC regressions for the 7th quarter. This may result because different persons are affected. In the 3rd quarter, some individuals may lose earnings while they are participating in vocational training. It is possible that other persons who were assigned to vocational training refused to participate (perhaps because they were already working at jobs in the underground economy) and, as a result, were dropped from the AFDC rolls by the 7th quarter.

The regression estimates imply that inclusion of financial incentives in a welfare-to-work program, typically by increasing the amount of earnings disregarded in calculating AFDC entitlement, shrinks its impact on reducing AFDC receipt by nearly 8 percentage points in both the 3rd and 7th quarters. This is probably because raising the earnings "disregard" increases the amount of earnings that AFDC recipients can receive and still remain on the rolls. Time limits on how long families could remain on AFDC, which some welfare-to-work programs in our sample included, apparently had a substantial 3rd quarter effect, increasing program impact on reducing the receipt of AFDC by around 6 percentage points, even though program participants had not actually reached the time limit in any of the programs in our sample by then.

None of the variables that measure the socio-economic characteristics of the program sites have a statistically significant relation with program effects on AFDC receipt. They were therefore dropped from the regressions. However, a number of participant characteristics were found to be associated with movements off AFDC. For example, the results imply that a welfare-to-work program with a caseload with fewer white participants, more female participants, and a higher proportion of pro-

Table 4. Values of the Explanatory Variables Used in the 3rd Quarter Earnings Regressions for a “Typical Site” and for Riverside and Portland.

Variable	“Typical Site”		Observed Values		Differences from “Typical Site”	
	Mean	Standard Deviation	Riverside	Portland	Riverside	Portland
Net Program Effect on % Receiving:						
Sanctions	8.69	9.74	6.00	14.00	-2.69	5.31
Job Search	21.18	11.91	36.60	32.20	15.42	11.02
Basic Education	6.13	10.33	18.20	5.30	12.07	-0.83
Vocational Training	2.77	4.55	-1.80	7.30	-4.57	4.53
Work Experience	3.09	5.36	-0.60	7.10	-3.69	4.01
Program Includes:						
Financial Incentives	0.27	0.42	0.00	0.00	-0.27	-0.27
Time Limits	0.13	0.32	0.00	0.00	-0.13	-0.13
Participant Characteristics						
% White	37.27	21.92	51.20	69.60	13.97	32.33
% Female	90.99	5.31	88.00	93.20	-2.99	2.21
% Employed in Year before R.A.	39.07	13.82	49.30	39.00	10.23	-0.07
Average Age < 30	0.39	0.47	0.00	0.00	-0.39	-0.39
% under Age 25	27.99	23.43	10.20	22.70	-17.79	-5.29
Site Characteristics						
Unemployment Rate	6.65	2.56	6.70	4.30	0.05	-2.35
Poverty Rate	15.53	4.76	10.80	14.30	-4.73	-1.23
Poverty Rate Squared	241.18	160.72	116.64	204.49	-124.54	-36.69

Notes: Net program effects on sanctions and receipt of services are estimated by deducting the relevant control group values from the relevant program group values. Site and participant characteristics are means. “Average Age < 30” is a dummy variable that equals one if the average age of the program group was under 30 and zero if the average age was above 30 for individual sites; it will assume a value between zero and one if averaged across sites.

gram participants who were employed at some time during the year prior to random assignment would have less effect on reducing the receipt of AFDC in both the 3rd quarter and the 7th quarter than would an otherwise similar program. The negative effect of previous work may seem surprising but could occur if persons with recent work experience are better able than others to leave the AFDC rolls without the aid of a welfare-to-work program.

How Do Riverside and Portland Differ from Other Program Sites?

Table 4 compares Riverside and Portland with a “typical site” in terms of the explanatory variables used to estimate the

3rd quarter earnings regression equations that appear in Table 2. Presumably, these differences contributed to Riverside’s and Portland’s exceptional performance. As previously mentioned, the values for a “typical site” are simply averages for all the observations used to estimate each of the regressions.⁷

⁷Because somewhat different sets of observations were used to estimate the 3rd and 7th quarter regressions and the earnings and AFDC receipt regressions, the values for the 7th quarter earnings regressions and the 3rd and 7th quarter AFDC receipt regressions vary slightly from those in Table 4. Moreover, the values reported for the unemployment and poverty rates in the 3rd quarter differ a bit from those reported for the 7th quarter because they change over time.

Table 4 indicates that the welfare-to-work program in Riverside, as compared to that in a "typical site," had a relatively low net sanction rate and relatively high net rates of participation in job search and basic education. Moreover, Riverside actually reduced participation in vocational training and work experience. Unlike the other statistics, the slightly below-average net sanction rate may appear inconsistent with Riverside's work-first approach. However, the program was notable for the clear message about finding a job quickly that it communicated to clients (Riccio, Friedlander, and Freedman 1994), and this may have reduced the need for sanctions. Participants in Riverside's welfare-to-work program tended to be considerably older and more likely to be employed during the year prior to random assignment than program participants at a "typical site." Moreover, a considerably larger proportion of Riverside's participant population was white and a somewhat smaller proportion was female. Finally, the unemployment rate in the Riverside area was a little higher than that in a "typical site" and the poverty rate was substantially lower.

Relative to a "typical site," Portland's welfare-to-work program, like Riverside's, was characterized by comparatively large increases in participation in job search. In contrast to Riverside, however, net sanctioning and participation in vocational training and work experience were higher in Portland than in a "typical site," while net participation in basic education was lower. Participants in the Portland program were similar to the norm in age, percentage female, and the likelihood of having been employed during the year before random assignment. However, the proportion that was white was considerably larger. Finally, Portland's unemployment and poverty rates were lower than average.

Neither the Riverside program nor the Portland program provided financial incentives or included time limits. As will be seen, the absence of those features helps explain why the program effects at the Riverside and Portland sites differed from those at a "typical site."

Explaining Riverside and Portland's Exceptional Performances

By multiplying each of the differences between Riverside and Portland and a "typical site" (last two columns of Table 4) by the corresponding regression coefficient (first column of Table 2), we can examine how each factor contributed to the exceptionally large earnings effects found in Riverside and Portland in the 3rd quarter. A similar procedure is used to examine the factors that account for Riverside's and Portland's exceptional performance in increasing earnings in the 7th quarter and in reducing the receipt of AFDC in the 3rd and 7th quarters. The findings from this exercise are shown in Tables 5 and 6 for earnings and the receipt of AFDC, respectively. These tables report both the absolute contributions and the percentage contributions of each factor to the total observed differences between program effects at Riverside and Portland and those at a "typical site," which are shown in the last two columns of Table 1.

Several factors appear to have made especially important contributions to Riverside's and Portland's exceptionally large effects on earnings. These include the strong emphasis at both sites on job search. In addition, Portland's relatively high use of sanctions and Riverside's limited use of vocational training also increased their effects on earnings. However, Portland's heavy use of vocational training reduced its impact. The fact that the caseloads in both sites were older and mostly white also contributed to the large effects on earnings. However, site characteristics, such as the local unemployment and poverty rates, do not seem to have been important.

As in the case of earnings, relatively high net rates of participation in job search at the Portland site and (especially) the Riverside site contributed to their above-average success in reducing the percentage of cases receiving AFDC. That neither program incorporated financial incentives was also important, as were certain caseload characteristics—for example, the relatively high

Table 5. Estimates of the Influence of Program and Contextual Characteristics on the Difference in Program Effects on Earnings between Riverside/Portland and a "Typical Site."

Variable	Quarter 3				Quarter 7			
	Riverside		Portland		Riverside		Portland	
	Absolute	%	Absolute	%	Absolute	%	Absolute	%
Program Effect on % Receiving:								
Sanctions	-12.00	-5.43	23.68	21.58	-16.43	-6.86	29.53	12.48
Job Search	48.08	21.78	34.36	31.32	30.73	12.82	21.97	9.28
Basic Education	-19.70	-8.92	1.36	1.24	-5.29	-2.21	0.90	0.38
Vocational Training	33.17	15.03	-32.88	-29.97	21.18	8.84	-19.06	-8.05
Work Experience	7.67	3.47	-8.33	-7.60	-4.64	-1.94	4.91	2.07
Program Includes:								
Financial Incentives	12.29	5.57	12.29	11.20	17.37	7.25	17.37	7.34
Participant Characteristics								
% White	22.72	10.29	52.73	48.06	9.24	3.85	27.93	11.80
Average Age < 30	18.68	8.46	18.68	17.02	31.27	13.05	31.27	13.21
Site Characteristics								
Unemployment Rate	0.16	0.07	-7.53	-6.86	-9.53	-3.98	24.66	10.42
Poverty Rate	175.94	79.71	45.75	41.70	87.77	36.62	42.04	17.75
Poverty Rate Squared	-147.83	-66.97	-43.55	-39.69	-67.48	-28.15	-34.23	-14.46

proportion of caseloads in both sites that were white. Riverside's relatively high proportion of male-headed cases also contributed to its success in reducing receipt of AFDC, while the relatively high proportion of clients who worked during the year before random assignment had the opposite effect.

The analysis so far suggests that both program design and participant and site characteristics contributed to the exceptional success of the Riverside and Portland programs in increasing earnings and reducing the receipt of AFDC, but does not reveal their relative importance. Table 7 addresses this issue.

The table is divided into two panels. The upper panel is based on the earnings regressions from Table 2; the lower, on the AFDC receipt regressions from Table 3. Row 1 in each panel shows the total observed differences between program effects in Riverside and Portland and those in a "typical site." These values also appear in Table 1. Row 2, computed by summing

each column in Tables 5 and 6, indicates the extent to which the regressions reported in Table 2 successfully "explain" these differences. Row 3, which is computed by summing the first six rows in Table 5 and the first seven rows in Table 6, indicates the part of each "explained" difference attributable to program design. Similarly, row 4, which is computed by summing the remaining rows in Tables 5 and 6, indicates the part due to participant and site characteristics. Finally, row 5 (row 1 minus row 2) shows what part of the observed difference cannot be explained by the regressions.

Row 2 in Table 7 indicates that, with the exception of the two 7th quarter predictions for Riverside, the regressions successfully explain between two-thirds and three-quarters of the total observed difference in program effects between Riverside or Portland and a "typical site." The "prediction error" is probably due, in part, to differences in program design that are inherently difficult to measure and, consequently, omitted from most of the study reports and

Table 6. Estimates of the Influence of Program and Contextual Characteristics on the Difference in Program Effects on AFDC Receipt between Riverside/Portland and a "Typical Site."

Variable	Quarter 3				Quarter 7			
	Riverside		Portland		Riverside		Portland	
	Absolute	%	Absolute	%	Absolute	%	Absolute	%
Program Effect on % Receiving:								
Sanctions	-0.054	-2.43	0.128	2.49	-0.556	-16.58	0.820	8.50
Job Search	0.966	43.12	0.683	13.29	1.293	38.58	0.926	9.60
Basic Education	0.110	4.90	-0.025	-0.49	-0.480	-14.33	0.109	1.13
Vocational Training	-0.332	-14.82	0.314	6.10	-1.788	-53.36	1.434	14.86
Work Experience	0.250	11.15	-0.196	-3.82	0.596	17.80	-0.482	-4.99
Program Includes:								
Financial Incentives	2.088	93.23	2.088	40.63	2.258	67.41	2.258	23.40
Time Limits	-0.631	-28.18	-0.631	-12.28	-0.181	-5.39	-0.181	-1.87
Participant Characteristics								
% White	0.812	36.25	2.010	39.11	0.784	23.40	2.213	22.94
% Female	0.843	37.63	-0.052	-1.00	0.807	24.08	-0.280	-2.90
% Employed in								
Year Prior to R.A.	-0.943	-42.09	0.139	2.70	-1.185	-35.37	0.185	1.91
% under Age 25	-1.469	-65.59	-0.437	-8.50	-0.229	-6.84	-0.079	-0.82

hence from the regressions—for example, staff attitudes toward training participants, or the message staff communicate to participants about whether they need to find a job as soon as possible or should instead wait until they can find a “good” job. It probably also results from client and site characteristics that are omitted from the regressions because they are not in our database and from sampling error that results because estimates of the effects of welfare-to-work programs rely on samples from the programs’ target populations and, hence, are not statistically precise. Because sampling error does not result from systematic differences among programs, even if no relevant explanatory variables were omitted from the regressions, the observed difference between program effects in Riverside and Portland and those in a “typical site” could not be fully explained. Considering that at least some of the unexplained ways in which the program effects for Riverside and Portland differ from those for a “typical site” are likely due to sampling error, with the exception of the 7th quarter for Riverside, the regressions seem to do a

pretty good job of “explaining” these differences.

The key implication of Table 7 is suggested by a comparison of row 3 with row 4. All four comparisons in the upper panel of the table imply that various observed contextual factors associated with client and site characteristics were at least as important as program design in accounting for Riverside’s and Portland’s extraordinary success in increasing the earnings of program participants. In contrast, all four comparisons in the bottom panel imply that most of Riverside’s and Portland’s exceptional success in reducing the percentage of program participants on the AFDC rolls is attributable to program design features. Indeed, there is a negative sign in row 4, implying that observed client and site characteristics actually tended to reduce program impact on AFDC receipt. Thus, it appears that the design of the programs in Riverside and Portland contributed considerably more to their relatively large effects on AFDC receipt than to their exceptionally sizable effects on earnings. Perhaps it is not surprising that those

Table 7. Summary Statistics for the Influence of Program and Contextual Characteristics on the Difference in Program Effects between Riverside/Portland and a “Typical Site.”

Variable	Quarter 3				Quarter 7			
	Riverside		Portland		Riverside		Portland	
	Absolute	%	Absolute	%	Absolute	%	Absolute	%
Earnings								
(1) Total Observed Difference	220.73	100.00	109.72	100.00	239.17	100.00	236.75	100.00
(2) Total “Explained” Difference	139.18	63.05	96.55	88.00	94.17	39.29	147.28	62.21
(3) Due to Program Design	69.52	31.49	30.47	27.77	42.91	17.90	55.63	23.50
(4) Due to Other Factors	69.66	31.56	66.08	60.23	51.26	21.39	91.65	38.71
(5) Unexplained Difference	81.55	36.95	13.17	12.00	145.00	60.71	89.47	37.79
AFDC Receipt								
(1) Total Observed Difference	2.24	100.00	5.14	100.00	3.35	100.00	9.65	100.00
(2) Total “Explained” Difference	1.64	73.17	4.02	78.21	1.32	39.39	6.93	71.76
(3) Due to Program Design	2.40	106.97	2.36	45.90	1.14	34.12	4.89	50.62
(4) Due to Other Factors	-0.76	-33.80	1.66	32.31	0.18	5.27	2.04	21.14
(5) Unexplained Difference	0.60	26.83	1.12	21.79	2.03	60.61	2.72	28.24

responsible for administering AFDC can exercise more control over whether recipients continue to receive benefits than over the earnings they receive.

Conclusions

This study has focused on two well-known welfare-to-work programs that have often been held up as models worthy of emulation: those that operated in Riverside and Portland.

These two programs have attracted widespread attention mainly because of well-conducted random assignment evaluations indicating that they were exceptionally effective. We have used meta-analytic techniques to investigate why these programs were found to produce more positive effects than most other such programs in terms of increasing the earnings of program participants and decreasing their dependence on AFDC. The meta-analysis provides several interesting insights.

It demonstrates, for example, that the strong showing of the welfare-to-work programs in Riverside and Portland was due, in part, to certain features of their design, such as the large effect that both programs had on participation in job search and Portland’s large effect on the use of sanc-

tions. Interestingly, these are attributes usually associated with the work-first approach, although, as discussed earlier, neither Riverside’s nor (especially) Portland’s program slavishly followed this model.

However, the findings from this study suggest that, at least in the case of earnings effects, the apparently superior performance of the Riverside and Portland programs was only partly attributable to the design of these programs and, hence, only partly due to factors under administrators’ control. Certain contextual factors—for example, the older and mostly white caseloads in Riverside and Portland—were perhaps more important. This insight is of consequence because it suggests that other sites that attempt to replicate the Riverside or Portland program models are unlikely to obtain identical effects if, as is likely, their contextual characteristics differ. Indeed, the findings imply that a single welfare-to-work model does not fit all circumstances. Although both Riverside’s and Portland’s programs were successful, as indicated by Table 4, the two cities ran different programs in different environments. Consequently, as implied by Tables 5 and 6, their success was due to different factors.

Moreover, some of the observed differ-

ences between Riverside and Portland and other sites are probably attributable to sampling error. This highlights the danger of simply comparing effects estimated in an evaluation of one program with those obtained from an evaluation of another, and then declaring one program a “winner” and the other a “loser,” even though both evaluations are based on random assignment and are otherwise of high quality. Meta-analysis provides more appropriate tools for making valid comparisons.

Finally, the results imply that program administrators can both improve the earnings of participants in welfare-to-work programs and reduce their dependence on welfare performance by increasing the use of job search and sanctions. However, the evidence suggests that program administrators have considerably more control over the ability of welfare-to-work programs to push families off the welfare rolls than they do over using these programs to increase the earnings of program participants, where contextual factors appear to play a much more important role. Viewed from a differ-

ent perspective, the results suggest that substantial reductions in welfare rolls, such as those that occurred in the 1990s, do not necessarily imply that those previously receiving welfare are better off. There is not necessarily a corresponding increase in earnings.

It is important to recognize the limitations of this study. Perhaps the most critical is that while we were able to obtain useful measures of program design, we did not have measures of program administration—for example, the degree to which the organizational units responsible for different program components cooperated, the quality of program leadership, or the message that program administrators communicated to program participants about the speed with which they were expected to find employment and the type of jobs that were considered acceptable. Such factors, only some of which are subject to administrative decision-making, may also have been important determinants of the success of the Riverside and Portland programs.

Appendix Table A
Welfare-To-Work Evaluations Included in the Database

<i>Program Title</i>	<i>Short Program Name</i>	<i>Evaluator/Author</i>	<i>Mid-Point of Random Assignment</i>
Community Work Experience Demonstrations	West Virginia	MDRC	1983
WORK Program	Arkansas	MDRC	1983
Employment Initiatives	Baltimore	MDRC	1983
San Diego Job Search and Work Experience Demonstration	San Diego	MDRC	1983
Employment Services Program	Virginia	MDRC	1984
Job Search and Work Experience in Cook County	Cook County	MDRC	1985
Saturation Work Initiative Model In San Diego	SWIM	MDRC	1985
Saturation Work Program	Philadelphia	PA Department of Public Welfare	1986
Teenage Parent Demonstration	Teenage Parents	Mathematica Policy Research	1988
Wisconsin Welfare Employment Experiment	Wisconsin	University of Wisconsin	1988
California's Greater Avenues for Independence Program	GAIN	MDRC	1989
Ohio Transitions to Independence Demonstration	Ohio	Abt Associates	1990
JOBS Program	Florida	MDRC	1991
National Evaluation of Welfare-to-Work Strategies	NEWWS	MDRC	1993
To Strengthen Michigan Families	Michigan	Abt Associates	1993
Family Transition Program	FTP (Florida)	MDRC	1994
Minnesota Family Investment Program	MFIP	MDRC	1994
Family Investment Program	Iowa	Mathematica Policy Research	1994
Vermont's Welfare Restructuring Project	Vermont	MDRC	1995
Indiana Welfare Reform Program	Indiana	Abt Associates	1995
Jobs First	Connecticut	MDRC	1996
Los Angeles Jobs-First GAIN Evaluation	Los Angeles	MDRC	1996
A Better Chance	ABC (Delaware)	Abt Associates	1996
Virginia Independence Program	VIEW	Mathematica Policy Research	1996

REFERENCES

- Ashworth, Karl, Andreas Cebulla, David Greenberg, and Robert Walker. 2004. "Meta-Evaluation: Discovering What Works Best in Welfare Provision." *Evaluation*, Vol. 10, No. 2 (April), pp. 193–216.
- Blank, Rebecca, David Card, and Philip K. Robins. 2000. "Financial Incentives for Increasing Work and Income among Low-Income Families." In Rebecca Blank and David Card, eds., *Finding Work: Jobs and Welfare Reform*. New York: Russell Sage, pp. 373–419.
- Bloom, Howard S., Carolyn J. Hill, and James Riccio. 2003. "Linking Program Implementation and Effectiveness: Lessons from a Pooled Sample of Welfare-to-Work Experiments." *Journal of Policy Analysis and Management*, Vol. 22, No. 4 (Fall), pp. 551–75.
- Friedlander, Daniel, David H. Greenberg, and Philip K. Robins. 1997. "Evaluating Government Training Programs for the Economically Disadvantaged." *Journal of Economic Literature*, Vol. 35, No. 4 (December), pp. 1809–55.
- Greenberg, David, Charles Michalopoulos, and Philip K. Robins. 2003. "A Meta-Analysis of Government-Sponsored Training Programs." *Industrial and Labor Relations Review*, Vol. 57, No. 1 (October), pp. 31–53.
- Hedges, Larry V. 1994. "Fixed Effects Models." In Harris Cooper and Larry V. Hedges, eds., *The Handbook of Research Synthesis*. New York: Russell Sage Foundation, pp. 285–300.
- Michalopoulos, Charles, and Christine Schwartz, with Diana Adams-Ciardullo. 2001. *National Evaluation of Welfare-to-Work Strategies: What Works Best for Whom? Effects of 20 Welfare-to-Work Programs by Subgroup*. Washington, D.C.: U.S. Department of Health and Human Services, Administration for Children and Families, Office of the Assistant Secretary for Planning and Evaluation, and U.S. Department of Education, Office of the Under Secretary, Office of Vocational and Adult Education.
- Raudenbush, Stephen W. 1994. "Random Effects Models." In Harris Cooper and Larry V. Hedges, eds., *The Handbook of Research Synthesis*. New York: Russell Sage Foundation, pp. 301–21.
- Riccio, James, Daniel Friedlander, and Stephen Freedman. 1994. *GAIN: Benefits, Costs, and Three-Year Effects of a Welfare-to-Work Program*. New York: MDRC.
- Scrivener, Susan, et al. 1998. *National Evaluation of Welfare-to-Work Strategies: Implementation, Participation Patterns, Costs, and Two-Year Effects of the Portland (Oregon) Welfare-to-Work Program*. Washington, D.C.: U.S. Department of Health and Human Services, Administration for Children and Families, Office of the Assistant Secretary for Planning and Evaluation, and U.S. Department of Education, Office of the Under Secretary, Office of Vocational and Adult Education.
- Shadish, William R., and C. Keith Haddock. 1994. "Combining Estimates of Effect Size." In Harris Cooper and Larry V. Hedges, eds., *The Handbook of Research Synthesis*. New York: Russell Sage Foundation, pp. 261–81.